Random mutagenesis of the human immunodeficiency virus type-1 *trans*-activator of transcription (HIV-1 Tat)

David P.Siderovski, Toshifumi Matsuyama, Elena Frigerio, Stephen Chui, Xia Min, Heather Erfle¹, Martin Sumner-Smith¹, Richard W.Barnett¹ and Tak W.Mak

The Ontario Cancer Institute and Departments of Immunology and Medical Biophysics, University of Toronto, 500 Sherbourne Street, Toronto, Ontario M4X 1K9 and ¹Allelix Biopharmaceuticals Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1, Canada

Received July 23, 1992; Accepted September 15, 1992

EMBL accession number X64650

ABSTRACT

A new method is described for the direct construction of randomly mutagenized genes by applying the polymerase chain reaction (PCR) to an oligonucleotide synthesized using doped nucleotide reservoirs. We have demonstrated the utility of this method by generating a library of mutant HIV-1 *tat* genes. Several arbitrarily selected, inactive *tat* clones were sequenced to evaluate the extent of the mutagenesis. Moreover, fourteen recombinants encoding varying levels of transcriptional *trans*-activator activity were isolated by transient transfection of sub-library pools into a HeLa cell line bearing an HIV-LTR-chloramphenicol acetyltransferase (CAT) reporter gene. Sequence data revealed a spectrum of alterations including nucleotide substitutions, insertions, and deletions, suggesting that mutations arose from both the doped DNA synthesis and the subsequent PCR 'rescue' of full-length product. Sequence comparison between inactive and active Tat clones revealed a selection pressure against amino-acid substitutions within the N-terminal domains of Tat, indicating the importance of this region to *trans*-activation competence. In addition, single and double missense mutations within the basic-rich, TAR RNA-binding domain were seen to be tolerated within active Tat clones.

INTRODUCTION

Replication of the human immunodeficiency virus type-1 (HIV-1) is critically dependent on the action of the viral regulatory protein Tat (1, 2), which activates efficient viral transcription by interacting with the *trans*-activation responsive (TAR) element, a structured RNA sequence present in the 5' non-coding region of all viral transcripts (reviewed in 3, 4). The structural requirements for Tat-TAR interaction have recently been described; a basic-rich region of the Tat primary sequence (RKKRRQRRR; aa 49–57) mediates direct binding of Tat to TAR RNA (5–9). However, an equally precise definition of the transcriptional activation domain(s) of Tat remains elusive. One candidate activation domain is the N-terminus of Tat, which contains three acidic residues (Glu2, Asp5, Glu9) in a proline-

rich domain predicted to form an acidic, amphipathic α -helix (10), common among activation domains of eukaryotic transcription factors (11). Yet replacement of these acidic residues with the corresponding acid amides (thereby removing negative charge without disrupting predicted α -helical character) fails to significantly affect Tat function (12). A second candidate, the 'core' domain (aa 32–47), has been identified by its evolutionary conservation among *trans*-activators of the human immunodeficiency, simian immunodeficiency, and equine infectious anemia lentiviruses (HIV, SIV, EIAV; refs. 13–15).

To explore the functional tolerance of Tat to random mutation(s) within all regions of the protein and to facilitate the isolation of mutant Tat proteins with novel characteristics (*e.g.*, dominant negative or temperature-sensitive phenotype), we decided to create a randomly mutagenized *tat* gene library. Chemical mutagens such as nitrous acid have been used to generate random mutations within pre-existing DNA segments; however, chemical mutagenesis often results in a high proportion of wild-type sequences, a low diversity of possible substitutions, and the appearance of mutational 'hot spots' (16–18). Therefore, we chose to mutagenize the *tat* gene using a wholly synthetic approach, in which gene segments are created by automated solid phase DNA synthesis using contaminated or 'doped' nucleotide precursors (19, 20). By doping each of the four nucleoside phosphoramidite mixes (A, C, G, and T) with small, defined amounts of the other three nucleotides, random substitutions can occur with equal frequency at each base of the synthetic oligonucleotide.

Synthetic gene construction has commonly involved the construction of a series of overlapping oligodeoxynucleotides comprising both DNA strands, which are subsequently annealed and ligated to form the complete gene (21, 22). However, random mutations within overlapping regions, as introduced by doped DNA synthesis, would likely decrease annealing efficiency, whereas avoiding mutations within overlapping regions would restrict the diversity of possible mutations. Recently, we have demonstrated an alternative to this 'block ligation' gene assembly approach. We have shown that gene-sized DNA fragments can be constructed in their entirety by automated DNA synthesis of one complete strand and subsequent amplification by the polymerase chain reaction (PCR) (23). Post-synthesis PCR is

required due to the low abundance of full-length oligonucleotides within the synthetic product, even with highly efficient base coupling (98–99%; ref. 24).

In this paper, we describe the creation of a randomly mutagenized *tat* gene library using complete strand synthesis in combination with doped phosphoramidites. Initial characterization of this library was performed by sequencing 12 arbitrarily selected *tat* clones. In addition, this library was screened for recombinants retaining HIV-1 LTR *trans*-activation function by a rapid sib selection/grid analysis screening assay. Analysis of both active and inactive *tat* clones from this one study was found to confirm and extend the results of several previous *Tat* mutagenesis studies, including the tolerance of the TAR-binding domain to missense mutations and the importance of N-terminal regions in transcriptional activator function.

MATERIALS AND METHODS

Reagents, cell lines, and plasmids

Reagents and solvents for automated DNA synthesis were purchased from Aldrich Chemicals, except for the nucleoside phosphoramidites (Applied Biosystems). The Sequenase ver. 2.0 kit and *Taq* polymerase were purchased from U.S. Biochemicals. GeneClean was obtained from BIO 101. *E. coli* SURE competent bacteria were purchased from Stratagene. TransfectACE liposome transfection reagent, Opti-MEM I reduced serum medium, restriction endonucleases, and T4 DNA ligase were purchased from Gibco-BRL. Qiagen anion-exchange columns were obtained from Diagen.

The HeLa TARCAT cell line bears stably integrated copies of both pSV2neo (25), conferring resistance to the antibiotic G418, and pU3RIII (26), a plasmid expressing the bacterial chloramphenicol acetyltransferase (CAT) gene under the control of the HIV-1 LTR. HeLa TARCAT cells were cultured at 37°C and 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), 100 µg/mL penicillin, 100 µg/mL streptomycin, and 250 µg/mL G418.

Plasmid pGEM5zf+ and phosphorylated *Bam*HI linkers were purchased from Promega. Plasmid pSV2hyg was generated by Andrew Smith by replacement of the neomycin-resistance gene of pSV2neo with the hygromycin-resistance (*hyg*) gene (25, 27). The *Eco*RV restriction site in the pGEM5zf+ multiple-cloning site was converted to a *Bam*HI site by digestion with *Eco*RV and insertion of a *Bam*HI linker. A 2.5 kb *Bam*HI fragment of pSV2hyg bearing the SV40 early promoter, *hyg* gene cassette, and SV40 polyadenylation signal was subsequently inserted into the *Bam*HI site of pGEM5zf+, creating plasmid pGEMSV2hyg. A wild-type *tat* cDNA was obtained by PCR amplification of pSVE183 (28) using the 5'TatATG and 3'TatTAG primers as described below. This *tat* cDNA cassette was digested with *Hind*III and *Bam*HI and substituted for the *Hind*III-*Bgl*II *hyg* gene cassette of pGEMSV2hyg, creating the eukaryotic *Tat* expression plasmid pGEMSV2tat^{wt}.

Oligonucleotide synthesis

Small scale syntheses of the PCR primers (0.15 µmol) and the doped 280-nt oligonucleotide (0.1 µmol) were performed on an Applied Biosystems 380B synthesizer using standard 2-cyanoethylphosphoramidite chemistry (29).

Doped amidite pots were prepared by mixing pre-weighed aliquots of the individual nucleoside CE-phosphoramidites. For

example, to prepare 'doped G' with an overall doping frequency of 4.5% (see Results for discussion), 767.6 mg of G-amidite (95.5% by moles) + 12.4 mg A-amidite (1.5%) + 11.8 mg C-amidite (1.5%) + 10.6 mg T-amidite (1.5%) were added to a clean, dry septum vial. The amidite mixture was dried under vacuum overnight, flushed with dry argon, and subsequently dissolved in 9.0 mL dry acetonitrile (MeCN) to give a 0.1 M stock solution (total amidite).

The doped 280-mer was synthesized on 'low-loaded', 100 angstrom controlled-pore glass (CPG, 4–7 µmole nucleoside/gram, ChemGenes Corp.) using the following synthesis cycle: (a) Detritylation with 2.5% dichloroacetic acid in dichloromethane (v/v) (7×13 second deliveries), (b) Coupling with 0.1 M CE-phosphoramidite (5 µmole) and 0.5 M tetrazole (25 µmole) in dry MeCN (5 second delivery + 35 second WAIT step), (c) Capping with 10% acetic anhydride in 8:1 tetrahydrofuran(THF)/2.6-lutidine (v/v/v) + 6.5% N,N-dimethylaminopyridine in THF (w/v) (9 second delivery + 2 second WAIT step), (d) Oxidation with 0.1 M iodine in 40:10:1 THF/lutidine/water (v/v/v) (15 second delivery + 30 second WAIT step). MeCN washes were performed following the detritylation, capping, and oxidation reactions (20–60 seconds) using MeCN freshly distilled from calcium hydride.

The protected oligonucleotides were cleaved from the CPG-support and fully deblocked by standard treatment with ammonium hydroxide (29). The deprotected oligonucleotides were evaporated to dryness under vacuum. The primers were PAGE-purified according to standard protocols (29).

The evaporated 280-mer was redissolved in 400 µL sterile water to yield a working stock solution of 10 µg/µL (total DNA). The crude 280-mer was size-fractionated as follows. Aliquots (15 µL) of crude stock solution were electrophoresed through a 4% agarose gel and bands with electrophoretic mobilities between 200–260 bp were excised. DNA was recovered by electroelution (IBI apparatus), butanol-extracted, ethanol-precipitated, and dissolved in 60 µL sterile water to yield a stock solution of ~0.4 µg/µL.

Polymerase chain reaction (PCR)

Aliquots of crude or fractionated 280-mer stock solutions were used directly or serially diluted (1/10 to 1/100,000) for PCR amplification (ref. 30). PCR was carried out on a programmable thermal cycler (Ericomp TwinBlock) in 50 or 100 µL PCR buffer (10 mM Tris-Cl, pH 8.3/1.5 mM MgCl₂/50 mM KCl) containing 200 µM dNTPs, 2 U *Taq* polymerase, and 1.25 µM each of primers 5'TatATG and 3'TatTAG, complementary to the 5' and 3' ends of the synthetic product, respectively. The 41 nt primer 5'TatATG (5'-GGGTTTGAAGCTTAGATCTCG-AGAGAGGAGAGCAAGAAATG-3') contained *Hind*III, *Bgl*II, and *Xho*I sites (underlined), whereas the 45 nt primer 3'TatTAG (5'-GGCCTTTGGATCCTAGGTCGACTATTCCTT-CGGGCCTGTCTGGGTC-3') contained *Bam*HI, *Avr*II, and *Sal*I sites (underlined). Reaction mixtures were subjected to 25 or 35 amplification cycles following these general conditions: denaturation 2 min, 90°C or 94°C; annealing 2 min, 50–66°C; extension 5 min 72°C; final incubation 30 min 72°C. Several attempts to enhance the amplification of crude product (by increasing the stringency of primer annealing in latter cycles) were also performed using the following amplification scheme: 5 cycles of [denaturation 2 min, 90°C; annealing 2 min, 60°C; extension 5 min, 72°C], followed by 25 cycles of [denaturation

2 min, 90°C; annealing and extension 5 min, 80°C]. One-tenth volume aliquots of the amplification mixtures were analyzed by electrophoresis on 2–4% agarose gels.

Cloning and sequencing of mutant *tat* genes

Residual primers and unincorporated dNTPs were removed from nine-tenth volume aliquots of PCR amplification mixtures using GeneClean directly, without prior preparative gel electrophoresis, as described by the manufacturer. DNA was redissolved in TE (10 mM Tris-Cl, 1 mM Na₂EDTA, pH 8.0) and digested with *Hind*III (100 U) and *Bam*HI (100 U) for 2 hr at 37°C. Short digestion fragments were removed using a second round of GeneClean and DNA was again redissolved in TE. 20 ng of digested *tat* PCR product was then added to 500 ng of pGEMSV2hyg vector, previously predigested with *Hind*III and *Bgl*II to remove the *hyg* gene cassette and purified by two consecutive rounds of preparative 1% agarose gel electrophoresis and GeneClean isolation. The insert-vector mixture was diluted to 250 μ L in a buffer containing T4 DNA ligase (2.5 U) and 1 $\times$ ligase buffer (Gibco-BRL) supplemented with an additional 0.6 mM ATP; forced-directional ligation was carried out at 14°C

for 18 hr. 50 μ L aliquots of resultant ligation mixture were used to transform competent *E. coli* SURE cells, and ampicillin-resistant bacteria were selected from YT + ampicillin (100 μ g/mL) plates and restreaked onto YT + ampicillin plates in numerical order.

Plasmid DNA from pools of colonies or single colonies (for screening for active *tat* clones) was prepared using a quick, alkaline-lysis miniprep protocol previously described (31). Plasmid DNA from selected single colonies (for sequencing and quantitative CAT assays) was prepared using Qiagen Pack-100 columns by the manufacturer-specified protocol, except eluted DNA was precipitated by 2 vol ethanol. Qiagen-purified plasmid DNA was sequenced by the double-strand Sequenase procedure as described by the manufacturer; each *tat* clone was sequenced over both strands.

Trans-activation assay

Transient transfection of 70–80% confluent HeLa TARGAT monolayers with saturating amounts of plasmid DNA/Transfect ACE liposome mixtures was performed exactly as per manufacturer's directions for 6-well (35-mm) tissue culture dishes. A modified dilution scheme was used for 24-well (15-mm) tissue culture dishes: 500 ng plasmid DNA diluted in 30 μ L Opti-MEM I reduced serum medium was mixed with 5 μ g TransfectACE reagent diluted in 30 μ L Opti-MEM and allowed to stand for 15 min at room temperature; this DNA/liposome mixture was then further diluted to a total volume of 300 μ L with Opti-MEM and plated over target cells for incubation at 37°C and 5% CO₂ for 18 hr. DNA/liposome-containing medium was then removed and replaced by DMEM with 20% FCS. Forty-eight hours after transfection, cell extracts were prepared (32), protein concentrations of cell extracts were measured using the method of Bradford (33), and CAT activity was determined using the radiolabel-sparing, fluorescence method of Aubin *et al.* (34). Both unacetylated and acetylated [¹⁴C]-chloramphenicol was measured using liquid scintillation counting to determine percentage transacetylation (32).

RESULTS

Synthesis of a randomly mutagenized *tat* gene library

The nucleotide sequence programmed for doped DNA synthesis represented a 280-nt subset (nts 18–297) of the top (5' $\rightarrow$ 3') strand of a *tat* gene encoding an 86 amino-acid protein conforming to the consensus Tat polypeptide sequence (EMBL accession # X64650; ref. 13). This 280-nt subset extended across 24 nts of the 5' untranslated region, included the initiator ATG codon and the entire open-reading frame of the *tat* gene, and ended at the first position of the terminator TAG codon. As automated solid phase DNA synthesis is carried out in the 3' $\rightarrow$ 5' direction (Figure 1), the first nucleotide of the termination codon (T, nt 297) was directly coupled to the controlled-pore glass (CPG) support and, therefore, was the sole invariant nucleotide of the synthetic product.

Monte Carlo simulations of the doped synthesis were performed using the computer program RAMHA (D.P. Siderovski, *submitted*). A phosphoramidite doping level of 4.5% was predicted to yield an average of 8 missense mutations per synthesized *tat* gene; moreover, this level of doping was predicted to be the minimum required to ensure that no synthesized *tat* genes would encode wild-type sequences. Hence, a doping level

Figure 1. Synthesis of a randomly mutagenized *tat* gene library. Sequence for the top strand of the HIV-1 *tat* gene (nts 18–297 of EMBL accession # X64650) was programmed into an automated DNA synthesizer containing doped nucleoside phosphoramidite reservoirs, including three nucleotide mixes designed to avoid random introduction of stop codons (as illustrated for the first base of lysine codon AAA; see text). Oligonucleotide synthesis proceeded in a 3' $\rightarrow$ 5' direction, yielding both the full-length (280 nt) sequence and (280-x) nt failure sequences, all of which have identical 3' ends. Aliquots of the resultant synthetic product were subjected to size-fractionation using agarose gel electrophoresis. Two specific primers (primer 1 = 3'TatTAG and primer 2 = 5'TatATG) were added, containing sequences overlapping those of the 3' and 5' ends of the synthetic product, respectively. PCR was performed for 25 cycles to amplify specifically the full-length product (320 bp). Restriction enzyme sites present in the 5' ends of the two PCR primers were then used for cloning the cDNAs into an eukaryotic expression vector for sequencing and functional studies.

Figure 2. Agarose gel electrophoresis of PCR amplification products. (A) PCR amplification (5 + 25 cycles) of crude synthetic product, as described in Methods. Lanes 1 and 2, PCR on 1 μ L of undiluted and 1/10-diluted crude oligonucleotide stock solution, respectively. (B) PCR amplification (25 cycles) of size-fractionated synthetic product using denaturation temperature of 94°C and different annealing temperatures. Lane 3, PCR on 1 μ L of 1/10-diluted fractionated stock solution; annealing at 66°C. Lanes 4 and 5, PCR on 1 μ L of undiluted or 1/10-diluted fractionated stock solution, respectively; annealing at 50°C. Lane M, *Hae*III-digested Φ X174 RF marker DNA (bp). The 320 bp product is indicated by the arrow.

of 4.5% was used for all four nucleoside phosphoramidite reservoirs (e.g., 'G'-reservoir = 95.5% G, 1.5% A, 1.5% C, 1.5% T). However, these same computer projections indicated that, without any attempt to circumvent the random introduction of stop codons (TAG, TGA, TAA), greater than one-third of the resultant *tat* genes would encode truncated polypeptides terminating before the TAR RNA-binding domain (aa 49–57). Therefore, three additional doped nucleoside reservoirs were used to avoid random placement of thymidine in the first position of codons xAG, xGA, or xAA, where x = A, C, G (see example for lysine codon AAA in Figure 1). Since the ABI 380B DNA synthesizer used in this study held a maximum of seven nucleotide reservoirs, we were unable to use an additional set of three 'anti-termination' nucleotide mixes required to avoid the random placement of adenine or guanine in the second or third position of codons Tx₂x₃, where either x₂ or x₃ = A or G (D.P. Siderovski, *submitted*).

Doped synthesis of the 280-nt sequence was performed as outlined in Materials and Methods. The average base coupling efficiency (C) observed over the first 100 synthetic cycles was 0.99 by DMT analysis. The theoretical yield (P) of full-length product of length N is given by $P = C^{(N-1)}$ (ref. 24); therefore, the estimated proportion of full-length 280-nt ssDNA in the resultant synthetic product was $P = (0.99)^{279} = 0.06$ or 6%, with $(1-P) = 0.94$ or 94% synthetic failure sequences. Given a 0.1 μ mole scale synthesis and assuming 1% recovery of 280-mer following cleavage/deprotection, the yield of full-length mutant *tat* oligonucleotides would be 60 pmoles.

We have previously demonstrated that low yield, full-length oligonucleotides greater than 200 nt long can be selectively and efficiently amplified directly from crude synthetic mixtures using the polymerase chain reaction (23). Such PCR amplification is complicated by the fact that the highly-abundant (N–x) synthetic failure sequences all contain equivalent 3' ends which compete with the full-length product for hybridization to the 3'-specific

PCR primer (Figure 1). Initial attempts to PCR amplify aliquots of the crude 280-mer synthesis mixture failed to give the desired 320 bp product, as analysed by agarose gel electrophoresis (e.g., Figure 2A, lanes 1 and 2). In order to reduce the amount of non-amplifiable, sub-280 nt failure sequences and thereby enhance the exponential amplification of full-length synthetic product relative to the linear replication of failure sequences, aliquots of the crude synthetic product were size-fractionated by agarose gel electrophoresis (see Methods). PCR amplification of the fractionated template yielded the expected 320 bp gene fragment (Figure 2B, lanes 3–5); sub-200 bp PCR products, resulting presumably from incomplete primer extension and/or mispriming, were substantially reduced by increasing the temperature during PCR primer annealing (Figure 2B, lane 3 vs 5).

One consequence of PCR amplification is the 'correction' of any randomly introduced mutations within the two primer-binding regions (see Figure 1). However, since the first 72 codons of *tat* (nts 39–275) are sufficient to encode full *trans*-activator function (35, 36), codons 80–86 within the 3' primer-binding site (nts 276–297) are not expected to encode an important functional domain of Tat. Furthermore, correction of the 5' primer-binding site ensures both preservation of the ATG initiation codon and conservation of the 5' untranslated regions for all mutant *tat* genes, including a Kozak motif 5' to the initiation codon (37).

Characterization of recombinants

The successfully amplified, mutant *tat* genes were isolated from residual primers and unincorporated dNTPs and cleaved with *Hind*III and *Bam*HI (sites present within the 5'TatATG and 3'TatTAG primers, respectively). Digested fragments were then cloned by forced-directional ligation between the SV40 early promoter and polyadenylation signal of the vector pGEMSV2hyg, as described in Materials and Methods. Ligation mixture aliquots were used to transform competent *E. coli* SURE cells and twelve *tat* clones were randomly selected from the resultant bacterial colonies (M5, M8, M9, M15, M21, M287, M307, M309, M319, M601, M605, M617). Plasmid DNA from these twelve recombinants was sequenced to determine the mutation frequency at the level of DNA (Figure 3) and protein sequence (Figure 6). All possible nucleotide substitutions were observed, dispersed across the 234 bp region (nts 42–275) between the two primer binding sites. Clones M21 and M605 represent members of a gene subfamily, sharing a similar, but not identical, collection of nucleotide substitutions, relative to the wildtype *tat* sequence (Figure 3). On average, there were 13.4 nucleotide changes per clone, representing an average substitution frequency of 5.7%. Most nucleotide changes altered a single position in a codon, although several codons possessed two nucleotide changes and one codon had three changes. Given the degeneracy of the genetic code, several nucleotide changes were silent, yielding an average 10.9 amino-acid changes per full-length protein. Not all *tat* clones are predicted to encode full-length (86 aa) proteins. Clone M319 is predicted to encode a 17 aa polypeptide due to a nonsense mutation (nts 90–92: CCT → TGA). Two clones (M309 and M617) possess insertions of an additional G, causing early frameshifts (Figure 6). Frameshifts caused by deletions were also evident: two clones possess single nucleotide deletions (M15 at nt 153, M601 at nt 244), whereas M287 has a deletion of 42 nucleotides (Δ 170–211), 30 nts downstream of a nonsense mutation (nts 138–140: TGC → TGA).

Figure 3. DNA sequences of 26 *tat* recombinants. Twelve randomly selected (M5, M8, M9, M15, M21, M287, M307, M309, M319, M601, M605, M617) and fourteen active (M13, M73, M296, M314, M626, M734, M741, M769, M866, M895, M910, M929, M936, M1001) *tat* clones were isolated and sequenced as described in Methods. Uppermost sequence and nucleotide numbering refer to wildtype *tat* gene (EMBL accession # X64650); single underline denotes codons encoding basic domain (see text), double underline denotes PCR primer-binding sites. Conserved nucleotides are indicated by a period (.), deletions by a dash (-); similar sequences are grouped together by a left brace.

All twelve *tat* clones were devoid of detectable HIV-1 LTR *trans*-activation activity, as assessed by transient transfection into HeLa TARCAT cells and measurement of resultant chloramphenicol acetyltransferase (CAT) activity (e.g., Figure 4). The loss of Tat function due to mutation could be caused by the disruption of a functional domain, a generalized disruption of protein folding, and/or increased protein/mRNA instability. The apparent loss of *trans*-activator function observed among these twelve *tat* clones can, for the most part, be rationalized by comparing the mutations encoded by each clone with those of previous site-directed mutagenesis studies. The inactive *tat* clones containing early truncations (M15, M287, M319, M617) or an early frameshift (M309) are expected to be functionally silent as all do not encode the basic-rich domain indispensable for binding to TAR (Arg49 to Arg57; refs. 5–9). In addition, inactive clones M5, M9, M21, and M605 all code for, among other mutations, at least two substitutions to cysteine residues each previously shown to be critical for Tat activity (Cys22, Cys25, Cys27, Cys30, Cys34, Cys37; refs. 38–40).

Replacement of His33, one of 10 missense mutations in inactive clone M8, was previously reported to reduce severely *trans*-activator function (41), in part by destabilizing the highly-folded, trypsin-resistant conformation of Tat (42). Hence, M307 is the only inactive clone isolated without a clear precedent for its complete lack of *trans*-activation activity; however, substitutions at Tyr47 have previously been shown to decrease Tat function 2- to 3-fold (10, 39).

Isolation of active *tat* clones

A rapid sib selection/grid analysis method (43) was used to screen the mutant *tat* gene library for individual clones encoding *trans*-activation activity (Figure 5). As cationic liposome-mediated transfection of eukaryotic cells is effective even with the use of impure DNA preparations (44, 45), a quick, alkaline-lysis protocol (33) was used to prepare plasmid DNA for transfection. In addition, a qualitative, fluorescence-based assay for chloramphenicol acetyltransferase activity was employed to circumvent the delay associated with autoradiography (34).

Figure 4. Relative *trans*-activation activities of *tat* recombinants. Plasmid DNA from active and randomly selected *tat* clones was isolated, transiently transfected into the indicator cell line HeLa TARCAT, and quantitatively assessed for LTR-directed *trans*-activator function by subsequent CAT assay as described in Methods. Autoradiographs A and B represent CAT assay reactions taken beyond the linear response range. (A) Average *trans*-activation values (relative to pGEMSV2tat^{wt} = 1.0) are derived from CAT assays run within the linear range from three independent transfection experiments (variation $\pm 10\%$). Transfection of the wild-type Tat expression plasmid pSVtat (ref. 5) demonstrated equivalent *trans*-activation as pGEMSV2tat^{wt} (data not shown). (B) Clone M769, although the weakest of the active *tat* clones (0.03), does exhibit *trans*-activation activity greater than that of the completely inactive *tat* clones M309, M319, M601, M605, and M617 (all <0.01). Cp = unacetylated [¹⁴C]-chloramphenicol, Ac = monoacetylated [¹⁴C]-chloramphenicol, Mock = mock transfection with pGEMSV2hyg parental vector, Tat^{wt} = pGEMSV2tat^{wt}.

Plasmid DNA from pools of 20 individual bacterial colonies each was transiently transfected into HeLa TARCAT cells (Figure 5(a–b)); out of 51 pools screened in this manner, fourteen pools were observed to *trans*-activate the HIV-1 LTR-CAT gene (e.g., Figure 5(c)). Each of these fourteen pools was then subjected to grid analysis (43) to identify the specific *tat* clone(s) within each pool harbouring *trans*-activation activity (e.g., Figure 5(d–e)). Hence, out of 1020 individual bacterial colonies tested, 14 unique, active *tat* clones were isolated (M13, M73, M296, M314, M626, M734, M741, M769, M866, M895, M910, M929, M936, M1001); two clones (M626 and M658) possessed identical DNA sequences, whereas one positive pool (M921–940) was found to contain two, distinct, active *tat* clones (M929 and M936) upon subsequent grid analysis.

DNA sequence obtained from these fourteen active *tat* clones revealed the presence of another two gene subfamilies (Figure 3). Clones M13, M296, and M314 share a set of eleven nucleotide substitutions, with M13 and M296 possessing additional A $\rightarrow$ G transitions not present in M314 (nts 80 and 156). Clones M769, M895, M929, M936, and M1001 form a more distantly related gene subfamily with the inactive *tat* clone M601. M769 and M601 share a near-identical set of nucleotide substitutions and deletions, with an additional 1 nt deletion present in M769 ($\Delta 297$); M929, M936, and M1001 contain various subsets of these mutations, and M895 possesses only three nucleotide changes found within the 3' region of M929 not homologous to M601/M769 (nts 194, 209, and 225). On average, there were 8.3 nucleotide changes per clone, representing an average substitution frequency of 3.5% and translating into an average 5.5 amino-acid changes per full-length protein (Figure 6). Clones M73 and M769 possess frameshifts caused by nucleotide deletions (M73: $\Delta 212$ – $\Delta 221$; M769: $\Delta 244$ and $\Delta 297$). Clones M13 and M296 are predicted to encode identical mutant Tat proteins.

Figure 5. Sib selection/grid analysis screening assay for active *tat* clones. (a) 200, sequentially numbered, individual bacterial colonies, obtained by transfection of *E. coli* SURE with the mutagenized *tat* gene library, were separately cultured overnight at 37°C in 1mL YT+ampicillin. 100 μ L aliquots were pooled into 10 cultures (5mL YT+amp) of 20 colonies each (e.g., M1–20, M21–40, ... M181–M200) and grown overnight. (b) Plasmid DNA was extracted from overnight cultures and lipofected into HeLa TARCAT cells in 6-well plates as described in Methods. (c) Protein extracts from transfected cells were assayed 48hr later for CAT activity using the non-radioactive, fluorescence method of Aubin *et al.* (34); Mk = Mock transfection of pGEMSV2hyg parental vector, lanes 1–7 = pools M621–M640 to M741–M760, lanes 8–10 = pools M821–840 to M861–M880, wt = pGEMSV2tat^{wt}. (d) Individual clones from CAT-positive pools were pooled into 9 cultures according to a 5 $\times$ 4 grid (example illustrated for M181–200), and plasmid DNA from overnight cultures were lipofected into HeLa TARCAT cells in 24-well plates as described in Methods. (e) Protein extracts were assayed for CAT activity to identify individual positive clones; lanes A–E and α – δ = isolation of positive clone M895 (E/ γ intersection) from pool M881–M900, lanes F–J and ϵ – θ = isolation of positive clone M910 (J/ ζ intersection) from M901–M920.

The relative biological activities of these fourteen active *tat* clones were measured by transient transfection into HeLa TARCAT cells and subsequent quantitative CAT assays. A wide spectrum of *trans*-activation function was observed, from activity similar to wild-type Tat (M910) to barely detectable activity (M769) (Figure 4). Clone M895 was observed to encode a mutant Tat protein ~5-fold more active than the related clone M929, which possesses three additional, N-terminal amino-acid substitutions (Figure 6). Clone M314 exhibited a *trans*-activation activity ~3-fold greater than that of related clones M13 and M296, which are predicted to encode an additional Thr40 → Ala replacement. Related clones M929, M936, and M1001 were seen to encode Tat proteins which, although weakly active, were each at least 5-fold more active than the closely related clones M601 and M769 (Figure 4A).

DISCUSSION

Two other synthetic *tat* gene constructions using long oligodeoxynucleotides (> 100 nt) have been described (46, 47). However, since both assembly methods rely on internal regions of oligonucleotide hybridization to generate the full-length gene, neither is directly amenable to random mutagenesis over the entire

coding region. We chose instead to synthesize the complete coding region of *tat* (280 nts) using precisely doped nucleotide mixes and, subsequently, to 'rescue' the low yield, full-length product using PCR amplification. We have previously shown that short PCR primers complementary to the 5' and 3' ends of a synthetic oligonucleotide can be used to amplify 179- and 234-mers from a crude synthesis mixture. Specific PCR amplification of the target DNA occurs even in the presence of synthetic failure sequences which share the same 3' end and which compete for the binding of the 3' PCR primer (23). This strategy failed to amplify the target 280-mer from crude *tat* oligonucleotide using a variety of PCR cycling conditions. This failure may reflect the lower yield of full-length 280-mer within the crude product vs prior 179- or 234-mer syntheses, a result of both the increased number of base couplings and slightly decreased coupling efficiencies of mixed phosphoramidites.

Recently, Ciccarelli *et al.* reported a similar PCR-based strategy for gene synthesis (48). In this approach, the inhibitory effect of failure sequences is overcome by synthesizing both the top and bottom strands of the target gene and performing PCR amplification on a mixture of both synthetic templates. We have found that, by size-fractionating the crude template prior to PCR amplification, the correct full-length gene product can be obtained

Figure 6. Protein sequences of *tat* recombinants and plot of relative Tat activity versus missense mutation position. Predicted polypeptide sequences of both active and inactive *tat* clones were translated from DNA sequence data (Figure 4). *Trans*-activation activity of active clones (relative to wild-type Tat) is denoted by a bar graph (=) to the left of the clone name. Wildtype protein sequence and amino-acid numbering refer to translation of *tat* gene sequence (EMBL accession # X64650). Underlined residues represent codons 'corrected' by 5'TatATG or 3'TatTAG PCR primer (see text). Conserved residues are indicated by a period (.) and termination codons by an asterisk (*). Conservative and non-conservative amino-acid substitutions are plotted as open circles (○) and closed squares (■), respectively; premature termination is plotted with a dash (-). Clone M13 (not listed) is predicted to encode the same Tat protein as that encoded by M296. Missense mutations of clones M15, M287, M309, M319, and M617 are not plotted as these clones are prematurely truncated or frameshifted before the TAR-binding domain.

in good yield, thereby obviating the time and expense of synthesizing the opposite strand. The ability to generate the full-length target DNA from a single template provides direct evidence that full-length 280 nt oligonucleotides (or nearly full-length (280-x) nt oligonucleotides, where $x < 24$ nt overlap of 5' PCR primer) were actually synthesized.

Initial characterization of the mutagenized *tat* gene library revealed no recurrent mutation common to all clones. However, in addition to the expected spectrum of random nucleotide substitutions, single nucleotide insertions and both point and gross deletions were observed. Furthermore, the average nucleotide substitution frequency among 12 arbitrarily selected *tat* genes (5.7%) was slightly higher than the 4.5% doping level used during synthesis. These additional mutations most likely arose during chemical DNA synthesis and/or PCR amplification (reviewed in 49). For example, G → A and A → G transitions, the most prevalent mutations arising from synthesis side reactions (50) and *Taq* polymerase misincorporation (49), respectively, were found to be over-represented among the 12 possible nucleotide substitutions, together accounting for 30% of all nucleotide changes observed (Figure 3). The presence of gene subfamilies among the isolated *tat* clones also suggests additional mutagenesis by PCR. For example, clones M13, M296, and M314 appear to be derived from the same primordial oligonucleotide and differ only in additional A → G transitions, whereas clones M601, M769, M895, M929, M936, and M1001 share only 5' or 3' regions of common mutations, suggestive of strand-switching or 'jumping PCR' (51).

The utility of this random mutagenesis method was explored by screening the resultant *tat* gene library for clones encoding HIV-1 LTR transcriptional *trans*-activator function. Fourteen active *tat* clones with widely varying activities were isolated out of 1020 individual clones tested, demonstrating that genes constructed in this manner can be translated into functional proteins. Of these fourteen active clones, none possess wild-type sequence at either the DNA or protein level; however, one clone (M910) has a sole, conservative, C-terminal missense mutation (Leu69 → Ile), and encodes essentially wild-type *trans*-activator activity. The lack of non-mutagenized *tat* genes among 1020 clones tested contrasts sharply with results frequently obtained using chemical mutagenesis procedures, in which up to 50% of resultant clones are found to contain no mutations (e.g., refs. 17, 18).

In comparison with 12 arbitrarily selected, inactive *tat* clones, the fourteen active clones possess a 2-fold lower average missense mutation frequency (5.5 mutations/active Tat vs. 10.9 mutations/inactive Tat). In addition, a plot of relative Tat activity versus missense mutation position clearly illustrates a non-random distribution of amino-acid substitutions in *trans*-activation-competent *tat* clones (Figure 6). There is an approximately equal number of missense mutations within the N-terminal (aa 1-40) and C-terminal (aa 41-80) halves of the six, full-length, inactive *tat* clones (34 N-half mutations vs. 37 C-half mutations). However, within the eleven, full-length, active *tat* clones, 75% of amino-acid changes are found in the C-terminal halves of the encoded proteins (15 N-half mutations vs. 46 C-half mutations). This selection pressure to avoid N-terminal mutations in active clones is statistically significant ($\chi^2 = 7.63$, df = 1, $p < 0.01$) and supports the notion that, in the context of full-length Tat, the N-terminal acidic/proline-rich (aa 1-21), cysteine-rich (aa 22-37), and 'core' (aa 32-47) domains all contribute to LTR-directed transcriptional activation (12, 15, 38, 39, 41). Indeed,

the scattergram plot in Figure 6 clearly maps the C-terminal border of the final N-end domain by illustrating a sharply increased tolerance of Tat to missense mutations past Ser46.

Two active clones, M866 and M1001, possess missense mutations to acidic residues within the putative acidic, amphipathic α -helix activation domain (aa 1-21; ref. 10), yet retain at least 20% of wildtype Tat activity. This finding is consistent with previous studies which have argued against the presence of a classical acidic activation domain within Tat; both Tiley *et al.* and Rice *et al.* have previously demonstrated that Glu2 or Asp5 substitutions which remove negative charge result in decreased activity but are not capable of abolishing Tat function (12, 42). Furthermore, substitution of Lys19 within active clones M734 and M866 suggests that the strong conservation of a basic charge at this position among different HIV-1 isolates (13, 52) does not reflect an essential structural feature. Most of the active *tat* clones contain one or two missense mutations within the TAR-binding domain (aa 49-57; Figure 6), confirming that this basic-rich domain is tolerant to substitution provided a threshold level of basic charge and arginine residues are present (5, 7, 8). In addition, the frameshift and early truncation of active clone M73 immediately following the basic-rich domain confirms that the critical functional domains of Tat are found within the first 58 amino-acids (9, 39).

Additional information regarding the functional domains/residues of Tat can be discerned by comparing the relative activities of *tat* clones sharing subsets of amino-acid substitutions. For example, M314 and M296 share five mutations, including Ile39 → Leu and Tyr47 → Asn within the core domain, and two substitutions within the TAR-binding domain. Tyr47 substitutions have been shown to decrease Tat function only 2- to 3-fold (10, 39), whereas leucine at position 39 represents a conservative change observed in HIV-1 strains RF and ELI (13); in addition, the loss of two arginine residues from the basic domain should decrease, but not ablate, *trans*-activation (5, 7, 53). In agreement with these predictions, M314 exhibits nearly 50% of wild-type *trans*-activation activity. However, M296 has an additional substitution within the core domain (Thr40 → Ala) and encodes only ~15% of wild-type activity. Substitution of Thr40 with alanine, in the context of full-length, wild-type Tat, was previously shown not to affect Tat function (10). The 3-fold relative decrease in activity of M296 versus M314 might therefore reflect a threshold of tolerance to multiple perturbations within the core domain.

The concept of a threshold of tolerance to mutations in the core domain might also explain the pronounced cluster of conservative substitutions observed between Ile39 and Ser46 in active Tat clones (see scattergram in Figure 6). For example, M626 encodes two missense mutations within the core domain, Ile39 → Val and Leu43 → Ile, the latter replacing a residue highly conserved among the core domains of HIV, SIV, and EIAV Tats (14). Replacement of Leu43 with the large aromatic phenylalanine was shown to decrease *trans*-activation at least 4-fold (53); as only a 2-fold decrease is observed for M626, the more conservative replacement of Leu43 with the branched aliphatic isoleucine, perhaps in concert with the conservative Ile39 → Val substitution, might be considered more compatible with the structure of Tat and/or the function of the core domain.

M895 and M929 share two missense mutations: Arg57 → Ser and Gln63 → Glu, the former not expected to decrease severely TAR RNA-binding (5, 7, 53) and the latter representing a conservative change seen in several, functional HIV-1 *tat* isolates

(13, 52). As only these two substitutions are found within M895, no dramatic loss of *trans*-activation activity is predicted and, indeed, M895 exhibits ~75% wild-type Tat activity. However, M929 contains three additional mutations (Pro14 → Gln, Lys29 → Ser, Thr40 → Ser) and exhibits a ~5-fold lower activity than M895. Replacement of Thr40 with serine can be considered even more conservative (*i.e.*, loss of one methyl group) than replacement with alanine (see above). This suggests that the two other replacement sites, Pro14 within the highly conserved Pro-X₃-Pro repeat sequence and Lys29 within the highly conserved, basic residue doublet (Lys28, Lys/Arg29) should be explored further as to their specific role(s) in Tat structure/function. To date, these two positions have only been examined by site-directed mutagenesis in combination with other substitutions (40, 41).

The remaining *tat* clones, M601, M769, M936, and M1001, are related by a complex series of common mutations and frameshifts (Figure 6). The observation that M936 and M1001 exhibit *trans*-activation activities similar to that of M929 indicates that the two additional substitutions not found in M929 (Ile45 → Leu and Glu2 → Met) are not capable of abolishing Tat function. However, M769 and M601, which also possess random C-terminal fusions, are severely deficient in *trans*-activation. Although C-terminal fusions of Tat with other mature proteins have previously demonstrated the ability to *trans*-activate appropriate chimeric promoters (*e.g.*, Tat-R17 coat protein, Tat-Rev; ref. 15, 54), the stability of the proteins encoded by M601 and M769 might be adversely affected by the random nature of their C-terminal fusions.

In this paper, we have described a novel, general strategy for random mutagenesis of a gene in which a long oligonucleotide encoding the entire gene is synthesized using doped nucleotide mixes, size-fractionated by agarose gel electrophoresis, and selectively amplified by the polymerase chain reaction. This approach has been used successfully to create a randomly mutagenized *tat* gene library. Rapid characterization of both inactive and active clones from this library has produced a general outline of the essential and expendable regions of Tat, both confirming the findings of several previous Tat mutagenesis studies and providing new insights into the structure/function relationship of this viral protein. The presence of additional point mutations, insertions, deletions, and mutant *tat* gene subfamilies within the library, a by-product of the chemical DNA synthesis and PCR 'rescue', was instrumental in providing some of these insights. Thus, the powerful and facile nature of this mutagenesis strategy should prove useful in identifying the functional domains of less-well-characterized proteins. In addition, it is hoped that the randomly mutagenized *tat* gene library described in this study will be helpful in the isolation of mutant Tat proteins exhibiting novel phenotypes.

ACKNOWLEDGEMENTS

We thank Raya Kuperman for the oligonucleotide syntheses, Neil Parkin and Nahum Sonenberg for the gift of HeLa TARCAT cells and pSVTat, Wai-Ping Fung-Leung for pSVE183, and Alexandra Ho and Lorenz Trümper for technical assistance. D.P.S. was supported by a Ph.D. Studentship from the Medical Research Council of Canada. This work was supported in part by a grant from the National Health Research and Development Program of Canada (T.W.M.), and a National Cooperative Drug Discovery grant (M.S.-S. and T.W.M.).

REFERENCES

- Dayton, A.I., Sodroski, J.G., Rosen, C.A., Goh, W.C. and Haseltine, W.A. (1986) *Cell*, **44**, 941–947.
- Fisher, A.G., Feinberg, M.B., Josephs, S.F., Harper, M.E., Marselle, L.M., Reyes, G., Gonda, M.A., Aldovini, A., Debouk, C., Gallo, R.C. and Wong-Staal, F. (1986) *Nature (London)*, **320**, 367–371.
- Rosen, C.A. (1991) *Trends Genet.*, **7**, 9–14.
- Haseltine, W.A. (1991) In Haseltine, W.A. and Wong-Staal, F. (eds.), *Genetic Structure and Regulation of HIV*. Raven Press, New York, Vol. 1, pp. 1–42.
- 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.
- Sumner-Smith, M., Roy, S., Barnett, R., Reid, L.S., Kuperman, R., Delling, U. and Sonenberg, N. (1991) *J. Virol.*, **65**, 5196–5202.
- Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D. and Frankel, A.D. (1991) *Science*, **252**, 1167–1171.
- Calnan, B.J., Biancalana, S., Hudson, D. and Frankel, A.D. (1991) *Genes Dev.*, **5**, 201–210.
- Kamine, J., Loewenstein, P. and Green, M. (1991) *Virology*, **182**, 570–577.
- Rappaport, J., Lee, S.-J., Khalili, K. and Wong-Staal, F. (1989) *New Biol.*, **1**, 101–110.
- Mitchell, P.J. and Tjian, R. (1989) *Science*, **245**, 371–378.
- Tiley, L.S., Brown, P.H. and Cullen, B.R. (1990) *Virology*, **178**, 560–567.
- Myers, G., Berzofsky, J.A., Korber, B., Smith, R.F. and Pavlakis, G.N. (1991) *Human Retroviruses and AIDS: A compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos National Laboratory, Los Alamos, NM.
- Dorn, P., DaSilva, L., Martarano, L. and Derse, D. (1990) *J. Virol.*, **64**, 1616–1624.
- Derse, D., Carvalho, M., Carroll, R. and Peterlin, B.M. (1991) *J. Virol.*, **65**, 7012–7015.
- Smith, M. (1985) *Ann. Rev. Genet.*, **19**, 423–462.
- Tymms, M.J., McInnes, B., Alin, P., Linnane, A.W. and Cheetham, B.F. (1990) *Genet. Anal. Techn. Appl.*, **7**, 53–63.
- Diaz, J.-J., Rhoads, D.D. and Roufa, D.J. (1991) *BioTechniques*, **11**, 204–210.
- Horwitz, B.H. and DiMaio, D. (1990) *Methods Enzymol.*, **185**, 599–611.
- Hutchison III, C.A., Swannstrom, R. and Loeb, D.D. (1991) *Methods Enzymol.*, **202**, 356–390.
- Wosnick, M.A., Barnett, R.W., Vicentini, A.M., Erfle, H., Elliott, R., Sumner-Smith, M., Mantel, N. and Davies, R.W. (1987) *Gene*, **60**, 115–127.
- Wosnick, M.A., Barnett, R.W. and Carlson, J.E. (1989) *Gene*, **76**, 153–160.
- Barnett, R.W. and Erfle, H. (1990) *Nucleic Acids Res.*, **18**, 3094.
- Gait, M.J. (1984) *Oligonucleotide Synthesis: A Practical Approach*. IRL Press, Oxford, UK.
- Southern, P.J., Berg, P. (1982) *J. Mol. Appl. Genet.*, **1**, 327–341.
- Rosen, C.A., Sodroski, J.G. and Haseltine, W.A. (1985) *Cell*, **41**, 813–823.
- Zalacain, M., Gonzalez, A., Guerrero, M.C., Mattaliano, R.J., Malpartida, F. and Jimenez, A. (1986) *Nucleic Acids Res.*, **14**, 1565–1581.
- Muesing, M.A., Smith, D.H. and Capon, D.J. (1987) *Cell*, **48**, 691–701.
- Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. and Tang, J.-Y. (1987) *Methods Enzymol.*, **154**, 287–313.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
- Morelle, G. (1989) *FOCUS*, **11**, 7–8.
- Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044–1051.
- Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
- Aubin, R., Fournier, R., Weinfeld, M. and Paterson, M.C. (1987) *Nucleic Acids Res.*, **15**, 10069.
- Cullen, B.R. (1986) *Cell*, **46**, 973–982.
- Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. and Haseltine, W. (1985) *Science*, **229**, 74–77.
- Kozak, M. (1987) *Nucleic Acids Res.*, **15**, 8125–8148.
- Ruben, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W.A. and Rosen, C.A. (1989) *J. Virol.*, **63**, 1–8.
- Kuppuswamy, M., Subramanian, T., Srinivasan, A. and Chinnadurai, G. (1989) *Nucleic Acids Res.*, **17**, 3551–3561.
- Garcia, J.A., Harrich, D., Pearson, L., Mitsuyasu, R. and Gaynor, R.B. (1988) *EMBO J.*, **7**, 3143–3147.
- Rice, A.P. and Carlotti, F. (1990) *J. Virol.*, **64**, 1864–1868.
- Rice, A.P. and Carlotti, F. (1990) *J. Virol.*, **64**, 6018–6026.
- McCormick, M. (1987) *Methods Enzymol.*, **151**, 445–449.

44. Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M. and Danielsen, M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7413–7417.
45. Karger, B.D. and Kormo, C. (1990) *FOCUS*, **12**, 25–27.
46. Adams, S.E., Johnson, I.D., Braddock, M., Kingsman, A.J., Kingsman, S.M. and Edwards, R.M. (1988) *Nucleic Acids Res.*, **16**, 4287–4298.
47. Ciccarelli, R.B., Loomis, L.A., McCoon, P.E. and Holzschu, D.L. (1990) *Nucleic Acids Res.*, **18**, 1243–1248.
48. Ciccarelli, R.B., Gunyuzlu, P., Huang, J., Scott, C. and Oakes, F.T. (1991) *Nucleic Acids Res.*, **19**, 6007–6013.
49. Eckert, K.A. and Kunkel, T.A. (1991) *PCR Methods Applic.*, **1**, 17–24.
50. Eadie, J.S. and Davidson, D.S. (1987) *Nucleic Acids Res.*, **15**, 8333–8349.
51. Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. (1990) *Nucleic Acids Res.*, **18**, 1687–1691.
52. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjo, B. and Wain-Hobson, S. (1989) *Cell*, **58**, 901–910.
53. Sadaie, M.R., Rappaport, J., Benter, T., Josephs, S.F., Willis, R. and Wong-Staal, F. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 9224–9228.
54. Southgate, C., Zapp, M.L. and Green, M.R. (1990) *Nature (London)*, **345**, 640–642.