© 1997 Oxford University Press
Nucleic Acids Research, 1997, Vol. 25, No. 20 **4013–4017**

The primer binding site on the RNA genome of human and simian immunodeficiency viruses is flanked by an upstream hairpin structure

Benjamin Berkhout

Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, PO Box 22700, 1100 DE Amsterdam, The Netherlands

Received July 16, 1997; Revised and Accepted August 26, 1997

ABSTRACT

Reverse transcription of retroviral genomes is primed by a tRNA molecule that anneals to an 18 nt primer binding site (PBS) on the viral RNA genome. Additional base pair interactions between the tRNA primer and the viral RNA have been proposed. In particular, base pairing was proposed between the anticodon loop of tRNA^{Lys3} and the 'A-rich' loop of a hairpin located immediately upstream of the PBS site in HIV-1 RNA. In order to judge the importance of this sequence/structure motif, we performed an extensive phylogenetic analysis of this genomic region in a variety of simian and human immunodeficiency viruses (SIV and HIV). Both the phylogeny of natural HIV/SIV sequences and the behaviour of U5-PBS mutant/revertant viruses support the idea that this RNA structure is critical for virus replication. Although this hairpin may play a role in tRNA annealing and/or initiation of reverse transcription, the proposed base pairing interaction between the A-rich loop of the HIV-1 hairpin and the anticodon of the initiator tRNA is not directly supported by this analysis.

INTRODUCTION

Reverse transcription of retroviral genomes is primed by a tRNA annealed to an 18 nt primer binding site (PBS). Different retroviruses use different tRNA species as primer; these natural primers include tRNA^{Pro}, tRNA^{Trp}, tRNA^{Lys1,2} and tRNA^{Lys3}. Multiple interactions between the tRNA primer and both viral protein and RNA components are involved in this specific primer usage (reviewed in 1). First, there is evidence for specific binding of the tRNA primer to reverse transcriptase (RT) enzyme, an interaction that leads to selective encapsidation of the tRNA primer into virion particles. Second, annealing to the viral genome is facilitated by the 18 nt complementarity between the 3′-end of the priming tRNA and the PBS site. Third, annealing of the tRNA may be stimulated by additional base pairing contacts between other parts of the tRNA and viral sequences flanking the PBS site. Such additional contacts were originally proposed for Rous sarcoma virus (2–4), but similar interactions have been reported for human immunodeficiency virus type 1 (HIV-1) (5), HIV-2 (6) and several retrotransposon elements (7–9). Consistent with the idea that regions of the viral genome outside the PBS sequence participate in selective tRNA primer usage is the observation that retroviruses mutated in the PBS are genetically unstable, as they revert to the wild-type PBS sequence after a few passages in cell culture (10–13). On the other hand, reasonable transduction efficiencies were obtained with murine leukemia virus vectors that use an unnatural or genetically engineered tRNA primer (14,15).

A structure has been proposed for the HIV-1 RNA–tRNA^{Lys3} complex which defined several sequences in the U5 region that interact with different parts of the tRNA^{Lys3} primer (5). According to this detailed RNA structure model, base pairing occurs between the anticodon loop of tRNA^{Lys3} and an 'A-rich' loop that is part of a hairpin in the U5 region. Recent studies with mutant HIV-1 genomes have analyzed the role of this additional base pairing interaction (16–19). Specifically, Morrow *et al.* tested whether PBS variants that use an unnatural tRNA primer could be stabilized by introduction of a compensatory mutation in the U5 loop region that is anticipated to contact the anticodon of the novel tRNA primer. These studies identified viruses that stably maintained the modified PBS site after extended growth in tissue culture. A major problem with this type of analysis is that multiple parameters can influence the genetic stability or reversion capacity of a mutated retrovirus. For instance, it cannot be excluded that genotypes are relatively stable because the corresponding viruses are severely replication impaired, resulting in a low reversion capacity. Alternatively, the U5 mutation may cause an unwanted defect in another replicative function of the HIV-1 leader RNA and revertants may have restored this function. Another obstacle to interpretation of a mutational analysis is that the effect on RNA secondary structure of the U5-PBS region is largely unknown. I therefore performed an extensive phylogenetic analysis of the U5-PBS RNA structures of all types of human and simian immunodeficiency viruses (HIV and SIV).

Tel: +31 20 566 4822; Fax: +31 20 691 6531; Email: b.berkhout@amc.uva.nl

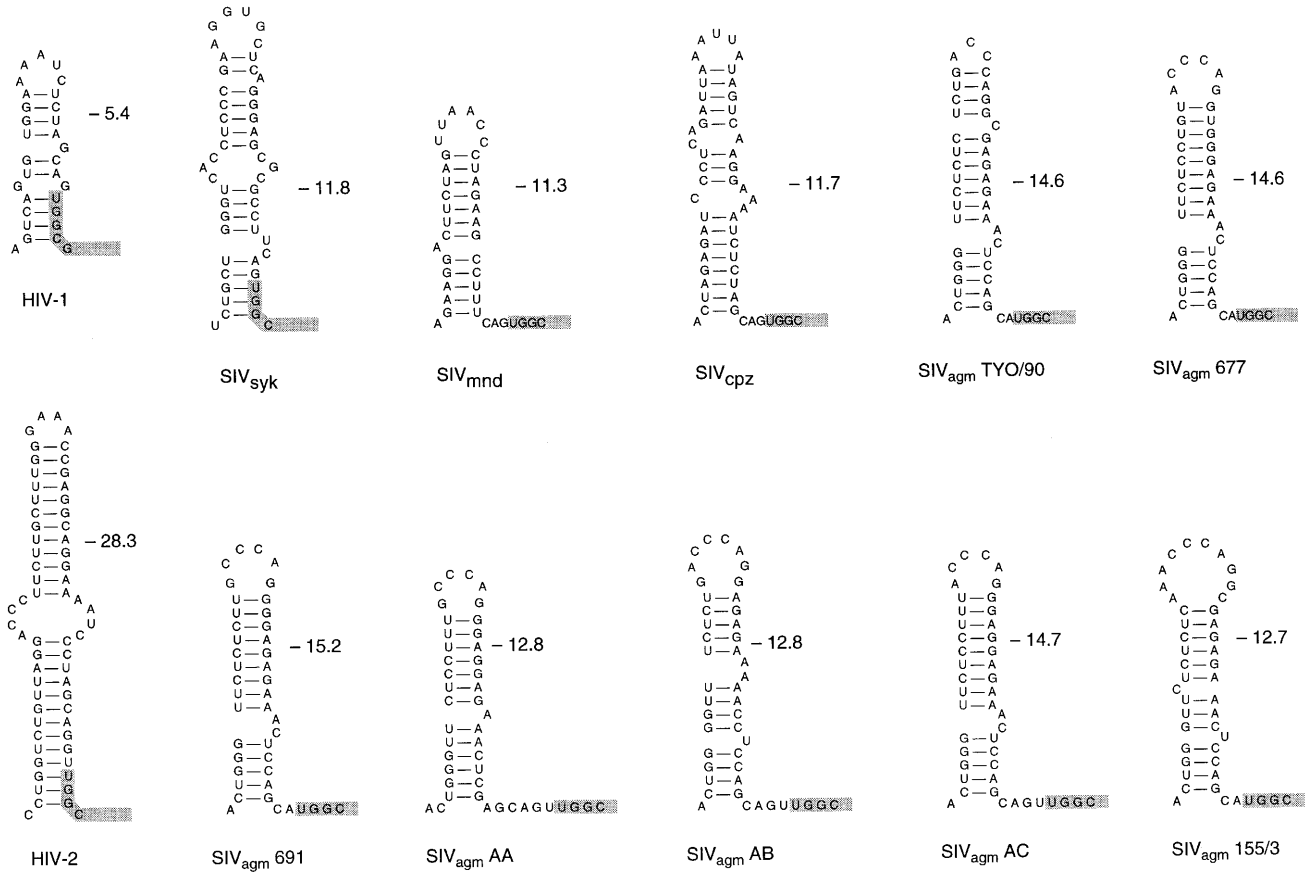

Figure 1. Phylogenetic comparison of the U5 RNA hairpin structure in different HIV/SIV isolates. The PBS sequences are marked by grey boxes. The sequences of all primate lentiviruses were derived from the AIDS database (33). The structure of the HIV-1 and HIV-2 RNA was presented previously (5,6,27). Structure prediction and free energy minimization were performed with the MFOLD program (34) in the GCG package, which uses the energy rules as defined by Freier *et al.* (35). Some additional base pairs were predicted for the RNA segments shown when a larger RNA fragment was used for RNA structure predictions. The structures shown were used for ΔG calculations (indicated next to the stem regions in kcal/mol, standard settings, including terminal stacks).

RESULTS AND DISCUSSION

Phylogenetic conservation of a U5 hairpin, but not of the A-rich loop

RNA secondary structure models for the U5-PBS region of both HIV-1 and HIV-2 have been reported previously. The HIV-1 hairpin contains the A-rich loop (Fig. 1) and is based primarily on RNA structure probing experiments (5,20). Little phylogenetic support is available for this structure because the sequence is relatively well conserved among different virus isolates (21). Biochemical RNA structure probing and phylogeny were used to model the HIV-2 RNA structure near the PBS site (6). An extended hairpin has been proposed for part of the U5 region immediately upstream of the PBS (Fig. 1) and sequence differences in 13 HIV-2 isolates yielded several co-variations that strongly support this base pairing scheme. Interestingly, this hairpin contains two single-stranded A-rich motifs in the internal and external loops that may participate in binding of the U-rich anticodon loop of $tRNA^{Lys3}$.

Phylogenetic analysis of more distantly related sequences of the SIV viruses has proven useful in identification of other structured RNA signals in the viral genome (21–23). I therefore analyzed the RNA secondary structure of all SIV species identified in a number of Old World monkey species: the mandrill (SIVmnd), Sykes monkey (SIVsyk), chimpanzee (SIVcpz) and African green monkey (SIVagm). A stem–loop structure can be folded for all these isolates in the U5 region immediately flanking the PBS^{Lys3} site (Fig. 1). The conservation of RNA structure, despite considerable divergence in nucleotide sequence, suggests a role for this RNA motif in virus replication. For HIV-1 it has been suggested that the presentation of an A-rich sequence in a single-stranded loop is required to contact the initiator tRNA. However, this phylogenetic analysis of SIV isolates does not support this particular mechanism because such an A-rich loop can be discerned only in a minority of the SIV hairpins (Fig. 1, SIVcpz and SIVagm155/3), even though all use $tRNA^{Lys3}$ as primer. Thus the proposed loop–loop interaction between the U5 region and the initiator tRNA cannot be formed for most SIV viruses. However, one cannot currently exclude the possibility that a similar interaction occurs between the tRNA primer and other regions of the SIV genome. It was recently reported that deletion of the A-rich loop in HIV-1 RNA does not interfere with annealing and extension of the natural $tRNA^{Lys3}$ primer (24). In addition, mutation of the $tRNA^{Lys3}$ anticodon has been demonstrated not to affect its capacity to prime reverse transcription on HIV-1 RNA templates (25). These combined results

imply that the A-rich loop is not as critical as previously suggested (5,20).

A striking feature of the wide variety of hairpins presented in Figure 1 is that the thermodynamic stability does not exceed a value of –15.2 kcal/mol, with the exception of the extended HIV-2 structure. Although the HIV-2 hairpin at –28.3 kcal/mol is relatively stable, the presence of a large internal loop and multiple G-U base pairs at the bottom part of the stem restrain this hairpin from becoming excessively stable. Furthermore, the lower stem segment of the extended HIV-2 structure has a stability comparable with that of the other hairpins (ΔG = –12.4 kcal/mol), which may be a more appropriate comparison. A similar tendency to modulate hairpin stability can be discerned for the different SIV sequences (Fig. 1), with ΔG values ranging from –11.3 to –15.2 kcal/mol. The tendency to limit the thermodynamic stability of structured RNA signals in the viral genome was described previously for other hairpin motifs in the HIV-1 leader region (26,27). It was suggested that excessively stable structures may interfere with viral replication, in particular the processes of mRNA translation or reverse transcription. A related issue is the predominant single-stranded character of the PBS site in most viral isolates, even though some viruses include part of the PBS sequence in the lower stem region of the U5 hairpin. Specifically, HIV-1, HIV-2 and SIVsyk have three to four PBS nucleotides involved in base pairing (Fig. 1). Occlusion of part of the PBS site seems to be counteracted by the presence of multiple weak G-U base pairs in the stem region encompassing the PBS signal. This is most clearly seen in the HIV-2 hairpin, which contains four consecutive G-U base pairs. In fact, this stem region showed some susceptibility to single-strand-specific reagents in structure probing experiments (6). These results suggest that the extent of PBS base pairing is confined in a strict manner, presumably to allow PBS–tRNA annealing during replication of the viral genome. This situation may change when mutations are introduced into the U5-PBS region (see below).

RNA structural effects of U5-PBS mutated viruses and revertants

The HIV-1 mutant His-AC was previously designed (18) and contains mutations in both the PBS site and the A-rich sequence of the U5 region to facilitate interaction with the $tRNA^{His}$ primer (Fig. 2). This mutant has a severe replication defect and upon prolonged culture a faster replicating revertant appears. I now propose that both the defect of this mutant and the phenotypic reversion can be explained by RNA structural changes in the flanking U5 hairpin. The U5 mutation is incompatible with formation of the wild-type U5 hairpin, because at least the upper two base pairs are affected by the mutation (Fig. 3, ΔG > 0 kcal/mol). Intriguingly, this mutant is expected to adopt an alternative structure that includes six PBS nucleotides (Fig. 3, hairpin His-AC*). This rearranged structure not only occludes a significant part of the PBS sequences, but is also more stable than the wild-type U5 hairpin (Fig. 3, ΔG values of –8.0 and –5.4 kcal/mol respectively). Thus the U5 mutation will lead to rearrangement of the U5-PBS structure and I propose that interference with annealing of the tRNA primer contributes to the severe replication defect of this mutant. This idea is supported by the structural changes observed in the revertant virus.

The His-AC mutant yielded a fast replicating revertant virus after 61 days culture (18). This His-AC(gac) revertant acquired

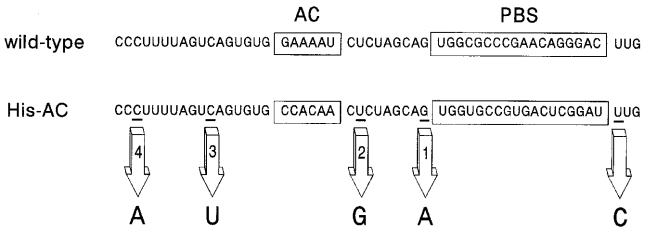

Figure 2. Schematic representation of the His-AC mutant and the appearance of reversion-based mutations. The partial nucleotide sequence of the U5 and PBS region is shown for the wild-type HIV-1 RNA genome (pHXB2), the His-AC mutant and the revertant virus. The sequence of the His-AC mutant and the revertant virus (His-AC-gac) were reported previously by Morrow *et al.* (17,18). Boxed are the PBS sites complementary to the 3′-end 18 nt of either $tRNA^{Lys3}$ (wild-type) or $tRNA^{His}$ (His-AC mutant). In addition, I have boxed the 'AC (anticodon) motif' in the U5 region that is complementary to the anticodon loop of the respective tRNA primers. The five reversion-based nucleotide substitutions are marked by underlining and arrows. The order of appearance of the four mutations within the U5 region is indicated within the arrow symbol.

five nucleotide changes in the U5-PBS region (Fig. 2). An RNA structural view of this reversion event can explain the appearance of all four mutations in the upstream U5 domain (Fig. 3). The first U5 mutation that appeared in the His-AC mutant genome is the G→A transition immediately upstream of the PBS site (17,18). This substitution has no effect on the wild-type U5 hairpin, but one obviously has to analyze the effect on the more stable, alternative hairpin (Fig. 3, His-AC*, ΔG = –8.0 kcal/mol). The first mutation seen in the revertant virus breaks the upper base pair, resulting in a significant loss of thermodynamic stability (Fig. 3, ΔG = –4.5 kcal/mol). Of the three subsequent reversion-based nucleotide changes (mutations 2, 3 and 4), only mutation 4 further decreases the stability of the His-AC* hairpin. Thus there must be another reason for the appearance of mutations 2 and 3. In fact, the sequence of the revertant 1234 is predicted to adopt a hairpin structure that is different from the conformation adopted by either the wild-type or His-AC mutant. This hairpin is relatively stable (Fig. 3, ΔG = –15.9 kcal/mol), with all four mutated U5 nucleotides involved in base pairing. The major advantage of this structure may be that only four PBS nucleotides are occluded by base pairing, which resembles the wild-type situation.

The U→C mutation immediately 3′ of the PBS^{His} site was not included in this analysis because this nucleotide substitution is not specific for HIV-1 mutants with an altered tRNA primer usage. We have found this U residue to be frequently mutated to either C or A in long-term cultures with a variety of virus mutants (26; data not shown). A likely mechanism for the frequent substitution of this nucleotide is suggested by its position in the viral genome. The position immediately 3′ of the PBS represents the first nucleotide that is incorporated after the second strand-transfer reaction during reverse transcription and there is accumulating evidence that these template switch reactions are highly mutagenic with regard to the next nucleotide incorporation, possibly because of the non-templated addition of an extra nucleotide when RT reaches the end of the template (28–31).

These results indicate that it is important to interpret the effect of mutations/reversions in the U5-PBS region in the context of RNA secondary structure. Does this RNA structural analysis also explain the stability of the PBS^{His} site in the context of this His-AC mutant? I propose that the stabilized hairpin as in His-AC* creates a new

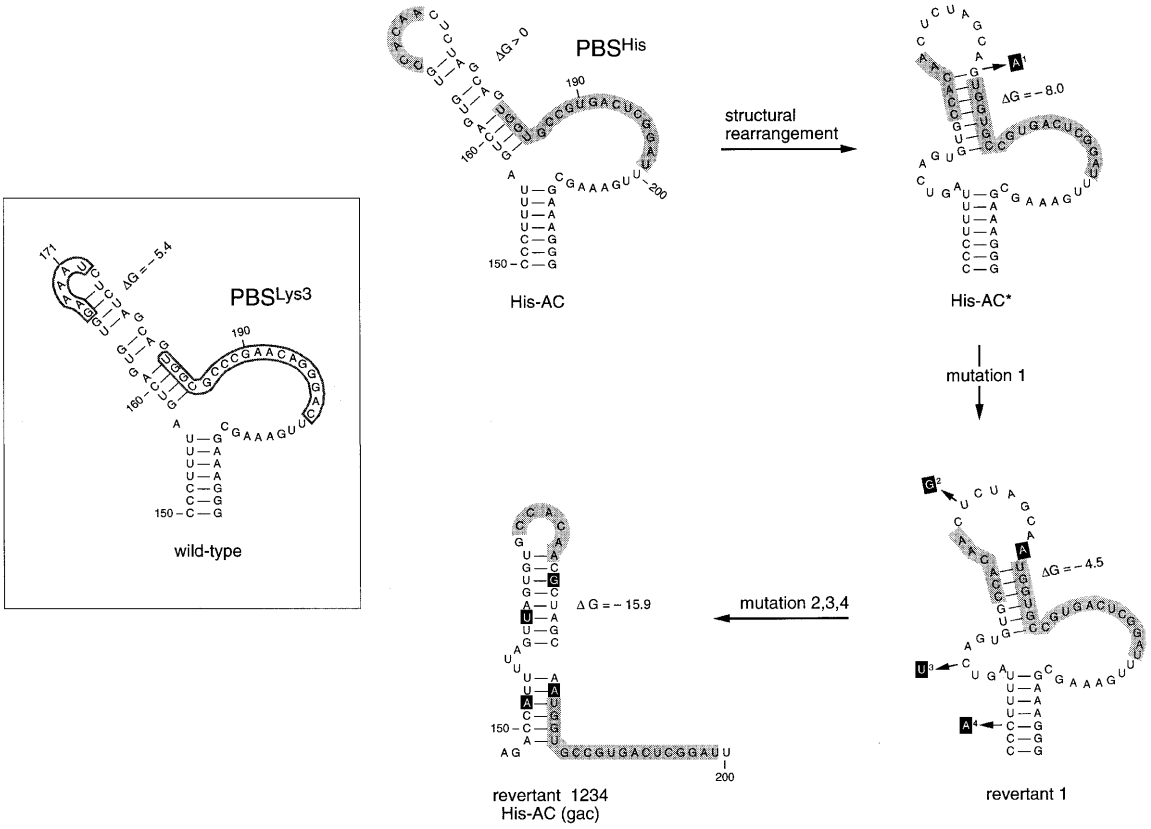

Figure 3. RNA secondary structure model of the His-AC mutant and the revertant viruses. Construction of the His-AC mutant and selection of the revertant virus has been reported previously by Morrow *et al*. (17,18). For comparison, I have included the structure model of wild-type HIV-1 RNA, with the PBS^{Lys3} site and A-rich loop motif marked by open boxes. Both sequence elements were changed in mutant His-AC (motifs marked by grey boxes). Structure prediction and free energy minimization were performed with the MFOLD program as described in Figure 1. This analysis predicts that the His-AC mutant will undergo a structural rearrangement, thereby occluding the 5′-part of the PBS site (structure His-AC*). Spontaneous mutations selected upon prolonged culture of this mutant are indicated by a black box and the number indicates their order of appearance (see also Fig. 2). Shown is the predicted RNA structure of the virus with one mutation (revertant 1) and the virus that eventually populates the culture [revertant 1234, which is termed His-AC(gac) in ref. 18].

bottleneck that limits virus replication, presumably at the level of initiation of reverse transcription. As a consequence, revertant viruses are likely to have increased their replicative capacity by selecting for an alternative RNA folding. It may therefore be of secondary importance for this mutant virus to change its PBS identity. The RNA structure analysis provides another explanation for maintenance of the PBS^{His} site. In fact, a $PBS^{His} \rightarrow PBS^{Lys3}$ reversion in the context of the inhibitory His-AC* hairpin is not expected to be beneficial because this structure will be stabilized further by exchange of a G-U for a G-C base pair (predicted stability increases from ΔG = –8.0 to –10.4 kcal/mol). Thus the structure of the His-AC* mutant not only creates a severe replication defect, but also obstructs reversion of the PBS site. It will be interesting to screen for stability of the PBS^{His} site in virus revertant 1234. This revertant has solved the RNA structure problem and reversion to the wild-type PBS^{Lys3} sequence is not expected to have any adverse structural effects in this context. Thus PBS reversion may eventually take place when this revertant is cultured for an extended period.

In conclusion, both the phylogeny of natural HIV/SIV sequences and the behaviour of U5-PBS mutant/revertant viruses support the idea that an important hairpin motif is located just upstream of the PBS site. The hairpin may play a role in tRNA annealing and/or initiation of reverse transcription, although the proposed base pairing interaction between the A-rich loop of the HIV-1 hairpin and the anticodon of the initiator tRNA is not supported by the extensive phylogenetic analysis presented in this study. Additional HIV-1 mutants should be tested to unravel the role of this structured RNA motif in virus replication. It is also important to realize that this U5 region encodes at least one other replication signal that is involved in chromosomal integration of the retroviral DNA. This step requires interaction of the viral integrase protein with sequences located at the ends of viral DNA, the so-called *att* sites in the U3 and U5 regions. The latter sequence motif has been suggested to encompass the terminal 16 nt of the U5 region, including the highly conserved CA dinucleotide (32). The presence of multiple, overlapping signals in this part of the HIV-1 RNA requires that future analyses should be performed with carefully designed mutants.

ACKNOWLEDGEMENTS

I thank several members of my laboratory for helpful discussions (Atze Das and Belinda Oude Essink) and Wim van Est for excellent artwork. Research in the Berkhout laboratory is sponsored by the Netherlands Organization for Scientific

Research (NWO), the Dutch Cancer Society (KWF), the European Community (grant 950675) and the Dutch AIDS Fund.

REFERENCES

1 Leis,J., Aiyar,A. and Cobrinik,D. (1993) In Skalka,A.M. and Goff,S.P. (eds), *Reverse Transcriptase*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 33–48.
2 Haseltine,W.A., Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 989–993.
3 Cobrinik,D., Soskey,L. and Leis,J. (1988) *J. Virol*., **62**, 3622–3630.
4 Aiyar,A., Cobrinik,D., Ge,Z., Kung,H.J. and Leis,J. (1992) *J. Virol*., **66**, 2464–2472.
5 Isel,C., Ehresmann,C., Keith,G., Ehresmann,B. and Marquet,R. (1995) *J. Mol. Biol*., **247**, 236–250.
6 Berkhout,B. and Schoneveld,I. (1993) *Nucleic Acids Res*., **21**, 1171–1178.
7 Friant,S., Heyman,T., Wilhelm,M.L. and Wilhelm,F.X. (1996) *Nucleic Acids Res*., **24**, 441–449.
8 Keeney,J.B., Chapman,K.B., Lauermann,V., Voytas,D.F., Astrom,S.U., von Pawel-Rammingen,U., Bystrom,A. and Boeke,J.D. (1995) *Mol. Cell. Biol*., **15**, 217–226.
9 Lin,J.-H. and Levin,H.L. (1997) *Genes Dev*., **11**, 270–285.
10 Wakefield,J.K., Wolf,A.G. and Morrow,C.D. (1995) *J. Virol*., **69**, 6021–6029.
11 Li,X., Mak,J., Arts,E.J., Gu,Z., Kleiman,L., Wainberg,M.A. and Parniak,M.A. (1994) *J. Virol*., **68**, 6198–6206.
12 Das,A.T., Klaver,B. and Berkhout,B. (1995) *J. Virol*., **69**, 3090–3097.
13 Whitcomb,J.M., Ortiz Conde,B.A. and Hughes,S.H. (1995) *J. Virol*., **69**, 6228–6238.
14 Lund,A.H., Duch,M., Lovmand,J., Jorgensen,P. and Pedersen,F.S. (1993) *J. Virol*., **67**, 7125–7130.
15 Lund,A.H., Duch,M., Lovmand,J., Jorgensen,P. and Pedersen,F.S. (1997) *J. Virol*., **71**, 1191–1195.
16 Kang,S.-M., Wakefield,J.K. and Morrow,C.D. (1996) *Virology*, **222**, 401–414.
17 Wakefield,J.K., Kang,S.-M. and Morrow,C.D. (1996) *J. Virol*., **70**, 966–975.
18 Zhang,Z., Kang,S.-M., LeBlanc,A., Hajduk,S.L. and Morrow,C.D. (1996) *Virology*, **226**, 306–317.
19 Kang,S.-M., Zhang,Z. and Morrow,C.D. (1997) *J. Virol*., **71**, 207–217.
20 Isel,C., Marquet,R., Keith,G., Ehresmann,C. and Ehresmann,B. (1993) *J. Biol. Chem*., **268**, 25269–25272.
21 Berkhout,B. (1992) *Nucleic Acids Res*., **20**, 27–31.
22 Berkhout,B., Klaver,B. and Das,A.T. (1995) *Virology*, **207**, 276–281.
23 Berkhout,B. and van Wamel,J.L.B. (1996) *J. Virol*., **70**, 6723–6732.
24 Li,X., Quan,Y., Arts,E.J., Li,Z., Preston,B.D., De Rocquigny,H., Roques,B.P., Darlix,J.-L., Kleiman,L., Parniak,M.A. and Wainberg,M.A. (1996) *J. Virol*., **70**, 4996–5004.
25 Huang,Y., Shalom,A., Li,Z., Wang,J., Mak,J., Wainberg,M.A. and Kleiman,L. (1996) *J. Virol*., **70**, 4700–4706.
26 Das,A.T., Klaver,B., Klasens,B.I.F., van Wamel,J.L.B. and Berkhout,B. (1997) *J. Virol*., **71**, 2346–2356.
27 Berkhout,B. (1996) *Prog. Nucleic Acid Res. Mol. Biol*., **54**, 1–34.
28 Peliska,J.A. and Benkovic,S.J. (1992) *Science*, **258**, 1112.
29 Patel,P.H. and Preston,B.D. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 549–553.
30 Wu,W., Blumberg,B.M., Fay,P.J. and Bambara,R.A. (1995) *J. Biol. Chem*., **270**, 325–332.
31 Kulpa,D., Topping,R. and Telesnitsky,A. (1997) *EMBO J*., **16**, 856–865.
32 Vicenzi,E., Dimitrov,D.S., Engelman,A., Migone,T.-S., Purcell,D.F.J., Leonard,J., Englund,G. and Martin,M.A. (1994) *J. Virol*., **68**, 7879–7890.
33 Myers,G., Korber,B., Hahn,B.H., Jeang,K.-T., Mellors,J.H., McCutchan,F.E., Henderson,L.E. and Pavlakis,G.N. (1995) *Human Retroviruses and AIDS. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
34 Zuker,M. (1989) *Science*, **244**, 48–52.
35 Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9373–9377.