Vol. 161, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 30, 1989 Page 134-142

UNUSUAL DNA STRUCTURES AT THE INTEGRATION SITE OF AN HIV PROVIRUS

Miguel A. Gama Sosa†, Jessica C. Hall, Karen E. Schneider, Gregory C. Lukaszewicz and Ruth M. Ruprecht*

Division of Cancer Pharmacology, Dana-Farber Cancer Institute, and Departments of †Pathology and *Medicine, Harvard Medical School, Boston, MA 02115

Received March 29, 1989

Supercoiled pHXBc2 DNA (containing the genome of the human immunodeficiency virus type 1 and human sequences) migrated more slowly than linear DNA in native and ethidium bromide agarose gel electrophoresis at 4.5 volts/cm, suggesting the presence of unusual DNA structures. S1 nuclease analysis of pHXBc2 revealed two S1 hypersensitive sites. Site I was located within a 25 bp direct repeat in host DNA 0.6 kB upstream from the 5' LTR. Site II was mapped 0.2 kB upstream from the vif gene start site. Sequence analysis showed that Site I sequences could assume different unusual DNA structures, whereas sequences at Site II could assume either slipped or H-DNA forms. Unusual DNA structures in host DNA may be associated with active chromatin regions and may favor proviral integration.

Unusual DNA structures may represent targets or control regions for replication, recombination, transcription, mutagenesis and repair (1). Such DNA structures include Z-DNA, H-DNA, cruciforms, Pur-Pyr structures, anisomorphic DNA, bent DNA and slipped DNA (1-13). Left handed Z-DNA and cruciform structures have been demonstrated in E. coli (1,2). With the exception of bent DNA, all the above structures are stabilized by negative supercoilicity as a result of DNA underwinding (1). In addition, the genomic DNA of eukaryotes is underwound as a result of chromatin structure (14,15). In general, unusual DNA structures are susceptible to S1 nuclease cleavage (1).

This study was undertaken to determine whether sequences in the HIV-1 genome could assume unusual DNA structures. Identification of such regions may potentially identify sequences with specialized regulatory functions. We have probed supercoiled pHXBc2 plasmid DNA, harboring a complete infectious provirus as well as neighboring sequences (16), for the presence of conformational transitions with S1 nuclease and linker insertion mutagenesis. We present evidence for the existence of unusual DNA structures within this clone.

0006-291X/89 $1.50

MATERIALS AND METHODS

Plasmid pHXBc2 (16), was grown in E. coli DH5 cells and purified by CsCl-ethidium bromide isopycnic centrifugation. Supercoiled pHXBc2 DNA was digested with S1 nuclease (400 units/μg of DNA) in 0.2 M NaCl, 1mM $ZnSO_4$, 0.5% glycerol, 50mM sodium acetate, pH 4.5, 0.2 mg/ml yeast tRNA for 7 minutes at 4°C. After a phenol-chloroform-isoamyl alcohol extraction, the DNA was purified by Elutip column chromatography (Schleicher and Schuell, NH) and ethanol precipitated. Digestions with different restriction endonucleases were performed as recommended by the manufacturer. DNA samples were analyzed by native (0.089 M Tris-borate, 0.089 M boric acid, 0.2mM EDTA), ethidium bromide (100 ng/ml) and alkaline agarose gel electrophoresis (17) at 4.5 or 0.8 volts/cm.

To map the S1 hypersensitivity within pHXBc2, S1 digested DNA was end-repaired with T4 DNA polymerase and ligated to phosphorylated ClaI linkers (d[CATCGATG]). After digestion with ClaI, pHXBc2 DNA was electrophoresed through agarose, purified by electroelution and Elutip column chromatography, self-ligated and transformed into E. coli DH5 cells. Plasmid DNA from recombinant clones harboring a ClaI site was double-digested with PstI and ClaI, and the corresponding ClaI-PstI fragments were subcloned into pBluescript KS (Stratagene, CA) and sequenced by the dideoxynucleotide method (18-20).

RESULTS

Supercoiled (RF I) pHXBc2 DNA was electrophoresed through 0.7% agarose, and its mobility was compared to that of linear (RF III) and of open circle forms (RF II). When electrophoresis was performed at 4.5 volts/cm, pHXBc2 RF III DNA migrated faster than the RF I DNA (Fig. 1A). Electrophoresis in the presence of ethidium bromide did not change this pattern (Fig. 1B), which reverted when the electrophoretic analysis was performed at 0.83 volts/cm. The identity of pHXBc2 RF III DNA was confirmed by alkaline gel electrophoresis (Fig. 1C).

Next, supercoiled pHXBc2 DNA was digested with S1 nuclease and treated with different restriction endonucleases to probe for the presence of unusual DNA structures. This resulted in distinct DNA fragments not present in single digests with the corresponding restriction endonucleases (Figs. 1A, 2, Lane 3). No specific S1 nuclease sensitivity was found in previously linearized pHXBc2 DNA (Fig. 3, Lane e), suggesting that formation of unusual, S1 sensitive DNA structures depends on DNA supercoilicity (1).

To map these S1 nuclease hypersensitive sites, ClaI linkers were introduced after digestion of supercoiled pHXBc2 with S1 nuclease. Supercoiled plasmid DNA was characterized by ClaI digestion in conjunction with other restriction endonucleases. With the exception of ClaI (which has no cleavage sites within native pHXBc2), the restriction maps of all 7 complete clones (pHSC clones) analyzed were identical to that of pHXBc2 (Figs. 2 and 4). A XhoI-ClaI double digestion of pHSC clones revealed DNA fragments of either 7.2 and 9.8 kB or 12.7 and 4.3 kB, respectively. These additional fragments were identical to those obtained when native pHXBc2 DNA was double-digested with S1 nuclease and the single cutter XhoI (Figs. 2,4). PstI-ClaI fragments from three pHSC clones were subcloned into pBluescript KS and sequenced. Figure 5 shows the sequence of the

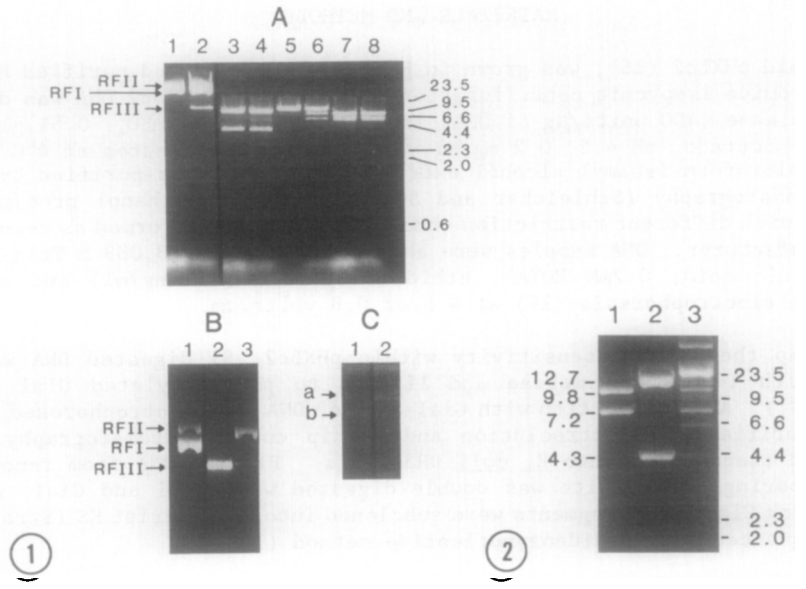

Figure 1. Agarose gel electrophoresis analysis of pHXBc2 DNA. A.Native gel. The DNA was digested with the following enzymes: 1) uncut, 2) S1 nuclease, 3) XbaI, 4) S1 nuclease followed by XbaI, 5) XhoI, 6) S1 nuclease followed by XhoI, 7) BamHI, 8) S1 nuclease followed by BamHI. B. Ethidium bromide gel. 1) uncut, 2) XhoI, 3) S1 nuclease. C. Alkaline gel. 1) uncut, 2) XhoI. a) represents linear single stranded DNA and b) represents RF V DNA, respectively.

Figure 2. ClaI linker insertion into the S1 nuclease hypersensitive regions of pHXBc2. pHSCI and pHSCII double digested with ClaI and XhoI (Lanes 1 and 2 respectively). pHXBc2 digested with S1 nuclease followed by XhoI (Lane 3).

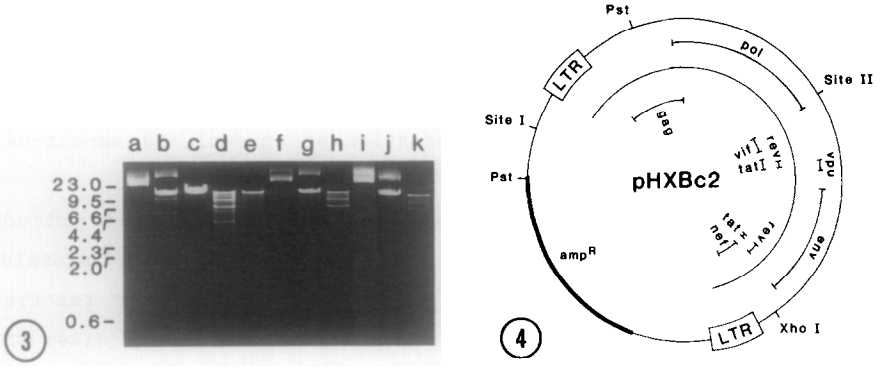

Figure 3. S1 nuclease sensitivity of pHXBc2, pHSCI and pHSCII DNAs. a) pHXBc2 uncut, b) pHXBc2 digested with S1 nuclease, c) pHXBc2 digested with XhoI, d) pHXBc2 digested with S1 nuclease followed by XhoI, e) pHXBc2 digested with Xho I followed by S1 nuclease, f) pHSCI uncut, g) pHSCI digested with S1 nuclease, h) pHSCI digested with S1 nuclease followed by XhoI, i) pHSCII uncut, j) pHSCII digested with S1 nuclease, k) pHSCII digested with S1 nuclease followed by XhoI.

Figure 4. Restriction map of pHXBc2 DNA. Site I and Site II represent the major S1 nuclease hypersensitive sites within the supercoiled plasmid.

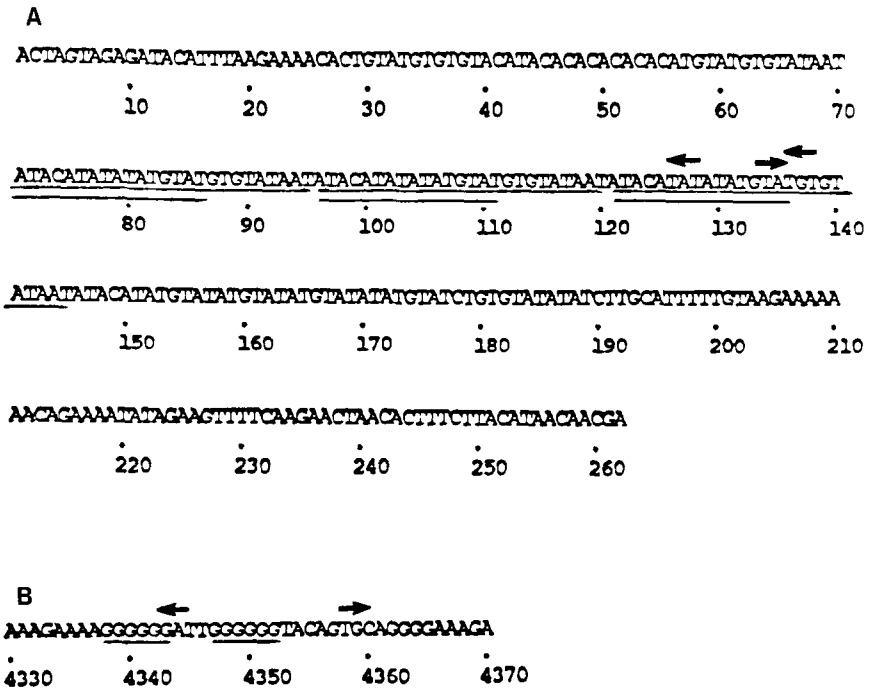

Figure 5. Sequence analysis of the S1 nuclease hypersensitive regions of pHXBc2 RF I DNA. Arrows represent the boundaries of the S1 nuclease hypersensitive regions and also indicate the direction of the nucleotide sequence analysis. A. Site I. Underlined sequences represent the perfect direct repeats and double-underlined sequences indicate inverted repeats contained within the direct repeats. B. Site II.

main S1 nuclease hypersensitive sites of pHXBc2. Site I maps about 0.6 kB upstream from the 5' LTR of HIV within a 25 bp direct repeat which could form either slipped DNA, cruciforms or both (Figs.6 and 7). Site II is found in the

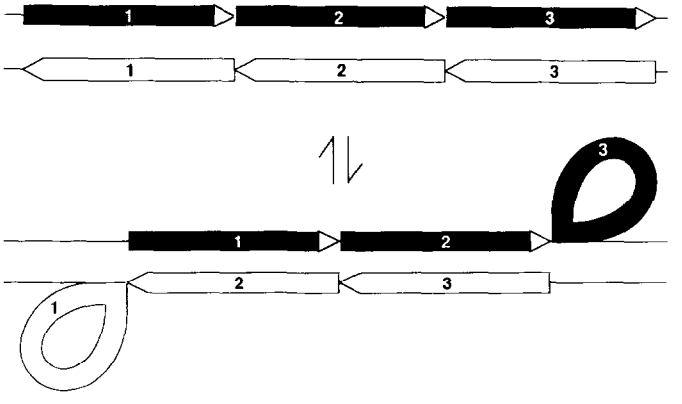

Figure 6. Slipped DNA at Site I in pHXBc2 RF I DNA. Schematic representation of the possible slippage of 2 direct repeats. Additional variations are possible (not shown).

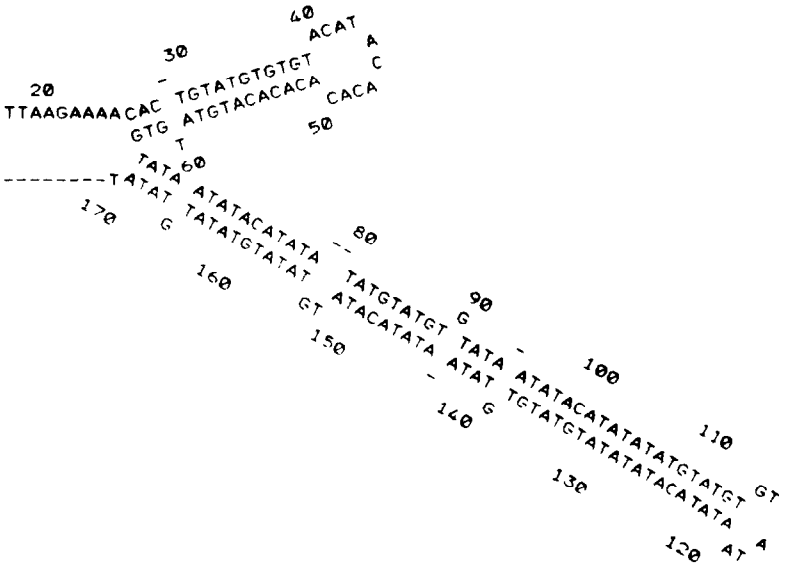

Figure 7. A computer model for unusual DNA structures at Site I in pHXBc2 RF I DNA. According to a computer program (22), the cruciform structure shown above exhibited the highest degree of thermodynamic stability. The program however, cannot calculate the free energy of unusual DNA structures in supercoiled DNA.

integrase gene, about 0.2 kB upstream from the ATG start site of vif and has the potential to form slipped DNA and/or H-DNA structures (5,11,12,21) (Fig. 8).

Supercoiled plasmid mutants carrying ClaI recognition sequences either at Site I (pHSCI) or Site II (pHSCII) showed slower mobility than their linear

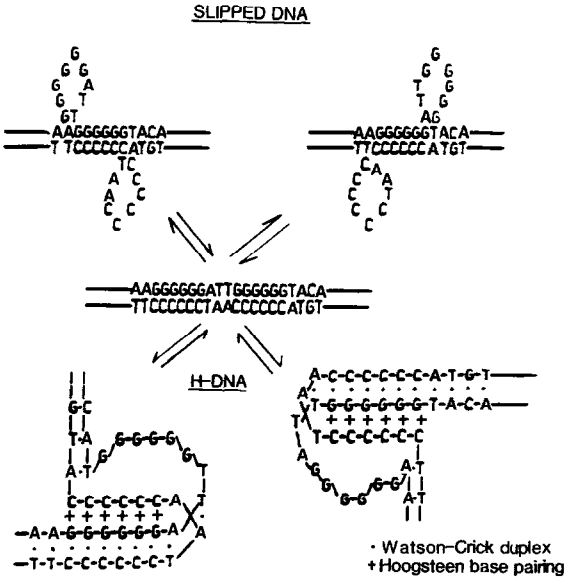

Figure 8. Unusual DNA structures at Site II in pHXBc2 RF I DNA.

counterparts in native agarose gel electrophoresis (4.5 volts/cm), as previously observed for native pHXBc2. S1 nuclease digestion of pHSCI followed by XhoI showed the same DNA fragments that were seen when native pHXBc2 DNA was treated identically. In contrast, S1 nuclease treatment followed by XhoI digestion of pHSCII revealed the loss of the 12.7 and 4.3 kB fragments seen in native pHXBc2 DNA following digestion with S1 nuclease and XhoI (Fig. 3, Lane k). Since Site II has been abolished in pHSCII, and since the supercoiled DNA form of this plasmid still migrates more slowly than its linear form in native agarose gel electrophoresis, unusual DNA structures localized at Site I seems to be primarily responsible for the unusual electrophoretic behavior of wild type pHXBc2.

DISCUSSION

We have analyzed pHXBc2 for the presence of unusual DNA structures by agarose gel electrophoresis, S1 nuclease sensitivity, linker insertion into the S1 hypersensitive sites and DNA sequence analysis. At high voltage electrophoresis, migration of the RF I DNA was retarded, suggesting that the presence of unusual DNA structures within supercoiled pHXBc2 DNA are responsible for the unusual plasmid migration. Digestion of S1 nuclease-treated pHXBc2 with different restriction endonucleases revealed the presence of two main S1 nuclease hypersensitive sites, which were not found in previously linearized pHXBc2 DNA. Site I was found about 0.6 kB upstream from the 5′ LTR of the HIV-1 provirus. By sequence analysis, the S1 nuclease sensitivity was mapped to a region containing 3 AT-rich identical direct repeat units composed mainly of alternating purines and pyrimidines. Each direct repeat contains within itself a 16 bp perfect, inverted repeat. In addition, these direct repeat units are contained within a larger AT-rich region, also rich in alternating purines and pyrimidines.

In other supercoiled DNA plasmids, S1 nuclease sensitivity has been found at direct repeats within promoter or enhancer regions, such as the heat shock promoter, chicken and mouse collagen promoters, as well as in the Moloney Murine Leukemia Virus (MoMuLV) LTR enhancer region (11,12,21). Our previous work showed that slipped DNA structures may be responsible for the S1 nuclease sensitivity of the MoMuLV enhancer (21). Slipped DNA structures within Site I of supercoiled pHXBc2 could explain the S1 nuclease sensitivity of this region. One direct repeat of one strand could base pair with a second direct repeat in the complementary strand, producing two staggered single stranded loops, susceptible to S1 cleavage (11,12) (Figs. 5 and 6). Alternatively, the nuclease sensitivity at Site I may be explained by the formation of cruciform structures within the direct repeat (Fig. 5). The formation of either slipped DNA or cruciform structures at Site I may not be mutually exclusive events since cruciform structures could arise within the staggered loops of the proposed slipped DNA structures. The fact that ClaI linker insertion into one of the direct repeats

did not result in the loss of S1 sensitivity may indicate that the formation of unusual DNA structures in the region is dependent on multiple structural elements. Computer analysis (22) of sequences at Site I predicts the formation of a complex cruciform as the most stable structure in this region (Fig. 7).

The sequence CACACACACA, a potential target for recombination events, is found upstream from site I (Fig. 5) and may be capable of adopting Z-DNA structures (32,33). Such structures may help stabilize the unusual DNA structures at Site I and may promote proviral integration in vivo (32).

Site II was mapped to sequences at position 4345-4357 within the HIV-1 proviral DNA. As seen in Figure 8, S1 nuclease sensitivity at this site may be explained by the formation of either slipped DNA or H-DNA. G residues within one repeat may slip and anneal to their complementary C residues in the second repeat generating two staggered single-stranded loops characteristic of slipped DNA. The second model implies the formation of triple-stranded DNA, or H-DNA structures, as a result of Hoogsteen base-pairing between $C.G.C^{+}$ residues (5). Formation of either slipped DNA or H-DNA is known to be stabilized by DNA supercoilicity as well as low pH (at which the S1 nuclease experiments were performed) (5,11,12). S1 nuclease digestion followed by ClaI linker insertion at Site II abolishes S1 hypersensitivity at this site but does not significantly affect the unusual migration of RF I versus RF III. This suggests that the formation of unusual DNA structures at Site I is mainly responsible for the unusual electrophoretic behavior of pHXBc2 RF I.

Unusual DNA structures at Site I may represent targets for the enzymatic machinery involved in proviral integration. Strong selectivity in the retroviral integration process has been found in other systems. Rohdewohld et al. found that MoMuLV preferentially integrates into chromatin regions located within a few hundred base pairs of DNase I hypersensitive sites (23). Shih et al. found that certain primary DNA sequences are present in provirus flanking regions at much greater frequency than expected solely by chance. In addition, they have postulated that the preferred target sites for integration may be correlated with structural aspects of the chromosomal DNA (24). Our data lends support for this hypothesis. Sequence analysis of Site I revealed AT-rich DNA sequences composed mainly of alternating purines and pyrimidines (simple DNA) similar to those found upstream of the Drosophila heat shock hsp70 gene (25). Therefore, Site I may constitute a host regulatory element which may facilitate activation of the HIV provirus under certain stimuli such as heat shock.

AT-rich sequences (ARS) have been shown to be origins of replication in vivo, which require an 11 bp consensus core sequence (TAAACATAAAA), in addition to variable AT-rich flanking sequences (26-31). DNA unwinding at yeast origins of replication is promoted by ARS (29-31). In addition, ARS sequences are cleaved by mung bean and other single-strand specific nucleases when present in

naked DNA (30, 31). Sequences at position 209-219 adjacent to Site I (Fig. 5) show a 9 bp homology to the 11 bp ARS consensus sequence (30). Therefore, one might speculate that ARS sequences around Site I may regulate local DNA replication and chromatin structure.

The significance of unusual DNA structures located within the HIV proviral genome (Site II) is unknown. They may constitute targets for effector proteins involved in the control of local chromatin structure and proviral transcription. At present, the influence of site II on HIV propagation is being investigated.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Sodroski for the pHXBc2 clone and Rita Antonelli for preparing this manuscript. Supported in part by grant #DMD-8509459 from the National Science Foundation and contract #FOD-0757 from the National Institute of Health. RMR acknowledges a Science Scholarship of the Mary Ingraham Bunting Institute of Radcliffe College and a Faculty Development Award from the Pharmaceutical Manufacturers Association Foundation.

REFERENCES

1. Wells, R.D. (1988) J. Biol. Chem. 263, 1095-1098.
2. Jaworski, A., Hsieh, W-T., Blaho, J.A., Larson, J.E., and Wells, R.D. (1987) Science 238, 773-777.
3. Rich, A., Nordheim, A., and Wang, H-J (1984) Annu. Rev. Biochem. 53, 791-846.
4. Singleton, C.K., Klysik, J., Stirdivant, S.M., and Wells, R.D. (1982) Nature 299, 312-316.
5. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A., and Kamenetskii, F. (1987) Nature 330, 495-497.
6. Sinden, R.R., and Pettijohn, D.E. (1984) J. Biol. Chem. 259, 6593-6600.
7. Lilley, D.M.J. (1980) Proc. Natl. Acad. Sci. USA 77, 6468-6472.
8. Kilpatrick, M.W., Torri, A., Kang, D.S., Engler, J.A., and Wells, R.D. (1986) J. Biol. Chem. 261, 11360-11354.
9. Wohtraub, F., McLean M.J., and Wells, R.D. (1987) J. Biol. Chem. 262, 6407-6416.
10. Wells, R.D., Larson, J.E., Grant, R.C., Shortle, B.E. and Cantor, C.R. (1970) J. Mol. Biol. 54, 405-497.
11. Mace, H.A.T., Pelham, H.R.B. and Travers, A.A. (1983) Nature 304, 555-557.
12. McKeon, C., Schmidt, A., and deCrombrugghe, B. (1984) J. Biol. Chem. 259, 6636-6640.
13. Pulleybank D.E., Haniford, D.B., and Morgan, A.R. (1985) Cell 42, 271-280.
14. Stillman, B. (1986) Cell 45, 555-565.
15. Zivanovic, Y., Goulet, I., Revet, B., LeBret, M., and Prunell, A. (1988) J. Mol. Biol. 220, 267-290.
16. Fisher, A.G., Collati, E., Ratner, L., Gallo, R.C. and Wong-Staal, F. (1985) Nature 316, 262-265.
17. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

18. Sanger, F., Miklen, S., and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
19. Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152, 232-238.
20. Chen, E.J. and Seeburg, P.H. (1985) DNA 4, 165-170.
21. Gama Sosa M.A., Hall, J.C., and Ruprecht, R.M. (1988) Biochem. Biophys. Res. Commun. 156, 417-423.
22. Zucker, M., and Stiegler, P. (1981) Nucleic Acids Res. 9, 133-148.
23. Rohdewohld H., Weiher, H., Reik, W., Jaenisch, R., and Breindl M. (1987) J. Virol. 61, 336-343.
24. Shih, C-C., Stoye, J.P., and Coffin, J.M. (1988) Cell 53, 531-537.
25. Mason, P.J., Torok, I., Kiss, I., Karch, T., and Udvardy, A., (1982) J. Mol. Biol. 156, 21-35.
26. Brenner, B.J.L. and Fangman, W.L. (1987) Cell 51, 463-471.
27. Hueberman J.A., Spotila, L.D., Nowotka, K.A., El-Assouli, S.M., and Davis, L.R. (1987) Cell 51, 473-481.
28. Koshland, D., Kent, J.C. and Hartwell, L.H. (1985) Cell 40, 393-403.
29. Broach, J.R., Li, Y-Y., Feldman, J., Jayaram, M., Nasmyth, K.A., and Hicks, J.B. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 1165-1173.
30. Umek, R.M. and Kowalski, D. (1987) Nucleic Acids Res. 15, 4467-4480.
31. Umek, R.M. and Kowalski, D. (1988) Cell 52, 559-567.
32. Hellman, L., Steen, M-L., Sundvall, M. and Pettersson, U. (1988) Gene 68, 93-99.
33. Haniford, D.B. and Pulleyblank, D.E. (1983) Nature 302, 632-634.