5020–5024 Nucleic Acids Research, 1993, Vol. 21, No. 22

In vivo selection of randomly mutated retroviral genomes

Ben Berkhout and Bep Klaver
Department of Virology, University of Amsterdam, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Received August 20, 1993; Revised and Accepted September 29, 1993

ABSTRACT

Darwinian evolution, that is the outgrowth of the fittest variants in a population, usually applies to living organisms over long periods of time. Recently, *in vitro* selection/amplification techniques have been developed that allow for the rapid evolution of functionally active nucleic acids from a pool of randomized sequences. We now describe a modification of the nucleic acid-evolution protocol in which selection and amplification take place inside living cells by means of a retroviral-based replication system. We have generated a library of HIV-1 DNA genomes with random sequences in particular domains of the TAR element, which is the binding site for the Tat *trans*-activator protein. This mixture of HIV genomes was transfected into T cells and outgrowth of the fittest viruses was observed within two weeks of viral replication. The results of this *in vivo* selection analysis are consistent with the notion that primary sequence elements in both TAR bulge and loop domains are critical for Tat-mediated *trans*-activation and viral replication.

INTRODUCTION

In vitro evolution of functional RNA/DNA molecules is possible in the absence of living cells by a combination of biochemical techniques (Fig. 1, left). First, genetic variation is introduced into the sequence of interest, usually by synthesizing randomized oligodeoxynucleotides. For RNA molecules, a randomized pool is generated by *in vitro* transcription. Second, the most fit sequence is selected, for example by its ability to bind a protein. Third, selected nucleic acids are amplified, usually by the polymerase chain reaction. The cycle of selection, amplification and *in vitro* transcription can be repeated, at the end of which individual molecules are sequenced. This technique has been successfully applied to RNA molecules either in solution or attached to a solid support (1–5). *In vitro* evolution becomes possible when an error-prone amplification step is included in the protocol (6,7).

A schematic outline of the intracellular RNA evolution protocol we tested in this report is shown in Fig.1 (right). Short patches of randomized sequence were introduced into a full-length, infectious DNA clone of the human immunodeficiency virus (HIV-1). This plasmid pool is transfected into HIV-susceptible cells containing the CD4-receptor and virus replication is continued for a prolonged period of time. Competition of the variants in this continuous replication cycle will lead to the preferential survival of the fittest viruses. The result may be the dominance of one or a few genotypes. Compared to the *in vitro* selection/amplification steps, much time and effort is saved in the *in vivo* approach. Furthermore, adaptive mutations may arise in the viral evolution system during reverse transcription, which is estimated to result in 1 in 10000 base substitutions per nucleotide per replication cycle (Figure 1: fine-tuning). Using this approach, we analyzed the sequence requirements for the single-stranded bulge and loop of the HIV-1 TAR RNA element.

MATERIALS AND METHODS

Construction of randomized HIV genomes and viral culturing

In order to reconstruct HIV genomes with randomized TAR sequences, sense oligonucleotides encompassing the TAR region were synthesized. Variable sequence was created in bulge or loop domains (R-bulge and R-loop) by mixed nucleotide incorporation at the positions underlined (GGCTCGAGCTGGGTCTCTCTG-GTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAA-CTAGGGAACCCACTGCTTAA). This sequence was extended and amplified in a PCR-reaction with the partially complementary, antisense oligo 3'PCR (GCCAAGCTTTATTG-AGGCTTAAGCAGTGGG) and sense oligo 5'PCR (GTAATA-CGACTCACTATAGGCTCGAGCTGG). The 5'PCR primer encodes the promoter for T7 RNA polymerase (not used in this study). We used a standard PCR protocol (35 cycles of 1 min 95°C, 1 min 55°C and 2 min 72°C) and a 50-fold molar excess of primers 5'PCR and 3'PCR (100 ng each) compared to the internal 78-mer in order to favour production of the full-length product.

The ds-DNA PCR product was digested with XhoI and HindIII (recognition sequences are underlined), subcloned in a modified Blue-5'LTR plasmid. This Bluescript-derived plasmid contains the complete HIV-1 LTR as *XbaI-ClaI* fragment. The *XhoI* site of the Bluescript polylinker was removed by ligation of blunted *SalI* and *KpnI* sites, and a unique *XhoI* site was introduced in the HIV-1 LTR at position −9 for cloning purposes. Furthermore, the Blue-LTR plasmid carried a defective LTR with a bulge-deletion in order to exclude rescue by wild-type input sequences. Subsequently, the randomized 5'LTR was inserted as XbaI-ClaI fragment into the full-length molecular clone pLAI (8). The corresponding wild-type plasmid contained the HXB-2 TAR sequences, that differ from the original pLAI molecular

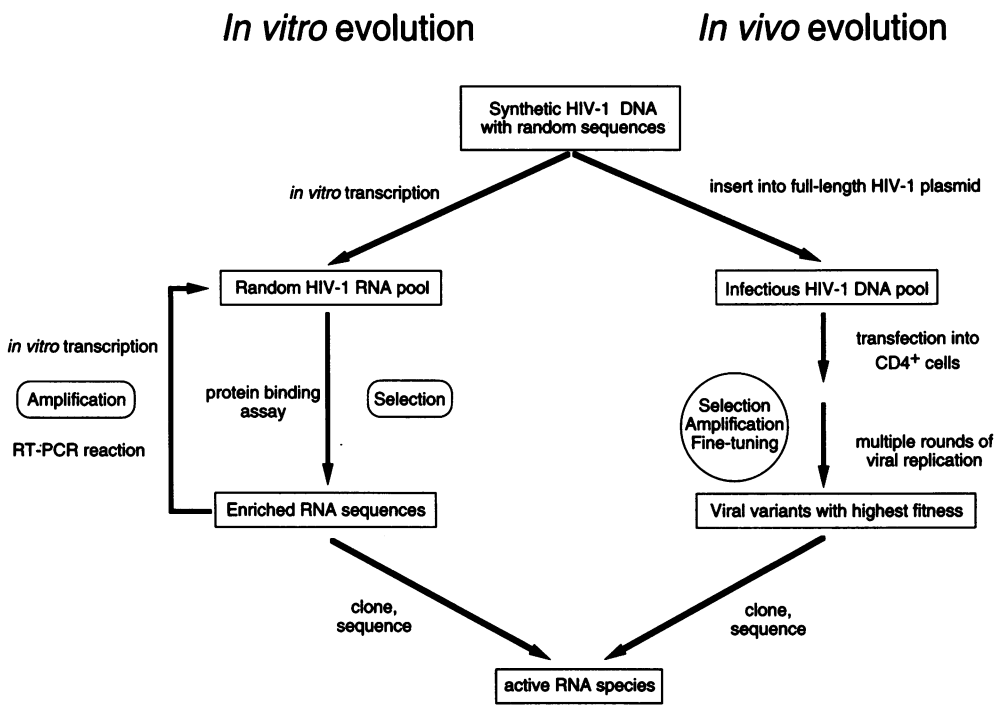

Figure 1. Schematic diagram of the *in vitro* and HIV-based *in vivo* RNA selection protocols. Details are discussed in the text. RT is Reverse Transcriptase, PCR is the polymerase chain reaction.

clone in the sequence of the TAR bulge (UCU and UUU, respectively). In order to avoid recombination between the randomly mutated 5'LTR and a wild-type 3'LTR, we used a modified, defective pLAI plasmid with a tri-nucleotide substitution in both 5' and 3' TAR bulges (Klaver and Berkhout, unpublished data).

The full-length HIV-1 DNA pool was transfected by electroporation into the SupT1 T cell line and viral replication was continued for up to 5 weeks. Cell-free culture supernatant was passed onto fresh SupT1 cells when massive HIV-induced syncytia were observed. Virus was passaged at least once before isolation of proviral DNA for sequence analysis in order to avoid contamination with the input HIV-1 DNA sequences. The proviral 3'LTR DNA was cloned by PCR using nef-primer 2 (CATGCGGCCGCAATAGAGTTAGGCAGGGATA) and LTR-U5 region primer C(N1) (CCAGACTCCCTAGAGAT-CAATGGTCTCAG), cloned into *Bam*HI-*Hind*III digested Bluescript KS$^+$ plasmid and sequenced on a Biosystems automated sequencer using the DyeDeoxy Terminator Cycle kit and LTR-U3 region primer 5'CE (CTACAAGGG-ACTTTCCGCTGG).

RESULTS

We have tested the potential of *in vivo* selection by creating a pool of HIV-1 genomes with random mutations in the long terminal repeat (LTR) that encodes the transcriptional promoter. Replication of the HIV-1 virus is critically dependent on the action of the viral regulatory protein Tat (9, reviewed in 10,11), 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. This TAR element (Fig. 2) has been extensively analyzed by classical mutagenesis in transient LTR-CAT transfection assays (10,11) and more recently by means of full-length infectious proviral constructs (12,13). Sequences capable of forming the RNA stem-loop secondary structure are critical for activity (14–16). In addition, both the sequence of the 3-nucleotide bulge (position 22–24) and the 6-nucleotide loop (position 29–34) are important for *trans*-activation of the viral LTR promoter (17,18). The critical role of the bulge is easily understood, since this is the region at which Tat contacts TAR (19,20). The hexanucleotide loop has been suggested to interact with cellular proteins involved in Tat-mediated LTR activation (21–23).

We synthesized two initial populations of mutant HIV genomes, either with a randomized bulge or with random substitutions at position 3–5 of the loop. The randomness was checked by sequence analysis of the two DNA pools (Fig. 2). We also sequenced twelve individual clones chosen at random from the two libraries in order to proof that no individual sequence predominated (Table 1, input sequences). The sequences obtained are indicative of the random nature of these regions (base composition of bulge/loop trinucleotides: G=28%, A=25%, U=28%, C=19%). The two randomized HIV-1 TAR pools were tested for the ability to generate replication-competent virus after introduction of the DNA into SupT1 cells. The spread of infectious virus through the T cell culture was monitored by the Gag p24 assay (Fig. 3) and the appearance of HIV-induced syncytia (not shown) at various days post-infection. The two HIV-1 DNA pools produced virus much delayed when compared to an equimolar amount of the fully wild-type HIV plasmid. This result supports the idea that only a limited subset of trinucleotides can serve as a functionally active TAR bulge or loop.

We consistently measured a more dramatic effect of randomization of the loop sequences on infectivity when compared to randomization of the bulge. Comparing the 5 μg

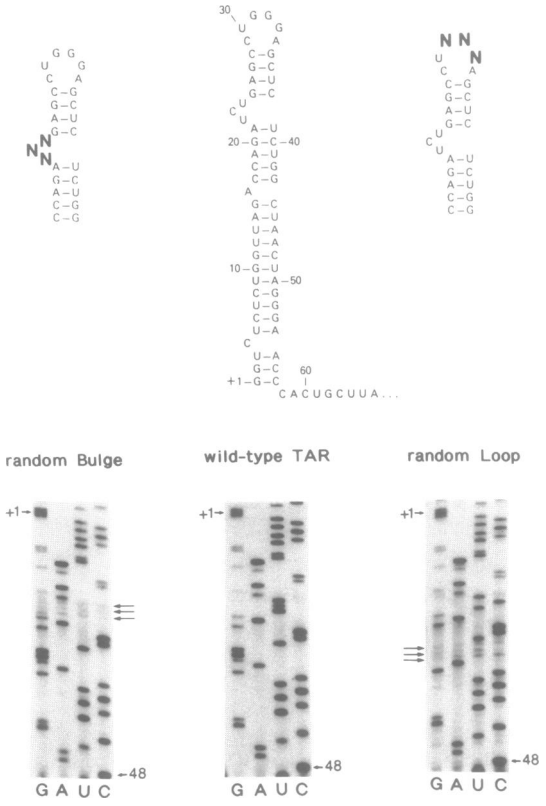

Figure 2. Secondary structure model of a natural HIV-1 TAR element (middle, HXB-2 isolate) and the bulge- or loop-randomized variants (left and right, respectively). Random nucleotides are indicated by N, +1 is the 5'end of all viral transcripts. The HIV-1 plasmid pLAI (containing an UUU bulge) and the two randomized pools were sequenced using the LTR-R region primer 80/50 (GCTTTATTGAGGCTTAAGCAGTGGGTTCCC) and the Sequenase enzyme. The position of the randomized triplets is indicated by arrows.

transfections of the R-Bulge and R-Loop DNAs, an approximately 10-day delay in virus virus production was seen for the latter (Fig. 3). We therefore performed an additional transfection with 25 μg of the R-loop DNA, resulting in virus production around day 8 (Fig. 3). This culture was used for the subsequent sequence analysis. The results so far strongly suggest that less sequence variation is allowed for the TAR loop when compared to the bulge. Interestingly, the initial virus production observed in cells transfected with the Random-bulge library (Figure 3, day 7–9) was consistently slower compared to the spread of wild-type HIV virus. This result confirmes that a substantial number of TAR bulge variants could support viral replication, albeit at a reduced efficiency compared to wild-type HIV-1.

To determine which TAR sequences were responsible for the replication-competent phenotype, we screened 10 individual clones from Random-loop transfected cells after 15 days of passage of the virus population onto fresh T cells. The relative complexity of the input DNA is dramatically reduced in that all but one clone contained the GGG trinucleotide (Table 1, right). Thus, only TAR loops with the GGG sequence seem compatible with high viral replication levels. The consensus loop derived from these experiments is strikingly similar to the nucleotide distribution of TAR loops in natural HIV-1 isolates (Fig. 4). A somewhat different picture emerged from the Random-bulge experiment. Here, some sequence heterogeneity was observed

Table 1. *In vivo* selection of TAR bulges (left) and TAR loops (right) from random sequences

Input bulge	day 13	day 19	day 25	day 33
CCU[1]	UAC	UCU	UCU	UCU
GCA[1]	UCG	UCU	UCU	UCU
UUC[1]	UAC	UAC	UCU	UCU
GAA[2]	UCA	UCU	UCC	UCU
UAA[2]	UAC	UCU	UCU	UCU
UGU[2]	UAC	UAG	UCU	
UCG[2]	UAU	UCU	UCG	
	UAC	UCU	UCU	
	UAC	UCU	UCU	
			UCU	

Input loop	day 15
AAA[1]	GGG
CGC[1]	GGG
UUG[2]	GGG
GAG[2]	GGG
GUU[2]	GUG
UGC[2]	GGG
	GGG
	GGG
	GGG
	GGG

Input sequences were derived from seven (left) and six (right) individual clones chosen at random from the subconstruct Blue-5'LTR[1] or the full-length, infectious HIV-1 plasmid[2].

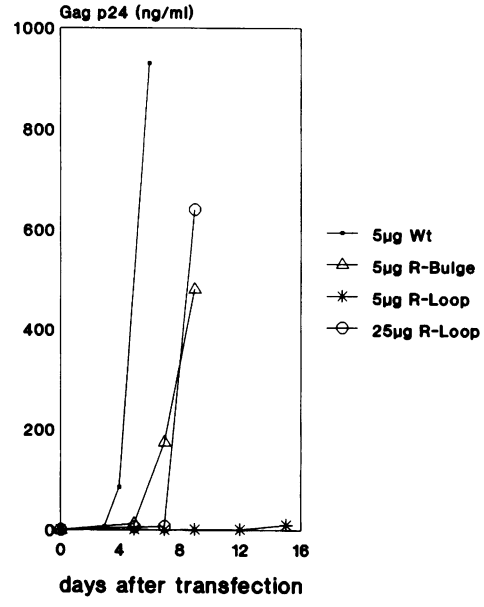

Figure 3. Transfection of the SupT1 cell line with wild-type HIV DNA (UCU bulge and CUGGGA loop) and the two randomized TAR libraries. Electroporation was used to transfect 5 μg cesium chloride-purified plasmid DNA into 5×10^6 T cells. For the Random-loop pool, we also used 25 μg DNA. Supernatant samples were collected at the indicated times and assayed for Gagp24. The culture was monitored for the presence of HIV-induced syncytia, at which time point 200 μl of the cell-free supernatant was transferred to a 10 ml fresh SupT1 culture (2×10^6 cells). The 5 μg Random-loop transfection did result in virus production upon prolonged incubation (not indicated).

at day 13 after transfection (Table 1, left). At this time point, however, the Bulge-1 position was found to be exclusively occupied by a U-residue. Apparently, this is the most critical bulge residue, consistent with previous mutagenesis studies. One intriguing aspect of *in vitro* selection, that also holds for our *in vivo* approach, is the ability to assess the changing nature of the virus population through time. For instance, we sequenced a total of 33 clones in the Random-bulge selection experiment at different time points (days 13,19,25,33, see Table 1 and Fig. 4). After 19 days of selective pressure in tissue culture, the UCU sequence was found in 7 out of 9 TAR bulges sequenced. Further propagation of the virus revealed that the pool converged

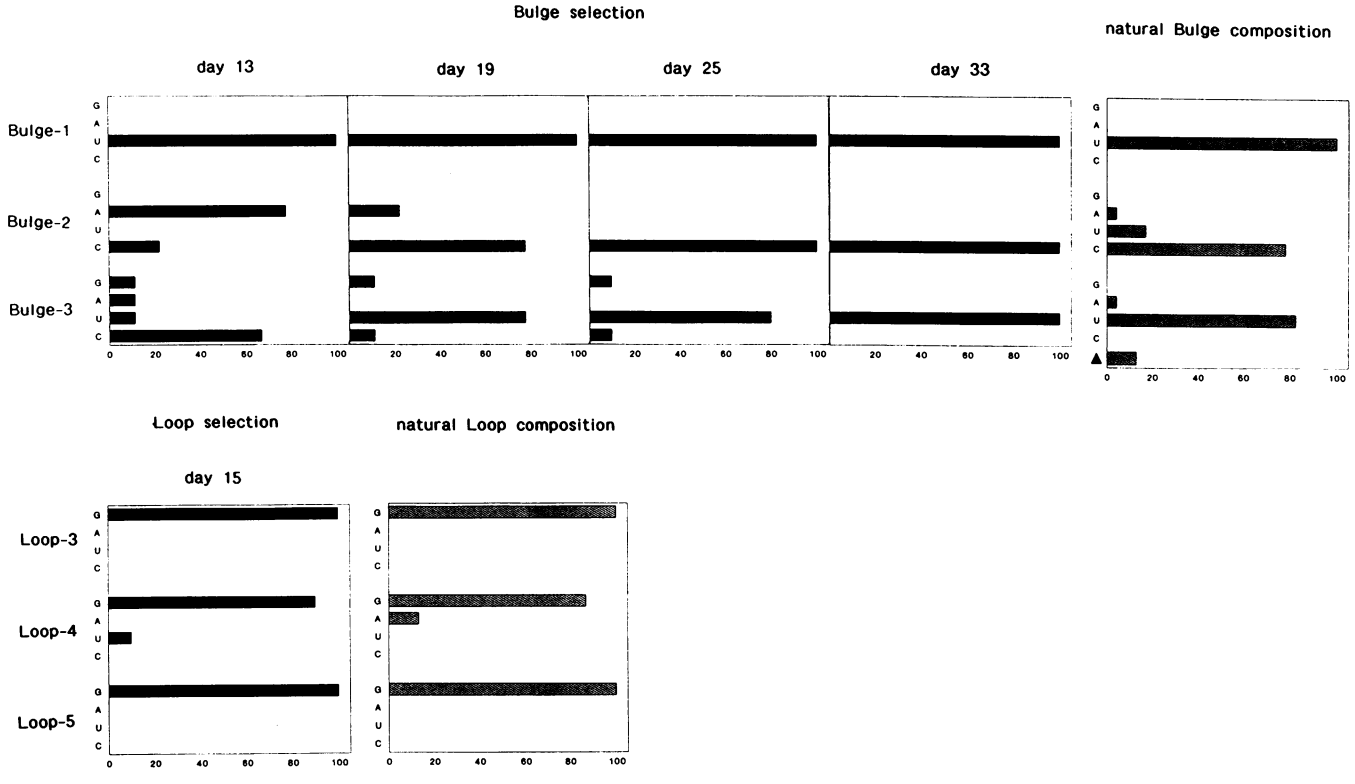

Figure 4. Schematic representation of the TAR loop-bulge sequences selected in tissue culture and a comparison with sequences from natural HIV-1 isolates. All experimental data are derived from Table 1, 23 HIV-1 isolates (26) were used to determine the natural TAR composition. Nucleotide deletions as observed in some natural TAR bulges are shown as ▲.

completely to the UCU consensus sequence. These results suggest that strong selective pressures are operating on TAR RNA in our HIV-1 replication strategy in T cell lines.

DISCUSSION

In order to understand the functional role of the TAR element in the life cycle of HIV-1, we subjected the single-stranded bulge and loop domains to random mutagenesis and selected replication-competent viruses from this library. We have shown that a very limited number of bulge and loop sequences were selected out of a pool of 64 different sequences and these sequences closely match the consensus TAR motif found in natural HIV isolates. It would be of interest to repeat this TAR-evolution experiment in cell types that have been reported to support HIV-replication in a TAR-independent manner, e.g. phorbol ester stimulated T lymphocytes (13) or astrocytic glial cells (24,25).

One limitation of viral evolution studies compared to *in vitro* evolution may be the limited number of bases that can be randomized without loss of progeny. In this study, we generated pools of 64 HIV clones and were able to outgrow one particular virus within two weeks of culture. When a larger pool is generated, the concentration of replication-competent virus may drop below a threshold for establishing a persistent infection. Our transfection protocol allows us to successfully initiate a productive infection with as little as 0.02 μg HIV DNA when mixed with 20 μg pBluescript carrier DNA (4000-fold molar excess). Thus, it is estimated that we can isolate 1 replication-competent genome from a mixture of 4000 defective HIV mutants. We therefore think it feasible to construct HIV libraries with randomized sequence at up to 6 nucleotide positions. Since regions other than TAR may be expected to display a more relaxed sequence-requirement, this will result in a relatively larger number of viable HIV constructs. Furthermore, while the HIV variants selected in this study were originally present in the DNA pool, it is likely that *in vivo* evolution, that is selection in combination with the generation of new variants by the error-prone HIV reverse transcriptase, will take place in experiments with more extensively randomized HIV genomes.

We believe HIV-driven evolution to be applicable not only to the functional analysis of HIV-specific sequences, but also to the analysis of cellular processes. The HIV-1 virus employs several proteins of cellular origin to regulate its gene expression, mRNA processing and translation, and it seems feasible to adapt the HIV-evolution system to study such cellular processes. For instance, a randomized HIV library could be constructed to study the nucleic acid requirements for binding of cellular transcription factors like TATAA-binding protein, Sp-1 or NF-κB to the HIV LTR promoter. Although HIV-evolution is a powerful new technique, we would like to add a cautionary note on its potential risk. For instance, in experiments on the HIV Env protein, one should be aware of the possibility that virions with an altered host-range or enhanced replication kinetics may be present in the randomized library.

ACKNOWLEDGEMENTS

We thank Jeroen van Wamel for oligonucleotide synthesis, and Rob Lutgerhorst and Wim van Est for artwork and photography. We are grateful to Jan van Duin, Thijs Tersmette and Jan van der Noordaa for critical reviews of the manuscript. This work was supported in part by the Netherlands Organization for Scientific Research (NWO) and the Dutch Cancer Society (KWF).

REFERENCES

1. Mills, D.R., R.L. Peterson, S. Spiegelman (1967) Proc. Natl. Acad. Sci. USA 58, 217–224.
2. Ellington, D. and J.W. Szostak (1990) Nature 346, 818–822.
3. Green, R., A.D. Ellington and J.W. Szostak (1990) Nature 347, 406–408.
4. Robertson, D.L. and G.F. Joyce (1990) Nature 344, 467–468.
5. Tuerk, C. and L. Gold (1990) Science 249, 505–510.
6. Beaudry A.A. and G.F. Joyce (1992) Science 257, 635–641.
7. Lehman N. and G.F. Joyce (1993) Nature 361, 182–185.
8. Peden, K., M. Emerman and L. Montagnier (1991) Virology 185, 661–672.
9. Dimitrov, D.S., R.I. Willey, H. Sato, L-J. Chang, R. Blumenthal and M.A. Martin (1993) J. Virol. 67, 2182–2190.
10. Cullen, B.R. (1992) Microbiol. Rev. 56, 375–394.
11. Jeang, K.-T., Y-N. Chang, B. Berkhout, M-L. Hammarskjöld and D. Rekosh (1991) AIDS 5 (suppl.) S3–S14.
12. Leonard, J., C. Parrott, A.J. Buckler-White, W. Turner, E.K. Ross, M.A. Martin and A.R. Rabson (1989) J. Virol. 63, 4919–4924.
13. Harrich, D., J.Garcia, R. Mitsuyasu and R. Gaynor (1990) EMBO J. 9, 4417–4423.
14. Berkhout, B. and K.-T. Jeang (1989) Cell 59, 273–282.
15. Weeks, K.M. and D.M. Crothers (1991) Cell 66, 577–588.
16. Berkhout, B. (1992) Nucl. Acids Res. 20, 27–31.
17. Feng, S. and E.C. Holland (1988) Nature 334, 165–167.
18. Berkhout, B. and K.-T. Jeang (1989) J. Virol. 63, 5501–5504.
19. Dingwall, C., I. Ernberg, M.J. Gait, S.M. Green, S. Heaphy, J. Karn, A.D. Lowe, M. Singh, M.A. Skinner and R. Valerio (1989) Proc. Natl. Acad. Sci USA 86, 6925–6929.
20. Churcher, M.J., C. Lamont, F. Hamy, C. Dingwall, S.M. Green, A.D. Lowe, P:J.G. Butler, M.J. Gait and J. Karn (1993) J. Mol. Biol. 230, 90–110.
21. Marciniak, R.A., M.A. Garcia-Blanco and P.A. Sharp (1990) Proc. Natl. Acad. Sci. USA 87, 3624–3628.
22. Sheline, C.T., L.M. Milocco and K.A. Jones (1991) Genes Dev. 5, 2508–2520.
23. Wu, F., J. Garcia, D. Sigman and R. Gaynor (1991) Genes Dev. 5, 2128–2140.
24. Bagasra, O., K. Khalili, T. Seshamma, J.P. Taylor and R.J. Pomerantz (1992) J. Virol. 66, 7522–7528.
25. Taylor, J.P., R. Pomerantz, O. Bagasra, M. Chowdhury, J. Rappaport, K. Khalili and S. Amini (1992) EMBO J. 11, 3395–3403.
26. Myers, G., J.A. Berzofsky, G.N. Pavlakis, B. Korber and R.F Smith (ed.) in Human Retroviruses and AIDS (1992) Los Alamos National Laboratory, Los Alamos, New Mexico, USA.