
Synthesis of a gene for the HIV transactivator protein TAT by a novel single stranded approach involving *in vivo* gap repair

Sally E.Adams, Ian D.Johnson, Martin Braddock¹, Alan J.Kingsman, Susan M.Kingsman¹ and R.Mark Edwards*

Department of Molecular Biology, British Bio-technology Ltd, Brook House, Watlington Road, Cowley, Oxford, OX4 5LY and ¹Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK

Received March 1, 1988; Revised and Accepted April 26, 1988

ABSTRACT

The synthesis of a gene for the HIV TAT protein is described using a novel approach that capitalises on the ability to synthesise oligonucleotides of greater than 100 bp in length. It involves the synthesis of large oligomers covering one strand of the desired gene in its entirety and the use of small complementary bridging and adapter oligonucleotides to direct the assembly and cloning of the large oligomers. After ligation to the cloning vector the partially single stranded intermediate is transformed directly into the recipient bacterial host where the plasmid is repaired. The synthetic *tat* gene has been expressed in HeLa cells and is shown to trans-activate TAR⁺ but not TAR⁻ HIV LTR-CAT constructs.

INTRODUCTION

In addition to the three principle genes *gag*, *pol* and *env* the Human immunodeficiency virus (HIV 1) possesses a number of shorter open reading frames with less well defined functions (1), see figure 1. The *tat* gene is composed of two exons and encodes the transactivator (TAT), a regulatory protein involved in the activation of the HIV long terminal repeat (LTR)(2,3). TAT is believed to exert its major effect on HIV gene expression by acting as an anti-terminator. Recent studies suggest that TAT interacts with a site in the transcribed R region of the LTR to prevent the premature termination of the RNA before it reaches the viral genes (4,5). Thus transcription beyond base 59 of the LTR derived transcript only occurs in the presence of TAT. The site at which TAT acts has been defined by deletion analysis and is known as TAR (for *trans*-acting response element)(6). It is not yet known whether TAT exerts its effect via the genomic TAR or an mRNA structural motif encoded by TAR but it is becoming clear that regulation of expression from the 5' LTR involves the interaction of a number of other transcription factors in addition to TAT (7).

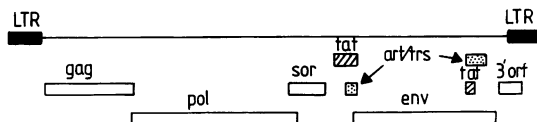


Figure 1. Map of HIV showing principal open reading frames.

We are interested in further delineating the mode of trans-activation by TAT, the use of TAT and TAR in the construction of inducible expression systems and the production of TAT protein for structural and immunological studies. We decided therefore to build a modular tat gene that would be readily adaptable for expression in a number of systems and that would be amenable to subsequent genetic manipulation including cassette mutagenesis.

Examples of total gene synthesis are becoming increasingly common as the reliability of oligonucleotide synthesis and the efficiency of assembly methods continue to improve. Gene synthesis is now an invaluable tool for the molecular biologist because of the total control it affords over restriction sites, codon usage and subsequent genetic manipulation and expression of the gene. This is particularly true where the gene is refractory to manipulation because of a lack of useful restriction sites or because it is derived from a spliced RNA, as is true in the case of the TAT protein.

Methods of oligonucleotide synthesis have been reviewed extensively elsewhere (8). A number of different methods for the assembly of oligonucleotides have been described which divide into two main groups. In the first pioneered by Khorana and co-workers (9), both strands of the desired sequence are divided such that adjacent pairs of complementary oligomers possess short (4-7 base) cohesive ends. The oligomers are then synthesised, kinased and annealed in pairs prior to ligation into a duplex corresponding to the intact gene. The ends of the gene are also endowed with cohesive ends to allow subsequent cloning of the gene in an appropriate vector. A recent development has been the successful solid phase assembly of a gene for cow colostrum trypsin inhibitor (10). The essential feature of these approaches is that both strands of the duplex are synthesised in their entirety.

The second strategy is based on the use of longer oligomers that share a complementary 3' end (11). Annealing a pair of such oligomers results in a short duplex region with two long single stranded extensions. Treating this partial duplex with Klenow fragment of DNA polymerase I in the presence of all four dNTP's results in the conversion of this structure to a complete duplex with blunt ends. This method has been successfully applied to the synthesis of a gene for Eglin C (12) and could in theory be extended to the construction of larger genes. It is attractive in that it reduces the amount of oligonucleotide synthesis required.

We now describe an adaptation of the partial synthesis approach that relies on in vivo gap repair, and its use in the synthesis of a gene for the TAT protein of HIV.

MATERIALS AND METHODS

Bacterial strains, mammalian cell lines and media.

E. coli strains HW87 (araD139 (ara-leu)del7697 (lacIPOZY)del74 galU galK hsdR rpsL srl recA56) and AKEC28 (C600, thrC, leuB6, thyA, trpC1117, hsdRK, hsdMK) were used for plasmid cloning and preparation. E. coli cultures were grown in Luria broth containing carbenicillin at 100 $\mu\text{g ml}^{-1}$ when required.

Hela cell monolayer cultures were maintained in DMEM supplemented with 10% foetal calf serum, penicillin (100 $\mu\text{g ml}^{-1}$) and streptomycin (100 $\mu\text{g ml}^{-1}$). Twenty-four hours prior to transfection, cells were trypsinized and seeded at a density of 1×10^5 cells/50mm dish.

DNA manipulation.

Molecular cloning was performed by standard protocols (13). Restriction enzymes, Klenow fragment of DNA polymerase I, T4 DNA ligase and polynucleotide kinase were purchased from BRL and New England Biolabs. BglII linkers (CAAAAGATCTTTTG) were obtained from Celltech Ltd.. The synthetic tat gene was sequenced by the dideoxy chain termination method of Sanger modified for alkali denatured plasmid DNA (14,15).

Oligonucleotide synthesis.

The oligomers were synthesised by automated solid phase phosphoramidite chemistry using cyanoethyl phosphoramidites on an Applied Biosystems 380B oligonucleotide synthesiser. All

syntheses were performed at the 0.2 μmol scale. Following de-blocking and removal from the controlled pore glass support the oligomers were purified on denaturing polyacrylamide gels as described (16). To improve the resolution of the full length product, the long oligomers were separated on a gel prepared from 10% acrylamide, 0.4% bisacrylamide that was run overnight at 500V. The bands were visualised by UV shadowing and those corresponding to the full length product cut out and eluted using standard techniques.

For kinasing, 250 pmole of oligomer was dried down and resuspended in 20 μl kinase buffer (70 mM Tris pH 7.6, 10 mM MgCl_2 , 1 mM ATP, 0.2 mM spermidine, 0.5 mM dithiothreitol). 10 u of T4 polynucleotide kinase was added and the mixture incubated at 37 $^\circ$ for 30 min. The kinase was then inactivated by heating at 85 $^\circ$ for 15 min.

Mammalian cell transfections.

HeLa cells were transfected by the calcium phosphate method (17) using 2 μg of HIVLTR-CAT plasmid (plus (pOGS210) or minus (pOGS209) the TAR region) in the presence or absence of 2 μg of hCMV-TAT plasmid (pOGS213). Plasmid pCH110 was also included in each transfection. Plasmid pCH110 contains the SV40 early promoter region driving the expression of the B-galactosidase gene and is used to correct for minor differences in transfection efficiency (18). Four hours after the addition of the precipitate, the cells were treated with 15% glycerol for 60 seconds (19), washed and incubated in complete media for 48 hours. A crude cell protein extract for the assay of chloramphenicol acetyl transferase (CAT) activity was prepared as described (20).

CAT and B-galactosidase assays.

CAT assays were performed as previously described (20). Crude cell protein extracts were incubated with ^{14}C -chloramphenicol (CM) (Amersham; 50 mCi mmol^{-1}) in 0.25 M Tris pH 7.8 and 0.5 mM acetyl-CoA (Sigma) for one hour at 37 $^\circ$. Conversion of ^{14}C -CM to the acetylated derivatives was monitored by thin layer chromatography and quantitated by liquid scintillation counting. B-galactosidase assays were performed as described (21).

Standard plasmids.

pTatenv5 was kindly provided by Dr. R. Jarrett and was used

as the source of HIV-LTR fragments. Plasmids pSV0CAT (20), pCMVCAT (22), pRSV-B-globin (23) and pUC18 (24) have been described previously. Plasmid pCH110 was purchased from Pharmacia.

Construction of a synthetic tat expression vector.

Plasmid pKV461 contains the promoter region of human cytomegalovirus (hCMV) (23), a unique BglII expression site, the splice and polyadenylation signals of SV40 and a pBR322 fragment containing the origin of replication and ampicillin resistance gene (Mark Sowden, per. comm.).

The synthetic tat gene was assembled by mixing 10 pmol of oligomers BB512 and BB513 with 20 pmol each of oligomers BB511, BB514 and BB515. BB513 had previously been kinased with T4 polynucleotide kinase by standard methods (13) to provide a 5' phosphate group. The mixture was then heated to 90°C for 10 minutes, cooled slowly to allow the oligomers to anneal, followed by ligation with HindIII-EcoRI digested pUC18. Plasmid DNA from carbenicillin-resistant colonies was subjected to restriction analysis and clones containing the expected 290bp insertion were sequenced. Plasmid pOGS212 which contains the correct tat gene sequence was then digested with NcoI and BamHI, the 270bp tat fragment end-repaired with Klenow fragment and blunt-end ligated into the filled-in BglII site of the expression vector pKV461 to generate plasmid pOGS213.

Construction of HIVLTR-CAT assay plasmids.

A 723bp XhoI-HindIII and a 509bp BglII fragment were isolated from plasmid pTatenv5. Both fragments contain the 3' LTR of HIV strain HTLV-III_B. The XhoI-HindIII fragment also contains the TAR region (-17 to +82 ;(7)) whereas the BglII fragment does not. The XhoI-HindIII fragment was filled in with Klenow fragment and inserted into the filled-in HindIII site of plasmid pSV0CAT to create plasmid pOGS210 (+TAR). The BglII fragment was inserted directly into the unique BglII site of a derivative of pSV0CAT in which the HindIII site has been converted to a BglII site. The resulting plasmid is designated pOGS209 (-TAR).

RESULTS

Gene Design.

The amino acid sequence of the TAT protein has been described (2,3). We assigned the appropriate codons to the synthetic gene

```

      M E P V D P R L E P W K H P G S
AAGCTTACCTGCCATGGAACCGATCGACCCCTAGACTGGAACCGTGGAACACCCGGGTTTC
HindIII BspM1NcoI Sali SmaI
TTCGAATGGACGGTACCTTGGTCAGCTGGGATCTGACCTTGGCACCTTTTGTGGGCCCAAG
      10      20      30      40      50      60

      Q P K T A C T N C Y C K K C C F H C Q V
CCAGCCGAAAACCTGCATGCACCAACTGTTACTGTAAAAAGTGTGCTCCACTGTCAAAT
      SphI
GGTCGGCTTTTGACGTACGTGGTTGACAATGACATTTTTCACAACGAAGGTGACAGTTCA
      70      80      90      100      110      120

      C F I T K A L G I S Y G R K K R R Q R R
TTGTTTCATACCAAGGCTTGGGTATCTCTACGGTCGTAAGAAACGTAGACAGCGCAG
AACAAAGTAGTGGTTCGGAACCCATAGAGGATGCCAGCATCTTTGTCATCTGTCCGGTC
      130      140      150      160      170      180

      R P P Q G S Q T H Q V S L S K Q P T S Q
ACGTCCACCGCAAGGTTCTCAGACTCATCAAGTTTCTTGTCCAAGCAACCGACCTCCCA
AatII
TGCAGGTGGCGTTCCAAGAGTCTGAGTAGTTCAAAGGAACAGGTTGTTGGCTGGAGGGT
      190      200      210      220      230      240

      S R G D P T G P K E * *
ATCTCGCGGTGACCCGACAGGTCTTAAGGAATAGTAAGGATCCGAATTC
      BstEII MstII BamHI EcoRI
TAGAGCGCCACTGGGCTGTCCAGGATTCCTTATCATTCTTAGGCTTAAG
      250      260      270      280

```

Figure 2. Design of Synthetic tat Gene Showing the Location of Useful Restriction Sites.

aiming for a compromise between E.coli and yeast codon bias using published tables of codon usage (24). In practice this presented difficulty in the assignment of codons for a number of amino acids where the favoured codon(s) in one organism was disfavoured in the other. For example, in the case of leucine codons E.coli strongly favours CTG whereas yeast favours TTG. For the codons where no compromise choice could be made we adopted the strategy of alternating the codon choice such that runs of suboptimal codons for one organism were avoided. In addition, a number of restriction sites were built into the sequence to facilitate the subsequent manipulation of segments of the tat gene. The codon selection was then randomised within this set of constraints by computer and the sequence checked finally to ensure that there were no regions of extensive direct or inverted repeats. To simplify the incorporation of the tat gene into expression vectors a number of flanking restriction sites were chosen including an upstream HindIII site and downstream BamHI and EcoRI sites. Provision was also made for the construction of tat

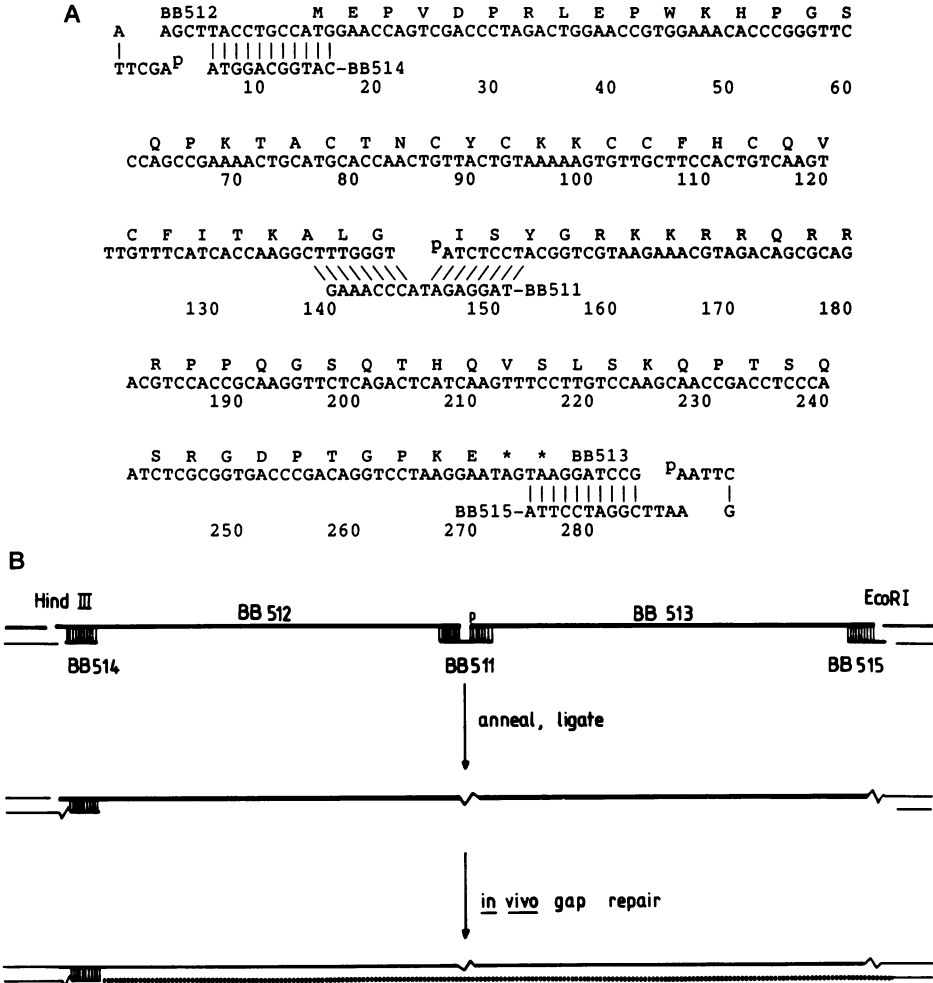


Figure 3. A) Sequence of synthetic *tat* gene showing linker and adapter oligonucleotides. Also indicated are the cohesive ends of the vector. Phosphorylated 5' ends are denoted with ^P. B) summary of assembly strategy.

fusion derivatives with or without the initiator methionine through the inclusion of *NcoI* and *BspMI* sites that encompass the initiator *ATG*. These sites allow the retention of a reasonable Kozak sequence (25) that may be important for applications involving the expression of the synthetic gene in mammalian cells. The final design of the gene is depicted in figure 2.

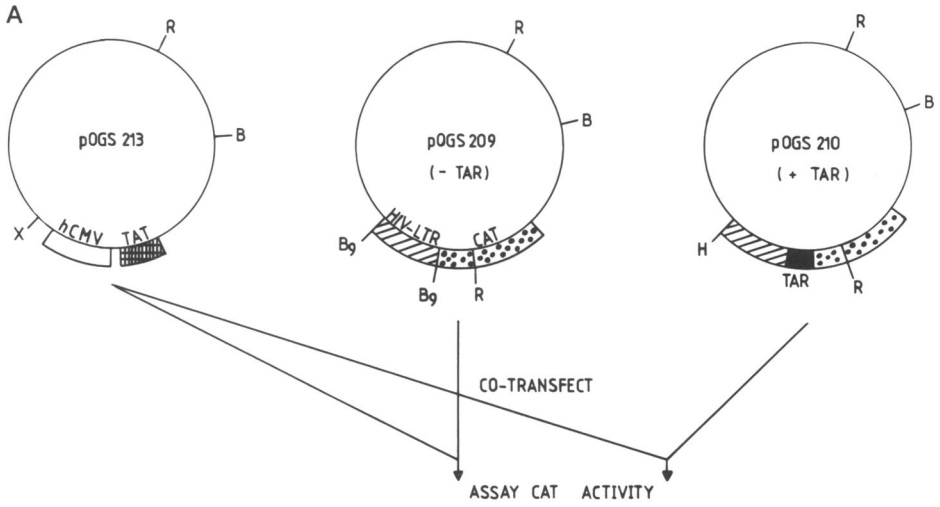
Gene Assembly.

The top (message) strand of the tat gene was synthesised as the two large oligomers BB512 (144mer) and BB513 (139mer), see figure 3a. The ligation of these two large oligomers, and their subsequent cloning into the plasmid vector pUC18 was then accomplished through the use of small complementary bridging oligomers as detailed in the materials and methods. Briefly, BB511 is a 16mer that is complementary to the 3' 8 bases of BB512 and the 5' 8 bases of BB513. BB514 is an 11 mer that serves as an adapter by annealing to the 5' end of BB512 in such a way as to leave a four base HindIII compatible cohesive end. Similarly, BB515 is a 14 mer that anneals to the 3' end of BB513 to provide an EcoRI cohesive end. The two end adapters were designed to be generic in that they can be used for the cloning of any gene synthesised by this method providing the flanking restriction sites chosen are the same.

To minimise the possibility of mis-ligation only BB513, the 3' of the two large oligomers, was kinased. All five oligomers were annealed with the three small adapter and bridging oligomers in two fold molar excess. The annealed mixture was then ligated to HindIII/EcoRI digested pUC18 DNA and transformed into HW87. This procedure allows the top (sense) strand to be ligated to the vector resulting in plasmid DNA carrying a single stranded gap covering the tat gene. We were relying, therefore, on in vivo gap repair to fill in the single-stranded region.

Of 48 clones analysed by restriction analysis, 7 released the expected HindIII-EcoRI fragment. Three clones were fully sequenced and one was found to be correct. One mutant clone carried a G→A transition at position 274, the other was found to carry a one base pair deletion at position 27 and an insertion of a G residue at position 221. We have insufficient evidence at present to comment on the mutagenicity of this approach compared to conventional gene synthesis methodologies. Expression of the synthetic tat gene in HeLa cells.

Expression of the tat gene was demonstrated in HeLa cells by a co-transfection assay in which the ability of TAT to activate expression of the CAT gene transcribed from the HIV 3' LTR is estimated. Similar approaches have been used by other groups (2-5,26-33). In this series of experiments the synthetic tat gene is expressed from the hCMV promoter in the plasmid pOGS213



B tat III activation of LTR-CAT

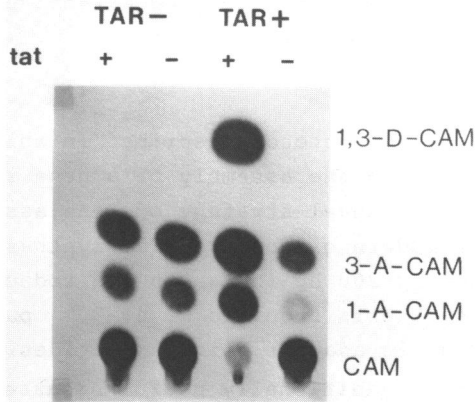


Figure 4. Analysis of synthetic *tat* gene expression using transient co-transfection assay in HeLa cells. A) Co-transfection strategy: B=BamHI, Bg=BglII, R=EcoRI, X=XhoI. See text for description of plasmids. B) Results of CAT assay: Track 1, pOGS210 alone (TAR⁺); Track 2: pOGS209 alone (TAR⁻); Track 3: pOGS210 + pOGS213 (TAR⁺ and TAT); Track 4, pOGS209 + pOGS213 (TAR⁻ and TAT). CAM = chloramphenicol, 1-A-CAM = 1-acetyl-chloramphenicol, 3-A-CAM = 3-acetyl-chloramphenicol, 1,3-D-CAM = 1,3-diacetyl-chloramphenicol.

and two different HIV-CAT constructs were employed that either contain (pOGS210) or lack (pOGS209) the TAR region (see figure 4a). The derivation of these plasmids is detailed in the materials and methods.

The results of a typical co-transfection assay are shown in figure 4b. In the absence of TAT, low but significant levels of CAT activity can be detected with both the TAR⁺ and TAR⁻ HIV-CAT constructs. It also appears that the presence of TAR causes a slight but reproducible diminution of CAT expression. The presence of TAT, however, whilst having no effect on the TAR⁻ HIV-CAT construct, results in a massive increase in expression from the TAR⁺ HIV-CAT construct. By repeating the assays with less cell extract and quantifying the conversion using liquid scintillation counting we estimate the degree of activation observed with the TAR⁺ construct pOGS210 in the presence as opposed to the absence of pOGS209 as about 300 fold.

This observation demonstrates conclusively that the synthetic tat gene is capable of expressing functional product and is an independent confirmation that it is the tat ORF that encodes the trans-activating function as opposed to another overlapping ORF since the design of the synthetic gene effectively scrambles the sense of the other two frames.

DISCUSSION.

The gene synthesis protocol described in this paper has been successfully applied to the assembly of a gene for the TAT protein of HIV. This novel strategy of gene assembly capitalises on the ability of modern oligonucleotide synthesisers to build oligomers approaching 200 bp in length and reduces the amount of oligonucleotide synthesis required by 40-50%, particularly if use can be made of generic adapter oligonucleotides. The small bridging oligomer is additionally useful because it can be employed as a sequencing primer.

The other partial synthesis approach to gene assembly that has been described is particularly prone to re-arrangements and deletions because of the in vitro polymerase step that occurs prior to ligation to the vector. For this reason it has not been found as reliable as the double stranded approach and has not been widely used. The single stranded cloning approach described in this paper avoids the problems associated with the polymerase

step because repair occurs subsequent to ligation of the insert to the vector. We have not found the approach to be significantly more mutagenic than conventional double-stranded approaches to gene synthesis. The use of large oligomers in gene assembly does however exacerbate the need for care in the gel purification step since we have noticed an increasing tendency for one base pair deletions in genes assembled from large oligomers (whichever strategy of assembly is adopted).

Current work is directed towards applying this methodology to the assembly of larger genes through the use of three or more large oligomers.

ACKNOWLEDGEMENTS

We thank Mark Sowden for providing plasmid pKV461 and Frank Caddick for excellent artwork.

*To whom correspondence should be addressed

REFERENCES

1. Wong-Staal, F and Gallo, R.C. (1985) *Nature* 317, 395-403.
2. Arya, S.K., Guo, C., Josephs, S.F. and Wong-Staal, F (1985) *Science* 229, 69-73.
3. Sodroski, J.G., Patarca, R., Rosen, C.R. and Haseltine, W.A. (1985) *Science* 229, 74-77.
4. Wright, C.M., Felber, B.K., Paskalis, H. and Pavlakis, G.N. (1986) *Science* 234, 988-992.
5. Kao, S-Y, Calman, A.F., Luciw, P.A. and Peterlin, B.M. (1987) *Nature* 330, 489-493.
6. Rosen, C.A., Sodroski, J.G. and Haseltine, W.A. (1985) *Cell* 41, 813-823.
7. Garcia, J.A., Wu, F.K., Mitsuyasu, R. and Gaynor, R.B. (1987) *EMBO Journal* 6, 3761-3770.
8. Gait, M.J. (1984) In M.J.Gait (ed), *Oligonucleotide Synthesis: a Practical Approach*, IRL Press, Oxford pp. 1-22.
9. Khorana, H.G. (1979) *Science* 203, 614-625.
10. Hostomsky, Z., Smrt, J., Arnold, L., Tocik, Z. and Paces, V. (1987) *Nucleic Acids Research* 15, 4849-4856.
11. Rossi, J.J., Kierzek, R., Huang, T., Walker, P.A. and Itakura, K. (1982) *J. Biol. Chem.* 257, 9226
12. Rink, H., Liersch, M., Sieber, P. and Meyer, F. (1984) *Nucleic Acids Research* 12, 6369-6387.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning*. Cold Spring Harbor laboratory, New York.
14. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) *P.N.A.S.* 80, 3963-3965.
15. Chen, E.J. and Seeburg P.H. (1985) *DNA* 4, 165-170.
16. *Applied Biosystems User Bulletin* (1987) 13, 18-27.
17. Graham, F.L. and Van der Eb, A.J. (1973) *Virology* 52, 456-467.
18. Harrison, S.M., Gearing, K.L., Kim, S.-Y., Kingsman, A.J. and Kingsman, S.M. (1987) *Nucleic Acids Research* 15, 10267-10284.

19. Frost, E. and Williams, J. (1978) *Virology* **91**, 39-50.
20. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Molecular and Cellular Biology* **2**, 1044-1051.
21. Gorman, C. (1985) In D.M. Glover (ed), *DNA Cloning: a Practical Approach*, IRL Press, Oxford, Volume II, pp. 143-165.
22. Vieira, J and Messing, J. (1982) *Gene* **19**, 259-268.
23. Foecking, M.K. and Hofstetter, H. (1986) *Gene* **45**, 101-105.
24. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) *Nucleic Acids Research* **9**, r43-r74.
25. Kozak, M. (1986) *Cell* **44**, 283-292.
26. Sodroski, J.G., Rosen, C.R., Wong-Staal, F., Popovic, M., Ayra, S., Gallo, R. and Haseltine, W. (1985) *Science* **227**, 171-173.
27. Cullen, B.R. (1986) *Cell* **46**, 973-982.
28. Dayton, A.I., Sodroski, J.G., Rosen, C.A., Goh, W.C. and Haseltine, W.A. (1986) *Cell* **44**, 941-947.
29. Feinberg, M.B., Jarret, R.F., Aldovini, A., Gallo, R.C. and Wong-Staal, F. (1986) *Cell* **46**, 807-817.
30. Fisher, A.G., Feinberg, M.B., Josephs, S.F. and Harper, M.E. (1986) *Nature* **320**, 367-371.
31. Gendelman, H.E., Phelps, W., Feigenbaum, L., Ostrove, J.M., Adachi, A., Howley, P.M., Khoury, G., Ginsberg, H.S. and Martin, M.A. (1986) *P.N.A.S.* **83**, 9759-9763.
32. Peterlin, B.M., Luciw, P.A., Barr, P.J. and Walker, M.D. (1986) *P.N.A.S.* **83**, 9734-9738.
33. Rosen, C.A., Sodroski, J.G., Goh, W.C., Dayton, A.I., Lippe, J. and Haseltine, W. (1986) *Nature* **319**, 555-559.