RECOMBINANT DNA RESEARCH AND VIRUSES

DEVELOPMENTS IN MOLECULAR VIROLOGY

Yechiel Becker, Series Editor
Julia Hadar, Managing Editor

Other books in the series:

Becker, Y. (ed.) *Herpesvirus DNA* (1981)
Becker, Y. (ed.) *Replication of Viral and Cellular Genomes* (1983)
Kohn A. and Fuchs, P. (eds.) *Mechanisms of Viral Pathogenesis: From Gene to Pathogen* (1984)
Becker, Y. (ed.) *Antiviral Drugs and Interferon: The Molecular Basis of their Activity* (1984)

RECOMBINANT DNA RESEARCH AND VIRUSES
Cloning and Expression of Viral Genes

edited by

Yechiel Becker
The Hebrew University of Jerusalem, Israel
Julia Hadar, Managing Editor

Martinus Nijhoff Publishing
a member of the Kluwer Academic Publishers Group
Boston/Dordrecht/Lancaster

Distributors for North America:
Kluwer Academic Publishers
190 Old Derby Street
Hingham, MA 02043

Distributors Outside North America:
Kluwer Academic Publishers Group
Distribution Centre
P.O. Box 322
3300 AH Dordrecht
The Netherlands

The figure on the cover is from Ratner, L., et al., "Cloning of Human Oncogenes". The figure appears on page 23 of this book.

Library of Congress Cataloging in Publication Data

Recombinant DNA research and viruses.

(Developments in molecular virology)
1. Viral genetics. 2. Recombinant DNA. 3. Cloning. 4. Gene expression. I. Becker, Yechiel. II. Series. [DNLM: 1. Cloning, Molecular. 2. DNA, Recombinant. 3. Genes, Viral. W1 DE998DG / QH 442 R311]
QR392.R44 1984 576'.64 84-16715

ISBN-13:978-1-4612-9611-9 e-ISBN-13:978-1-4613-2565-9
DOI:10.1007/978-1-4613-2565-9

CONTENTS

LIST OF CONTRIBUTORS

JOSE M. ALMENDRAL, Centro de Biologia Molecular (CSIC-UAM), Facultad de Ciencias, Universidad Autonoma, Canto Blanco, Madrid-34, SPAIN

SHOSHY ALTUVIA, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

DANIELE BENICHOU, Unit of Molecular Virology, Department of Virology, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, FRANCE

F. BROWN, Wellcome Biotechnology Ltd., FMD Division, Ash Road, Pirbright, Woking, Surrey GU24 ONQ, UNITED KINGDOM

A. BURNY, Department of Molecular Biology, Faculty of Agriculture, University of Brussels, rue des Chevaux 67, 1640 Rhode-Saint-Generse, Brussels, BELGIUM

AMIKAM COHEN, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

ANDREW DAVISON, MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UNITED KINGDOM

J. DESCHAMPS, Department of Molecular Biology, University of Brussels, rue des Chevaux 67, 1640 Rhode-Saint-Generse, Brussels, BELGIUM

RENATE DEURING, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

L. K. DIXON, Friedrich Miescher Institut, Post Office Box 2543, CH-4002 Basel, SWITZERLAND

WALTER DOERFLER, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

MICHEL DREANO, Unit of Molecular Virology, Department of Virology, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, FRANCE

RUDOLF EICHENLAUB, Lehrstuhl Biologie der Mikroorganismen, Ruhr-Universitat Bochum, 4630 Bochum 1, FEDERAL REPUBLIC OF GERMANY

DIRK EICK, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

LYNN W. ENQUIST, Molecular Genetics, Inc., 10320 Bren Road East, Minnetonka, Minnesota 55343, USA

REINHOLD GAHLMANN, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

ROBERT C. GALLO, Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

MARC GIRARD, Unit of Molecular Virology, Department of Virology, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, FRANCE

LUTZ GISSMANN, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 6900 Heidelberg, FEDERAL REPUBLIC OF GERMANY

LEON HIRTH, Laboratoire de Virologie, Institut de Biologie Moleculaire, et Cellulaire du CNRS, 15 rue Descartes, 67000 Strasbourg, FRANCE

T. HOHN, Friedrich Miescher Institut, Post Office Box 2543, CH-4002 Basel, SWITZERLAND

ALIK HONIGMAN, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

HANA HYMAN, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

AHUVA ITIN, Department of Virology, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

ROLF JESSBERGER, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

NURIT KATZIR, Unit of Immunology and Microbiology, University Center for Health Sciences, Ben Gurion University of the Negev, Beer Sheba, ISRAEL

ELI KESHET, Department of Virology, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

R. KETTMANN, Department of Molecular Biology, Faculty of Agriculture, University of Brussels, rue des Chevaux 67, 1640 Rhode-Saint-Generse, Brussels, BELGIUM

MICHAEL C. KIEFER, Department of Botany, University of Maryland, College Park, Maryland 20742, USA

SIMI KOBY, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

HELENA KOPECKA, Unit of Molecular Virology, Department of Virology, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, FRANCE

REINER LEISTEN, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

URSULA LICHTENBERG, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

HILLA LOCKER-GILADI, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

JAMAL MAHAJNA, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

AMOS B. OPPENHEIM, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

ROBERT A. OWENS, Plant Virology Laboratory, U.S. Department of Agriculture, Room 252, Building 011A, BARC West, Beltsville, Maryland 20705, USA

LEE RATNER, Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

FRAZER RIXON, MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UNITED KINGDOM

GALIT ROTMAN, Department of Virology, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

D. R. ROWLANDS, Wellcome Biotechnology Ltd., FMD Division, Ash Road, Pirbright, Woking, Surrey GU24 ONQ, UNITED KINGDOM

JOHN S. SALSTROM, Molecular Genetics, Inc., 10320 Bren Road East, Minnetonka, Minnesota 55343, USA

MANFRED SCHULZ, Institute of Genetics, University of Cologne, 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

ELISABETH SCHWARZ, Institut fur Virologie, Zentrum fur Hygiene, Hermann-Herder-Str. 11, 7800 Freiburg, FEDERAL REPUBLIC OF GERMANY

SILVIA STABEL, Institute of Genetics, University of Cologne 5000 Cologne 41, FEDERAL REPUBLIC OF GERMANY

ANTONIO TALAVERA, Centro de Biologia Molecular (CSIC-UAM), Facultad de Ciencias, Universidad Autonoma, Canto Blanco, Madrid-34, SPAIN

DINA TEFF, Department of Molecular Genetics, The Hebrew University, Hadassah Medical School, Jerusalem, ISRAEL

HOWARD M. TEMIN, McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706, USA

ELADIO VINUELA, Centro de Biologia Molecular (CSIC-UAM), Facultad de Ciencias, Universidad Autonoma, Canto Blanco, Madrid-34, SPAIN

ROGER J. WATSON, Molecular Genetics, Inc., 10320 Bren Road East, Minnetonka, Minnesota 55343, USA

JOHN H. WEIS, Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

SYLVIE VAN DER WERF, Unit of Molecular Virology, Department of Virology, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, FRANCE

FLOSSIE WONG-STAAL, Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

PREFACE

The development of recombinant DNA technology has made a marked impact on molecular virology. The cleavage of viral DNA genomes with restriction enzymes and the cloning of such DNA fragments in bacterial plasmids has led to the amplification of selected viral DNA fragments for sequencing and gene expression. RNA virus genomes which can be transcribed to their cDNA form were also cloned in bacterial plasmids, facilitating the study of RNA virus genes. With the elucidation in recent years of the promoter sequence of various viral genes and the expression of these genes in bacteria or yeast, the understanding of many viral gene functions has made great progress. Cloning and expression of viral genes in mammalian cells was made possible by the construction of shuttle plasmid vectors which carry the origins of DNA replication from bacteria and/or mammalian viruses. The expression of viral genes in bacteria, yeast and eukaryotic cells gives reason to hope that it will be possible to produce viral antigens in large quantities for use as human or animal vaccines.

The present volume attempts to capture for the reader some of the highlights of recombinant DNA research in the field of animal and plant viruses. The isolation and characterization of genes such as oncogenes, as well as genes coding for viral antigens, are presented, together with strategies for the transfer of viral genes to new hosts (which can be either cells or viruses). The development of approaches for the efficient expression of different viral genes in foreign hosts are described. Current studies on plant viruses and their future use in gene transfer in plants are presented along with developments in gene research of animal viruses.

I wish to thank all authors for their fine contributions and to express my appreciation especially to those who sent their manuscripts on time. My thanks to Mrs. Esther Herskovics for her excellent secretarial help.

Yechiel Becker
Jerusalem, June 1984

CLONING AND TRANSFER OF VIRAL GENES

1

DEVELOPMENTS IN MOLECULAR VIROLOGY: CLONING OF RETROVIRUS DNA IN BACTERIA AND CLONING OF OTHER DNA IN RETROVIRUSES

HOWARD M. TEMIN
McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI 53706 U.S.A.

SUMMARY

Retroviruses are natural vectors for the insertion of foreign DNA into the cell genome. With the use of recombinant DNA techniques retrovirus structure has been studied and retroviruses have been made into vectors to introduce cloned DNA into the cell genome.

INTRODUCTION

Retroviruses are a family of RNA-containing animal viruses whose replication is through a DNA intermediate that integrates into the cell genome. Because of this integration, retroviruses are natural vectors for the insertion of foreign DNA into the cell genome. Recent work using recombinant DNA technologies has shown that retroviruses are evolutionarily related to cellular movable genetic elements and that reverse transcription of cellular nucleotide sequences and integration into germ-line DNA has occurred repeatedly (1-3). Thus, studies of retroviruses are relevant to genetics in general. Moreover, some retroviruses rapidly cause some cancers in vertebrates (4). Cellular genes related to genes of these retroviruses have been implicated in non-viral mouse and human cancers (5). Thus, study of retroviruses is relevant to oncology in general.

Although hypotheses were proposed relating to these areas over a decade ago, it was not until the use of recombinant DNA technologies, as well as DNA transfection, that direct evidence was secured supporting these hypotheses. These technologies also made possible construction of retrovirus vectors to introduce cloned DNA into the cell genome.

It is necessary to know a little about the retrovirus life cycle to understand the types of cloning strategies which have been used in the study of retroviruses (4). Retrovirus RNA consists of two identical genomic molecules and associated transfer RNA molecules. (The transfer RNA is used as a primer for viral DNA synthesis.) The viral genomic RNA has a small

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

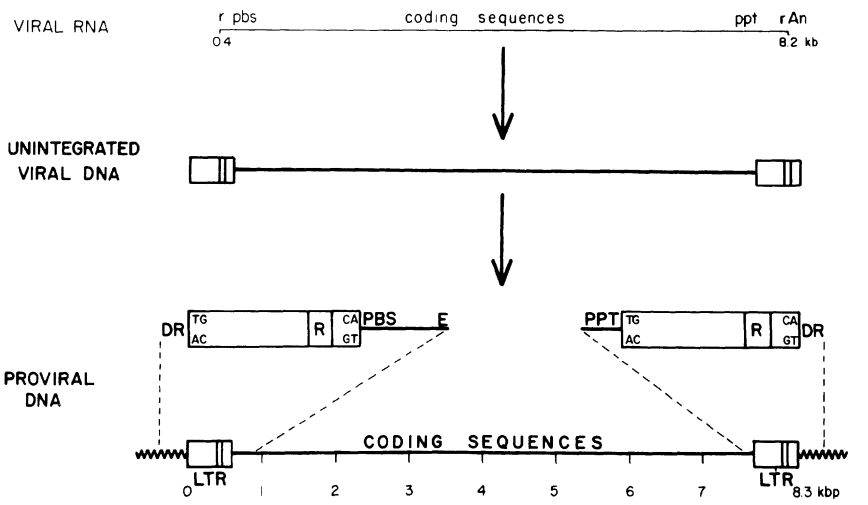


FIGURE 1. Formation of retrovirus provirus. In the virus particles, there are two molecules of viral RNA and a tRNA primer. In the infected cell there is also unintegrated circular viral DNA. r is repeat in viral RNA; pbs and PBS are primer binding site in viral RNA and DNA, respectively; ppt and PPT are polypurine track required for viral DNA synthesis in viral RNA and DNA, respectively; DR is direct repeat of cell DNA around provirus; E is encapsidation sequence; LTR is long terminal repeat; zig-zag line is cell DNA.

terminal direct repeat. After infection, this RNA is reverse transcribed by the viral reverse transcriptase to give double-stranded linear unintegrated viral DNA containing a large terminal repeat (LTR) at both ends (Figure 1). Thus, the molecular weight of unintegrated viral DNA is a little greater than two times the molecular weight of one molecule of viral genomic RNA. Closed circular DNA molecules containing one, two, or even three copies of the LTR are also found in smaller numbers.

Integration of viral DNA into the cell genome is a normal part of the viral life cycle. Viral DNA integrated into the host chromosomal DNA is co-linear with unintegrated linear viral DNA with the exception of two base pairs lost from each end. The integrated viral DNA or provirus is the template for synthesis of viral mRNAs and progeny RNA (Figure 2), although unintegrated viral DNA can also be transcribed at lower efficiency (6).

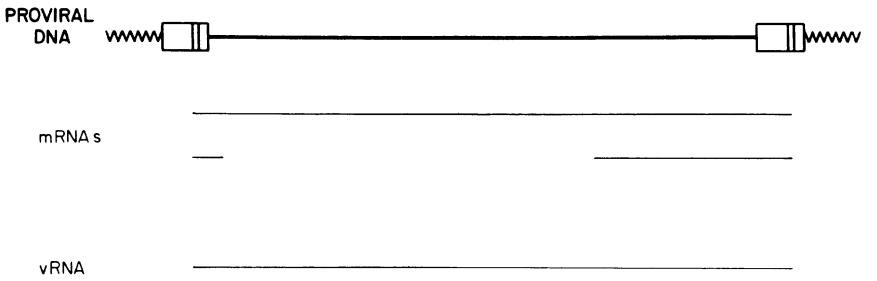


FIGURE 2. Synthesis of viral RNAs. A provirus of a non-defective retrovirus is shown. The mRNAs are for gag and gag-pol (full length) and for env (subgenomic). There may be different mRNAs for gag and gag-pol.

Cloning of Proviruses

Enzymes that do not digest viral DNA. The earliest cloning of retrovirus proviruses took advantage of the fact that the commonly used restriction endonuclease EcoRI does not cut DNA of murine leukemia virus or spleen necrosis virus. Since phage vectors capable of cloning EcoRI cut DNA fragments of 10 to 20 kbp and methods for screening for unique molecules of vertebrate cell DNA had been developed, it was relatively easy to clone provirus DNA (once legal restrictions were removed) (7,8). The integrity of the cloned DNA was validated by recovery of infectious virus after transfection of permissive vertebrate cells by the cloned DNA.

Cloned provirus DNA was recovered at frequencies expected for a small number of copies per cell. Non-infectious molecules were recovered at a frequency similar to that of infectious molecules. These molecules are probably not the result of an artefact of the cloning process, since non-infectious proviruses can be demonstrated in vertebrate cells and retroviruses are known to have a high frequency of genetic variation.

Two kinds of genetic variant do appear during the growth of the proviral clones in bacteria (Figure 3) (7,9). Both probably are the result of homologous recombination of the viral LTR. When the molecule containing the provirus is large, near the carrying capacity of the vector, the provirus frequently is deleted leaving only one LTR and surrounding cellular DNA. Alternatively, when the molecule containing the provirus is small, the provirus frequently duplicates resulting in the structure, vector DNA - cell

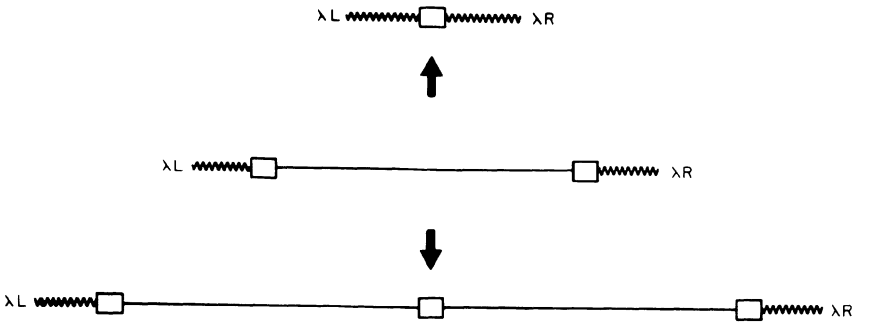


FIGURE 3. Deletion and duplication in provirus clones grown in bacteria. The original clone is in the center. The deleted clone containing cell DNA and one LTR is shown at the top. The duplicated clone containing three LTRs and two coding sequences is at the bottom.

DNA - LTR - viral coding sequences - LTR - viral coding sequences - LTR -cell DNA - vector DNA.

The latter class of molecules is a convenient source of viral molecules free of cell DNA to clone in plasmids. Digestion with a restriction endonuclease that cuts once in viral coding sequences yields a permuted molecule of viral DNA with one LTR. Such molecules are easily subcloned in plasmid vectors. Upon digestion of these subclones with the same enzyme and ligation, infectious viral DNA molecules are recovered in a concatemer.

Enzymes that digest viral DNA. Sometimes it has been necessary to use an enzyme that digests viral DNA. This requirement can be the result of the unavailability of a suitable enzyme that does not digest viral DNA or of a desire to clone partial molecules, for example to avoid "poison" sequences in mouse mammary tumor virus DNA (10,11).

We have cloned reticuloendotheliosis virus strain T (Rev-T) using EcoRI and a strategy similar to that described above even though there are two EcoRI cleavage sites in Rev-T DNA (11). After partial digestion with EcoRI, DNA molecules of 10 to 30 kbp were selected and cloned in a phage vector. Several complete proviral clones were secured, even though later restriction enzyme analysis revealed the presence of internal EcoRI cleavage sites.

Cloning Unintegrated Viral DNA

Cloning unintegrated circular DNA. As discussed above, some closed circular viral DNA is found in infected cells. This DNA can be partially purified using CsCl ethidium bromide gradients and then digested with an enzyme that cuts only once in the viral DNA and ligated to a suitable vector (12). (This digestion and ligation strategy is very similar to that used in cloning a permuted copy of viral DNA from a clone with a duplicated provirus discussed above.) This procedure results in the cloning of circularly permuted viral DNA. To recover infectious virus from such a clone, it is digested with the same single cut enzyme to release the viral DNA from the vector and ligated to form a concatemer including complete viral DNA genomes, and permissive cells are transfected. In some cases when the permuted viral DNA contains an expressible selectable marker, for example, a viral oncogene, it can be assayed directly without release from the vector (13).

To make such clones more useful, we have inserted an extra LTR with surrounding viral sequences into such a permuted clone to complete the viral DNA molecule (Figure 4) (14). Infectious virus can be recovered from this clone without formation of concatemers.

Frequently, the permuted clones recovered in this way have been aberrant. In one sample of 250 clones, we recovered only one infectious one (15,16). In some cases analyzed by others these cloned molecules had intramolecular inversions (17).

Cloning unintegrated linear DNA. Although linear viral DNA is the major species of unintegrated viral DNA, only one report has appeared of cloning such molecules (18). In this case, synthetic linkers were added to the ends of the molecule, and it was cut with the appropriate enzyme and cloned in a phage vector.

Structure of Viral DNA

The availability of cloned and infectious molecules of viral DNA has enabled detailed mapping of the viral genome by nucleotide sequencing and biological techniques. Figure 1 shows viral DNA with an enlargement of the ends. Almost all the *cis*-acting sequences of the virus are found near (within a few hundred bp) of the ends of the viral LTRs. The only exception is the splice acceptor(s) for sub-genomic mRNA(s).

Construction of Recombinant Viruses

There were many genetic experiments done with retroviruses before the advent of recombinant DNA technology. These experiments involved both mutations, esp. deletions, and virus recombination (4). However, these

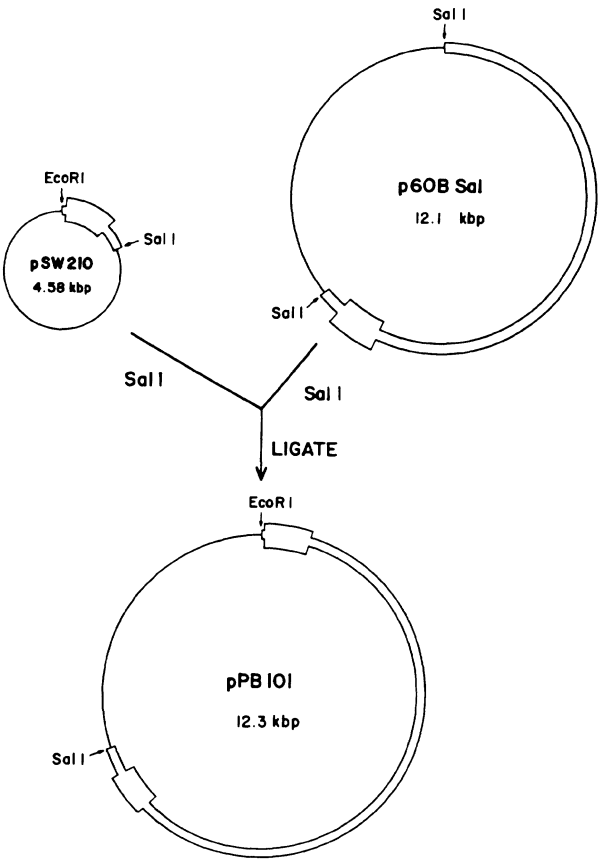


FIGURE 4. Formation of infectious plasmid clone containing entire viral DNA molecule. p60BSal was derived from a duplicated clone (see Figure 3, bottom). pSW210 was also derived from such a clone. pPB101 is as infectious as a provirus clone after transfection of sensitive chicken cells. (p60BSal is only infectious after digestion with SalI and ligation.)

experiments were somewhat limited by the lack of specific mutations and the lack of a knowledge of the mechanism of recombination.

The availability of recombinant DNA techniques resolved both of these problems. It is now possible to make specific desired mutants and recombinants.

Recombinants made with single cut enzymes. The simplest recombinants were made using restriction enzymes that cut once in cloned viral DNA. Then the two parts (5' and 3') of the parental molecules were separated, mixed with the appropriate part from the other parent, ligated, recloned, and cells were transfected to recover recombinant virus.

We used such a technique to map spontaneous mutations resulting in loss of infectivity in some provirus clones (19). We also used this technique to "repair" a deletion in a highly oncogenic retrovirus (20).

In fact, the technique can even be simplified by leaving out the ligation step. Co-transfection of permissive cells with separately cloned 5' and 3' parts of viral DNA results with high efficiency in the appearance of infectious progeny (21).

A further modification of the co-transfection technique is to transfect cells with a complete clone containing a mutation and a fragment (subclone) that overlaps the mutation -- marker rescue technique (19). In this case, two recombination events are required. However, a small (less than 100 bp) overlap is sufficient for recombination, and marker rescue is easily observed.

Another modification is to transfect two complete non-infectious proviruses both with deletions. If there is overlap between the sequences remaining (even 40 bp), recombination occurs at an efficiency of 0.1% to give infectious progeny virus (22) (see Figure 5).

Thus, the availability of cloned viral DNAs allows, after transfection, easy isolation of recombinant virus when the recombinant can be selected, for example, the recombinant is infectious and the parental viruses are not.

Recombination and deletion in vitro. Although recombination of transfected DNA after transfection is very efficient, it is only useful if the resultant recombinants can be selected. Thus, it is useful to construct recombinants in vitro and then assay them in cells.

We find it easiest to work with DNAs cloned in plasmids. We use standard recombinant DNA technology. In constructing recombinant viruses we maintain the organization as LTR - coding sequences - LTR. To recover infectious

FIGURE 5. Recombination between deleted virus clones. When co-transfection is carried out with two DNA molecules containing non-overlapping deletions (SW280 and SW279), infectious virus is recovered at high yield. Numbers are co-ordinates in kbp. Deletions are shown by absence of a line.

virus we need to preserve also PBS and E near the 5' LTR and PPT near the 3' LTR (see Figure 1).

Deletions in these regions or in the LTR may prevent recovery of infectious virus. However, such non-infectious DNA can be studied directly after transfection, that is, early expression can be studied.

Presence in recombinant viral DNA of control sequences between LTRs. Sequences for the 3' terminus of mRNA inserted between the LTRs may greatly reduce viral yield. For example, the yield of virus containing herpes simplex virus-thymidine kinase gene without its poly(A) addition sequences is much higher than the yield of virus containing those sequences (23). Presumably, little full length viral RNA is formed as a result of premature termination of transcription. The degree of inhibition differs for different terminal sequences -- different strengths of termination.

The presence of promoters in the same orientation in DNA inserted in a retrovirus vector has not been found to affect virus production, but the presence in inserted DNA of promoters in the opposite orientation to viral transcription can severely depress yield of virus (24).

Presence of intervening sequences in recombinant viral DNA. Intervening sequences in DNA inserted in SNV vectors in the same orientation as viral transcription are normally spliced out of full-length viral RNA (4,25). If the encapsidation sequences are not present in the spliced-out sequences, the resultant sub-genomic viral RNA is not packaged, and, so, does not replicate further, for example, env mRNA (26). However, if the encapsidation sequences are still present in the resultant sub-genomic viral RNA, it is packaged and then the spliced RNA becomes the predominant species of virus (Figure 6).

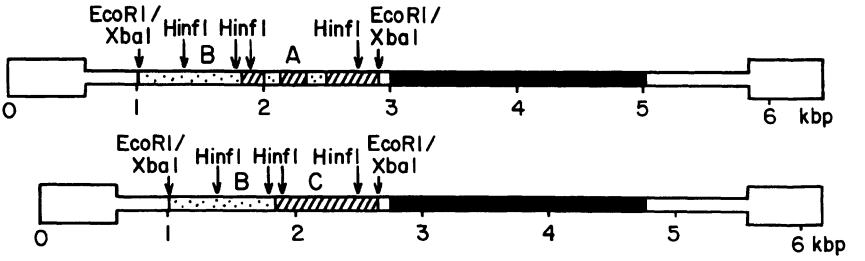


FIGURE 6. Loss of intervening sequences in genomic DNA inserted in retrovirus vector. Parental DNA is shown at the top; progeny virus DNA at the bottom. Open bases are retrovirus sequences; stippled bases are mouse α-globin sequences that are not exons; slashed bars are mouse α-globin exon sequences; solid bars are herpes simplex virus-thymidine kinase gene without the poly(A) addition sequences.

Size of recombinant viral DNA. Generally, smaller species of virus replicate faster than larger ones when molecules over 8 kbp are considered (27). Molecules as small as 2 kbp are replicated although they also appear to be selected against (18). Considering the requirements for cis-acting sequences in and near the LTRs, 2 kbp is close to a lower limit for continued viral replication. The largest constructed molecules have been 14 kbp, but spontaneous variants of over 16 kpb have been seen (28).

Since intervening sequences are spliced out of early transcripts, constructs of much larger size can be made as long as after removal of intervening sequences the size of the resultant virus is less than about 15 kbp.

Stability of recombinant virus. As mentioned above there is selection against viruses larger than 8 kbp. Furthermore, there frequently are deletions in such viruses, for example, the loss of src from avian sarcoma viruses (4,27). We have found another type of frequent deletion when two genes each containing its own promoter are inserted in a retrovirus vector -- SNV α-globinΔterR thymidine kinase ΔterR (Figure 7). Most of the virus recovered from thymidine kinase-negative cells transformed to thymidine kinase positive by this virus has a deletion of all of the α-globin coding sequences (28). One possible explanation for this result is that the

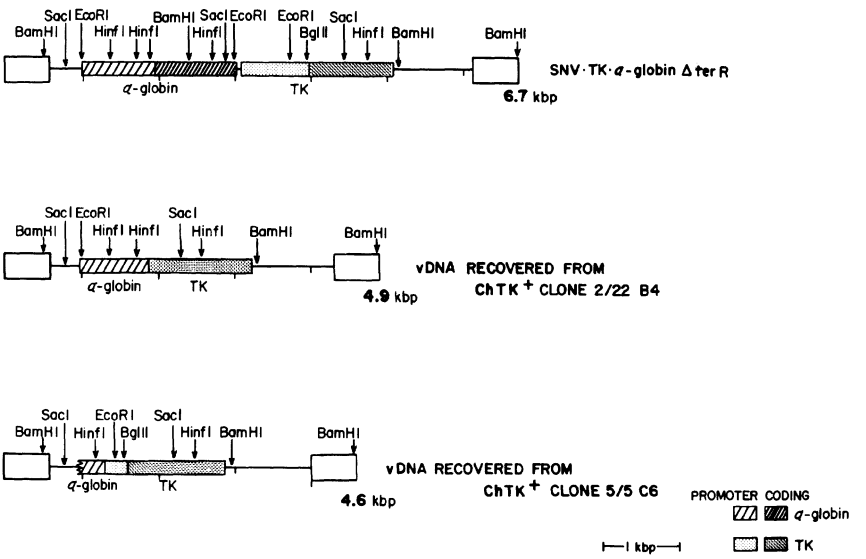


FIGURE 7. Variation in recombinant retroviruses. Parental DNA is on the top and progeny virus DNAs are in the middle and bottom. Open bars are LTRs; thin lines are retrovirus coding sequences; slashed bars are mouse α-globin sequences; dotted bars are herpes simplex virus–thymidine kinase gene sequences.

thymidine kinase promoter which is rather weak is replaced by a stronger promoter.

Expression of inserted DNA in recombinant virus. As discussed above inserted DNA can be expressed from its own promoter. Inserted genes can also be expressed from the strong promoter in the LTR (24). If the poly(A) addition sequences of the inserted sequences inhibit viral replication, the viral poly(A) addition signals in the LTR can be used. Encapsidation sequences must not be deleted in these constructions or infectious recombinant virus cannot be recovered.

Helper cell. All of the cis-acting sequences required for the formation of infectious recombinant virus are located in or near the LTRs (see Figure 1). For the propagation of defective virus, the trans-acting factors, viral proteins, are supplied by a "helper" virus. Helper cells are cells that do not produce virus but constitutively express all the proteins needed to form

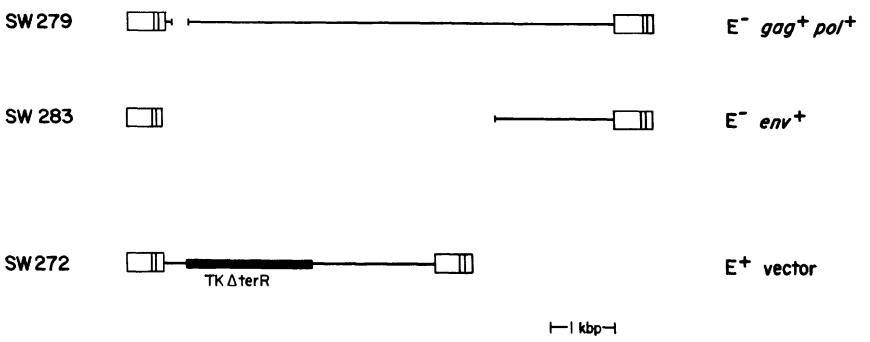


FIGURE 8. DNAs in helper cell after transfection with spleen necrosis virus vector. SW279 supplies gag and pol functions but is deleted for E (sd^- E^- gag^+ pol^+ env^-); SW283 supplies env function, but is deleted for E (sd^- E^- gag^- pol^- env^+); SW272 has all cis-functions and herpes simplex virus-thymidine kinase DNA but is deleted for gag, pol and most of env and for sequences before gag also missing in SW279 and SW283 (overlapping deletion).

infectious virus. Such a cell was constructed for our vectors by co-transfecting cells with a dominant selectable marker and viral DNA that is E^- and env^+ (SW283) and viral DNA that is E^- gag^+ pol^+ (SW279) (24) (Figure 8). Transfection of these cells with replication-defective retrovirus DNA (for example, SW272) results in the formation of infectious progeny able to infect cells but unable to form progeny virus.

CONCLUSIONS

Retroviruses can be made into excellent vectors to study a variety of problems in biology and also perhaps, to introduce active genes into cells with genetically damaged genes.

ACKNOWLEDGEMENTS

The work from my laboratory was supported by grants from the National Cancer Institute. I am an American Cancer Society Research Professor.

REFERENCES

1. Temin, H.M. Cell 21: 599-600, 1980.
2. Temin, H.M. J. Cell. Biochem. 19: 105-118, 1982.
3. Sharp, P.A. Nature 301: 471-472, 1983.
4. Weiss, R., Teich, N., Varmus, H. and Coffin, J. RNA Tumor Viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982.

5. Cooper, G.M. Science 217: 801-806, 1982.
6. Panganiban, A. and Temin, H.M. Nature 306: 155-160, 1983.
7. O'Rear, J.J., Mizutani, S., Hoffman, G., Fiandt, M. and Temin, H.M. Cell 20: 423-430, 1980.
8. Lowry, D.R., Rownds, E., Chattopadhyay, S.K., Garon, C.F. and Hager, G.L. Proc. Natl. Acad. Sci., U.S.A. 77: 614-618, 1980.
9. McClements, W.L., Enquist, L.W., Oskarsson, M., Sullivan, M. and Vande Woude, G.F. J. Virol. 35: 488-497, 1980.
10. Majors, J.E. and Varmus, H.E. Nature 289: 253-258, 1981.
11. Chen, I.S.Y., Wilhelmsen, K.C. and Temin, H.M. J. Virol. 45: 104-113, 1983.
12. Ju, G. and Skalka, A.M. Cell 22: 379-386, 1980.
13. Goldfarb, M.P. and Weinberg, R.A. J. Virol. 38: 125-135, 1981.
14. Watanabe, S. and Temin, H.M. Proc. Natl. Acad. Sci., U.S.A. 79: 5986-5990, 1982.
15. Mizutani, S. Personal communication.
16. Shimotohno, K. Personal communication.
17. Shoemaker, C., Hoffman, J., Goff, S.P. and Baltimore, D. J. Virol. 40: 164-172, 1981.
18. Scott, M.L., McKereghan, K., Kaplan, H.S. and Fry, K.E. Proc. Natl. Acad. Sci., U.S.A. 78: 4213-4217, 1981.
19. O'Rear, J.J. and Temin, H.M. J. Virol. 39: 138-149, 1981.
20. Chen, I.S.Y. and Temin, H.M. Cell 31: 111-120, 1982.
21. Miller, C.K. and Temin, H.M. Science 220: 606-609, 1983.
22. Bandyopadhyay, P., Watanabe, S. and Temin, H.M. Proc. Natl. Acad. Sci., U.S.A. 81: 3476-3480, 1984.
23. Shimotohno, K. and Temin, H.M. Cell 26: 67-77, 1981.
24. Bandyopadhyay, P. and Temin, H.M. Mol. Cell. Biol. 4: 749-754, 1984.
25. Shimotohno, K. and Temin, H.M. Nature 299: 265-268, 1982.
26. Watanabe, S. and Temin, H.M. Proc. Natl. Acad. Sci., U.S.A. 79: 5986-5990, 1982.
27. Estis, L.F. and Temin, H.M. J. Virol. 31: 389-397, 1979.
28. Emerman, M. and Temin, H.M. J. Virol. 50: 42-49, 1984.
29. Watanabe, S. and Temin, H.M. Mol. Cell. Biol. 3: 2241-2249, 1983.

2

CLONING OF HUMAN ONCOGENES

LEE RATNER, ROBERT C. GALLO, AND FLOSSIE WONG-STAAL
Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20205, U.S.A.

SUMMARY

Human oncogenes homologous to v-fes, v-myb, v-myc, and v-sis have been cloned in recombinant phage vectors. Their structures have been elucidated by restriction enzyme and heteroduplex analyses, and in the case of c-sis by nucleotide sequencing as well. The complexity of these human genomic sequences is greater than that of the retroviral oncogenes, including multiple intervening sequences. In the cases of c-myb, c-myc, and c-sis, additional exons have been found by hybridization to mRNA or cDNA samples, compared to those characterized by hybridization to the retroviral oncogenes. In addition, a cDNA-containing plasmid with sequences of human c-sis mRNA has been isolated which can transform 3T3 fibroblasts. By somatic cell hybrid analysis and in situ hybridization, the chromosomal location of each of these oncogene homologues was determined, and compared to the chromosomal breakpoints in translocations in specific malignancies.

INTRODUCTION

In order to gain a greater understanding of the molecular basis of normal development and the aberrancies associated with malignancy, we have analyzed the structure and function of cellular homologues of retroviral oncogenes (c-onc genes). Recombinant DNA technology was used to generate cloned sequences of genomic DNA and complementary DNA (cDNA) of cellular oncogenes from normal and transformed cells. This has provided the tools for investigating the qualitative and quantitative alterations in these genes and their expression in human cancers. It has also permitted the development of

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

model systems using a DNA transfection assay for studying the biochemical pathways associated with the initiation and maintenance of the transformed state. We will present here some studies from our laboratory on four c-onc genes, namely, the homologues of the transforming genes associated with feline sarcoma virus (fes), avian myeloblastosis virus (myb), avian myelocytomatosis virus, MC29 (myc) and simian sarcoma virus (sis).

MATERIALS AND METHODS

A Charon 4A library of EcoRI partial digest of human adult liver DNA and a Charon 28 library of human placenta DNA (partial Mbo I digest) were gifts of T. Maniatis (1), and screened with ^{32}P-labeled probe containing v-fes, v-myb, v-myc, or v-sis sequences, using standard recombinant DNA technology (2).

A cDNA clone of c-sis was obtained from a cDNA library of HUT102 cell mRNA constructed as described previously (3).

RESULTS AND DISCUSSION

c-fes

Feline sarcoma virus (FeSV) has been isolated from fibrosarcomas of domestic cats (4-6). Three isolates which have been characterized include the Snyder-Theilden (ST) strain, the Gardner-Arnstein (GA) strain, and the McDonough-Sarma (SM) strain. The first two strains have acquired the same oncogene, fes, from cats (7) which is homologous to the chicken oncogene acquired by the Fujinami and PRCII avian sarcoma viruses (8).

Using cloned ST FeSV DNA as a probe, we found a unique locus for c-fes in both the human and chicken genomes (9). A human genomic DNA library was screened and clones homologous to c-fes, constituting more than 20 kilobases (kb) of DNA sequences were isolated (9, 10). Though v-fes is 1.4 kb in size (11), c-fes spans 3.4 kb and contains at least three intervening sequences (Fig. 1). The finding of intervening sequences between the v-onc-related sequences of c-fes is distinct from that described for c-mos for which complete and uninterrupted homology with v-mos was found by restriction enzyme analysis and heteroduplex

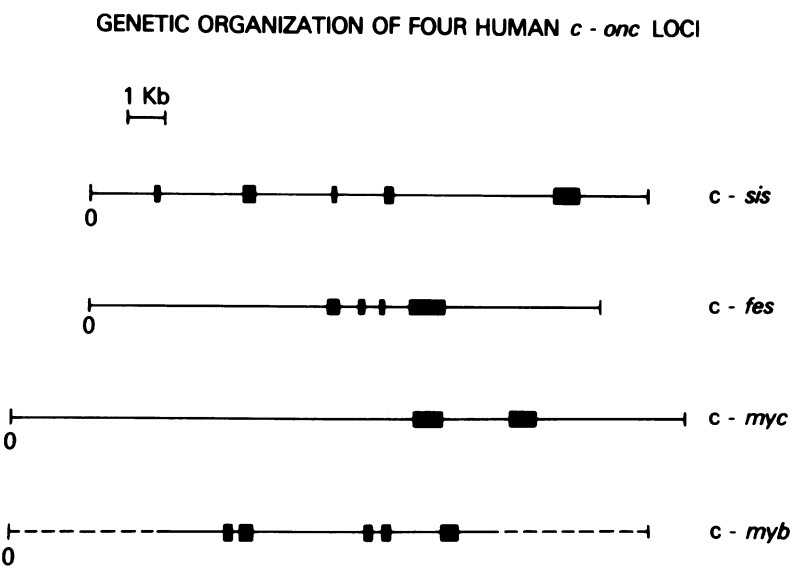


FIGURE 1. Organization of the c-fes, c-myb, c-myc, and c-sis oncogenes. The relative sizes and positions of v-onc homologous sequences and intervening sequences are shown. These maps are derived by comparison of genomic phage clones with c-onc sequences by restriction enzyme analysis and heteroduplex mapping.

mapping (12). However, a similar structure with multiple intervening sequences has been described by others for c-abl (13), c-src (14), c-rel (15), c-Ha-ras and c-Ki-ras (16-18), and here for c-myb, c-myc, and c-sis. This would suggest that the acquisition of most v-onc sequences from cellular DNA was a result of recombination with a reverse transcript of a spliced c-onc mRNA.

By somatic cell hybrid analysis, c-fes was assigned to human chromosome 15 (19, 20) and sublocalized by in situ hybridization to the 15q25-26 region (21). In a high percentage of acute promyelocytic leukemia cells, a 15:17 reciprocal translocation has been described (22); however, the breakpoint in the distal part of the long arm of chromosome 15 must be more accurately defined.

c-myb

Avian myeloblastosis virus (AMV) causes myeloid leukemias in chickens (23). The transforming sequence, v-myb is 1.1 kb in length (24). To characterize the homologous sequence in normal human genomic DNA, clones constructed using a phage vector with overlapping sequences homologous to v-myb were isolated encompassing 35 kb (25). Restriction enzyme digests and Southern blot hybridization and heteroduplex mapping demonstrate the v-myb homologous region to span 6.2 kb with four intervening sequences (Fig. 1). DNA probes generated from these clones which span 30 kb and include 5' flanking sequences, myb-related sequences, an intervening fragment lying between two myb-containing fragments and 3' flanking sequences all hybridized to a 4.5 kb mRNA from MOLT 4 acute lymphoblastic leukemia (ALL) cells, the same as detected with a v-myb probe (25). The c-myb mRNA is therefore encoded by a genomic DNA spanning 30 kb with transcriptional initiator and terminator sequences outside the 6.2 kb region of v-myb homology. The presence of a fragment between v-myb-related sequences with homology to human c-myb RNA suggests either the loss of at least one exon in the acquisition of myb-related sequences by AMV or a difference in splicing of c-myb transcripts in humans compared to chickens.

By somatic cell hybrid analysis (19) and more refined mapping by in situ hybridization (21), c-myb was mapped to human chromosome region 6q22-24. It is of interest that a high percentage of acute lymphoblastic leukemia (ALL) cells exhibit a deletion in 6q21-qter (26); high levels of c-myb transcripts are found in primary ALL lymphoblast cells and cell lines, as well (27). A 6;14 translocation has also been described in papillary serous adenocarcinoma of the ovary, and the breakpoint has been assigned to band 6q21 (28, 29). Deletions and translocations of the distal half of 6q have also been described in malignant lymphomas and melanomas (30). The role of the chromosomal alterations in activation of c-myb gene activity require further study.

c-myc

Avian myelocytomatosis virus strain MC29 is a replication-defective retrovirus capable of inducing myelocytomatosis, sarcomas, and hepatic and renal carcinomas in chickens (23). This virus has acquired an oncogene, v-myc, which encodes a nuclear protein (31). Recombinant phage clones containing normal human genomic DNA were screened with a v-myc probe; two positive clones with greater than 17 kb of human DNA were isolated (32). Restriction enzyme analysis and heteroduplex mapping revealed two exons homologous to v-myc separated by a 1.8 kb intron (Fig. 1). However, more recent experiments have demonstrated the presence of a third exon homologous to myc mRNA (33, 34) but lacking v-myc-related sequences. Nucleotide sequence analysis has shown that this exon consists of untranslated sequences (33, 35). Additional sequences were detected and isolated containing human c-myc sequences homologous to the central portion of the v-myc gene (32). These sequences lack an intervening sequence and are more divergent from the viral sequences than the complete gene. One such sequence has been found to be amplified in human neuroblastomas and referred to as N-myc (36).

In collaboration with the laboratory of C. Croce, we found by analysis of somatic cell hybrids between mouse cells and normal human lymphocytes, that c-myc is located on chromosome 8 (37). Analysis of somatic cell hybrids between mouse cells and Burkitt's lymphoma cells containing a reciprocal translocation between chromosomes 8 and 14, sublocalized the c-myc gene to the region q24-qter of chromosome 8 which was involved in this translocation.

Despite considerable further investigation on this problem, the relationship of the c-myc gene structure and function to chromosomal translocation still remains unclear. In some cases of 8;14 translocation, though the c-myc gene has translocated to the distal part of chromosome 14, no DNA rearrangement is detectable by Southern blot analysis (38). This suggests that in some cases the breakpoint may be quite distant, i.e., more than 20 kb from the c-myc gene. In the case of some of the

less common 2;8 and 8;22 translocations in Burkitt's lymphoma cells, though a part of the light chain immunoglobulin locus has translocated to chromosome 8, c-myc is not translocated (39-41). In cases in which c-myc recombination with the heavy chain immunoglobulin locus have been characterized, the breakpoint lies outside the protein coding sequence, upstream of exon 2 (42, 43). In one of these cases, several point mutations were found in exon 2 (43), but in the others which have been sequenced the predicted protein product was found to be unchanged from that inferred from the normal c-myc sequence (44). Therefore, a qualitative change in the myc product does not seem to be necessary in Burkitt's lymphoma. There also remains a controversy as to whether or not the level of c-myc transcription is altered in association with the chromosomal translocation (45-48). Two pieces of evidence suggest an alternative pattern of expression in this situation. First, it was found that most of the c-myc RNA transcribed in Burkitt's lymphoma cells utilizes a different promoter in or preceding exon 1 or within the downstream intron depending upon the chromosomal breakpoint (44), compared to that utilized in phytohemagglutin-stimulated normal lymphocytes. Second, transcription occurs almost exclusively from the rearranged c-myc allele and not from the unrearranged allele (48).

Though the details involved in relating the chromosomal translocations to oncogene function and initiation or maintenance of the transformed state still require elucidation, the univeral finding of a specific translocation in Burkitt's lymphoma in man and the analogous translocation (or rarely interstitial deletion) in the mouse plasmacytoma (49-53) point to the fundamental importance of this finding in these lymphoid malignancies. The role of translocations in other human malignancies, which are not universal findings in any particular tumor type, and the involvement of oncogenes are even less well defined at this point.

Simian sarcoma virus genome

Simian sarcoma virus (SSV) has been isolated from a fibrosarcoma in a pet woolly monkey (54). SSV and the simian

sarcoma-associated virus (SSAV) genomes have been cloned in phage vectors from either: a) closed circular viral DNA intermediates after cleavage with a one-cut restriction enzyme (55) or b) genomic DNA of a SSV-transformed nonproducer cell after cleavage in the flanking sequences (56). The resultant SSAV clone was compared to two SSV clones by restriction enzyme and heteroduplex mapping. In the permuted SSAV genome the two long terminal repeat (LTR) sequences are in the middle of the 9.0 kb genome. In comparison to SSAV, the SSV clones include a 0.2 kb deletion in the gag gene, a 1.9 kb deletion in the pol gene, a 1.5 kb deletion in the env gene, and a 1.0 kb substitution of SSV-specific sequences (v-sis) in the latter site (55, 57).

A single-stranded DNA probe was constructed in M13 phage which contains v-sis sequences (58). By liquid hybridization analysis, it was shown to be more closely homologous to primate than non-primate DNA, and to DNA of New World monkeys than to that of Old World monkeys. SSAV on the other hand is highly homologous to various gibbon ape leukemia (GaLV) isolates, derived from Old World apes. SSV was isolated from a fibrosarcoma in a pet woolly monkey, which had cohabitated with a gibbon ape. This suggests that SSV arose by transmission of a GaLV from the gibbon ape to the woolly monkey, and subsequent recombination of this retrovirus with woolly monkey DNA to give rise to the acutely transforming defective retrovirus.

V-sis represents a distinct oncogene, 1006 bp in length (59). The amino acid sequence inferred from the nucleotide sequence predicts a 28 kilodalton (Kd) protein which is processed to 20 kd and 24 kd proteins (60). Amino acids derived from the env gene are found at the amino-terminus (59).

Recent comparison of the amino acid sequence determined for human platelet-derived growth factor (PDGF) revealed it to be highly homologous to that predicted for v-sis (61, 62). Two species of PDGF have been identified, PDGF I (32 kd) and PDGF II (28 kd), which differ only in their carbohydrate content (63, 64). Upon reduction, a series of peptides ranging from 11,000-17,500 daltons is generated (65). These are thought to represent

proteolytic fragments derived from two peptide chains, A and B. The size of these fragments is dependent on the age of the platelets from which PDGF is purified. The amino acid sequence of peptide 2a of Antoniades and Hunkapiller (66) is identical in 23 of 28 amino acids to that predicted by the v-sis nucleotide sequence, and peptide I of Waterfield and coworkers (62) in 29 of 31 amino acids. The differences between the predicted woolly monkey v-sis amino acid sequence and that of human PDGF chain A could represent either species-specific differences, or alterations determining transformation potential. We therefore sought to characterize the c-sis sequence from normal and malignant human cells to answer this question.

c-sis proto-oncogene

A clone of c-sis was isolated from a recombinant phage library (67); the location of exon/intron boundaries was determined by: a) restriction enzyme digests and hybridization to the v-sis probe and to probes from a c-sis cDNA clone (see subsequent section), b) heteroduplex mapping with v-sis, and c) comparison of nucleotide sequences of v-sis (59), parts of genomic c-sis DNA (68), and c-sis cDNA (Figs. 1 and 2). This revealed that the c-sis gene contains 6 exons over 12.5 kb with one or more additional upstream exons homologous to the 5' portion of c-sis mRNA (69).

The nucleotide sequence of the 6 regions of v-sis homology with the normal human c-sis gene was determined (68). No candidate ATG initiator codon or promoter sequence analogous to TATAAA and no conserved CCAT sequence could be found in the 247 nucleotides upstream of the first region of v-sis homology. Each v-sis homologous region, except region 6, is bounded by acceptor and donor splice sites and resembles exons. Region 6 does not have a 3' donor splice site and terminates -5 base pairs (bp) from the 3' v-sis-helper viral junction. Comparison of this region to the partial sequence data of c-sis cDNA shows that region 6 represents part of an exon with at least an additional 900 bp. The open reading frames for the v-sis and c-sis gene products coincide with the stop codon of the c-sis gene located 123 bp into the 5th region of homology. There is

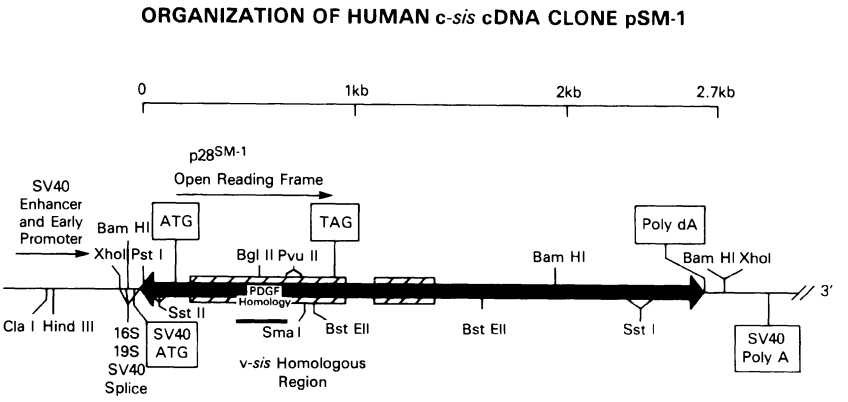


FIGURE 2. Restriction enzyme map of c-sis cDNA plasmid clone pSM-1. The restriction enzyme map was determined by Southern blot hybridization (14) to a v-sis probe. The dark area shows the cDNA insert, and the cross-hatched area indicates the v-sis homologous region. The potential initiator and termination codons of this sequence were obtained from nucleotide sequence data.

91% homology between the nucleotide sequences of v-sis and the corresponding region of c-sis with substitutions mainly in the third codon position in the open reading frame, and the greatest divergence in the 3' untranslated portion of the sequences.

The predicted protein product for human c-sis is identical in all of 31 amino acids to that determined for one of the peptides of PDGF (62, 68). Further amino acid sequence data of PDGF will allow a more complete comparison. However, the data strongly suggest that c-sis encodes chain A of PDGF. The locus of genetic information for chain B of PDGF is unknown, but the significant homology (about 60%) between chains A and B (62, 66) suggests that they have been derived from a common ancestral gene.

The c-sis gene has been mapped to chromosome region 22q12.3-13.1 by somatic cell hybrid analysis (70, 71) together with in situ hybridization (72). Though this gene is translocated to chromosome 9 in patients with Philadelphia chromosome (Ph^1)-positive chronic myelogenous leukemia (CML), this region is somewhat distant from the breakpoint at 22q11 (72). Furthermore, transcripts of c-sis were detected in only 1 of 10 cases of CML examined by our group (M. Blick, unpublished data). Of interest, however, is the finding that c-abl is localized to chromosome region 9q34 which corresponds to the breakpoint seen in CML (73). The c-abl gene was found to be translocated in all Ph^1-positive CML patients and in a cell line derived from a CML patient, K562, it was also found to be amplified and actively transcribed (74, 75). Thus, if either oncogene plays an important role in the pathogenesis of CML, c-abl is more likely to be important.

Translocations involving chromosome 22 have also been reported in several aneuploid Ewing's sarcoma cell lines (76, 77). This is of interest in light of the findings already discussed of transcription of c-sis in several sarcoma lines. Further characterization of both fresh and cultured Ewing's sarcoma cells must be done to determine if c-sis is translocated in this situation, if a DNA rearrangement is detectable in the vicinity of the c-sis gene, and if the gene is expressed in these cells at the RNA and protein levels.

Expression of cellular onc gene homologues

We investigated, together with the laboratories of S. Aaronson and T. Papas, the expression of c-onc genes in various solid and hematopoietic malignancies (78, 79). Poly A(+) RNA was isolated from various malignant cell lines and fresh tissues and examined by Northern blot analysis by hybridization with various v-onc probes (Table 1).

This survey revealed no expression of c-fes, and relatively low levels of c-Ha-ras and c-abl in all cells examined. The multiple different transcripts detected for c-abl and c-Ha-ras raise questions about their source from one or more different genes and/or alternative patterns of post-transcriptional

Table 1. Expression of oncogenes in human hematopoietic cells

Cell Type	Source	Stage of Differentiation	mRNA Species Detected With v-abl (kb) 7.2, 6.4 3.8, 2.0	v-myc (kb) 2.7	v-myb (kb) 4.5	v-Ha-ras (kb) 6.5 5.8	v-sis (kb) 4.3	v-fes ?
Myeloid	KG-1	Myeloblast	++	++	++	+	-	-
	HL60	Promyelocyte	++	++++	++	+	-	-
	HL60 + DMSO, RA	Granulocyte	++	±	-	+	-	-
	Fresh AML cells (4 patients)	Myeloblast	++	++	++	+	-	-
Erythroid	K562	(Immature erythroid precursor)	++	++	++	+	-	-
Lymphoid								
T-cells	CEM	Immature T cell	++	++	+++	+	-	-
	MOLT4	Immature T cell	++	++	+++	+	-	-
	HUT78	Mature T cell	++	++	-	+	-	-
	HUT102	Mature T cell	++	++	-	+	+	-
B-cells	Raji	Burkitt lymphoma line	++	++	-	+	-	-
	Daudi	Burkitt lymphoma line	++	++	-	+	-	-
	NC37	EBV transformed normal B cell line	++	++	-	+	-	-
Normal peripheral lymphocytes			NT	++	-	NT	NT	NT
Normal peripheral lymphocytes + PHA			NT	++	-	NT	NT	NT

NT = not tested

processing. The findings of conservation of oncogenes in diverse vertebrate species and the active transcription of some in many normal and/or neoplastic cells provides evidence for a functional role of these genetic elements.

The levels of expression of c-myc ranged from about 2-20 copies/cell in the various hematopoietic and solid tumor cells examined, except for the HL60 promyelocytic cell line containing an amplification of the c-myc gene. The amplification was also demonstrated in the fresh tumor tissue of the patient from whom the cell line was derived (80), showing that the gene alteration was not an artifact of in vitro culture conditions. While the site of the amplified myc gene in the fresh promyelocytic leukemia tissue was most likely in double minute chromosomes, the HL60 cell line has lost the DM chromosomes and developed an abnormally-banded region in chromosome 8, presumably at the site of the normal c-myc locus (81). Agents that induce differentiation give rise to a decline in the level of c-myc transcription without any significant alteration in the level of gene amplification. The amplification of c-myc gene could not be found in several other promyelocytic leukemias. Subsequently amplification of myc has been demonstrated in cell lines of more aggressive variants of small cell carcinomas of the lung (82) and a human colon carcinoma cell line (83), amplification of a related gene, N-myc in neuroblastomas (36), amplification of abl in the K562 erythroblastic cell line from a patient with CML in blast crisis (74, 75), and amplification of Ki-ras in a murine adrenocortical tumor (84).

The expression of c-myb was somewhat more restricted; 4.5 kb transcripts were found in immature T-lymphoid, myeloid, and erythroid cells but not in B-lymphoid cells, mature T-lymphoid, or myeloid cells, or solid tumor cell lines (27). It would be of interest to determine if the c-myb gene product plays a general role in the maturation of hematopoietic cells.

Transcripts of c-sis were not detected in any hematopoietic cells except some mature T-cell leukemia lines, transformed by human T-cell leukemia virus (HTLV) (78). A single 4.2 kb transcript was detected in these cells as well as in 5 of 6

sarcomas and 1 of 2 glioblastomas (78, 79). No c-sis expression was found in carcinomas or melanomas.

c-sis cDNA

To characterize the transcribed sequences involved in malignancy, we constructed a cDNA library from an HTLV-transformed cell line (85). This provided a reference bank to study oncogene transcripts, viral RNAs, and other specific mRNAs associated with the transformed phenotype. The vector chosen was that described by Okayama and Berg (3) which includes the pBR322 origin of replication and ampicillin-resistance gene, and simian virus 40 (SV40) enhancer, promoter, splice, and poly-adenylation signals. This vector allows cloning in bacteria and ready transfection and expression in eukaryotic cells. Three clones have been isolated after screening about 300,000 colonies for hybridization to a v-sis probe. These have been denoted pSM-1, pSM-2, pSM-3 which contain cDNA inserts of 2.7, 2.7, and 1.8 kb, respectively. All of these represent incomplete transcripts compared to the full-length 4.2 kb mRNA detected in these cells. pSM-1 was selected for further analysis. Restriction enzyme digests and Southern blot hybridization to a v-sis probe showed the area of homology to lie in the 5' portion of the cDNA insert (Fig. 2). There are about 180 bp 5' to the area of v-sis homology which represent an additional upstream exon(s) to those already described in the c-sis gene (L. Ratner and S. Josephs, unpublished data).

Using pSM-2 as a probe, we compared the c-sis gene structure of HUT-102 cells to that of an uninfected B-lymphoid cell line, CB-B (85). The digestion fragments with a number of enzymes were the same in the two cell lines (not shown), providing no evidence for a rearrangement of the c-sis gene.

Restriction maps of pSM-2 and pSM-3 are the same as that of pSM-1 except for deletions at the 5' portion of the cDNA insert (our unpublished data with M. Reitz). Thus, these two clones likely represent shorter reverse transcripts of the same mRNA sequences. The restriction map of the cDNA insert in pSM-1 is identical to that of the corresponding regions of the

normal c-sis gene. Nucleotide sequence analysis of pSM-1 reveals a potential initiation ATG codon 64 bp upstream from the region of v-sis homology (our unpublished data with S. Josephs). No env gene sequences were found 5' to the v-sis homologous sequences. There are also about 1130 bp downstream from the v-sis homologous open reading frame.

Upon transfection of NIH-3T3 cells with pSM-1, 200 transformed foci/microgram DNA were detected at 18 days (85). These foci appeared different morphologically from those typically seen upon transfection with ras; these cells are less spindle-shaped but retain the ability to pile up in a focus. Digests with BamHI and ClaI revealed the entire sequence of pSM-1 to be present in all the transfectants. Digests with a no-cut enzyme, EcoRI showed multiple different integration sites. RNA transcripts homologous to c-sis in these transformed 3T3 cells were 3.5 kb which is consistent with the use of the SV40 promoter and polyadenylation signals (our unpublished data with E. Westin).

While the nucleotide sequence of the open reading frame of the v-sis homologous region is identical to that of the corresponding regions of the normal c-sis gene, external application of PDGF to these same cells does not induce transformation (86). This raises questions as to the differences between the protein product of this truncated c-sis sequence in 3T3 cells versus that of the complete c-sis gene in normal and transformed human cells. Is the transformation of 3T3 cells with this cDNA clone due to: a) loss of regulatory sequences in the c-sis mRNA, b) higher levels of intracellular PDGF expression, c) differences in post-translational processing of the c-sis gene product, d) alteration in subunit structure, i.e., homodimer rather than heterodimer and/or e) differences in interaction with cellular receptors? Is the c-sis gene product exported in these cells and if so does it play any role in the malignant state? It would thus be of interest to test the effect of antibodies to PCGF on the establishment and growth of transformed cells in culture.

Preliminary analysis of the 3T3 fibroblasts transformed by the c-sis cDNA clone has revealed significantly increased expression of the mouse c-myc gene (our unpublished data with E. Westin). This is similar to the activation of the myc gene reported by Kelley and coworkers with 3T3 cells treated with PDGF, or of lymphocytes treated with different mitogens (87). The interaction of these oncogenes in transformation is reminiscent of the cooperativity of myc and ras in transformation of various primary fibroblasts.

With the identification of an oncogene protein product and characterization of at least some of its biologic functions, questions arise as to the usefulness of these reagents in the diagnosis, staging, and treatment of human malignancies.

Mechanisms of oncogene activation

It has been speculated that cellular protooncogenes are genes involved in cell proliferation and/or differentiation. The direct link of these genes to growth factors and growth factor receptors provided strong evidence for this speculation. Therefore, it is conceivable that inappropriate expression of these genes will lead to a cell that is blocked in differentiation or kept in constant proliferation, in other words, a transformed cell. The mechanism of activation of a normal gene to a transforming gene may be based on either a qualitative or a quantitative change.

Qualitative alterations have been demonstrated with c-Ha-ras, c-Ki-ras, and N-ras with specific point mutations in codons 12 and 61 being associated with transforming ability in the 3T3 transfection assay (88-97). Several models of quantitative alteration in oncogene expression have been studied. One is insertional mutagenesis by chronic leukemia viruses in bursal lymphomas induced by avian leukosis virus after a long latent period; c-myc is activated 50-100 fold (98-102). In some cases this is the result of integration of the ALV genome upstream of the c-myc gene in the correct orientation for utilization of the 3' LTR as the transcriptional promoter. In other cases in which ALV is integrated downstream from c-myc or in the wrong orientation upstream

of c-myc, gene activation may be due to utilization of enhancer sequences in one of the LTRs. Similar transcriptional activation of c-myc by chicken syncytial virus (103) and c-erb-B by avian leukosis virus (104, 105) have been described. Furthermore, the cellular mos gene, which is closely homologous to v-mos can transform NIH-3T3 cells when linked to an active viral promoter (106).

Another mechanism of over production of c-onc mRNA which has already been discussed is via gene amplification. The processes associated with this gene amplification, however, are poorly characterized. Surveys for such an amplification, however, have revealed that it is not a common process in human cancer prior to the institution of therapy (our unpublished data).

The role of chromosomal alteration and involvement of oncogenes in specific human malignancies is intriguing though not understood. Definition of transcriptional and translational changes of oncogene information will require further study.

REFERENCES

1. Maniatis, T., Madison, P.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sin, G.K. and Esfratiadis, A. Cell 15:687-701, 1978.
2. Maniatis, T., Fritsch, E.F. and Sambrook, J. Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, 1982.
3. Okayama, H. and Berg, P. Mol. Cell. Biol. 2:161-170, 1982.
4. Gardner, M.B., Rongey, R.W., Anstein, P. Estes, J.D., Sarma, P., Huebner, R.J. and Richard, G. G. Nature 226:807-809, 1970.
5. McDonough, S.K., Larsen, S., Brodey, R.S., Stock, N.D. and Hardy, W.D., Jr. Cancer Res. 31:953-956, 1971.
6. Snyder, S.P. and Theilen, G.H. Nature 221:1074-1075, 1969.
7. Frankel, A.C., Gilbert, J.H., Porzig, K.J., Scolnick, E.M. and Aaronson, S.A. J. Virol. 30:821-827, 1979.
8. Shibuya, M., Hanafusa, T., Hanafusa, H. and Stephenson, J.R. Proc. Natl. Acad. Sci. USA 77:6536-6540, 1980.
9. Franchini, G., Gelmann, E.P., Dalla Favera, R., Gallo, R.C. and Wong-Staal, F. Mol. Cell. Biol. 2:1014-1019, 1982.
10. Franchini, G., Even, J., Scherr, C.J. and Wong-Staal, F. Nature 290:154-157, 1981.
11. Scherr, C.J., Fedele, L.A., Oskarsson, M., Maizel, J., Vande Woude, G.F. J. Virol. 34:200-202, 1980.
12. Van Beveren, C., Galleshaw, J.A., Jonas, V., Berns, A.J.M., Doolittle, R., Donoghue, D.J. and Verma, I.M. Nature 289:258-262, 1981.

13. Goff, S.P., Gilboa, E., Witte, O.N. and Baltimore, D. Cell 22:777-785, 1980.
14. Takeya, T. and Hanafusa, H. Cell 32:881-890, 1983.
15. Chen, I.S.Y., Wilhelmsen, K.C. and Temin, H.M. J. Virol. 45:104-113, 1983.
16. Dhar, I.R., Ellis, R.W., Oroszlan, S., Shapiro, B., Maizel, J., Lowy, D. and Scolnick, E. Science 217:934-936, 1982.
17. Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, D.H. and Goeddel, D.V. Nature 302:33-37, 1983.
18. Ellis, A.W., Lowy, D.R. and Scolnick, E.M. In: Advances in Viral Oncology (Ed. G. Klein), Raven Press, New York, 1982, Vol. 1, pp. 107-126.
19. Dalla Favera, R., Franchini, G., Martinotti, S., Wong-Staal, F., Gallo, R.C. and Croce, C.M. Proc. Natl. Acad. Sci. USA 79:4714-4717, 1980.
20. Heisterkamp, N., Groffen, J., Stephenson, J.R., Spurr, N.K., Goodfellow, P.N., Solomon, E., Carritt, B. and Bodmer, W.F. Nature 299:747-750, 1982.
21. Harper, M.E., Franchini, G., Love, J., Simon, M.I., Gallo, R.C. and Wong-Staal, F. Nature 304:169-171, 1983.
22. Rowley, J.D., Golomb, H.M. and Doughterty, C. Lancet i: 549-550, 1977.
23. Graf, T. and Beug, H. Biochim. Biophys. Acta. 516:269-299, 1978.
24. Rushlow, K.M., Lautenberger, J.A., Papas, T.S., Baluda, M.A., Perbal, B., Chirikjian, J.G. and Reddy, E.P. Science 216:1421-1423, 1982.
25. Franchini, G., Wong-Staal, F., Baluda, M.A., Lengel, C. and Tronick, S.R. Proc. Natl. Acad. Sci. USA 80:7385-7389, 1983.
26. 3rd International Workshop on Chromosomes in Leukemia. Cancer Res. 43:868-873, 1983.
27. Westin, E.H., Gallo, R.C., Arya, S.K., Eva, A., Souza, L.M., Baluda, M.A., Aaronson, S.A. and Wong-Staal, F. Proc. Natl. Acad. Sci. USA 79:2194-2198, 1982.
28. Wake, N., Hreshchyshyn, M.M., Piver, S.M., Matsui, S. and Sandberg, A.A. Cancer Res. 40:4512-4518, 1980.
29. Trent, J.M. and Salmon, S.E. Cancer Genet. Cytogenet. 3:279-291, 1981.
30. Mitelman, F. and Levan, G. Hereditas 95:79-139, 1981.
31. Hann, S.R., Abrams, H.D., Rohrscheider, L.R. and Eisenman, R.N. Cell 34:789-798, 1983.
32. Dalla Favera, R., Gelmann, E.P., Martinotti, S., Franchini, G., Papas, T.S., Gallo, R.C. and Wong-Staal, F. Proc. Natl. Acad. Sci. USA 79:6497-6501, 1982.
33. Watt, R., Stanton, L.W., Marcu, K.B., Gallo, R.C., Croce, C.M. and Rovera, G. Nature 303:725-728, 1983.
34. Saito, H., Hayday, A.C., Wiman, K., Hayward, W.S. and Tonegawa, S. Proc. Natl. Acad. Sci. USA 80:7476-7480, 1980.
35. Stanton, L.W., Watt, R. and Marcu, K. Nature 303:401-406, 1983.
36. Kohl, N.E., Kanda, N., Schreck, R.R., Bruns, G. Latt, S.A., Gilbert, F. and Alt, F.W. Cell 35:359-367, 1983.

37. Dalla Favera, R., Brogni, M., Erikson, J., Patterson, D., Gallo, R.C. and Croce, C.M. Proc. Natl. Acad. Sci. USA 79:7824-7827, 1982.
38. Rabbitts,T.H., Forster, A., Baer, R. and Hamlyn, P.H. Nature 306:806-809, 1983.
39. Croce, C.M., Thierfelder, W., Erikson, J., Nishikura, K., Finan, J., Lenoir, G. and Nowell, C. Proc. Natl. Acad. Sci. USA 80: 6922-6926, 1983.
40. Erikson, J., Nishikura, K., Ar-Rushdi, A., Finan, J., Emanuel, B., Lenoir, G., Nowell, P.C. and Croce, C.M. Proc. Natl. Acad. Sci. USA 80:7581-7585, 1983.
41. Hollis, G.F., Mitchell, K.F., Battey, J., Potter, H., Taub, R., Lenoir, G.M. and Leder, P. Nature 307:752-755, 1984.
42. Gelmann, E.P., Psallidopoulos, M.C., Papas, T.S. and Dalla Favera, R. Nature 306:799-803, 1983.
43. Rabbitts, T.H., Hamlyn, P.H. and Baer, R. Nature 306: 760-768, 1983.
44. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. and Leder, P. Cell 34:799-789, 1983.
45. Hamlyn, P.H. and Rabbitts, T.H. Nature 304:135-139, 1983.
46. Erikson, J., Ar-Rushi, A., Drwinga, H.L., Nowell, P.C. and Croce, C.M. Proc. Natl. Acad. Sci. USA 80:820-824, 1983.
47. Klein, G. Nature 294:313-318, 1981.
48. Nishikura, K., Ar-Rushdi, A., Erikson, J., Watt, R., Rovera, G. and Croce, C.M. Proc. Natl. Acad. Sci. USA 80:4822-4826, 1983.
49. Shen-Ong, G.L.C., Keath, E.J., Piccoli, S.P. and Cole, M.D. Cell 31:443-452, 1982.
50. Adams, J.M., Gerondakis, S., Webb, E., Mitchell, J., Bernard, O. and Cory, S. Proc. Natl. Acad. Sci. USA 79:6966-6970, 1982.
51. Wiener, F., Onno, S., Babonitz, M., Sumegi, J., Wirschubsky, Z., Klein, G., Mushinski, J.F. and Potter, M. Proc. Natl. Acad. Sci. USA 81:1159-1163, 1984.
52. Calame, K., Kim, S., Lalley, P., Hill, R., Davis, S. and Hood, L. Proc. Natl. Acad. Sci. USA 79:6994-6998, 1982.
53. Taub, R., Hirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S. and Leder, P. Proc. Natl. Acad. Sci. USA 79:7837-7841, 1982.
54. Theilen, G.H., Gould, D., Fowler, M. and Dungworth, D.L. J. Natl. Cancer Inst. 47:881-889, 1971.
55. Gelmann, EP., Wong-Staal, F., Kramer, R.A. and Gallo, R.C. Proc. Natl. Acad. Sci. USA 78:3373-3377, 1981.
56. Robbins, K.C., Devare, S.G. and Aaronson, S.A. Proc. Natl. Acad. Sci. USA 78:2918-2922, 1981.
57. Gelmann, E.P., Petri, B., Cetta, A. and Wong-Staal, F. J. Virol. 41:593-604, 1982.
58. Wong-Staal, F., Dalla Favera, R., Gelmann, E.P., Manzari, V., Szala, S., Josephs, S.F. and Gallo, R.C. Nature 294:273-275, 1981.

59. Devare, S.G., Reddy, E.P., Law, D.J., Robbins, K.C. and Aaronson, S.A. Proc. Natl. Acad. Sci. USA 80:731-735, 1983.
60. Robbins, K.C., Antoniades, H.N., Devare, S.G., Hunkapiller, M.W. and Aaronson, S.A. Nature 305:605-608, 1983.
61. Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins, K.C., Aaronson, S.A. and Antoniades, H. Science 221:275-277, 1983.
62. Waterfield, M.D., Scrace, G.T., Stroobant, P., Wasteson, A., Westermark, B., Helden, C-H., Huang, J.S. and Deuel, T.F. Nature 304:35-39, 1983.
63. Deuel, T.F., Huang, J.S., Proffitt, R.T., Chang, D. and Kennedy, B.B. J. Supramolec. Struct. Cell Biochem. Suppl. 5:128, 1981.
64. Deuel, T.F., Huang, J.S., Proffitt, R.T., Baenziger, J.U., Chang, D. and Kennedy, B.B. J. Biol. Chem. 256: 8896-8899, 1981.
65. Johnsson, A., Heldin, C-H., Westermark, B. and Wasteson, A. Biochem. Biophys. Res. Commun. 104:66-74, 1982.
66. Antoniades, H.N. and Hunkapiller, M.W. Science 220:963-965, 1983.
67. Dalla Favera, R., Gelmann, E.P., Gallo, R.C. and Wong-Staal, F. Nature 292:31-35, 1981.
68. Josephs, S.F., Guo, C., Ratner, L. and Wong-Staal, F. Science 223:487-491, 1984.
69. Josephs, S.F., Dalla Favera, R., Gelmann, E.P., Gallo, R.C. and Wong-Staal, F. Science 219:503-505, 1983.
70. Dalla Favera, R., Gallo, R.C., Giallongo, A. and Croce, C.M. Science 218:686-688, 1982.
71. Swan, D.C., McBride, O.W., Robbins, K.C., Keithley, D.A., Reddy, E.P. and Aaronson, S.A. Proc. Natl. Acad. Sci. USA 79:4691-4695, 1982.
72. Bartram, C.R., deKlein, A., Hagemeijer, A., Grosveld, G., Heisterkamp, N. and Groffen, J. Blood 63:223-225, 1984.
73. deKlein, A., Geurts van Kessel, A., Grosveld, G., Bartram, C.R., Hagemeijer, A., Bootsma, D., Spurr, N.K., Heisterkamp, N., Groffen, J. and Stephenson, J.R. Nature 300:765-767, 1982.
74. Collins, S. and Groudine, M.T. Proc. Natl. Acad. Sci. USA 80:4813-4817, 1983.
75. Selden, J.R., Emanuel, B., Wong, E., Cannizzaro, L., Polumbo, A., Erikson, J., Nowell, P.C., Rovera, G. and Croce, C.M. Proc. Natl. Acad. Sci. USA 80:7289-7292, 1983.
76. Aurias, A., Rimbaut, C., Buffe, D., Dubosset, J. and Mazabraud, A. N. Engl. J. Med. 309:496-497, 1983.
77. Turc-Carel, C., Philip, I., Berger, M.P., Philip, T. and Lenoir, G.M. N. Engl. J. Med. 309:497-498, 1983.
78. Westin, E.H., Wong-Staal, F., Gelmann, E.P., Dalla Favera, R., Papas, T.S., Lautenberger, J.A., Eva, A., Reddy, E.P., Tronick, S.R., Aaronson, S.A. and Gallo, R.C. Proc. Natl. Acad. Sci. USA 79:2490-2494, 1982.

79. Eva, A., Robbins, K.C., Andersen, P.R., Srinivasan, A., Tronick, S.R., Reddy, E.P., Ellmore, N.W., Galen, A.T., Lautenberger, J.A., Papas, T.S., Westin, E.H., Wong-Staal, F., Gallo, R.C. and Aaronson, S.A. Nature 295: 116-119, 1982.
80. Dalla Favera, R., Wong-Staal, F. and Gallo, R.C. Nature 299:61-63, 1982.
81. Nowell, P., Finan, J., Dalla Favera, R., Gallo, R.C., Ar-Rushdi, A., Ramaczuk, H., Rovera, G. and Croce, C.M. Nature 306:494-497, 1983.
82. Little, C.D., Nau, M.M., Carney, D.N., Gazdar, A.F. and Mann, J.D. Nature 306:194-196, 1983.
83. Alitalo, K., Schwab, M., Lin, C.C., Varmus, H.E. and Bishop, J.M. Proc. Natl. Acad. Sci. USA 80:1707-1711, 1983.
84. Schwab, M., Alitalo, K., Varmus, H.E. and Bishop, J.M. Nature 303:497-501, 1983.
85. Clarke, M.F., Westin, E., Schmidt, D., Josephs, S.F., Ratner, L., Wong-Staal, F., Gallo, R.C. and Reitz, J.S., Jr. Nature 308:464-467, 1984.
86. Stiles, C.D. Cell 33:653-655, 1983.
87. Kelley, K., Cochran, B.H., Stiles, C.D. and Leder, P. Cell 35:603-610, 1983.
88. Land, H., Parada, L.F. and Weinberg, R.A. Nature 304: 596-602, 1983.
89. Shimizu, K., Birnbaum, D., Ruley, M.A., Fasano, O, Suard, Y., Edlund, L., Taparowsky, E., Goldfarb, M. and Wigler, M. Nature 304:497-500, 1983.
90. Nakano, H., Yamamoto, R., Neville, C., Evans, D., Mizuno, T. and Perucho, M. Proc. Natl. Acad. Sci. USA 81:71-75, 1984.
91. Tabin, C.J., Bradley, S.M., Bargmann, C.I., Weinberg, R.A., Papageorge, A.G., Scolnick, E.M., Dhar, R., Lowy, D.R. and Chang, E.H. Nature 300:143-149, 1982.
92. Reddy, E.P., Reynolds, R.K., Santos, E. and Barbacid, M. Nature 300:149-152, 1982.
93. Taparowsky, E., Suard, Y., Fasano, O., Shimizu, K., Goldfarb, M. and Wigler, M. Nature 300:762-765, 1982.
94. Capon, D.J., Seeburg, P.H., McGrath, J.P., Hayflick, J.S., Edman, U., Levinson, A.D. and Goeddel, D.V. Nature 304:507-513, 1983.
95. Taparowsky, E., Shimizu, K., Goldfarb, M. and Wigler, M. Cell 34:581-586, 1983.
96. Santos, E., Reddy, E.P., Pulciani, S., Feldmann, R.J. and Barbacid, M. Proc. Natl. Acad. Sci. USA 80:4679-4683, 1983.
97. Santos, E., Martin-Zanca, D., Reddy, E.P., Pierotti, M.A., Della Porta, G. and Barbacid, M. Science 223:661-664, 1984.
98. Hayward, W.J., Neel, B.G. and Astrin, S.M. Nature 290: 475-480, 1981.
99. Neel, B.G., Hayward, W., Robinson, M.L., Fang, J. and Astrin, S.M. Cell 23:323-334, 1981.
100. Fung, Y.K., Faelly, A.M., Crittenden, L.B. and Kung, H.J. Proc. Natl. Acad. Sci. USA 78:3418-3422, 1981.

101. Payne, G.S., Bishop, J.M. and Varmus, M.E. Nature 295:209-214, 1982.
102. Fung, Y.K.T., Crittenden, L.B. and Kung, H.J. J. Virol. 44:742-746, 1982.
103. Noori-Dalooi, M.R., Swift, R.A., Kung, H.J., Witter, R.L. and Crittenden, L.B. Nature 294:574-576, 1981.
104. Vennstrom, B. and Bishop, J.M. Cell 28:135-143, 1982.
105. Fung, Y-K.T., Lewis, W.G., Crittenden, L.B. and Kung, H.J. Cell 33:357-368, 1983.
106. Blair, D.G., McClements, W.L., Oskarsson, M., Fischinger, P. and Vande Woude, G.F. Proc. Natl. Acad. Sci. USA 77:3504-3508, 1980.

GENETIC ENGINEERING STRATEGY

3

DEVELOPMENT OF PLASMIDS AND CLONING PROCEDURES

RUDOLF EICHENLAUB

Lehrstuhl Biologie der Mikroorganismen, Ruhr-Universität Bochum, 4630 Bochum 1, Federal Republic of Germany

SUMMARY

The investigation of the F factor of Escherichia coli is chosen as an example for the recent developments in the field of plasmid research. F-derived mini-plasmids were constructed and their properties with respect to replication and copy number control were studied, applying numerous experimental approaches including electron microscopy, mutagenesis, cloning of subfragments, expression of gene products and DNA sequencing. Thus, insights into the mechanism of replication, copy number control and partitioning were achieved which finally allowed construction of special purpose cloning vectors based on plasmid F.

INTRODUCTION

Bacterial plasmids have played an important role since the very beginning of genetic engineering. However, the interest in studying their biology has started much earlier and was initiated by genetic and physiological phenomena for which they were responsible. These were the production of colicins by the Col plasmids (1), fertility of bacteria conferred by the F factor (2) and antibiotic resistance due to the presence of R factors (3). Intensive research by many groups on these three types of plasmids has culminated in the first successful cloning experiments demonstrating that the in vitro recombination of a DNA segment with a plasmid vector resulted in a genetically stable hybrid molecule.

This tremendous progress in molecular genetics was, however, only possible due to the fact that the enzymology for in vitro recombination and the methods for the introduction of DNA into the bacterial cell (transformation) were developed. The fact that numerous restriction endonucleases and other DNA modifying enzymes

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

are now commercially available has made DNA cloning an every day routine in many laboratories.

The cloning of DNA into an organism depends largely on the carrier DNA molecule, the vector. The most important prerequisite of a vector is that it represents a replicon, an autonomously replicating entity of DNA. Furthermore the vector has to carry markers that allow selection and screening for the presence of hybrid plasmids in the cell. Two types of plasmids which fulfill these requirements, the R factors carrying antibiotic resistance and the Col factors responsible for colicin production and colicin immunity were the first plasmids to be used in *in vitro* recombination experiments.

In 1973 Cohen and coworkers (4) who had also established a procedure to transform *E. coli* with plasmid DNA (5) found that by shearing DNA of a large R factor and transformation with the DNA fragments a small plasmid could be obtained which carried tetracycline resistance. This plasmid, designated pSC101, later served in the first cloning experiment with eukaryotic DNA (6).

Plasmid ColE1 is another example for a cloning vector (7,8). As pSC101, plasmid ColE1 has a single recognition site for the endonuclease *EcoRI*. Insertion of a DNA fragment into this site is accompanied by loss of colicin production in the transformed clone. Since such clones are still colicin immune a direct selection for transformants is possible. Using this strategy ColE1 was used to clone part of the tryptophane operon of *E. coli* (9). Although the use of ColE1 as a cloning vector was difficult with regard to its markers, it had the definite advantage over other plasmids in that its copy number could be amplified by treating the cells with chloramphenicol (10). The large increase in copy number up to 1000 per cell facilitated plasmid isolation. Furthermore, Hershfield and coworkers isolated a spontaneously generated mini-ColE1 plasmid, pVH51 which was only 2.3 Megadalton (Mdal) in size and exhibited an elevated copy number of about 60-80 as compared to 20 of the parental ColE1 plasmid (11). In addition mini-ColE1 showed a very low rate of cotransfer in the presence of a transfer factor, an important fact for safety considerations (11). Later the joining of antibiotic resistance

markers to plasmid ColE1 generated a number of vectors like pCR1, pML21, and pMB9 (11,12,13) and the now generally used vector pBR322 and its derivatives (14, 15).

Progress in the cloning technology is basically achieved on two levels (i) by the improvement of the biochemical methodology in nucleic acid chemistry and (ii) by the construction of special cloning vectors. The latter is only possible by the detailed study of the biology and genetics of potential vector molecules.

In this communication I do not want to elaborate on the enzymatic and methodical details of gene cloning, numerous reviews already exist on this subject (see i.e. 16, 17). Instead I prefer to present our studies on the F factor of E. coli and the construction of mini-F derived cloning vectors.

RESULTS

The F factor of E. coli and the construction of a mini-F plasmid.

The F factor of E. coli is a 64 Mdal (94.5kb) plasmid which confers to the bacterial cell the ability for DNA transfer (conjugation) (2). During conjugation either the F factor itself or when F is integrated into the host chromosome (Hfr) the chromosome is transferred into the recipient cell. Fertility of E. coli due to the F factor enabled geneticists to establish the detailed genetic knowledge we have on E. coli. The F factor itself was also intensively studied to understand the mechanism of DNA transfer and vegetative replication. While DNA transfer could be studied by a genetic approach (2) the investigation of the mechanism of vegetative replication and maintenance of F was strongly hampered by the size of F and its low copy number of only 1-2 per chromosome (18). Therefore, a considerable progress was only achieved when a mini-derivate was obtained. F'lac DNA was fragmentated by restriction endonuclease EcoRI and the fragments were joined in vitro with another EcoRI fragment carrying an antibiotic resistance marker. After transformation of E. coli with such randomly recombined DNA molecules and selection for kanamycin resistance clones harbouring a plasmid consisting of a 6 Mdal F fragment and the antibiotic resistance fragment were obtained (19, 20). Heteroduplex studies showed that the 6 Mdal (9kb) segment present

in the mini-F plasmid represented the region 40.3-49.3F of the F factor (21,22,23) (Fig. 1). This demonstrated, that all functions involved in replication and its control were closely linked on a relatively small segment of DNA. The mini-F plasmid further behaved like the parental F factor with respect to copy number of 1-2 per chromosome and incompatibility against IncFI group plasmids.

Incompatibility refers to the inability of two plasmids of the same incompatibility group to stably coexist in a cell. As a result of incompatibility segregation of the two plasmids is observed (24,25). As we will see later incompatibility is caused by the replication control and partitioning mechanisms.
By a similar approach mini-plasmids have been generated from various large plasmids like R6-5, R1, and R6K (26,27,28,29).
The map of mini-F shown in Figure 1 summarizes the data of several laboratories (30,31,32,33).

Mapping of replication origins.

One of the first steps in the characterization of a replicon is the mapping of replication origins and the determination of the direction of replication. Since in a growing bacterial population only few plasmids are replicating at a given time it is necessary to enrich for replicative intermediates in a plasmid preparation. A method which has been applied successfully for a variety of plasmids (34,35,36,37) uses thymine starvation followed by a short pulse with thymidine in a thy mutant of E. coli, resulting in a synchronization of replication. The lysis procedure also effcts the yield of plasmid replicative intermediates, since it has been shown that during replication plasmids are attached to the cell membrane (38). Lysis of the cells using a modified SDS-salt procedure (39), originally described by Hirt (40) for preparation of polyoma DNA, gives satisfying results.

Replicative intermediates can be purified on CsCl-dye density gradients or sedimentation in sucrose gradients and then analysed by electron microscopy. Typical examples as observed in the electron microscope are shown in Fig. 2.
Statistical evaluation of the molecules allowed determination of the location of the startpoint (origin) and the direction of replica-

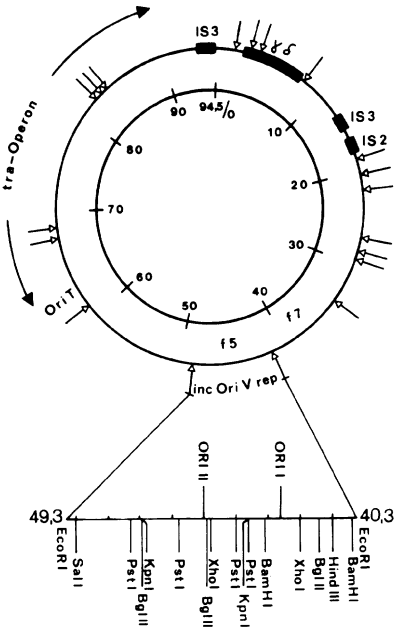


Figure 1. Map of the F factor and plasmid mini-F.
Arrows ——▷ indicate recognition sites of EcoRI. The tra-operon and OriT are required for mating pair formation and DNA transfer. IS2, IS3, and γδ are used for integration into the host chromosome (2). The inner circle indicates F coordinates in kilobases (kb). The EcoRI f5-fragment present in plasmid mini-F encoding functions for incompatibility (inc), replication (rep) and origins of replication (ori) is enlarged.

tion. For plasmid mini-F an origin of replication was observed at F coordinate 42.6F from which replication proceeded bidirectionally (41).

Since this origin was located on a 2.5kb BamHI fragment (see Fig. 1) it was examined whether this BamHI fagment carried all the information for replication and whether deletion of this region from mini-F abolished replication. It was found that the BamHI fragment 40.4-43.1F could not replicate while a deletion of the BamHI fragment (40.4-43.1F) from mini-F did not affect

replication of the plasmid (32,33). This result was surprising in two ways, first it showed that the origin of replication at 42.6F by itself was not sufficient for replication. Secondly, these findings implied that mini-F carried another origin of replication in the region 43.1-49.3F, which was activated when the primary origin (ori I) at 42.6F was deleted. The analysis of replicative intermediates from such a deleted mini-F derivative indeed revealed a second origin located at 45.1F (32,42).

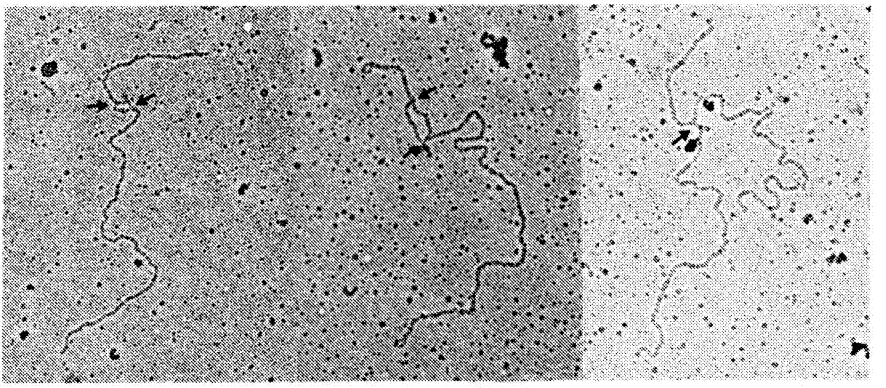


Figure 2. Replicative intermediates of plasmid mini-F.
Replicative intermediates were purified by CsCl-dye density gradient centrifugation, linearized by digestion with EcoRI and after spreading, examined in the electron microscope as described (41). Arrows indicate the position of the replication forks.

Expression of plasmid-encoded genes.

There are two methods for the analysis of plasmid-encoded proteins: (i) synthesis of proteins in a cell-free system (43) and (ii) expression of plasmid-encoded genes in bacterial mini-cells (44). These procedures are especially useful when plasmid products cannot be detected directly in cell lysates. Both systems allow for the exclusive labeling of proteins synthesized de novo from the plasmid DNA template.

In the cell-free system a cell extract from a plasmid-free E. coli strain is prepared (S-30 extract). After preincubation the

extract is virtually free of chromosomal template activity. Addition of CCC plasmid DNA or purified mRNA then leads to the synthesis of protein from the exogenously added template. It is possible to identify the proteins either by activity tests or by PAGE when synthesis was performed in the presence of radioactively labeled amino acids.

Using this method colicin E1 was the first plasmid protein synthesized *in vitro* (45). Meanwhile the cell-free system has been employed in numerous studies for the identification and characterization of plasmid-encoded proteins (46,47).

In *E. coli* and some other bacterial species, mutants producing mini-cells exist. These mini-cells do not contain chromosomal DNA (44). In *E. coli* two mutations *minA* and *minB* are responsible for mini-cell production. When a plasmid is present in such a strain, and this is especially valid for small, high copy number plasmids, plasmids are trapped in the mini-cells. Since mini-cells dispose of a functional transcription and translation they effectively express plasmid-encoded genes.

The study of plasmid mini-F in the mini-cell system has revealed the presence of six proteins which could be resolved on PAGE (Fig. 3) (31,48,49). Since plasmid mini-F does not segregate into mini-cells, mini-F was cloned into high copy number vectors like pBR322 or pACYC184. In such hybrids replication is under the control of the high copy number plasmid resulting in a effective segregation into the mini-cells. Introduction of deletions into the mini-F genome followed by the analysis of the protein patterns allowed the mapping of gene loci with respect to the deleted regions (31,49)(Figure 4).

In order to confirm the protein loci we tried to locate the respective transcripts and promoters. For this we chose the R-loop analysis (50). In this procedure plasmid DNA is incubated with purified RNA-polymerase and rNTP's under conditions optimized for transcription. The transcription products can be hybridized with the partially denatured template DNA to form R-loop molecules. These are then examined in the electron microscope (Fig. 5). The statistical evaluation of such R-loop molecules provides all transcriptional parameters i.e. the number of startpoints and

direction of transcription and the size of the transcripts. Investigation of plasmid mini-F has identified the transcripts corresponding to the mini-F proteins (51) (Fig. 4). However, at present there is no consensus on the direction of transcription obtained by R-loop analysis as compared to other methods (49).

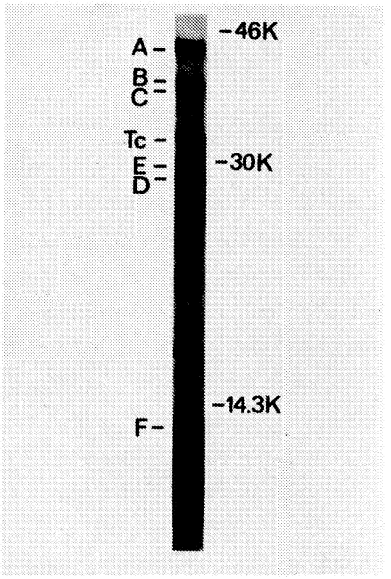


Figure 3. Electrophoretic separation of ^{35}S-methionine labeled mini-F proteins on SDS-polyacrylamide gel. Mini-cells were prepared from E. coli DS410 harboring a mini-F:pACYC184 hybrid plasmid as described (31). Mini-F proteins were labeled with ^{35}S-methionine and separated by electrophoresis on a SDS-polyacrylamide gel (31). Six mini-F specific proteins A-F can be identified. The Tc-protein originates from pACYC184. Molecular weight standards used were: ovalbumin (46K), carbonic anhydrase (30K), and lysozyme (14.3K).

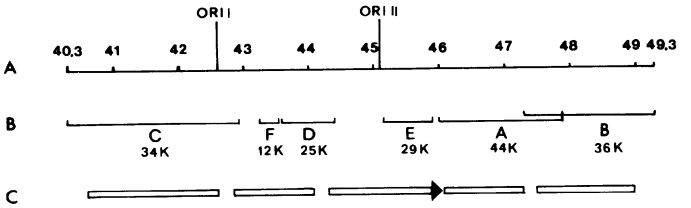


Figure 4. A) Map of plasmid mini-F with F coordinates in kb. B) Map positions of the six mini-F proteins with molecular weights in kilodaltons (K)(31,49). C) Position of transcripts as identified by R-loop mapping (51). For one transcript the direction is indicated by an arrow (51,52).

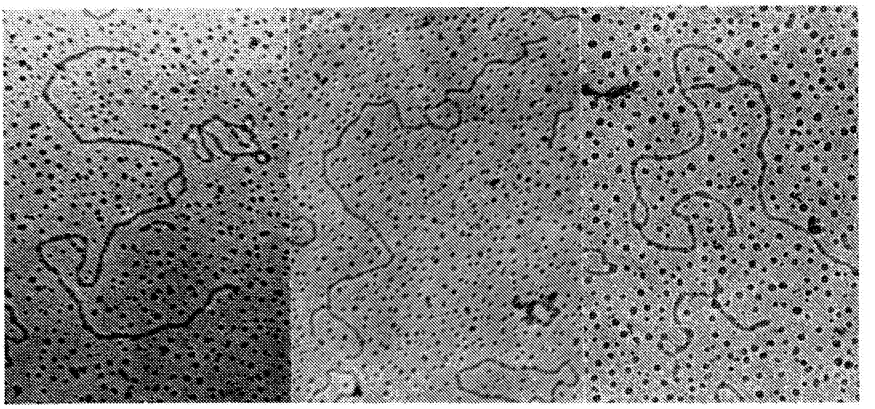


Figure 5. R-loop molecules of plasmid mini-F.
R-loop molecules were prepared as described (50) and visualized by electron microscopy (51).

Diminuation of the mini-F replicon.

The identification of mini-F encoded proteins and transcripts raised the question which of these were essential for replication and copy number control. Numerous experiments finally showed that the smallest mini-F plasmid which could be established in a cell was a 2248 bp segment bordered by recognition sites for the endonucleases PstI (44.1F) and AluI (46.35F) (52). This region which has been sequenced contains a functional origin of replication, encodes a polypeptide of 29K (protein E) and has two regions flanking the E protein containing 4 and 5 repeated nucleotide sequences of 19 bp (Fig. 6) (52,53).

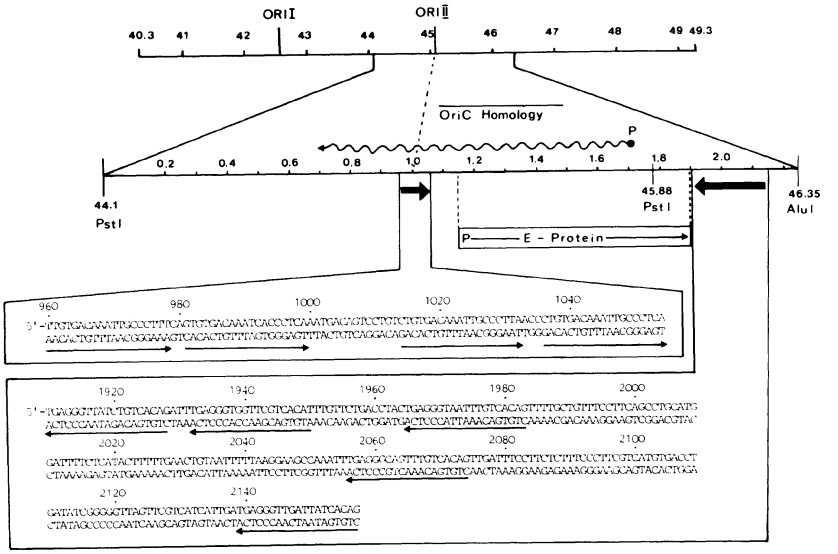


Figure 6. The basic replicon of F.

The region 44.1-46.35F is enlarged. The coding region of the E protein is indicated with its direction of transcription from the promoter p (52). The location of the repeated sequences is indicated by heavy arrows and the nucleotide sequence is given below (52). The wavy line indicates a possible transcript with its promoter which might serve as a primer in the initiation of replication at ori II (51). Ori C homology refers to an extensive nucleotide sequence homology with the E. coli chromosomal origin of replication (52).

It has been shown that these repeats when cloned into pBR322 express incompatibility against another F plasmid (52,53). A current explanation for this is that the repeats may function as recognition sites for a protein essential for the initiation of replication. The protein is then diluted below a critical concentration by binding to the numerous copies of its binding site. The specific protein in question could be the E protein which was shown to be essential for mini-F replication (52,54). Complementation of a replication-deficient plasmid which is deleted for protein E is possible when the E protein is supplied by another plasmid *in trans* (54,70).

Many mini-derivatives of larger plasmids are not completely stable and show a significant rate of segregation. This is more pronounced when the plasmid has a low copy number. A slight decrease in the rate of initiation of replication or a defective partitioning mechanism will immediatedly cause segregation and plasmid-free cells will occur in the bacterial population. This is also observed in many mini-F derivatives.

It has been found some time ago, that there is another region in mini-F exhibiting incompatibility versus another F plasmid or members of the IncFI group (55,56).This region was originally detected by hybridization studies as a homology shared by some plasmids of the IncFI group and located between F coordinates 47.6-49.3F (23). This incompatibility locus,termed *incD*, has recently been discussed as being involved in partitioning of the F factor (57,58). It was observed, that plasmids derived from the *E. coli* chromosome (OriC plasmids) which are highly unstable can be stabilized by joining them with the *incD* region of F (59). It is not clear yet whether a protein of 36K (protein B) or even another one of 44K (protein A) mapping in the *incD* region are involved in the partitioning reaction. Incompatibility exerted by *incD* is then explained as a competition reaction for a partitioning mechanism which may involve plasmid-encoded proteins and a limited number of membrane sites similar to a model originally suggested by Jacob, Brenner,and Cuzin (60).

Absence of the incD region in many mini-F derivatives may cause the observed instability.

Replication mutants of mini-F.

The isolation and characterization of mutations is a powerful tool for the identification of functions residing in a plasmid genome. The first replication mutants of a plasmid were isolated by Cuzin and Jacob who described an F'lac mutant thermosensitive in replication (61).

Plasmid mutants can be obtained by mutagenesis of the cells harbouring a plasmid. However, mutations induced in the chromosome which may be either lethal or affect plasmid replication complicate the isolation of plasmid mutants. In vitro mutagenesis of purified plasmid DNA by suitable agents like hydroxylamine circumvents this problem (62). By using hydroxylamine which reacts with cytosine residues in the DNA and causes transitions from C to T (63) a number of mini-F mutants were isolated carrying conditional mutations in replication (48,64). The following types of mini-F mutants were obtained (i) mutants which did not replicate at elevated temperature (42°C) but behaved normally at 28-30°C, termed rep_{ts} (64),(ii) rep_{am} mutants which do not replicate in a suppressor negative host strain (48); we used an E. coli strain with a thermosensitive suppressor tRNA ($supF_{ts}$), (iii) mutants with an elevated copy number of 60-80 per cell, designated cop (Ebbers and Eichenlaub, in prep.) and (iiii) mutants with an elevated copy number of 60-80 copies per cell at 28-30°C but a low copy number of 5-7 at 42°C, termed cop_{ts} (Ebbers and Eichenlaub, in prep.).

Meanwhile some of the mutations have been mapped precisely and identified by DNA sequencing. The rep_{am} mutations reside either in protein C or in protein E. The mutation in the C protein was identified through the absence of this protein in suppressor-negative cells (48).Mutations in the C protein seem only to effect replication starting at origin I and are not bypassed by replication from origin II. Only when origin I is deleted in such a mutant (mini-Fam1 Δ 40.4-43.1F) start of replication at origin II seems possible (48).

Mutations within the E protein were identified by DNA sequencing of the structural gene (Helsberg, Ebbers and Eichenlaub, in prep.). In the $\underline{rep}_{am}$ mutant effecting the E protein codon 54 of the gene is changed from CAG (Glu) to UAG. Three cop mutations residing in the E protein have been identified as changes of glutamic acid to lysine at position 101, aspartic acid to asparagine at position 102, and histidine to tyrosine at position 130 of the amino acid sequence.

These findings support the notion that the E protein plays a central role in replication and copy number control. The exact position of the $\underline{rep}_{ts}$ mutations has not been obtained so far. The Cop_{ts} phenotype seems to result from the simultaneous occurrence of a cop and a $\underline{rep}_{ts}$ mutation in the mini-F genome (Ebbers and Eichenlaub, in prep.).

Construction of cloning vectors based on plasmid mini-F.

Of what use is it to develop a vector from a plasmid which has only a copy number of 1-2 per cell? Such a copy number results in meager yields in DNA preparations and expression of cloned genes is always at the lowest possible level. However, especially the last point is in favour of using F as a cloning vector in certain cases. At a high gene dosage certain gene products may reduce cell viability as for example membrane and some regulatory proteins. Such protein genes may successfully be cloned in a low copy number vector. Based on the data obtained on mini-F it was possible to construct a single copy cloning vector.

Vector pRE435 carries antibiotic resistance genes against tetracycline (Tc^r), ampicillin (Ap^r), and chloramphenicol (Cm^r). The Tc^r and Cm^r genes have been transferred to mini-F in form of a PstI-AvaI fragment from plasmid pBR325 (15). Prior to this manipulation the BamHI fragment 40.4-43.1F was exchanged against a BglII fragment from plasmid pHC79 (65) carrying the λ-cos site. Furthermore the SalI site at 49.1F was destroyed by cleavage with SalI, treatment with nuclease S1 which removed the unmatched bases at the 5'end, and consecutive ligation of the blunt ends by polynucleotide ligase. For stability the vector has the incD region required for partitioning.

Cloning of fragments derived from restriction digests with HindIII, BamHI, SalI and EcoRI can be inserted into the Tc^r and the Cm^r genes, respectively. Insertions cause inactivation of the respective antibiotic resistance allowing an effective screening for hybrid plasmids. In Figure 7 a physical map of the mini-F vector pRE435 is shown.

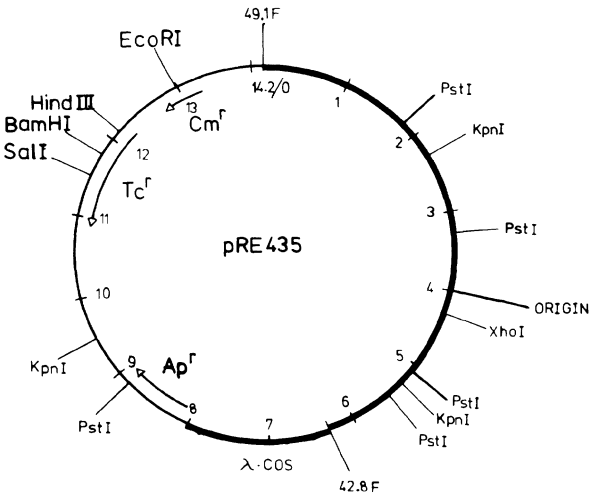


Figure 7. Map of single copy number vector pRE435.
Heavy line indicates mini-F derived sequences and λ-cos sequences from pHC79 (65). The genes for resistance against tetracycline (Tc^r), chloramphenicol (Cm^r), and ampicillin (Ap^r) are given with their direction of transcription. There are single restriction sites for EcoRI, HindIII, BamHI, and SalI. pRE453 has a size of 14.2 kb.

An additional feature of vector pRE435 is the possession of the λ-cos site. Thus pRE435 can be used as a cosmid to establish gene banks. Since cosmid cloning by packaging of *in vitro* recombined DNA into λ heads selects for the insertion of large fragments into the vector, the number of clones representing a complete gene bank of an organism will not be too high (65,66).

Another mini-F vector has been constructed from a replication

mutant with a Cop_{ts} phenotype. In vector pJE253 copy number can be modulated by changing the growth temperature of the strain. At 28°C pJE253 is present in 60-80 copies, while 42°C reduces the copy number to 5-7 copies per cell. In vector pJE253 the mini-F genome is reduced to the region 44.1-47.3F. Although pJE253 does not contain partitioning function incD , it is stably inherited at all temperatures, due to its higher copy number. The construction of pJE253 followed the same strategies as used for pRE435. A map of vector pJE253 is shown in Figure 8.

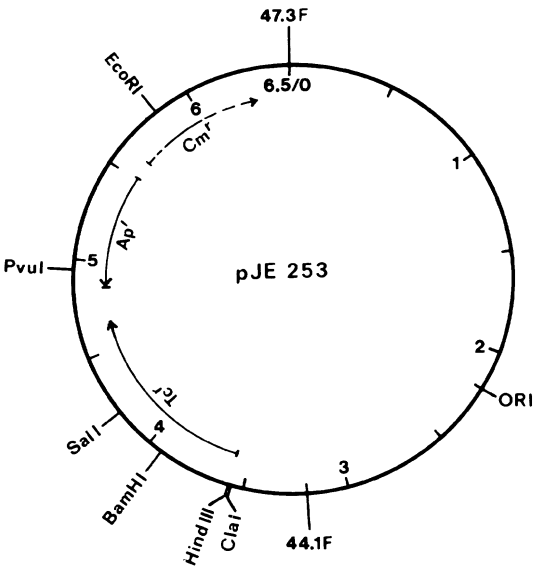


Figure 8. Map of vector pJE253 with temperature-dependent copy number (cop_{ts}). The right part of the molecule between coordinates 47.3-44.1F represent sequences derived from F. The antibiotic resistances are from plasmid pBR325 (15). The ampicillin resistance(Ap´) is not functional. The plasmid has a size of 6.5 kb. The single restriction sites which can be used for cloning are indicated.

DISCUSSION

The study of the replication of large plasmid genomes has been facilitated by the construction of mini-plasmids, exhibiting identical replication properties as the parental plasmids. Establishment of a detailed map of recognition sites for restriction endonucleases allows further diminution of the replicon to the essential functions for replication - the basic replicon. The origin of replication is identified by the isolation of replicative intermediates and their analysis in the electron microscope. In some cases more than one origin of replication may be found and the direction of replication may be either unidirectional or bidirectional. The isolation of plasmid mutants and the mapping of genes for polypeptides helps further to identify functions involved in replication, copy number control or functions naturallly carried by the plasmid or those cloned into it. Generation of random mutations by treating the plasmid genome with a mutagen is possible as well as using currently developed methods for site-directed mutagenesis (67). Proteins can be analysed in a cell-free system or in mini-cells. The cell-free system may be more useful for those proteins which are normally repressed for the most part of the cell cycle.

Although plasmid encoded proteins have been identified to be required for replication, the plasmid is also using most of the host functions involved in chromosome replication. The basic replicon confined usually to a 1-2 kb segment of DNA contains an origin of replication and structural genes for a positive function rep and in some cases another gene for negative control cop, regulating the copy number (for reference see 68). In mini-F only a positively acting function exerted by the E protein has been identified. It seems that this protein also regulates the copy number, since cop mutations have been mapped within the structural gene of the E protein. A negative control function acting as an inhibitor of replication has not been found so far in F.

The repeated sequences in the vicinity of the origin II, expressing incompatibility versus another F plasmid may function as binding sites for the E protein. A similar case has been observed in bacteriophage λ where four repeats in the origin re-

gion bind the O protein which is required for initiation of replication (69).

Separated from the origin region are a second group of gene loci involved in the partitioning of the plasmid. While in eukaryotes a complex mechanism has evolved for the distribution of the genetic material in mitosis and meiosis, bacteria with their small haploid genome seem to have a much simpler mechanism to ensure equal partitioning of the chromosomes during cell division. Nevertheless, almost nothing is known so far on the bacterial partitioning mechanism. In F a *cis*-acting function *incD* and two proteins of 44K and 36K (proteins A and B) acting in *trans* may be involved. Also required are at least two proteins supplied by the host and a membrane attachment site (58).

Further research has to show how these proteins interact with the DNA and membrane sites to achieve equal partitioning. Also, *in vitro* replication of plasmid DNA with purified polypeptides will eventually answer the questions on the exact functions of the proteins identified to play a role in this process.

REFERENCES

1. Konisky, J. *In*: The bacteria (Ed. L.C. Gunsalus) Vol. VI Academic Press, New York, 1978, pp. 71-136.
2. Willets, N. and Skurray, R. Ann. Rev. Genet. 14:41-76, 1980.
3. Falkow, S. Infectious multiple drug resistance. Pion Limited, London, 1975.
4. Cohen, S.N. and Chang, A.C.Y. Proc. Nat. Acad. Sci. USA 70:1293-1297, 1973.
5. Cohen, S.N., Chang, A.C.Y. and Hsu, L. Proc. Nat. Acad. Sci. USA 69:2110-2114, 1972.
6. Cohen, S.N. and Chang, A.C.Y. Mol. Gen. Genet. 134:133-141, 1974.
7. Staudenbauer, W.L. Curr. Top. Microbiol. Immunol. 83:93-156, 1978.
8. Collins, J. Curr. Top. Microbiol. Immunol. 78:121-170, 1977.
9. Hershfield, V., Boyer, H.W., Yanofsky, C., Lovett, M.A. and Helinski, D.R. Proc. Nat. Acad. Sci. USA 71:3455-3459, 1974.
10. Clewell, D.B. J. Bacteriol. 110:667-676, 1972.
11. Hershfield, V., Boyer, H.W., Chow, L. and Helinski, D.R. J. Bacteriol. 126:447-453, 1976.
12. Covey, C., Richardson, D. and Carbon, J. Mol. Gen. Genet. 145:155-158, 1976.
13. Rodriguez, R.L., Bolivar, F., Goodman, H.W., Boyer, H.W. and Betlach, M. *In*: Molecular Mechanisms in the Control of Gene Expression (Eds. D.P. Nierlich, W.J. Rutter and C.F. Fox) Academic Press, New York, 1976, pp. 471-478.

14. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.B. Crosa, J.H. and Falkow, S. Gene 2:95-113, 1977.
15. Bolivar, F. Gene 4:121-136, 1978.
16. Timmis, K.N. In: Genetics as a tool in Microbiology (Eds. S.W. Glover and D.A. Hopwood) Cambridge University Press, Cambridge, 1981, pp. 49-109.
17. Methods in Enzymology, Vol. 68 (Ed. R. Wu) Academic Press, New York, 1979.
18. Lane, H.E.D. Plasmid 5:100-126, 1981.
19. Timmis, K., Cabello, F. and Cohen, S.N. Proc. Nat. Acad. Sci. USA 72:2242-2246, 1975.
20. Lovett, M.A. and Helinski, D.R. J. Bacteriol. 127:982-989, 1976.
21. Skurray, R.A., Guyer, M.S., Timmis, K.N., Cabello, F., Cohen, S.N., Davidson, N. and Clark, A.J. J. Bacteriol. 127:1571-1575, 1976.
22. Guyer, M.S., Figurski, D. and Davidson, N. J. Bacteriol. 127:988-997, 1976.
23. Palchandhuri, S. and Maas, W.K. Proc. Nat. Acad. Sci. USA 74:1190-1194, 1977.
24. Maas, R. Proc. Nat. Acad. Sci. USA 50:1051-1055, 1963.
25. Datta, N. In: Plasmids of Medical, Environmental and Commercial Importance (Eds. K.N. Timmis and A. Pühler) Elsevier/North-Holland, Biomedical Press, Amsterdam, 1979, pp. 3-12.
26. Andres, J., Slocombe, P.M., Cabello, F., Timmis, J.K., Lurz, R., Burkhardt, H.-J. and Timmis, K.N. Mol. Gen. Genet. 168:1-25, 1979.
27. Kollek, R., Oertel, W. and Goebel, W. Mol. Gen. Genet. 162:51-58, 1978.
28. Crosa, J.H., Luttropp, L.K. and Falkow, S. Cold Spring Harb. Symp. Quant. Biol. 43:111-120, 1978.
29. Kolter, R. and Helinski, D.R. Plasmid 1:571-580, 1978.
30. Kahn, M.L., Figurski, D., Ito, L. and Helinski, D.R. Cold Spring Harb. Symp. Quant. Biol. 43:99-103, 1979.
31. Wehlmann, H. and Eichenlaub, R. Molec. Gen. Genet. 180:205-211, 1980.
32. Figurski, D., Kolter, R., Meyer, R., Kahn, M., Eichenlaub, R. and Helinski, D.R. In: Microbiology - 1978 (Ed. D. Schlessinger) Am. Soc. Microbiol., Washington D.C., 1978, pp. 105-109.
33. Manis, J.J. and Kline, B.C. Mol. Gen. Genet. 152:175-182, 1977.
34. Lovett, M.A., Sparks, R.B. and Helinski, D.R. Proc. Nat. Acad. Sci. USA 72:2905-2909, 1975.
35. Cabello, F., Timmis, K. and Cohen, S.N. Nature 259:285-290, 1976.
36. Lovett, M.A., Katz, L. and Helinski, D.R. Nature 251:337-340, 1974.
37. Eichenlaub, R., Sparks, R.B. and Helinski, D.R. J. Bacteriol. 138:257-260, 1979.
38. Sparks, R.B. and Helinski, D.R. Nature 277:572-575, 1979.
39. Guerry, P., Le Blanc, D.J. and Falkow, S. J. Bacteriol. 116:1064-1066, 1973.
40. Hirt, B. J. Mol. Biol. 26:365-369, 1967.

41. Eichenlaub, R., Figurski, D. and Helinski, D.R. Proc. Nat. Acad. Sci. USA 74:1138-1141, 1977.
42. Eichenlaub, R., Wehlmann, H. and Ebbers, J. In: Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids (Eds. S. Levy, E. Koenig and R. Clowes) Plenum Press, New York, 1981, pp. 327-336.
43. Zubay, G. Ann. Rev. Genet. 7:267-287, 1973.
44. Frazer, A.C. and Cortiss III, R. Curr. Top. Microbiol. Immunol. 69:1-84, 1975.
45. Eichenlaub, R. J. Bacteriol. 120:1476-1477, 1974.
46. Yang, H.-L., Zubay, G. and Levy, S.B. Proc. Nat. Acad. Sci. USA 73:1509-1512, 1976.
47. Kennedy, N., Bentin, L., Achtman, M., Skurray, R., Rahmsdorf, U. and Herrlich, P. Nature 270:580-585, 1977.
48. Eichenlaub, R. and Wehlmann, H. Mol. Gen. Genet. 180:201-204, 1980.
49. Komai, N., Nishizawa, T., Hayakawa , Y., Murotsu, T. and Matsubara, K. Mol. Gen. Genet. 186:193-203, 1982.
50. Brack, C. Proc. Nat. Acad. Sci. USA 76:3164-3168, 1979.
51. Wehlmann, H. and Eichenlaub, R. Plasmid 5:259-266, 1981.
52. Murotsu, T., Matsubara, K., Sugisaki, H. and Takanami, M. M. Gene 15:257-271, 1981.
53. Tolun, A. and Helinski, D.R. Cell 24:687-694, 1981.
54. Tolun, A. and Helinski, D.R. Mol. Gen. Genet. 186:372-377, 1982.
55. Kline, B.C. and Palchandhuri, S. Plasmid 4:281-291, 1980.
56. Kline, B.C. and Lane, D. Plasmid 4:231-232, 1980.
57. Gardner, R.C., Malcolm, L., Bergquist, P.L. and Lane, H.E.D. Mol. Gen. Genet. 188, 345-352, 1982.
58. Ogura, T. and Hiraga, S. Cell 32:351-360, 1983.
59. Ogura, T., Miki, T. and Hiraga S. Proc. Nat. Acad. Sci. USA 77:3993-3997, 1980.
60. Jacob, F., Brenner, S. and Cuzin, F. Cold Spring Harb. Symp. Quant. Biol. 28:329-348, 1963.
61. Cuzin, F. and Jacob, F. Ann. Inst. Pasteur, Paris, 112:398-418, 1967.
62. Humphreys, G.O., Willshaw, G.A., Smith, H.R. and Anderson, E.S. Mol. Gen. Genet. 145:101-108, 1976.
63. Freese, E., Bautz, E. and Bautz-Freese, E. Proc. Nat. Acad. Sci. USA 47:845-855, 1961.
64. Eichenlaub, R. J. Bacteriol. 138:559-566, 1979.
65. Collins, J. and Brünning, H.J. Gene 4:85-107, 1978.
66. Hohn, B. and Murray, K. Proc. Nat. Acad. Sci. USA 74:3259-3265, 1977.
67. Shortle, D., DiMaio, D. and Nathans, D. Ann. Rev. Genet. 15:265-294, 1981.
68. Nordström, K. Plasmid 9:1-7, 1983.
69. Tsurimoto, T. and Matsubara, K. In: "ICN-UCLA Symp. on Mol. and Cellular Biol." Vol. 22, 1981, pp. 263-275.
70. Ebbers, J. and Eichenlaub, R. J. Bacteriol. 147:736-743, 1981.

4

CLONED DNA AS A SUBSTRATE OF BACTERIAL RECOMBINATION SYSTEM

AMIKAM COHEN
Department of Molecular Genetics, The Hebrew University-Hadassah Medical School, Jerusalem, Israel

INTRODUCTION

Once a DNA fragment is cloned in a bacterial plasmid, it becomes an object for all the bacterial systems that control plasmid replication, maintenance, gene expression and recombination. While the first systems ensure faithful amplification of the cloned DNA, and if applicable, its expression in the host bacteria, bacterial recombination systems may lead to duplications, deletions and rearrangements of cloned DNA fragments. Conservation of the nucleotide sequence during cloning procedure and during propagation of the chimera plasmids is an absolute requirement for any investigation of the structure and function of the cloned fragment at its place of origin. Measures must therefore be taken to minimize deletions and rearrangements of cloned DNA in the host bacteria. The nature of the elements which affect the integrity of cloned DNA is the scope of this manuscript.

In several studies in which deletions in cloned DNA were observed, the nature of the reaction and the role of some host functions which were involved were determined. Some systems were further developed, to be used as convenient probes for the investigation of bacterial recombination systems. These studies have led to characterization of the substrates and the products of the recombination reaction and to the determination of bacterial metabolic activities involved in recombination pathways that affect plasmids (Table 1).

Recombination provides the mechanism for deletions and rearrangement of cloned DNA. However, a newly formed recombination product must replicate and segregate in an environment consisting of other plasmids -

Table 1. Deletions in cloned DNA fragments

Cloned DNA	Cloning vehicle	Proposed mechanism(s)	Dependency on host functions		Reference
			recA	recB	
Adenovirus DNA	λ	Intraplasmidic recombination	ND[a]	Independent	(1)
Retrovirus-like repeated mouse gene family	λ	Intramolecular recombination	ND	ND	(2)
Murine sarcoma virus DNA	λ	Intramolecular recombination or intermolecular unequal crossover	Independent	ND	(3)
Highly repeated satellite Drosophila DNA	pSC101, pCR1	Unequal intermolecular recombination, intramolecular recombination, recombination between daughter DNA helices in replicating fork	Independent	Independent	(4)
Balbiani ring DNA	pBR322, pHC79 cosmid	Specific sequence elimination	Independent	Independent	(5)
Yeast ribosomal DNA	pSC101, pMB9	Intramolecular recombination	Partially dependent	Independent	(6-8)
Palindromic DNA (Tn5 derivatives)	pBR322, pBR325	"Snapback" during replication	Independent	Independent	(9)
Staphylococcus plasmid pC194	pBR322	Intramolecular recombination between short repeats or slipped mispairing	Partially dependent	Independent	(10,11)
Repeated *tet* gene fragment	pBR322	Intramolecular recombination	Partially dependent	Independent	(12,13)
E. coli lac operon DNA	pBR322	Slipped mispairing	Partially dependent	ND	(14,15)

[a] Not done.

the recombination substrates - which belong to the same plasmid incompatibility group. The survival of the recombination product is therefore affected also by selective pressure, copy number control and incompatibility. The influence of these factors on the proportion of the recombination product in plasmid population is discussed.

PLASMID AS RECOMBINATION SUBSTRATE

Two distinct classes of recombination systems function in the bacterial cell: a) the general recombination system, which induces recombination between homologous DNA sequences and depends, in most cases, on a functional recA gene product (for review see 16-20) and b) site-specific recombination systems which depend on nucleotide sequences in transposable elements and on functions which are coded by these elements (for review see 21, 22). Since transposable elements can be inserted into any DNA fragment, all plasmids may be considered as substrates for site-specific recombination. In fact, transposable elements have been used for in vivo insertion mutagenesis of cloned DNA fragment (23, 24).

One class of transposable elements, the insertion sequences (IS), is represented in all Escherichia coli cells (21, 22). Therefore, IS elements must be considered potential rearrangement agents affecting plasmids propagating in bacteria of any genetic background. These elements can mutate cloned DNA sequences by insertion. In addition, insertion sequences induce deletions, inversions and transpositions in a region close to the site of insertion (22).

Recombination events which do not involve transposable elements depend on the activity of bacterial gene products and on plasmid DNA sequences. Thus, in bacteria with a functional recombination system (rec^+), repetitive DNA sequences are substrate to intraplasmidic recombination. This process may lead to deletions, in case of direct repeats, or to inversions, in case of inverted repeats. Analysis of plasmid sequences which are "hot spots" per deletion activity indicates that most recombination events occur between repetitive sequences (10, 14, 15). The length of the repeats where recombinations were observed varied between several

hundreds (1-3, 6-8, 12, 13) to as little as five nucleotides (10, 14, 15). The occurrence of intraplasmidic recombination between very short repeats, makes a large number of cloned DNA fragments potential substrates for this reaction, since the probability of a sequence of five nucleotides to be present twice in a random sequence of one thousand nucleotides is about 60% and most cloned fragments are longer than that.

The dependency of the intraplasmidic recombination reaction on repetitive sequences has been demonstrated in several systems, including plasmids carrying yeast ribosomal DNA sequences. Yeast ribosomal DNA is arranged as a long chain of tandem repeats (25). When a fragment the length of one, or less than one, repeat is cloned, little or no deletion activity is observed. On the other hand, cloning of a fragment longer than one repeat, leads to the formation of a cloned repetitive sequence, and subsequently to intraplasmidic recombination and deletion of one repat (7). Extensive recombination activity is observed also in clones carrying highly repeated Drosophila satellite DNA (4). In this system recombination activity leads to the loss of about 90% of the cloned DNA and the formation of a plasmid population which is not homogenous in size.

Repetitive nucleotide sequences are present on linear viral and pro-viral DNA molecules. These repeats make cloned viral DNA sequences a substrate for intramolecular recombination events, leading to deletions and rearrangement. Deletion products have been characterized for coliphage λ-murine sarcoma virus recombinants (3) and retrovirus-like repeated mouse gene family (2). In both systems, a fragment consisting of one of the repetitive sequences and the region between the repeats was deleted. This observation is consistent with the occurrence of an intraplasmidic recombination event of the type presented in Fig. 1.

The length of the repeated sequence and the degree of homology determine the frequency of recombination, while repeats of several hundred nucleotides lead to recombination frequency of about 10^{-4} (12, 13), repeats of 5-7 nucleotides lead to recombination frequency of 10^{-7} to 10^{-8} deletions per cell (10, 14). Induction of mutations at the

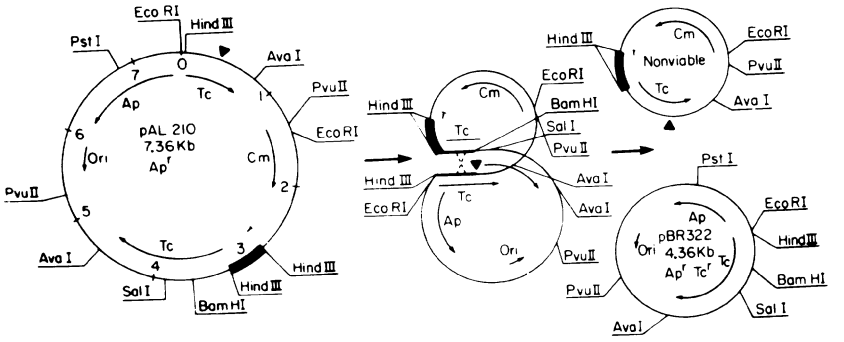


FIGURE 1. Regeneration of a functional tet gene by intraplasmidic recombination of two mutated tet genes. A functional tet gene is generated by intraplasmidic recombination in pAL210(B), if the crossing-over site is between the mutations on the tet genes. The proposed structure of the recombination product is presented. Heavy line indicates DNA insertion and the sequence between mutations. ▲ site of deletion.

repeated sequence allows the influence of the degree of homology on recombination frequency to be determined. When homology is reduced from 9 out of 10 bp to 8 out of 10 bp, reduction of the frequency of recombination by one order of magnitude is observed. On the other hand, base changes which do not affect the degree of homology have little or no effect on recombination frequency (14)

Intramolecular recombination between inverted repeats should lead to an inversion of the DNA fragment between the point of crossover. Nevertheless, large palindromes, which are highly unstable in recA and recA$^+$ genetic background, yield, in most cases, deletants, and the deletion occurs around the palindromic axis of symmetry (9). The independence of this deletion process of the recA function led to the proposal of a model which does not involve the general recombination system. According to this model, a snapback of the palindromic region, which occurs during replication, could bring the repeats into a juxtaposition and subsequently leads to a "slippage" deletion (9) The independence of the deletion process of the recA gene function does not exclude the possibility that

other functions, which are part of the general recombination pathway, are involved. The center of the cruciform structure, which may be formed by palindromic sequences is indistinguishable from the center of the "Holliday junction" which is a key intermediate in the general recombination pathway (17-19). This structure has been shown to be a substrate for specific endonucleases that resolve the Holliday junction into recombination products (26). In fact, palindromic sequences are being used as substrates for this endonuclease (27).

This heteroduplex joint is a key intermediate in the process of general recombination (17-19). This structure is synthesized in vitro, by the recA protein, through a process which involves synopsis and polar branch migration (20). The effect of bracketing a short genetic interval by nonhomologous sequences on the migration of branch points into the interval has been recently investigated in an interplasmidic recombination system (Laban and Cohen, in press). Results indicate that when recombination occurs within a region which is bracketed by major nonhomologous sequences, branch migration from outside the region and mismatch repair is not involved, and recombination proceeds as presented in Fig. 1. This process leads to dimer formation when interplasmidic recombination takes place, and to deletion of the sequence between the recombination sites when intraplasmidic recombination occurs. Branch migration and mismatched repair processes are involved in recombination, when the region where recombination occurs is not bracketed by major nonhomologous sequences or when it is bounded by very short nonhomologous sequences (4-24 bp). In this case, recombination products will differ from the ones presented in Fig. 1. They may include monomers or dimers in the case of interplasmidic recombination and in the case of intraplasmidic recombination not all recombination events will lead to deletions.

THE INVOLVEMENT OF HOST GENE FUNCTIONS IN PLASMIDIC RECOMBINATION

In order to determine the degree of dependency of plasmidic recombination on host gene functions, plasmids which facilitate direct and con-

venient analysis of intraplasmidic and interplasmidic recombination proficiencies in bacteria of various genotypes were constructed (12, 13). For analysis of intraplasmidic recombination, two copies of the tetracycline resistance (tet) gene, each of which mutated at a different site, were inserted in tandem duplication into one plasmid. In this system, an intraplasmidic recombination event, at a site between the mutations, leads to deletion and reconstruction of a functional tet gene (Fig. 1). Using a similar approach, interplasmidic recombination proficiencies are determined in cells harboring two compatible plasmids, each one carrying a mutation at a different site of the tet gene. In both systems, recombination proficiency is expressed as the ratio of tetracycline resistant (Tc^r) to tetracycline sensitive (Tc^s) cells in the culture (13).

Table 2 represents the relative recombination proficiency values for intraplasmidic recombination for E. coli K-12 cells of various genotypes. Relative values of postconjugational recombination is presented as a reference. Like postconjugational recombination and other recombination systems, intraplasmidic recombination proceeds via alternative pathways. However, intraplasmidic recombination differs from postconjugational recombination by the degree of its dependency on host functions. RecA functional gene product is an absolute requirement for postconjugational recombination. Mutations in the recA gene lower the frequency of transfer of genetic markers from Hfr to F^- cells by five orders of magnitude and the residual activity does not represent recombination (16). On the other hand, the residual plasmidic recombination activity in recA mutants is 1-2 percent of that in $recA^+$ cells and plasmidic recombination products in recA and $recA^+$ cells are indistinguishable from each other (13).

Postconjugational recombination depends on a functional recBrecC gene product - exonuclease V. Mutations in these genes lower the postconjugational recombination proficiency by about hundredfold (16). On the other hand, plasmidic recombination is independent of exonuclease V activity and recBrecC mutations do not affect its proficiency. The difference between plasmidic recombination and postconjugational recombination with

Table 2. Relative recombination proficiencies in intraplasmidic and postconjugational recombination system

Bacterial relevant genotype (representative strain)	Relative recombination proficiencies	
	Intraplasmidic[a]	Postconjugational[b]
rec^+ (AB1157)	1.0	1.0
recA (JC2926)	1.0×10^{-2}	10^{-5}
recB (JC5519)	1.0	10^{-2}
recF (JC9239)	2.6×10^{-2}	1.0
recFrecB (JC3881)	1.0	10^{-4}
recBsbcB (JC2623)	ND[c]	0.5
sbcB (JC1145)	1.0	0.5
recBsbcA (JC5183)	32	0.5
recBsbcArecA (DR107)	33	10^{-5}
recBsbcArecF (JC8691)	0.9	10^{-2}

[a] Relative intraplasmidic recombination proficiencies were calculated from references (12, 13, 28, 29).

[b] Relative postconjugational recombination proficiencies were calculated from references (16, 30, 31, 32).

[c] Plasmidic recombination proficiencies cannot be determined for recBsbcB strains due to poor maintenance of plasmids in this genetic background.

respect to the degree of its dependency on exonuclease V activity may be due to structural differences between the substrates of recombination, or, as in the case of phage λ recombination (33), to the absence of chi sequences (34) from the substrate. In order to decide between the two alternatives, chi sequences were introduced by point mutation or insertion into the tet gene of pBR322 and the effect of the presence of these sequences on the recombination proficiency was determined in $recB^+$ and recB cells. Results indicate that the introduction of chi sequences into the plasmid substrate has no effect on recombination in recB or $recB^+$ cells (A. Laban, unpublished data). This finding is consistent with the

notion that the recBrecC gene product does not participate in plasmidic recombination.

Postconjugational recombination in wild type E. coli is recF independent; the involvement of recF gene function in postconjugational recombination is apparent only after inactivation of exonuclease V and exonuclease I by recBrecC and sbcB mutations (30). On the other hand, plasmidic recombination depends on recF gene activity in wild-type cells. recF mutations lower both interplasmidic and intraplasmidic recombination to a level close to that observed in recA mutants (28, 29).

Surprisingly, the requirement for recF activity in intraplasmidic recombination is alleviated by recBrecC mutations. This may be due to the functioning of two intraplasmidic recombination pathways, one recF-dependent exonuclease V resistant and the other recF-independent exonuclease V sensitive. Alternatively this observation may be due to an inhibitory effect of exonuclease V and protective activity of recF products.

In the postconjugational recombination system, sbcA mutations supress recBrecC mutations by activating the RecE recombination pathway (30). Thus, recBrecCsbcA mutants have postconjugational recombination activity similar to that of wild-type cells. Activation of the recF pathway by sbcA mutation leads to a dramatic increase in plasmidic recombination proficiency (12, 13). Even more striking is the observation that intraplasmidic recombination via RecE recombination pathway is recA independent (12, 13). Similar proficiency values are obtained in recArecBrecCsbcA as in recBrecCsbcA strains.

recA and recB gene functions that are independent of intraplasmidic recombination have also been observed in systems in which very short repeats were involved. recA-independent deletion activity has been demonstrated in all systems tested (10, 11, 14, 15), but in systems where activity could be determined quantitatively, the presence of a functional recA gene product led to a 10-12 fold increase in recombination proficiency (10, 14). Like intraplasmidic recombination between long repeats, intraplasmidic recombination between short repeats is recB independent (10, 11).

It has not been determined whether the bacterial systems which are functioning in intraplasmidic recombination between long repeats are also involved in deletions between short repeats. If the same systems are involved in both reactions, one would expect that deletions between short repeats would be partially dependent on recF activity. It should also proceed at high efficiency and be recA independent in the recBsbcA genetic background. A model which relates deletion between short repeats to "slipped mispairing" during replication, has been recently proposed (14).

FACTORS WHICH DETERMINE THE PROPORTION OF DELETION PRODUCTS IN PLASMID POPULATION

The frequency of intraplasmidic recombination ranges between 10^{-7} events per cell generation for repeats of less than 10 bp in the recA genetic background to 10^{-2} events per cell generation for repeats of several hundred nucleotides in recBrecCsbcA cells. Nevertheless, it is quite common to observe with certain DNA fragments, that following transformation, a major part of plasmid population consists of deletion products. The recombination process provides the mechanism which induces variability in the plasmid population and subsequently in the bacterial population. However, selective pressure which may either favor the recombination product during plasmid replication and maintenance, or discriminate against cells which harbor plasmids carrying DNA sequences deleterious to the cell, will determine the proportion of deletants in the plasmid population of the culture.

In some cases, the reason for the selective pressure resides in the product coded for by the deleted fragment. Even certain bacterial products which are essential for the cell when present in minute amounts become deleterious to the cell when their concentration increases due to amplification of their respective genes. pACYC184 has a copy number of about 20 (35) and its derivatives carrying the lacY (lactose permease) gene are stable under noninductive growth conditions. On the other hand, derivatives of pBR322 with a copy number of 50-70 (36) carrying the same lacY gene fragment, are unstable under all growth conditions and tend to undergo

deletions of various lengths (37, 38). Induction of lacY gene expression in cells harboring pACYC184 derivatives carrying this gene, leads to the enrichment of mutants, some of which carry chromosomal mutations which interfere with the synthesis of lactose permease, its transport to the membrane or its function (37).

Selective pressure against a nucleotide sequence is not necessarily a result of its expression as a protein product. Some sequences may have deleterious effects on plasmid replication and maintenance regardless of their coding capacity. Long palindromes have lethal properties for plasmids; such sequences may release free energy for supercoiling into the cruciform structure affecting plasmid transcription and replication (39). While inverted repeats of more than 100 nucleotides have a lethal effect and cannot be cloned, repeats of less than 13 nucleotides are stable (9, 39). The selective pressure against plasmids having inverted repeats of more than 13 nucleotides may lead to enrichment of deletion products. Since the stability of the cruciform structure is directly proportional to the stem length, and inversely proportional to the loop size, the integrity of cloned inverted repeats may be a function of the length of the repeats and the distance between them.

Inverted repeats may be part of the genomic (40) or viral gene bank (41). Therefore, one should consider the integrity of plasmids carrying inverted repeats when constructing such gene banks or when cloning DNA fragments with palindromic structures.

In most cases where rapid deletion processes take place, the reason for the selective advantage of the deletion product is not clear. Such is the case of plasmids carrying yeast ribosomal DNA. As stated above, plasmids having an insert of yeast rDNA, which is longer than one rRNA gene, contain a tandem duplication. An intraplasmidic recombination event between the repeats leads to a deletion of one copy of the rRNA gene. Deletion products are apparent in cultures of rec$^+$ cells and to a lesser extent in recA cultures (7). For an unknown reason, there is a strong selective pressure against the parent plasmid in minicell-producing strains. This pressure leads to a rapid displacement of parent plasmids

Table 3. Instability of cloned yeast rDNA with direct repeats in *E. coli* K-12

Strains	Relevant genotype	Number of clones tested	Clones with nondetectable deletion products	Clones with only deletion products	Percent of deletion product in plasmid population	Intraplasmidic recombination proficiency Tc^r/Ap^r cells[a]
DR104	*minA*$^+$ *minB* *recA*$^+$	16	0	0	18.5	2×10^{-4}
DR105	*minA*$^+$ *minB* *recA*$^-$	12	75	0	4.7	2×10^{-6}
DS410	*minA* *minB* *recA*$^+$	14	0	50	98.9	2×10^{-4}
DS498	*minA* *minB* *recA*$^-$	33	0	66	87.7	2×10^{-6}

[a] Intraplasmidic recombination proficiency was determined in the respective host bacteria by determining the proportion of tetracycline-resistant clones in bacterial population harbouring pA210 (13).

regardless of the host cell *recA* genotype. Apparently, this selective pressure is associated with the *minB* mutation. Following transformation of *minB* mutants with yeast rDNA plasmids, carrying an rDNA fragment longer than one repeat, rapid displacement of the parent plasmid by its deletion product is observed. This process is observed in *minB* mutants having *recA* or *rec*$^+$ genetic background (Table 3). Since intraplasmidic recombination proficiency is not affected by the *minB* mutation, and the *recA* background has a minor effect on the process, one may conclude that the reason for the rapid increase in the proportion of deletion products resides in the selective pressure against the parent plasmid or against cells harbouring this plasmid.

The presence of repetitive sequences in cloned viral DNA, makes it a substrate for intraplasmidic recombination. Other features of viral or proviral DNA, such as palindromic structures, may lead to selective pressure, favouring the deletion products. Rapid displacement of cloned viral sequences by deletion products have been documented in several systems, including Moloney murine sarcoma virus (3) and retrovirus-like repeated mouse gene family (2). Thus, while in most cases a cloned viral DNA fragment represents faithful amplification of viral DNA sequences,

some cloned fragments may represent an enrichment of viral DNA rearrangement products.

REFERENCES

1. Perricaudet, M., Fritsch, A., Pettersson, U., Philipson, L. and Tiollais, P. Science 196:208-210, 1977.
2. Keshet, E. and Shaul, Y. Nature 289:83-85, 1981.
3. McClements, W.L., Enquist, L.W., Oskarsson, M., Sullivan, M. and Vande Woude, G.F. J. Virol. 35:488-497, 1980.
4. Brutlag, D., Fry, K., Nelson, T. and Hang, P. Cell 10:509-519, 1977.
5. Case, S.T. Gene 20:169-176, 1982.
6. Cohen, A., Ram, D., Paiss, A. and Friedman, A. Contr. Microb. Immunol. 6:146-158, 1979.
7. Cohen, A., Ram, D., Halvorson, H.D. and Wensink, P.C. Gene 3:135-147, 1978.
8. Cramer, J.H., Farrelly, F.W., Barnitz, J.T. and Rownd, R.H., Mol. Gen. Genet. 148:233-241, 1976.
9. Collins, J., Cold Spring Harbor Symp. Quant. Biol. 45:409-416, 1980.
10. Jones, I.M., Primrose, S.B. and Ehrlich, D. Mol. Gen. Genet. 188: 486-489, 1982.
11. Primrose, S.B. and Ehrlich, S.D. Plasmid 6:193-201, 1981.
12. Fishel, R.A., James, A.A., Kolodner, R. Nature 294:184-186, 1981.
13. Laban, A. and Cohen, A. Mol. Gen. Genet. 184:200-207, 1981.
14. Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J. Cell 29: 319-328, 1982.
15. Farabaugh, P.J., Schmeissner, U., Hofer, M. and Miller, J. J. Mol. Biol. 126:847-857, 1978.
16. Clark, A.J. Annu. Rev. Genet. 7:67-86, 1973.
17. Holliday, R. Genetics 78:273-287, 1974.
18. Dressler, D. and Potter, H. Ann. Rev. Biochem. 51:727-761, 1982.
19. Radding, C.M. Ann. Rev. Genet. 16:405-437, 1982.
20. Radding, C.M. Cell 25, 3-4, 1981.
21. Calos, M.P. and Miller, J.H. Cell 20:579-597, 1980.
22. Kleckner, N. Cell 11:11-23, 1977.
23. Ausubel, F.M. and Ruvkun, G.B. In: Microbiology - 1981 (Schlessinger, D. Ed.) Am. Soc. Microbiol. Washington, D.C. pp. 124-127.
24. Maas, K.W., Maas, R., Mazaitis, A.J. In: Microbiology - 1981 (Schlessinger, D. Ed.) Am. Soc. Microbiol. Washington D.C. pp.133-136.
25. Cramer, J.H., Bhargava, M.M. and Halvorson, H.G. J. Molec. Biol. 21:11-20, 1972.
26. Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R. Cell 29:357-365, 1982.
27. Kemper, B., Vepka, M.V., Jensch, F. and Fritz, H.J. J. Cellular Biochem. Supp. 7B, 71, 1983.
28. Cohen, A. and Laban, A. Mol. Gen. Genet. 189:471-474, 1983.
29. James, A.A., Morrison, P.T. and Kolodner, R., J. Mol. Biol. 160: 411-430, 1982.

30. Barbour, S.D., Nagaishi, H., Templin, A. and Clark, A.J., Proc. Natl. Acad. Sci. USA 67:128-135, 1970.
31. Horri, Z. and Clark, A.J. J. Mol. Biol. 80:327-344, 1973.
32. Gillen, J.R., Willis, D.K. and Clark, A.J. J. Bacteriol. 145:521-532, 1981.
33. Stahl, F.W. and Stahl, M.M. Genetics 86:715-725, 1977.
34. Smith, G.R., Schultz, D.W. and Craseman, J.M. Cell 19:785-793, 1980.
35. Chang, A.C.Y. and Cohen, S.N. J. Bacteriol. 134:1141-1156, 1978.
36. Johnson, A., Willets, N.S. Plasmid 4:292-304, 1980.
37. Padan, E., Arbel, T., Rimon, A., Bar-Shira, A. and Cohen, A. J. Biol. Chem. 258:5666-5673, 1983.
38. Teather, R.M., Müller-Hill, B., Abrutsch, U., Aichele, G. and Overath, P. Molec. Gen. Genet. 159:239-248, 1978.
39. Lilley, D.M.J. Nature 292:380-382, 1981.
40. Thomas, C.A., jr., Pyeritz, R.E., Wilson, D.A., Dancis, B.M. Lee, L.S., Bick, M.D., Huang, H.L. and Zimm, B.H. Cold Spring Harbor Symp. Quant. Biol. 38:353-370, 1974.
41. Spear, P.G. and Roizman, B. In: DNA tumor viruses. (Tooze, J. ed.) Cold Spring Harbor U.S.A. 1980, pp. 615-746.

5

UTILIZATION OF λ CONTROL ELEMENTS FOR GENE EXPRESSION STUDIES IN ESCHERICHIA COLI

AMOS B. OPPENHEIM, JAMAL MAHAJNA, SHOSHY ALTUVIA, SIMI KOBY, DINA TEFF, HILLA LOCKER-GILADI, HANA HYMAN AND ALIK HONIGMAN
Department of Molecular Genetics, The Hebrew University - Hadassah Medical School, Jerusalem, Israel

SUMMARY

A family of plasmids designed for studying gene expression in bacteria is described. The system employs λ regulatory elements. Transcription is regulated by the thermolabile cI repressor. Transcription termination signals can be introduced between the λ P_L promoter and several structural genes. This system permits qualitative and quantitative studies of termination signals.

INTRODUCTION

The level of synthesis of a gene product in E. coli is affected by the efficiency of two major processes: transcription and translation. Efficient expression of a specifically tailored gene depends on a complex number of factors. Of special interest are those signals situated at the beginning of the gene. Transcription is initiated at specific signals, promoters, that differ in efficiency (1). A specific ribosome binding site must be present to allow efficient and correct protein synthesis (2). Transcription or translation termination signals may

prevent the synthesis of a full size gene product. Additional factors such as rapid RNA cleavage, and degradation of foreign proteins in E. coli (3,4) can greatly influence the level of specific gene products.

Lambda (λ) phage has been used extensively for cloning purposes (5). Recently, several plasmid expression vectors that make use of λ control elements have been developed (6,7,8,9,10). In these vector plasmids the efficient P_L or P_R promoters direct the transcription of the cloned gene.

The expression of λ prophage early genes is under the negative control of the cI repressor. By binding to the operator sequence the repressor prevents transcription from the P_L and P_R promoters. The availability of a temperature-sensitive repressor mutation permits rapid and simple induction. The cI repressor can be provided either by cloning the cI gene or by a cryptic prophage deleted for a large part of its genome. The remaining functional prophage genes are limited to the cI and N genes. The N gene product acts as an antitermination factor by modifying RNA polymerase (11,12). The modification of RNA polymerase by the N gene product requires a nut site (N utilization site) located downstream of the promoter. A schematic representation of the system is presented in Fig. 1.

The plasmids described below can be used for the following purposes:

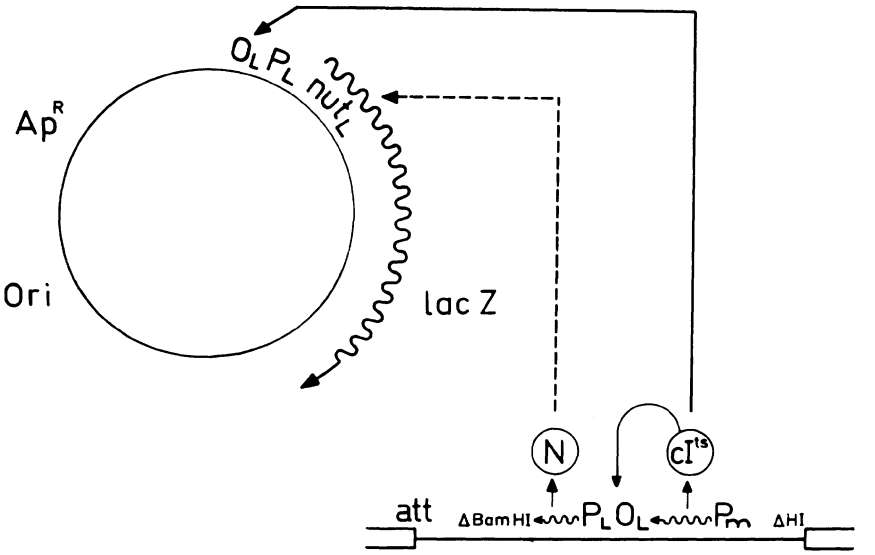


Fig. 1. Use of expression system. The cells carry a cryptic λ prophage. At low temperature, the cI repressor is active, repressing P_L of the prophage and that cloned on the plasmid. At 42°C repressor is inactivated allowing expression of P_L leading to synthesis of β - galactosidase and N gene product. The presence of the nut$_L$ site allows the antitermination factor to modify RNA polymerase.

1. Expression of foreign genes directed by the P_L promoter.
2. Controlled expression of genes that may be lethal to the E. coli host.
3. Selection of clones carrying transcription termination signals.
4. Studying the influence of the N gene product and host factors on transcription termination signals.

5. Assaying termination signals *in vivo* and *in vitro*.
6. Mutant isolation.

We have concentrated on using the *lacZ* gene coding for β-galactosidase, the *cat* gene, whose product, chloramphenicol acetyltransferase, confers chloramphenical resistance, and the *galK* gene coding for the galactokinase enzyme. The function of these genes can be monitored with great simplicity either as genetic markers affecting colony morphology or by quantitative enzymatic assays.

RESULTS AND DISCUSSION

A small λ DNA fragment of the early regulatory region of phage was cloned into pBR322 to yield pKC30N/2 (received from Dr. D. Court, National Cancer Institute, N.I.H., Fredrick, Maryland). The λ DNA fragment carries the regulatory signals O_L P_L nut_L and part of the *N* gene. The cloned fragment replaced the segment between the HindIII and BamHI restriction sites of pBR322 (13). Plasmids carrying strong promoters, such as the P_L promoter, are unstable. This plasmid, however, can be maintained in cells carrying an active λ cI repressor. The λ regulatory region of this plasmid was joined to the *lacZ* gene derived from plasmid pMLB1010 (14) (received from Dr. M. Berman, National Cancer Institute, N.I.H., Frederick, Maryland). Both plasmids were cut with

the enzymes PstI and BamHI and ligated together. In the recombinant plasmid (pOLZ21) the *bla* gene conferring resistance to ampicillin was reconstructed and the promoterless *lacZ* gene was placed under the control of the λ P_L promoter (Fig. 2). Transferring a culture of bacteria carrying this plasmid from 30°C to 43°C induced rapid synthesis of β-galactosidase (Fig. 3). These results were corroborated by analysis of protein samples on SDS polyacrylamide gel, where a protein band of β-galactosidase was observed (Fig. 4). This plasmid permits the introduction and study of transcription termination signals at a unique BamHI recognition site located between the promoter and the *lacZ* gene.

Introduction of the *cat* gene (promoterless gene carrying BamHI sites at both ends (15)) permits regulated transcription of both genes from the P_L promoter (plasmid pOLCZ14, Fig. 2). Cells carrying the above plasmid are sensitive to chloromphenicol at 30° but resistant at 38° (10 µg/ml of chloramphenicol). Even at low temperatures transcription from P_L is sufficient for the colonies to score as *lac*$^+$ on McConkey lactose plates. This property interferes with the direct selection of transcription termination signals.

We have introduced the *galK* gene by similar constructions (Fig.5). In the first stage the *cat* gene was introduced into pKG1800 (16). The EcoRI DNA fragment containing the *gal* promoter and part of the *cat* gene was

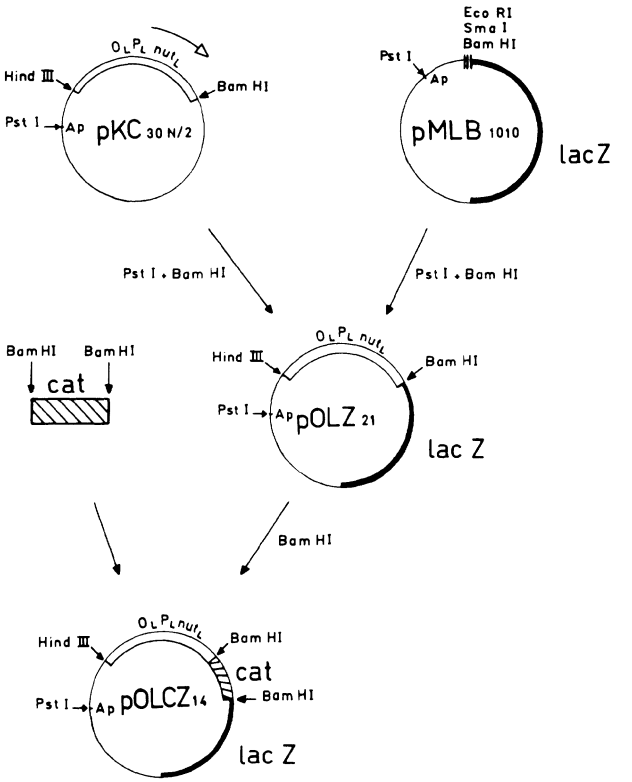


Fig. 2. Construction of plasmids pOLZ21 and pOLCZ14. The steps in the construction of pOLZ21 and pOLCZ14 are described in the text. The restriction sites used in the various steps are given.

replaced by the EcoRI fragment isolated from the plasmid pOLCZ14 carrying the λP_L promoter. The resulting plasmid,pOLCK18, conferred resistance to ampicillin. The plasmid pOLCK18R9, carrying the cI repressor, was constructed as described in Fig. 5. The cat and galK genes in the plasmids are heat inducible (Fig. 6). These plasmids were used for the identification and analysis of

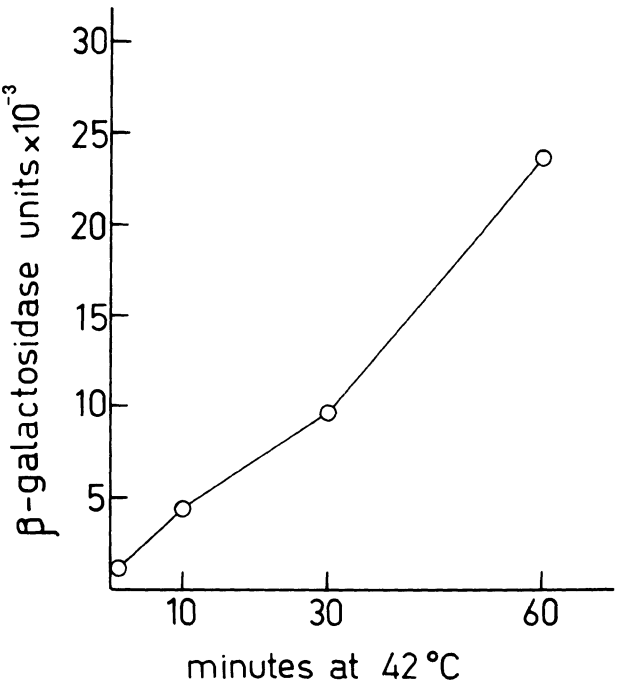


Fig. 3. Synthesis of β-galactosidase from P_L. Lysogens carrying the pOLZ21 plasmid were induced at 42°C and the β-galactosidase level was determined as described in this volume (18).

termination signals . Transcription terminators were introduced upstream or downstream from the cat gene (17). The translation termination signal at the end of the cat gene eliminates possible influence of the translating ribosomes on transcription termination signals that are introduced between the cat and gal genes. The three translation termination signals located upstream of the galK gene (16) prevent interference of translation from the inserted DNA with the expression of the galK gene.

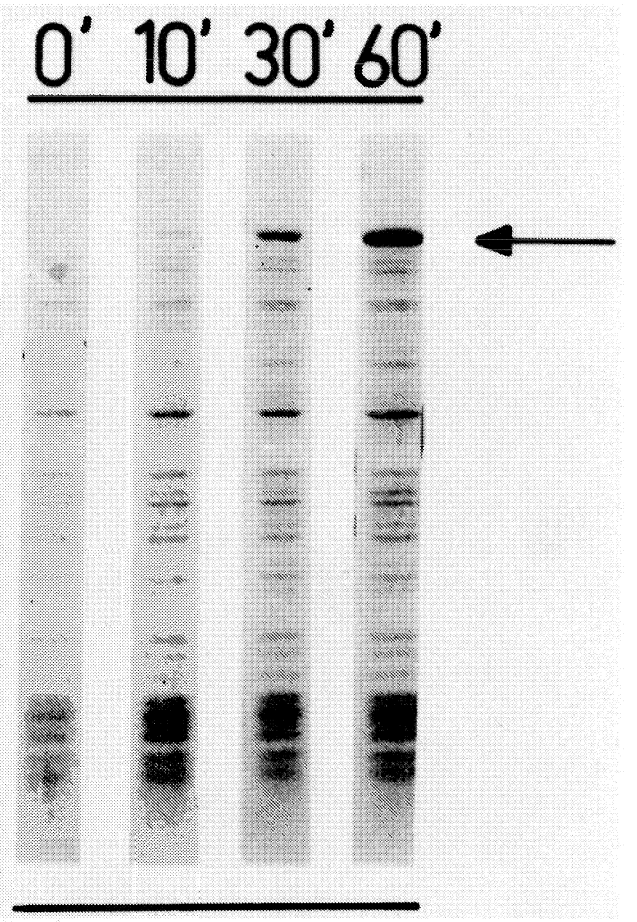


Fig. 4. Synthesis of β-galactosidase from pOLZ21. Cells containing pOLZ21 were grown at 32°C to OD_{650}=0.3 and then transferred to 43°C for induction. At time intervals, 0, 10, 30,and 60 min, 1 ml samples were centrifuged, boiled and loaded onto a 10 to 26% SDS polyacrylamide gradient gel which was stained with coomassie blue. The arrow denotes the β-galactosidase band.

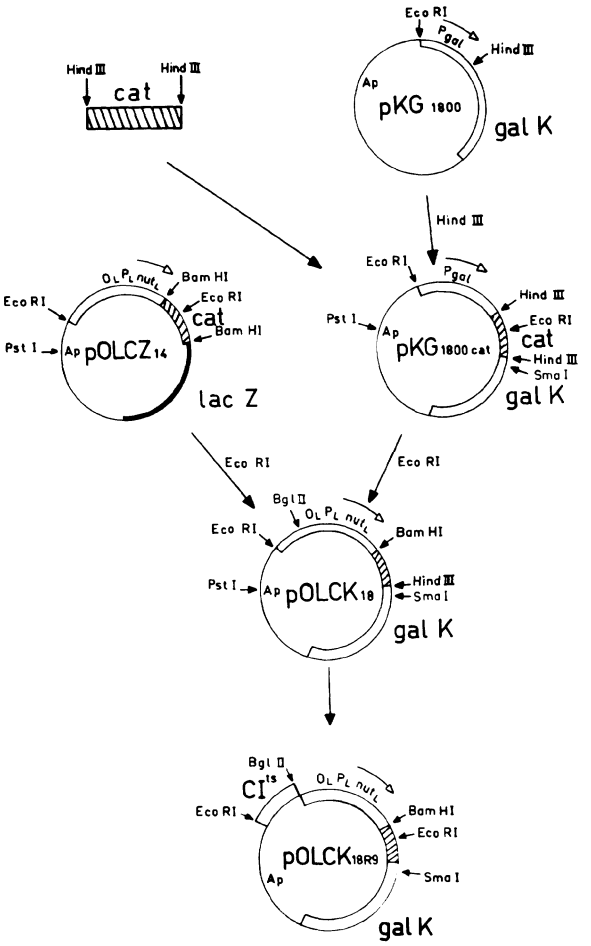


Fig. 5. Construction of plasmids pOLCK18 and pOLCK18R9. The steps in the construction of pOLCK18 and pOLCK18R9 are described in the text. The restriction sites used in the various steps are given. Unique restriction sites are also presented.

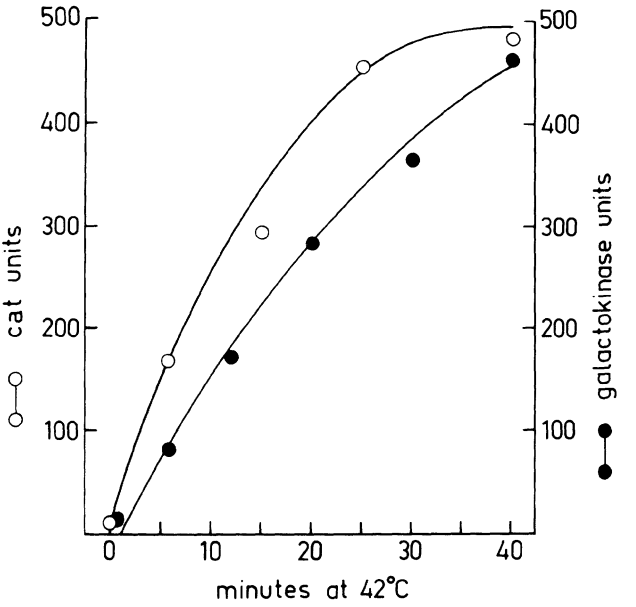


Fig. 6. Expression of CAT and galK from P_L. The *E. coli* strain W3102galKΔH_1ΔBamTn10::GalE carrying the plasmid pOLCK18 was grown at 30°C to log phase. At time zero the culture was transferred to 42°C and samples were taken at various times for OD_{650} measurement, and determination of galactokinase and chloramphenicol acetyltransferase (CAT) enzyme activity. The assay for galactokinase was performed as described (16). Galactokinase units are presented here as the number of nanamoles of galactose phosphorylated per min per ml of cells at OD_{650}=1. CAT was measured as described (18). CAT units are presented as moles of chloramphenicol acetylated per min per ml of cells at OD_{650} = 1.

Two further modifications were introduced (Fig.7). First, a unique blunt end restriction site (SmaI) was introduced between the *cat* gene and the *lacZ* gene (pOLCZ6). Second, the cI temperature-sensitive repressor gene was cloned between the ampicillin resistance gene and

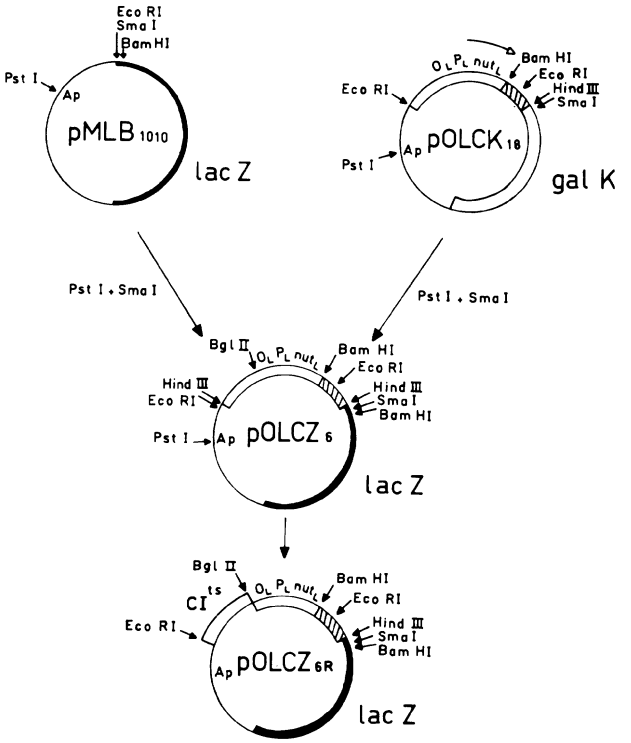


Fig. 7. Construction of pOLCZ6 and pOLCZ6R. The steps in plasmid construction are shown together with restriction sites used. Note that the cI repressor fragment carries two HindII recognition sites and one PstI recognition site. The HindIII site of pBR322 was deleted during the construction of pOLCZ6R.

P_L. The improved plasmid permits transferring to various bacterial backgrounds. Effective synthesis of β-galactosidase and chloramphenicol acetyltransferase was observed following heat induction (Fig. 8).

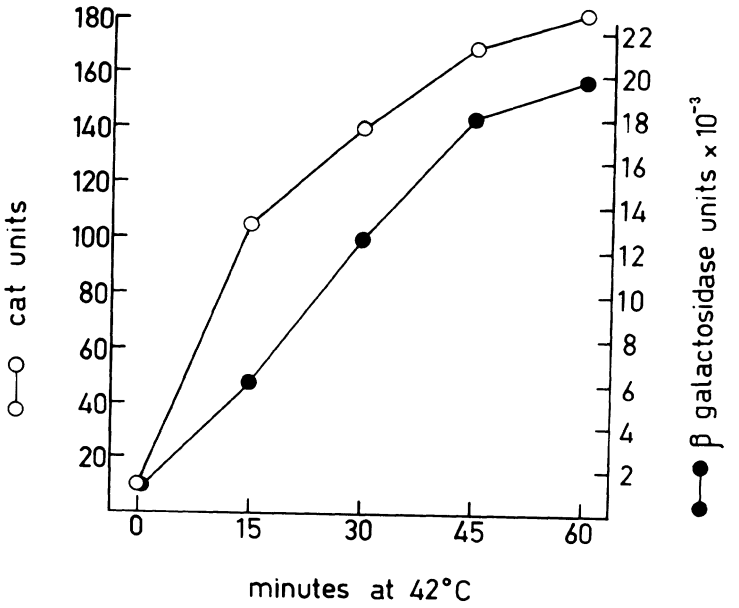


Fig. 8. Synthesis of β-galactosidase and chloramphenicol acetyltransferase from pOLCZ6R. Cells containing the pOLCZ6R plasmid were grown at 32°C to OD_{650}=0.3 on LB containing 50 µg ampicillin/ml. The cultures were induced at 42°C and β-galactosidase and chloramphenicol acetyltransferase were determined as described in this volume (18).

Introducing a terminator between the two genes allows one to measure, by simple and rapid enzymatic assays, the relative level of transcription before and after the termination sequences. Placing a transcription terminator in front of the *cat* gene permits rapid isolation and characterization of mutants by selection for chloromphenicol-resistant colonies. We have recently used this system to develop a similar cloning vector designed to look at translation-initiation signals.

ACKNOWLEDGMENTS

We thank the National Academy of Sciences and the United States-Israel Binational Science Foundation for their support. We thank Dr. Mike Berman for providing us with pMLB1010, and Dr. Max Gottesman and Dr. Donald Court for the bacterial strains and plasmids. We thank Alisa Shilor for her help in bringing this manuscript to its final form.

REFERENCES

1. Rosenberg, M. and Court,D., Ann. Rev. Genet. 13 : 319-353, 1979.

2. Gold, L., Pribnow, D., Schneider T., Shinedling, S., Singer, B.S. and Stormo, G., Ann. Rev. Microbiol. 35 : 3653-403, 1981.

3. Gottesman, M., Oppenheim, A. and Court, D. Cell 29 : 727-728, 1982.

4. Gottesman, S. Gottesman, M., Shaw, J. and Pearson, M., Cell 24 :225-233, 1981.

5. Maniatis, T., Fritsch, E.F. and Sambrook, J., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1982.

6. Hedgpeth, J., Ballivet, M. and Eisen, H., Molec. gen. Genet., 163 : 197-203, 1978.

7. Bernard, H., Remault, E., Hershfield, M., Das, H., Helinski, D., Yanofsky, C. and Franklin, N., Gene 5 : 59-76, 1979.

8. Remaut, E., Stanssens, P. and Fiers, W., Gene 15 : 81-93, 1981.

9. Derom, C., Gheysen, D. and Fiers, W., Gene 17 : 45-54 ,1982.

10. Queen, C. J. Molec. Appl. Genet., 2 : 1-10 ,1983.

11. Herskowitz, I. and Hagen, D., Ann Rev.Genet. 14 : 399-445, 1980.

12. Hershey, A.,(ed.), The Bacteriophage Lambda, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1981.

13. Bolivar, F., Rodriguez, R.L., Greene, P.Y., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, Y.H. and Falkow, S, Gene 2 :95-113, 1977.

14. Casadaban, M., Chou,I. and Cohen, S., J. Bact. 143 : 971-980, 1980.

15. Close, T. and Rodriguez, R.L., Gene 20 : 305-316, 1982.

16. McKinney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M., in J.G. Chirikjian and T.S. Papas,(eds.), Gene Amplification and Analysis, vol.2., Elsevier-North Holland, Amsterdam, 1981 pp. 383-415.

17. Honigman, A., et al, this volume.

18. Oppenheim, A.B., Katzir, N., Koby, S. and Locker-Giladi, H., this volume.

CLONING OF DNA VIRUS GENES

6

CLONING AND MAPPING OF AFRICAN SWINE FEVER VIRUS DNA

JOSE M. ALMENDRAL, ANTONIO TALAVERA AND ELADIO VIÑUELA
Centro de Biología Molecular (CSIC-UAM). Facultad de Ciencias.
Universidad Autónoma. Canto Blanco. Madrid-34. Spain.

SUMMARY

African swine fever virus DNA (about 170 kbp) was cleaved with the restriction endonuclease Eco RI and the resulting fragments, before or after cleavage with the endonuclease Sal I, were cloned in plasmid or phage vectors. The two terminal Eco RI fragments were cloned after removal of the crosslinks with nuclease S1 and addition of Eco RI linkers to the fragment ends. The order of the restriction fragments produced by the restriction endonucleases Sal I, Eco RI, Kpn I, Pvu I and Sma I was established by identifying the crosslinked Eco RI and Sal I terminal fragments and overlapping fragments.

INTRODUCTION

African swine fever (ASF) virus is a cytoplasmic icosahedral deoxyvirus which infects only domestic pigs and related species (family Suidae) and ticks of the genus Ornithodoros (family Argasidae). In swine, the virulent forms of the virus produce a fulminating and highly lethal disease with a mortality close to 100 per cent of the infected animals. Attenuated virus forms also exist and they may not produce any disease symptoms in pigs, which become virus carriers. No vaccine is available for ASF, because the sera from chronically infected animals do not neutralize the virus (for reviews, see references 1 and 2).

The genome of ASF virus is a linear duplex DNA with a molar mass of about 100×10^6 g mol^{-1} (3) and covalently closed ends (4), similar to those present in poxvirus DNA (5,6). The large size of ASF virus DNA and the relatively low virus production in tissue culture have made it difficult to study ASF virus

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

and the disease at the molecular level. The availability of clones with defined DNA fragments and a knowledge of the order of those fragments in the DNA will be useful for studies of regions coding for viral polypeptides of immunological importance and for analysis of the genetic heterogeneity of the virus.

MATERIALS AND METHODS

Viruses and cells

ASF virus, adapted to grow in VERO cells (CCL81, American Type Culture Collection) was cloned by plaque purification(7).

DNAs

ASF virus DNA was isolated and labeled as described by Almendral et al. (8). Recombinant DNA was obtained as described by Ley et al. (9).

Cloning and mapping techniques, as well as recombinant analysis have been described previously (8,9).

Materials

Most of the materials used in the work reported here have been described elsewhere (8,9).

Recombinant DNA nomenclature

Recombinants were named with the prefixes p2, p5 or L according to the vector employed (pBR322, pBR325 or Lambda, respectively), followed by a letter indicating the restriction nuclease used (R; Eco RI; S; Sal I), a letter designating the fragment inserted, and the isolation number of the recombinant.

Recombinants with fragments produced by digestion with two nucleases were preceded by the same prefixes as before, whereas the insert was designated with two pairs of symbols, separated by a tilted bar. The first pair of symbols indicates the restriction site on the left and the second the restriction site on the right side of the insert. Thus, fragment (or insert) RA/SC stands for the overlapping part of fragments Sal I-C and Eco RI-A (see Fig. 9).

RESULTS

Selection of restriction endonucleases for mapping

ASF virus DNA, uniformly labeled with ^{32}P, was digested with different restriction nucleases. Fig. 1. shows the results obtained with some of the endonucleases used. Those producing a low number of fragments were chosen for mapping purposes, namely Sma I, Pvu I, Sal I and Kpn I. We also selected Eco RI as most of the radioactive probes used in the ordering of fragments were cloned Eco RI fragments (see later). DNA bands were designated by capital letters in the order of decreasing size. Densitometric scannings of the autoradiographs shown in Fig. 1 revealed that the following bands contained two fragments each: Pvu I-A and C, Sal I-F and I, Eco RI-C,D,E,K,N and Q. These fragments were named with the same letter with or without a prime superscript. Later experiments showed that bands Kpn I-P and Eco RI-U and X also contained two fragments. Table I shows the size of ASF virus DNA restriction fragments as determined from their electrophoretic mobilities.

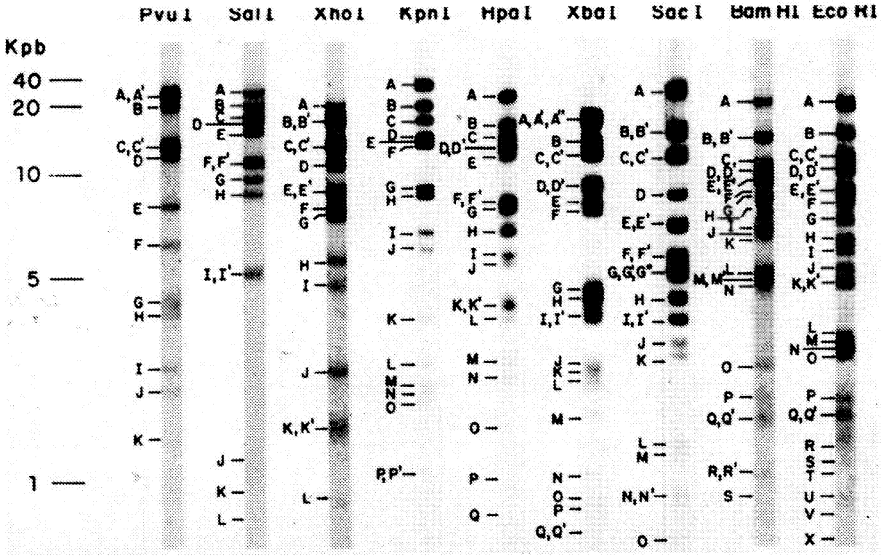


FIGURE 1. Restriction fragments of ASF virus DNA. Autoradiographs of dried agarose gels showing the bands obtained after digestion of uniformly labeled ASF virus ^{32}P-DNA with the restriction nucleases indicated. Aproximate size values are given on the left side. Reprinted with permission from (8).

Table 1. Sizes of ASF virus DNA restriction fragments, kbp

Band	Sma I	Pvu I	Sal I	Kpn I	Eco RI
A	50.0	$32.0^{2(a)}$	32.3	39	21.2
B	28.0	25.0	25.0	24.6	14.5
C	16.0	10.0^{2}	21.0	17.2	11.5^{2}
D	15.3	12.5	18.7	14.3	10.7^{2}
E	10.2	7.9	16.0	13.6	8.8^{2}
F	9.1	6.1	11.7^{2}	12.8	8.4
G	3.7	4.5	10.3	8.9	7.5
H	2.9	4.0	8.9	8.4	6.6
I	2.5	2.8	5.0^{2}	6.2	5.6
J	2.1	2.5	1.4	5.7	5.3
K	1.9	1.8	0.9	3.6	4.8^{2}
L			0.6	2.7	3.3
M				2.4	3.0
N				2.2	2.9^{2}
O				2.1	2.7
P				1.3^{2}	2.2
Q					1.9^{2}
R					1.6
S					1.5
T					1.3
U					0.8^{2}
V					0.6
X					0.4^{2}
Y					0.3
Z					0.2
Fragments	11	13	14	17	33

(a) The number 2 as a superscript indicates the existence of two fragments in the corresponding band. Reprinted with permission from (8).

Cloning of ASF virus DNA

Internal Eco RI restriction fragments were cloned in vector λWES.λB (10) or pBR325 (11). Phage recombinants were obtained that contained fragments Eco RI-C to K as single inserts (Fig. 2a).

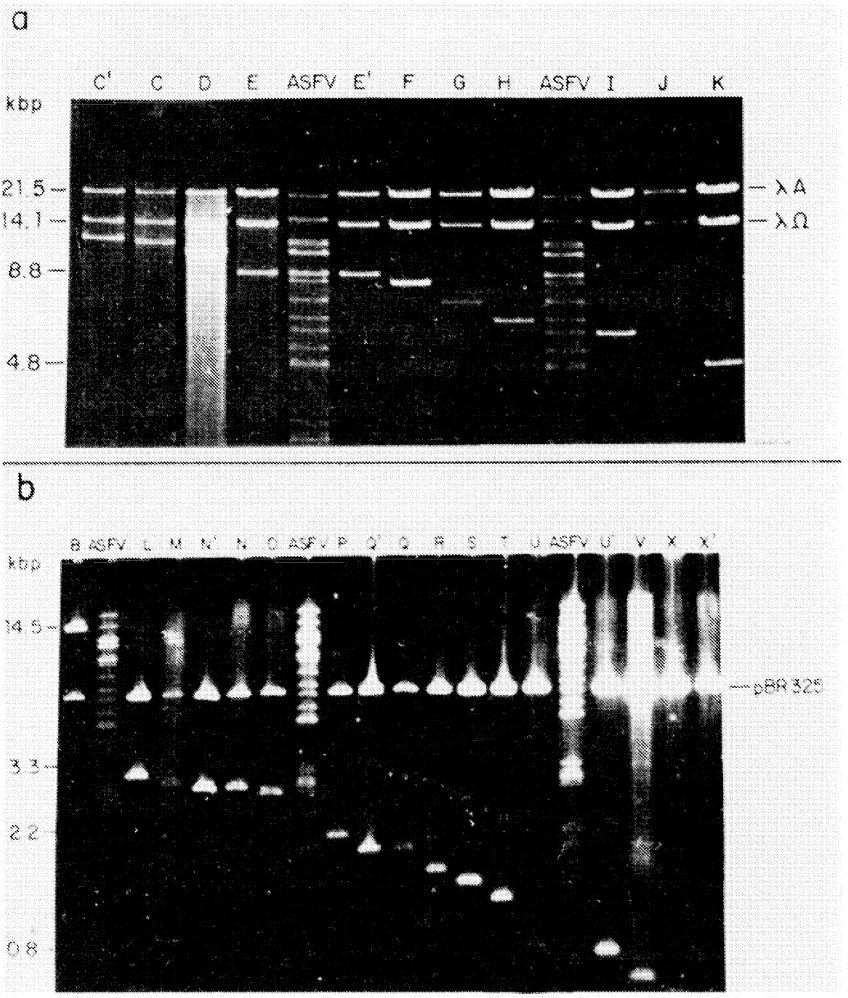


FIGURE 2. Eco RI fragments cloned in a) λWES.λB; b) pBR325. Recombinants carrying each of the cloned fragments were digested with Eco RI and electrophoresed in the presence of ethidium bromide. Lanes labeled ASF were loaded with Eco RI digests of ASF virus DNA. Other lanes are labeled with the letter corresponding to the cloned fragment. λA and λΩ indicate the left and right arms of the phage vector, respectively. Reprinted with permission from (8).

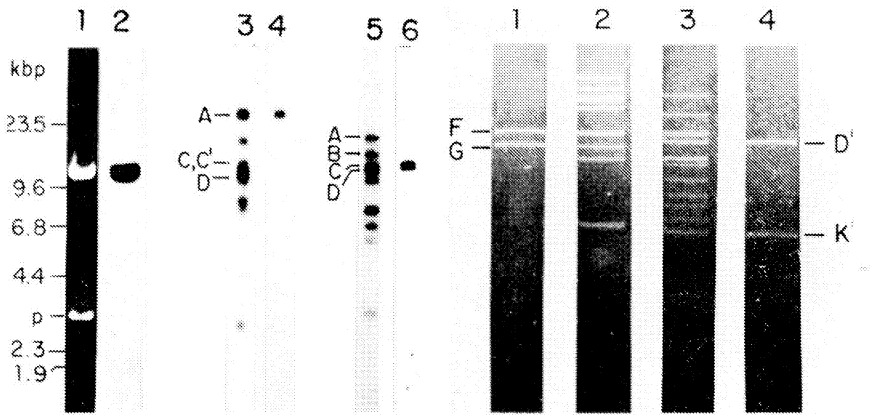


FIGURE 3. (left). Characterization of recombinant p2RA/SC3. Lane 1: the recombinant was digested with Eco RI and Sal I, and subjected to agarose electrophoresis in the presence of ethidium bromide. p indicates the position of the vector. Lane 2: DNA in lane 1 was transferred to nitrocellulose and hybridized to ASF virus ^{32}P-DNA. Lanes 3 and 4: nitrocellulose strips with immobilized Eco RI fragments of ASF virus DNA hybridized to ASF virus ^{32}P-DNA (lane 3) or ^{32}P-labeled DNA from the recombinant (lane 4). Lanes 5 and 6: nitrocellulose strips containing immobilized Sal I fragments of ASF virus DNA hybridized as in lanes 3 and 4, respectively. Reprinted with permission from (9).

FIGURE 4. (right). Sal I and Eco RI terminal fragments of ASF virus DNA. ASF virus DNA was digested with either Sal I (lane 2) or Eco RI (lane 3) and the resulting fragments subjected to electrophoresis. A portion of Sal I (lane 1) or Eco RI (lane 4) digests was heat-denatured, ice-cooled and digested with nuclease S1 before electrophoresis. Reprinted with permission from (8).

Fragments shorter than Eco RI-K and also Eco RI-B were cloned individually in plasmid pBR325 (Fig. 2 b) after electroelution from agarose gels. Neither fragment Eco RI-A, Y and Z nor terminal fragments Eco RI-D' and K' (see later), were obtained in these collections.

Upon digestion of fragment Eco RI-A with Sal I, three pieces were produced (see Fig. 9). The middle one, Sal I-I' was cloned in the Sal I site of pBR322 (12). The left and right ones (RA/SC and SB/RA, respectively) were inserted between the Sal I and Eco RI sites of pBR322. Fig. 3 shows the characterization of fragment RA/SC.

Internal fragments Y and Z, not detected originally in the autoradiographs (8), were cloned as part of a 8.2 kbp Hind III fragment (see later).

Identification of terminal fragments

The ends of ASF virus DNA are covalently closed by single-strand DNA (4) as those of vaccinia virus (5,6). DNA was cleaved by restriction nucleases and the fragments denatured and quickly renatured (13,14). In these conditions only the terminal fragments became S1 nuclease-resistant duplex molecules that could be identified by gel electrophoresis. Fig. 4 shows the identification of fragments Sal I-F and G and Eco RI-D' and K' as terminal fragments.

Eco RI linkers were added to the Eco RI terminal fragments, obtained by the technique outlined above, and then the fragments were cloned in pBR325 or λWES.λB. Fig. 5 shows the identification of two of these clones corresponding, respectively, to the left and right terminal fragments. In both cases, the inserts were shorter than the original fragments, due, presumably, to the S1 treatment. Identification of these fragments was, nevertheless, possible by an analysis of restriction sites inside them, revealing that most of the deletions produced were towards the inner end of both fragments (9). Fig. 5 shows also that either terminal fragment hybridized with both terminal fragments Eco RI K' and D'. This indicated the presence of repetitions in both terminal Eco RI fragments of ASF virus DNA (15).

Restriction site mapping

The order of restriction fragments produced by the selected endonucleases was deduced, after identification of the terminal fragments, from a) hybridization of fragments produced by one enzyme to individual fragments produced by a second enzyme(16) and b) analysis of the products of partial digestion by one enzyme, of fragments produced by another enzyme (17,18). Figs. 6 and 7 illustrate the order of Sal I and Eco RI fragments. Except where otherwise indicated, the Sal I probes used consisted of fragments electrophoretically separated and electroeluted from agarose gels. The Eco RI fragments used as probes were the ones previously cloned. The existence of two double bands in the

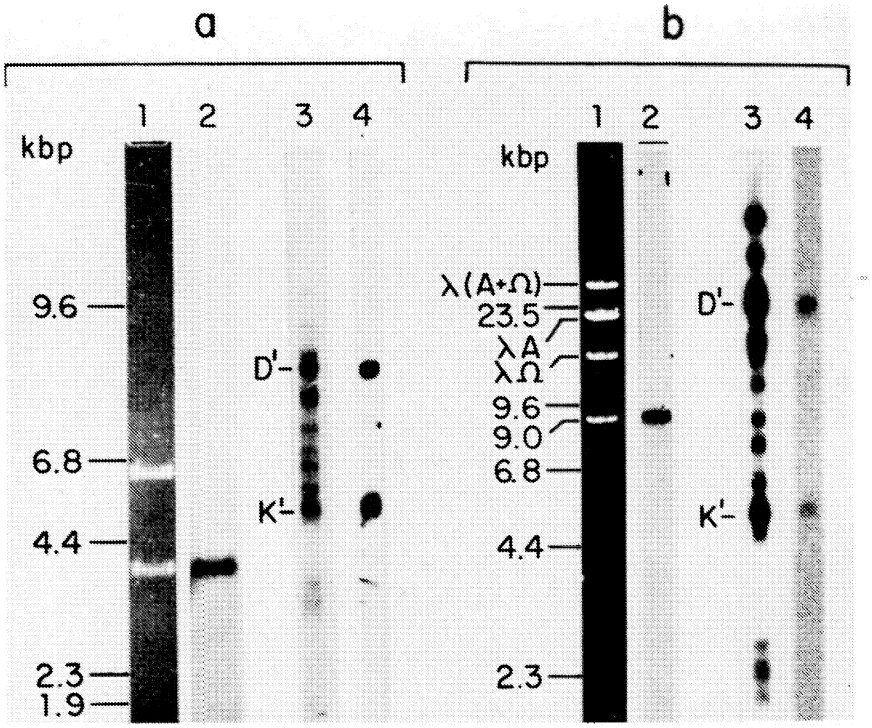


FIGURE 5. Characterization of recombinants carrying terminal Eco RI fragments. a) Lane 1: recombinant plasmid p5RK'1 was digested with Eco RI and electrophoresed. Lane 2: DNA in lane 1, transferred to nitrocellulose and hybridized to ASF virus ^{32}P-DNA. Lanes 3 and 4, nitrocellulose strips containing immobilized Eco RI fragments of ASF virus DNA hybridized to ASF virus ^{32}P-DNA (lane 3) or ^{32}P-labeled p5RK'1 (lane 4). b) Lane 1: recombinant phage LRD'16 was digested with Eco RI and electrophoresed. Lane 2: DNA in lane 1, transferred to nitrocellulose and hybridized to ASF virus ^{32}P-DNA. Lane 3: as in a). Lane 4: as in a) but using ^{32}P-labeled LRD'16 DNA as a probe. λA, λΩ and λ(A +Ω) indicate the vector arms, alone or linked through the cohesive ends. Reprinted with permission from (9).

Sal I restriction pattern, F-F' and I-I' made it difficult to determine, at first sight, a definite order of Sal I fragments. The F-F' indetermination could be solved taking into account that one of these fragments (F) had been previously identified as a terminal fragment. Fragments Sal I-I and I' were cloned and distinguished by hybridization (Fig. 7). These results

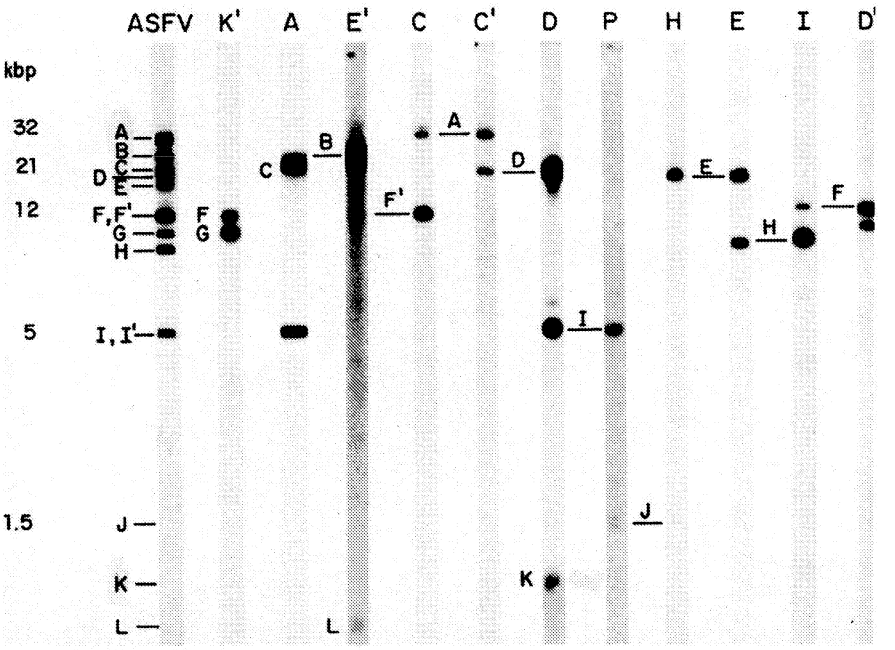


FIGURE 6. Hybridization of ^{32}P-labeled Eco RI fragments to Sal I fragments of ASF virus DNA. DNA was digested with Sal I, subjected to agarose electrophoresis and transferred to a nitrocellulose sheet. Strips of the blot were hybridized to the ^{32}P-labelled Eco RI fragments indicated. The DNA in the lane labeled ASF was hybridized with ASF virus ^{32}P-DNA. Reprinted with permission from (8).

showed unequivocally the order of all Sal I fragments and that of some Eco RI fragments, except the groups shown in parentheses. These uncertainties were solved either by similar experiments done with Kpn I, Sma I and Pvu fragments or by partial restriction analysis of clones that overlapped these groups of fragments (8).

Fig. 7 shows that no Eco RI fragment was found that linked fragments Sal I-G and C. Hybridization analysis of a Hind III digest of ASF virus DNA revealed the existence of an 8.2 kbp-long Hind III fragment that hybridized with both Sal I-G and C fragments. The Hind III fragment was cloned into plasmid pUC8

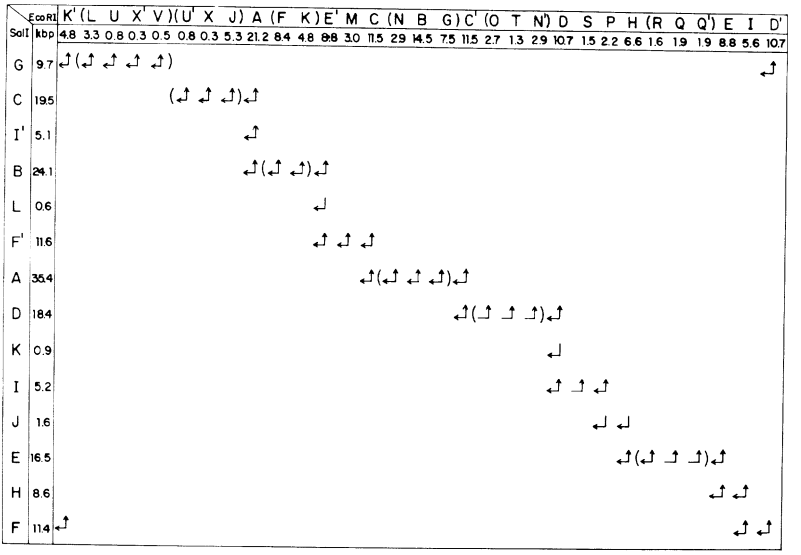


FIGURE 7. Reciprocal hybridization pattern between Eco RI and Sal I fragments of ASF virus DNA. The groups of Eco RI fragments in parentheses are those not ordered from these data. The arrows indicate a positive hybridization and point from the radioactive to the unlabeled fragment. Double arrows indicate that the hybridization was done in both directions. Reprinted with permission from (8).

(19) and, after digestion with Eco RI, showed to contain, apart from fragments U to X', two new fragments Y and Z; the first of which was further cleaved upon digestion with Sal I (unpublished results).

Experiments similar to those shown before allowed to establish the order of Kpn I, Sam I and Pvu I fragments.

Once the restriction site map for each nuclease was obtained, a composite map showing the relative locations and distances between all the restriction sites was established by mapping these sites within the Eco RI cloned fragments. Fig. 8 shows the maps deduced from those experiments. Once the individual maps were determined it was necessary to orient each of them along the overall map. In most instances, this was done from hybridization data. For instance, fragment Eco RI-C' was divided in two pieces by both Sal I and Kpn I. Since this fragment

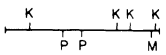
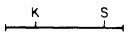
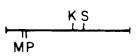
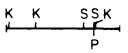
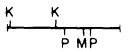
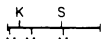
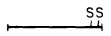
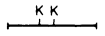
Eco RI, kbp	Fragments, kbp				Map
	Sal I	Kpn I	Pvu I	Sma I	
B, 14.5	N.S.	Dp-H-P'-N-Op 2.1-8.4-13-2.2-0.5	Fp-K-Ep 5.4-1.7-7.4	Dp-Ep 13.6-0.9	
C, 11.5	F'p-Ap 8.3-3.2	Bp-Dp 2.8-8.7	N.S.	N.S.	
C', 11.5	Ap-Dp 7.3-4.2	Gp-Fp 6.3-5.2	Hp-Bp 1.8-9.7	Ep-Bp 1.5-10.0	
D, 10.7	Dp-K-Ip 7.3-0.9-2.5	Fp-L-J-Cp <0.1-2.7-5.6-2.4	Bp-Ap 8.2-2.5	N.S.	
D', 10.7	N.S.	Ep-K-I <0.1-4.5-6.2	Ap-J-I 5.4-2.5-2.8	Cp-G 7.0-3.7	
E, 8.8	Ep-Hp 5.1-3.7	Cp-Ep 1.1-7.7	N.S.	Fp-J-H-Cp 0.1-2.2-2.9-3.6	
E', 8.8	Bp-L-F'p 7.9-0.6-0.3	N.S.	N.S.	N.S.	
F, 8.4	N.S.	Ap-P-Bp 3.0-1.3-4.1	N.S.	N.S.	
G, 7.5	N.S.	Op-M-Gp 1.5-2.4-3.6	Ep-G-Hp 0.6-4.5-2.4	N.S.	
H, 6.6	Jp-Ep 0.6-6.0	N.S.	N.S.	Kp-I-Fp 0.3-2.6-3.7	
I, 5.6	Hp-Fp 4.9-0.7	N.S.	N.S.	N.S.	
J, 5.3	N.S.	N.S.	Cp-A'p 2.6-2.7	N.S.	
K, 4.8	N.S.	N.S.	A'p-Dp 3.5-1.3	N.S.	
M, 3.0	N.S.	N.S.	Dp-C'p 2.2-0.8	N.S.	
N, 2.9	N.S.	N.S.	C'p-Fp 2.3-0.6	Ap-Dp 2.0-0.9	
P, 2.2	Ip-Jp 1.2-1.0	N.S.	N.S.	Bp-Kp 0.5-1.7	
Y, 0.2	Gp-Cp 0.1-0.1	N.S.	N.S.	N.S.	

FIGURE 8. Restriction site maps of Eco RI fragments of ASF virus DNA. The maps were obtained from an analysis of the digestion products of phage or plasmid recombinants with the indicated enzymes : K, Kpn I; M, Sma I; P, Pvu I; S, Sal I. Reprinted with permission from (8).

connects fragments Sal I-A and D (Fig. 7) on the one hand, and fragments Kpn I-G and F (see Fig. 9), on the other, the orientation shown in Fig. 8 predicts that the segment between both sites is common to fragments Sal I-A and Kpn I-F. The reverse

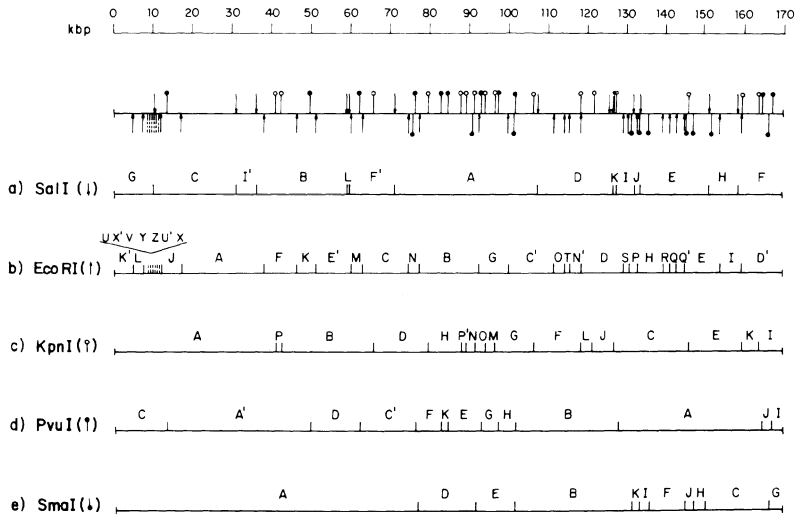


FIGURE 9. Restriction site map of ASF virus DNA. Lines a-e show the Sal I, Eco RI, Kpn I, Pvu I and Sma I individual maps, respectively. The upper line, below the length scale, shows the composite map for all the restriction sites indicated in the individual maps. Reprinted with permission from (8).

orientation would imply that Sal I-A and Kpn I-F had no sequences in common. Hybridization data (not shown) indicated that there was a common sequence between these fragments; this allowed fragment Eco RI-C' to be orientated correctly.

In other cases, the orientation of a fragment was deduced from the distances between successive restriction sites in this and the next fragment. For example, the orientation of fragment Eco RI-B, shown in Fig. 8, leaves on the right a space of 0.5 kbp that added to the left part of fragment Eco RI-G (1.5 kbp), gives a total of 2.9 kbp for fragment Kpn I-O. The reverse orientation of Eco RI-B would substitute a 2.1 kbp subfragment for the 0.5 kbp long one, which, in its turn, would give a total legnth of 3.6 kbp for Kpn I-O, in disagreement with the 2.1 kbp length calculated for this fragment by electrophoresis. Fig. 9 shows the composite map deduced from these orientations as well as the individual site maps for each of the restriction nucleases used.

DISCUSSION

Recombinant DNA techniques as well as the use of nucleic acid hybridization procedures have been employed for the establishment of restriction site maps of ASF virus DNA. Although in some cases the contiguity of two restriction fragments was found by using labeled fragments that had been previously extracted from agarose gels, in general, clones were used for the establishment of the order of fragments, by hybridization or partial restriction analysis. The availability of cloned ASF virus DNA fragments has allowed the precise mapping of viral transcripts and gene products (manuscripts in preparation). The cloned fragments may help to avoid, in some cases, the mapping of some necessary restriction sites along the whole DNA, by limiting this procedure to a particular region of the genome.

The sizes of the fragments obtained upon digestion with the nucleases chosen for mapping ranged between 0.2 (Eco RI-Z) and 76.6 (Sma I-A) kbp. The lengths of the largest fragments were obtained by addition of the smaller ones contained in them, whereas the lengths of the elementary fragments were obtained by restriction site mapping of the Eco RI clones, as shown in Fig. 8. The overall length of ASF virus DNA calculated by the addition of consecutive fragments yielded a value of about 170 kbp, which agrees with the ones reported by Enjuanes et al. (3).

The reciprocal hybridization of terminal fragments shown in Figs. 5-7 indicated the existence of repetitions in those fragments. Further experiments showed that these repetitions are inverted and located at the ends of ASF virus DNA (15), similarly to those found in vaccinia virus DNA (14,20).

REFERENCES

1. Viñuela E. 1984. In press. In "Iridoviruses". Current Topics in Microbiology and Immunology, Springer Verlag, New York.
2. Wardley RC, Andrade CM, Castro-Portugal FL, Enjuanes L, Hess WR, Mebus C, Ordás A, Rutili D, Sánchez Vizcaino, J, Vigario JO, Wilkinson PJ, Moura Nunes JF, Thomson G. 1983. Archives of Virology 76: 73-90.
3. Enjuanes L, Carrascosa JL, Viñuela E. 1976.J. gen. Virol. 32: 479-492.

4. Ortín J, Enjuanes L, Viñuela E. 1979. J. Virol. 31 : 579-583.
5. Geshelin P, Berns KI. 1974. J. Mol. Biol. 88: 785-796.
6. Baroudy BM, Venkatesan S, Moss B. 1982. Cell 28 : 315-324.
7. Enjuanes L, Carrascosa AL, Moreno MA, Viñuela E. 1976. J. gen. Virol. 32 : 471-477.
8. Almendral JM, Blasco R, Ley V, Beloso A, Talavera A, Viñuela E. In press. Virology.
9. Ley V, Almendral JM, Carbonero P, Beloso A, Viñuela E, Talavera A. In press. Virology.
10. Leder P, Tiemeier D, Enquist L. 1977. Science 196: 175-177.
11. Bolivar F. 1978. Gene 4: 121-136.
12. Bolivar F, Rodriguez RL, Greene AJ, Betlach MC, Heyneker HL, Boyer HW. 1977. Gene 2 : 95-113.
13. Jaureguiberry G. 1977. FEBS Letters 83 : 111-117.
14. Wittek R, Menna A, Muller HK, Schumperli D, Boseley PG, Wiler R. 1978. J. Virol. 28 : 171-181.
15. Sogo JM, Almendral JM, Talavera A, Viñuela E. In press. Virology.
16. McDonell M, Simon MH, Studier FW. 1977. J. Mol. Biol. 110: 119-146.
17. Danna JJ, Sack,Jr. GM, Nathans D. 1973. J. Mol. Biol. 78: 363-376.
18. Smith MO, Birnstiel ML. 1976. Nucl. Acids Res. 3: 3287-3298.
19. Vieira J, Messing J. 1982. Gene 19: 259-268.
20. Garon CF, Barbosa E, Moss B. 1978. Proc. Natl. Acad. Sci. USA 75: 4863-4867.

7

CLONING OF THE DNA OF ALPHAHERPESVIRINAE

ANDREW DAVISON and FRAZER RIXON
MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, U.K.

SUMMARY

DNA fragments of herpes simplex virus types 1 and 2, varicella-zoster virus and pseudorabies virus were cloned in bacterial plasmids either by direct ligation or by annealing via short homopolymer tracts. The clones have been used to investigate specific structural and functional aspects of these herpesvirus genomes.

INTRODUCTION

The herpesviruses are complex enveloped icosahedral viruses possessing a large linear double-stranded DNA genome. They have been isolated from a wide range of animals, and are classified into three groups according to their biological properties and genome structures: the alpha-, beta- and gamma-herpesvirinae (1). Included among the alphaherpesvirinae are herpes simplex virus types 1 and 2 (HSV-1 and HSV-2), varicella-zoster virus (VZV) and pseudorabies virus (PRV).

The genome of each of these herpesviruses consists of two segments, L and S (Figure 1). In HSV-1 and HSV-2, L and S each comprise a unique region (U_L,U_S) flanked by inverted repeats (IR_L,TR_L,IR_S,TR_S) (2-10). Four types of molecule are present in equimolar amounts in virion DNA as a consequence of inversion of L and S about their joint. They are defined arbitrarily as the P (prototype), I_L (L inverted), I_S (S inverted) and I_{SL} (L and S inverted) genome arrangements (11). The HSV-1 and HSV-2 genomes are terminally redundant, possessing a sequence of 250-500 base pairs (bp), known as the a sequence, as a direct repeat at each terminus and as an inverted repeat at the L-S joint (2,12-17). The genome structures of VZV (18-20) and PRV (21-23) are similar

to that of HSV-1 except that inverted repeats flanking U_L have not been detected and L does not invert. Therefore, virion DNA populations contain only two types of molecule (P and I_S).

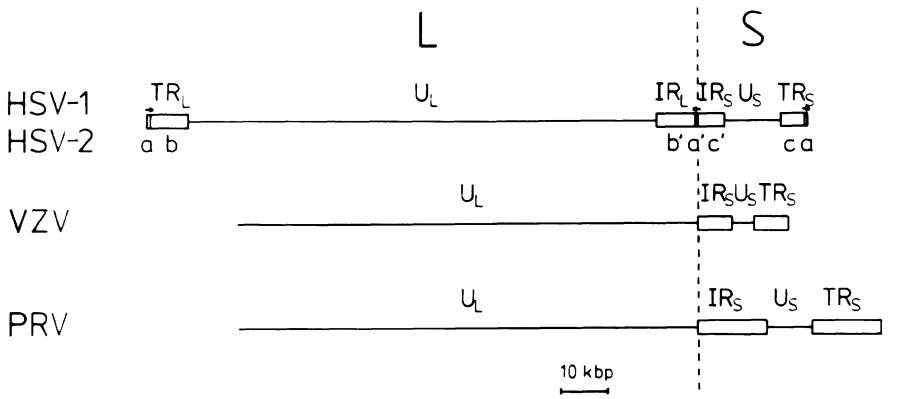


FIGURE 1. Genome structures of HSV-1, HSV-2, VZV and PRV. Inverted repeats are shown as rectangles, and orientations of a sequences in the HSV-1 and HSV-2 genomes are indicated by arrows.

Two techniques were used to clone subfragments of herpesvirus DNA, since the genomes are too large to clone as intact molecules in bacterial plasmids. Firstly, herpesvirus DNA was cleaved with a restriction endonuclease and ligated into a plasmid vector which had been linearized with the same enzyme. Secondly, restriction endonuclease fragments of herpesvirus DNA and linearized vector were extended at the 3' termini with complementary homopolymer tracts and annealed. Transformed bacteria were selected for resistance to the appropriate antibiotic and clone libraries were compiled.

MATERIALS AND METHODS

Cells and viruses

HSV-1 strain 17 (Glasgow), HSV-1 strain USA-8 (24), HSV-2 strain HG52 (25) and PRV (26) were grown and titrated at 37° in baby hamster kidney (BHK C13) cells as described previously (27,28). VZV was grown by passage of human foetal lung cells

(established by Dr B. Carritt, Institute of Genetics, Glasgow) infected with a VZV strain (18) supplied at passage 8 by Dr J.L.M.C. Geelen, University of Amsterdam.

Preparation of virus DNA

HSV-1 and HSV-2 DNA were prepared by phenol-extraction of virus isolated from infected BHK C13 cells by treatment with the detergent Nonidet P40 (29). DNA was purified by isopycnic banding in CsCl gradients. PRV DNA was isolated from virus purified by sedimentation in sucrose gradients (30). VZV DNA for cloning was supplied by Dr J.L.M.C. Geelen, University of Amsterdam, who had isolated the DNA from cells infected with VZV at passage 6. DNA for clone analysis was prepared from virus purified by sedimentation in sucrose gradients (31). DNA concentrations were estimated from absorbance at 260 nm.

Construction of recombinant plasmids

Insertion via ligation of termini. HSV-2 HindIII fragments were ligated into the bacterial plasmid vector pAT153 (32) essentially as described by Tanaka and Weisblum (33). HSV-2 DNA and pAT153 were separately digested with HindIII then heated at 70^{o} for 10 min to inactivate the endonuclease. Aliquots of the two reactions were mixed to give a final DNA concentration of 20-50 μg/ml, with HSV-2 HindIII sites in three- to five-fold excess over pAT153 HindIII sites, and appropriate solutions were added to give final concentrations of 0.02 M tris-HCl pH 7.5, 0.01 M $MgCl_2$, 0.0005 M ATP, 0.01 M dithiothreitol and 10 U/ml T4 DNA ligase. Incubation was at 4^{o} for 16 hr. HSV-1 and PRV BamHI fragments were ligated similarly into the BamHI site of pAT153, except that incubation was at 15^{o}. VZV BglII fragments were ligated into the BglII site of vector pKC7 (34). Specific BamHI fragments of HSV-1 and HSV-2 DNA were isolated by agarose gel electrophoresis of cleaved DNA followed by hydroxylapatite chromatography (7), and were then ligated into the BamHI site of pAT153 as described above.

The ligated mixtures were used to transform (35) Escherichia coli K12 strain HB101 (36) cells and colonies were grown on agar plates containing L-broth (0.17 M NaCl, 10 g/l Difco bactotryptone, 5 g/l yeast extract) and 0.1 mg/ml ampicillin. All procedures involving living bacteria subsequent to ligation were

performed formerly under Category II conditions and latterly under conditions of good microbiological practice, as advised by the U.K. Genetic Manipulation Advisory Group and the local safety committee. Colonies bearing recombinant plasmids with HSV-1 or HSV-2 inserts were identified by hybridization to ^{32}P-labelled virus DNA or restriction enzyme fragments (37). Individual inserts were identified by hybridization of ^{32}P-labelled plasmid DNA to Southern blots of HSV-1 or HSV-2 DNA, and by restriction endonuclease analysis. BamHI clones were identified by sensitivity to 10 μg/ml tetracycline hydrochloride and subsequently by restriction endonuclease analysis of plasmid DNA harvested from minicultures (38). Bacterial stocks of analysed clones were prepared after a further step of colony purification and stored at -20° in 1% Difco bactopeptone, 40% glycerol.

Bulk amounts of plasmid DNA were prepared from 1-2 l of bacterial cultures by a "soft" lysis procedure using the detergent Nonidet P40 (39,40). Closed circular DNA was purified by two steps of isopycnic banding in CsCl gradients containing 0.5 mg/ml ethidium bromide (41).

Insertion via annealing of homopolymer tracts. HSV-1, HSV-2 or PRV DNA was cleaved with KpnI and VZV DNA with KpnI or SstI, and pAT153 was linearized with PstI. Terminal deoxynucleotidyl transferase was used to add short homopolymer "tails" of deoxycytidine residues to KpnI and SstI sites and deoxyguanosine residues to PstI sites (42). "Tailed" PRV or VZV DNA fragments were annealed with "tailed" pAT153, each at 2 μg/ml in 0.01 M tris-HCl pH 7.6, 0.1 M NaCl, 0.001 M EDTA, by heating to 70° and cooling slowly to room temperature. Colonies of E. coli K12 strain HB101 transformed by the annealed DNA were selected in the presence of 10 μg/ml tetracycline hydrochloride, and the resulting HSV-1, PRV and VZV clones were analysed by restriction endonuclease digestion of plasmid DNA harvested from minicultures (38). Purified HSV-1 or HSV-2 DNA fragments were hybridized to HSV-2 KpnI clones in order to identify specific recombinant plasmids. Those colonies chosen to form the clone libraries were colony-purified once, bacterial stocks were prepared, and plasmids were further characterized by molecular hybridization as described above.

The cloning method involving annealing of homopolymer tracts offered two advantages over direct ligation. Firstly, ligation led to a large proportion of colonies containing recircularized vector, which had to be distinguished from recombinant plasmids by selection of a second drug-resistance marker or by colony hybridization. This background could have been reduced by treating the cleaved vector DNA with bacterial alkaline phosphatase prior to ligation. "Tailed" vector could not recircularize by self annealing, and thus all transformed colonies possessed recombinant plasmids. Secondly, ligation allowed the efficient cloning only of DNA fragments possessing the appropriate restriction endonuclease sites at both ends, whereas annealing of homopolymer tracts allowed any fragment to be cloned, including those containing the genome termini. The experimental design resulted in the reconstruction of PstI and KpnI or SstI sites flanking the insert (Figure 2), thus allowing the insert to be excised precisely. Plasmids containing a genome terminus lacked a KpnI or SstI site at the junction between the terminus and vector sequences.

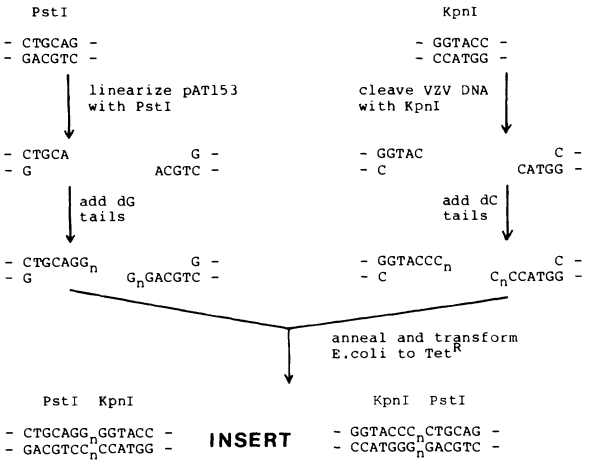


FIGURE 2. Reconstruction of restriction endonuclease sites flanking the inserts in recombinant plasmids generated by annealing of homopolymer tracts.

Clone libraries

Figure 3 shows the restriction endonuclease fragments of HSV-1, HSV-2, VZV and PRV DNA which form the clone libraries. The overlapping fragments HSV-1 KpnI v and HSV-1 BamHI v were both approximately 150 bp smaller than the corresponding virion DNA fragments. This may indicate the presence of a palindrome in this region of the genome, since such a structure would be unstable in plasmids. No clones of VZV KpnI e were isolated even though several clones of fragments of equal size (VZV KpnI f and g) were identified. The reason for this is not known, although sequence rearrangements in EcoRI clones originating from this region of the VZV genome, which is variable in size (43), have been reported (20).

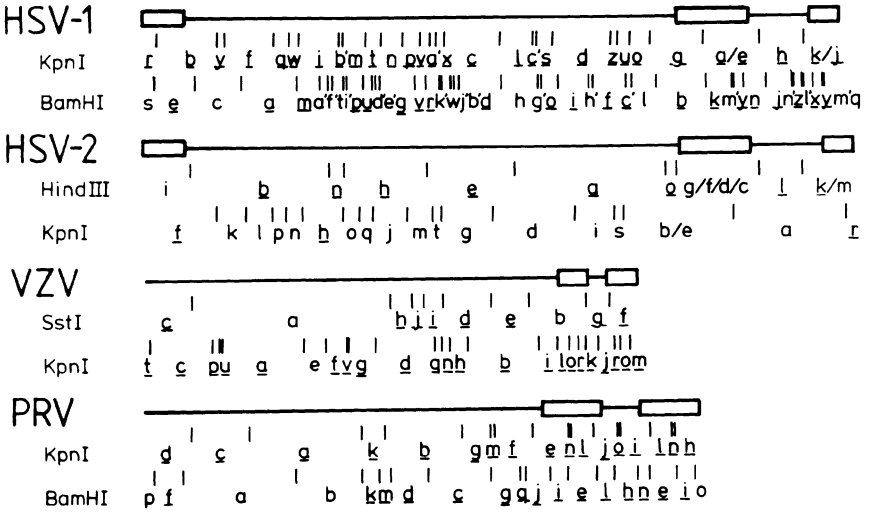


FIGURE 3. Summary of cloned DNA fragments (underlined) shown with respect to restriction endonuclease maps of HSV-1 (44,17), HSV-2 (10,44), VZV (45) and PRV DNA (22). Alternative fragments mapping at the termini and joint, due to segment inversion, are indicated.

RESULTS AND DISCUSSION

Restriction endonuclease site mapping

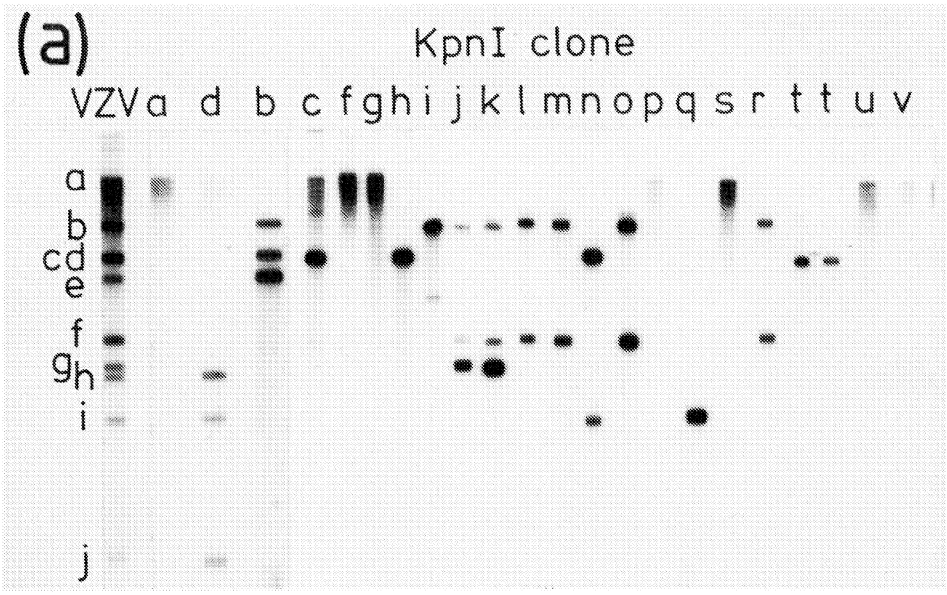
Maps for those restriction endonucleases which were used to clone HSV-1, HSV-2 and PRV DNA have been published previously, and consequently each clone was characterized readily by restriction endonuclease analysis and molecular hybridization. Moreover, analysis of the PRV clones allowed the correction of the published order of BamHI f and p (22,23) and the mapping of two additional fragments, KpnI n and o.

Maps of VZV DNA for PstI, XbaI, BglII, EcoRI and HindIII were also available (18-20), but the KpnI and SstI maps were not known. Therefore, data from the clones themselves were used to derive these maps. The relative order of KpnI fragments was deduced by hybridizing KpnI fragments from virion DNA to Southern blots of PstI fragments, and the KpnI clones were identified similarly and located precisely in the genome by restriction endonuclease analysis using PstI. The SstI map was determined by hybridizing KpnI clones to Southern blots of SstI fragments and by SstI cleavage of the KpnI clones (Figure 4). The SstI map was then used to identify the SstI clones. Maps for XhoI, PvuII, EcoRI and SalI were derived using the VZV clones (45), bringing the number of available VZV maps to eleven. As has been the case for the other herpesviruses, these maps will be of fundamental importance in molecular epidemiological and genetical studies of VZV.

Homology between herpesvirus genomes

The observations that HSV-1 and HSV-2 cross-neutralize (46), recombine (47), share multiple antigens (48) and possess extensive homology in 50% of their DNA sequences (49) demonstrate the close relationship between the two viruses. HSV-1 is related immunologically to PRV (50) and VZV (51) to a far smaller extent. PRV DNA is 8-10% homologous to HSV-1 DNA (52), and the homology is distributed throughout the PRV genome (53). The degree of homology between HSV-1 and VZV DNA is of the order of 1% (54).

Cloned DNA fragments, being free from contamination by sequences derived from elsewhere in the genome, have proved useful tools in evaluating the sequence relatedness of alphaherpesvirinae. To study this, cloned fragments from one herpesvirus were ^{32}P-labelled *in vitro* and hybridized to Southern



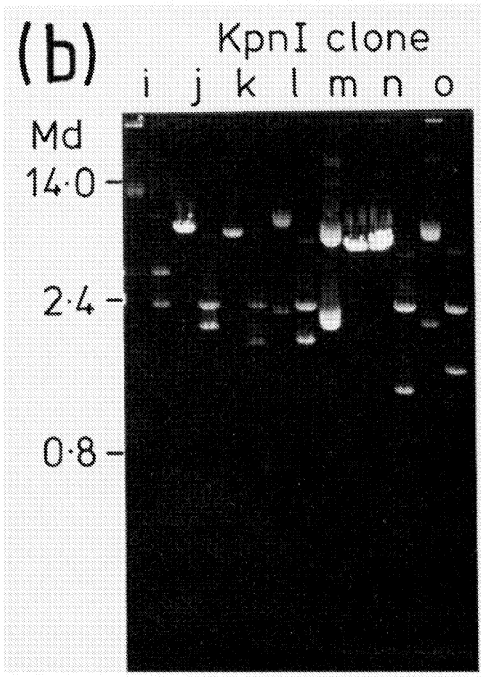


FIGURE 4. Some of the data used to derive the VZV SstI map. (a) Autoradiograph of Southern blots containing SstI fragments to which *in vitro* ^{32}P-labelled KpnI clones were hybridized. (b) Photograph of a UV-illuminated agarose gel containing KpnI clones cleaved with SstI (left track of each pair) or SstI plus KpnI (right track of each pair).

blots of virion DNA of a heterologous herpesvirus. The location and orientation of homologous regions in the HSV, PRV and VZV genomes are summarized graphically in Figure 5. The analysis was extended to include equid herpesvirus 1 (EHV-1), also a member of the alphaherpesvirinae.

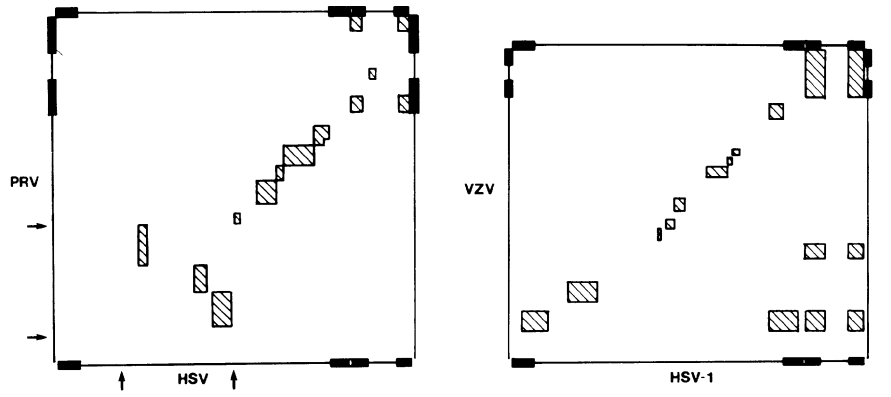


FIGURE 5. Summary of relative orientations of homologous regions in the HSV, PRV and VZV genomes (from ref. 54). Genome structures define the axes and cross-hatching denotes homologous regions. The HSV genome is shown in the I_L arrangement. Arrows define the inverted region in the HSV and PRV genomes. (Reprinted with permission).

The VZV genome is colinear with the I_L or I_{SL} genome arrangement of HSV-1. Non-colinear homology involving the inverted repeats of HSV-1 is probably due to the presence of multiple reiterations of short sequences of high guanine plus cytosine content in these regions (17,55, see Figure 10), rather than to genetic non-colinearity. The PRV genome is essentially colinear with the I_L genome arrangement of HSV-1 except for an inverted region in U_L. The results of similar experiments showed that the genome of EHV-1 is colinear with the I_L or I_{SL} genome arrangement of HSV, and that the P arrangements of the HSV-1 and HSV-2 genomes are colinear.

Several specific regions of the genomes are more highly conserved than others. The best candidates for conserved genes, based on existing knowledge of HSV-1 gene location, are those

encoding the major DNA binding protein, the major capsid protein, the DNA polymerase, the immediate-early gene V_{mw} IE 175 and one or both of the early proteins V_{mw} 136'(143) and V_{mw} 38. The major DNA binding protein and, to a lesser extent, the major capsid protein are antigenically conserved among several members of the alphaherpesvirinae (56,57). A degree of sequence conservation between the protein coding region of the major capsid protein genes of HSV-1 and PRV was demonstrated by hybridizing subfragments of HSV-1 KpnI i to subfragments of PRV BamHI d. The results of this experiment are summarized in Figure 6. This shows that the highest homology is within known protein coding regions of the HSV-1 gene.

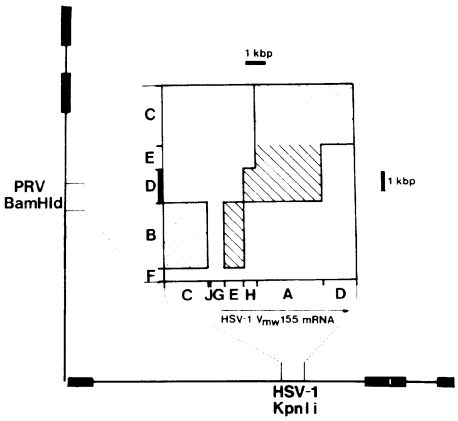


FIGURE 6. Summary of the location and orientation of homologous sequences between the HSV-1 major capsid protein (V_{mw} 155) gene, which is located in KpnI i, and PRV BamHI d (from ref. 54). The genome locations of the fragments are indicated on the outer axes, with the HSV-1 genome in the I_L arrangement. Homologous regions are shown by cross-hatching, and regions of weaker homology by stippling. (Reprinted with permission).

An important implication of these findings is that HSV-1 gene probes could be used to locate corresponding genes in heterologous genomes, provided that there is sufficient homology between the respective DNA sequences. This was illustrated by the localization of the PRV major capsid protein gene.

We propose the hypothesis that, taking into account the inverted region in the PRV genome, the genomes examined are colinear also with respect to genes of similar function for which no homology was detected, and that they share a similar genetic design of ancient origin. The structural and size differences between the genomes imply that it is unlikely that they have

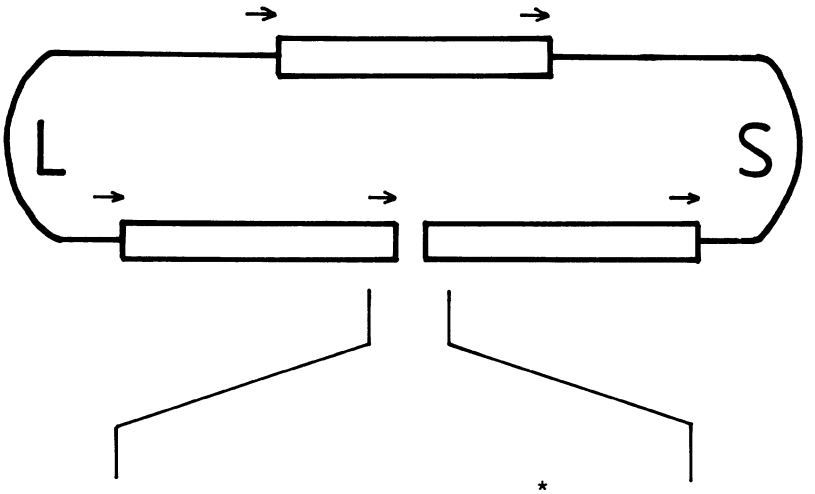
identical genetic organizations. However, many gene locations may be predicted on the basis of existing knowledge of HSV-1 genetic organization. We are currently testing these predictions by carrying out DNA sequence comparisons of extensive regions of the HSV-1 and VZV genomes.

Sequences at herpesvirus genome termini

Experimental evidence indicates that HSV-1 DNA replication includes circularization of linear input genomes, inversion of the L and S segments, production of concatemers by rolling-circle replication, and cleavage and encapsidation of unit-length DNA molecules. Sequences at the genome termini are involved in circularization and in cleavage and encapsidation. Indeed, studies involving the analysis of sequences required for encapsidation of defective DNA (N. Stow, personal communication) and the transfer of genome terminal sequences to a second site in the HSV-1 DNA molecule (58) show that cleavage and encapsidation signals are contained within the a sequence. Furthermore, the latter studies (58), in addition to our analysis of sequences at the termini of HSV-1/HSV-2 intertypic recombinants (59) show that the signal responsible for segment inversion also resides in the a sequence.

These studies point out the importance of the a sequence in DNA replication. Therefore the DNA sequences of the termini and L-S joint in HSV-1 clones were determined, and showed the presence of a short direct repeat (DR) of 17-21 bp at the b-a and a-c junctions (Figure 7). A DR of 17 bp which is present in the same locations in HSV-2 has a different sequence. The HSV-1 and HSV-2 termini are located close to one end of the DR sequences. The 3' nucleotide at the HSV-1 S terminus was identified unambiguously from the autoradiograph shown in Figure 8, but that at the L terminus could be one of three owing to the presence of the homopolymer "tail". However, recent evidence suggests that the genome of HSV-1 strain F possesses a single 3'-protruding nucleotide at each terminus (58), so this has been assumed in Figure 7 to be the case for HSV-1 strains 17 and USA-8. Although circularization of HSV-1 DNA has been assumed for several years to occur by annealing of complementary strands of the a sequence

exposed by the action of an exonuclease (60), our careful analysis of the termini of HSV-1/HSV-2 intertypic recombinants (59) and the discovery of a 3'-protruding nucleotide at each terminus (58) provide two separate lines of evidence for circularization by direct ligation of termini. Clearly, termini could be produced by cleavage of concatemeric DNA between two adjacent a sequences. The sequence at the cleavage site is somewhat variable (Figure 7), but it is likely that one function of DR is as part of the cleavage signal.



strain 17	CCGCGGGGGGCCCGGGCTGCC	CGCCGCCGCCGCTTTAAA
	GGCGCCCCCCGGGCCCGACG	GGCGGCGGCGGCGAAATTT
strain USA-8	CCGCGGGGGGCCCGGGCCGCC	-GCCGCCG-CGCTTTAAA
	GGCGCCCCCCGGGCCCGGCG	G-CGGCGGC-GCGAAATTT
strain F	CCGCGGGGGGCCCGGGCTG	--CGCCGCCG-CGCTTTAAA
	GGCGCCCCCCGGGCCCGA	C--GCGGCGGC-GCGAAATTT

FIGURE 7. Location of DR sequences (arrows) at the termini and L-S joint of the HSV genome. The a sequences are shown as rectangles. Sequences at the genome termini of three HSV-1 strains are aligned below. The DR sequences are underlined, and dashes indicate absent nucleotides. The asterisked G/C base pair at the S terminus of strain 17 was present as an A/T base pair in the L-S joint fragment sequenced. The data for strain F are from Mocarski and Roizman (58,62).

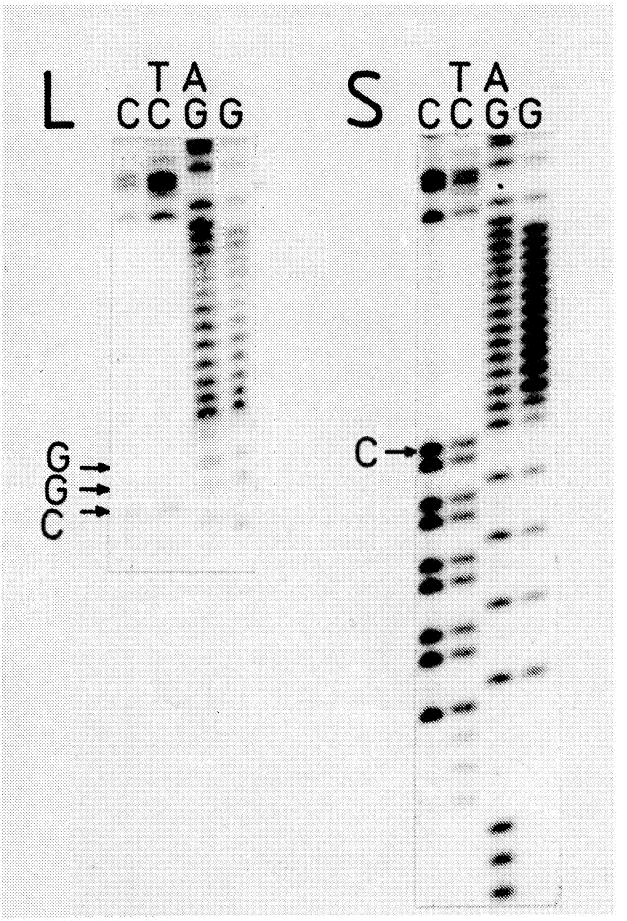


FIGURE 8. Autoradiographs showing DNA sequences at the L and S termini of HSV-1 strain 17. The sequences were determined from KpnI r and k by the chemical degradation method (61), and run 3'-5' across the genome termini, through the homopolymer "tail" and into vector DNA. Arrows indicate the terminal nucleotide in each case.

Our sequence analysis of PRV and VZV clones has provided a useful comparison with HSV-1 and HSV-2. Circularization of the DNA of these viruses probably occurs by direct ligation of termini, since neither is terminally redundant. It is not yet known whether these genomes also have a 3'-protruding nucleotide at each terminus. HSV-1, HSV-2, VZV and PRV possess a structurally conserved sequence approximately 30 bp from the S

terminus (Figure 9), which may function as part of the cleavage or inversion signal. Further analyses of the VZV and PRV genomes, in which only one segment inverts, will complement studies on the HSV-1 and HSV-2 genomes, in which both segments invert.

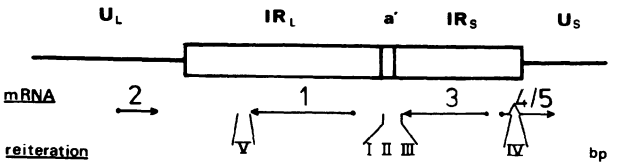
```
                                    *
HSV-1     CCCCGGGGGGTGTG---TTTT-GGGGGGGGG
HSV-2     CCCCGGGGGGGGTGTT-TTTT-GGGGGGGGG
VZV       AACCGGGGGGGG-GTTATTTTCGGGGGGGGG
PRV       CCCCGGGGGCCGCG-AAAAAA-GGGGGCGGG

                          <------>
                       AT-rich region
```

FIGURE 9. Sequence of conserved structure in the HSV-1, HSV-2, VZV and PRV genomes approximately 30 bp from the S terminus (to the left). Totally conserved sequences are underlined. The asterisked T residue in HSV-1 strain 17 is a C residue in strains USA-8 and F (17,62).

Tandem reiterations in the HSV-1 genome

Restriction enzyme analysis of DNA from independent isolates of HSV-1 has revealed several fragments which exhibit considerable variation in their mobility in agarose gels (63). A sub-set of these fragments was also found to be variable when DNA from individually plaque purified stocks of a single virus strain was examined (63,17). When fragments which demonstrated intra-strain variation were subjected to detailed restriction endonuclease analysis, it became clear that the size differences were a result of very great variability within a small region rather than general changes throughout the fragments (15,17). Overall size variation within a particular fragment was sometimes the result of variability at more than one locus. Thus, the L-S joint fragment BamHl k from HSV-1 strain 17 is highly heterogeneous as a consequence of variation in at least two separate loci (17).

Analysis of the DNA sequence of cloned HSV-1 DNA fragments has revealed a common cause for intra-strain variability. All such regions examined to date contain tandem reiterations of short sequences of between approximately 10 and 30 bp. A number of the



I	C C G C T C C T C C C C	12
	C C G C T C C T C C C *	11
	C C G C T C C T C C C *	24
	T C T G T G G G T G G G G	
II	C C C T C C C C A G C C C C A G	16
	C C C T C C C C G G C C C C A G	16
III	C C G C C C C T C G C C C C C T C	17
IV	C C C C C T C C T C C G C C C C C G C G T C	22
	C C C C C T C C T C C A C C C C C G C G T C	22
V	A C T C C C A C G C A C C C C C	16

FIGURE 10. Reiterated sequences present in HSV-1 strain 17 DNA. The locations of five tandem reiterations are shown with respect to IR_L and IR_S and the five major IE mRNAs. Reiteration I lies within the a sequence and reiteration IV lies within the intron of IE mRNAs-4 and -5. When more than one sequence is given it indicates that more than one form of the reiteration has been described (17,55). Asterisked sequences indicate reiterations present in strain USA-8.

best characterized examples and their locations in the HSV-1 genome are shown in Figure 10. Those shown are from IR_L/TR_L and IR_S/TR_S, but others have been identified in U_S (D. McGeoch, personal communication). Size variation in these regions is caused by changes in the number of copies of the short sequences. The variation in copy number is probably a consequence of recombination between misaligned families of reiterations. Our recent identification of a reiteration of a 27 bp unit in IR_S/TR_S of VZV suggests that such sequences are common to the genomes of other alphaherpesvirinae.

The function of tandem reiterations is unknown. Mocarski and Roizman (62) have suggested that they may function as binding sites for viral proteins, and that this could play a role in segment inversion. However, it is unlikely that reiteration I, which is located in the a sequence, functions thus, since the HSV-2 strain HG52 a sequence contains no equivalent reiteration (17). Reiterations have the potential to act as "hot spots" for recombination which could facilitate rapid exchange of genetic information and would increase the apparent genetic distance between loci flanking such sequences. At present insufficient genetic markers are available in regions containing reiterations to test whether this is so. Reiterations which are present at equivalent locations in the HSV-2 genome also might provide sites for high frequency intertypic recombination. Evidence against this hypothesis stems from the observation that HSV-1 reiteration 4 and the reiteration in an equivalent location in the HSV-2 genome are markedly dissimilar in sequence (J.L. Whitton and J.B. Clements, personal communication). Rapid divergence of reiterations may constitute a mechanism by which intertypic recombination is minimized.

Smith (64) has suggested that tandem reiterations will accumulate as a result of non-homologous recombination in any region of DNA which is not subject to selection pressure. The reiterations shown in Figure 10 lie outside coding regions and may fall into this category.

Transcript mapping

The large size of herpesvirus genomes is an indication of their considerable coding capacity, which makes the analysis of their expression a considerable undertaking. Mapping of HSV mRNA was done initially by Southern blot analysis, and permitted mRNA from different temporal and size classes to be ascribed to approximate locations in the genome. This method proved particularly useful in examining the HSV-1 immediate-early (IE) class of mRNA (viral mRNA made in the absence of de novo protein synthesis), and allowed the map locations and orientations of the five major IE mRNAs to be determined (65,66). However, Southern blot analysis is of limited use when large numbers of transcripts are present as it does not allow characterization of individual mRNAs. It is particularly limited where overlapping and spliced mRNAs are concerned.

The Northern blot technique (67), in which RNA is separated by gel electrophoresis and immobilised on nitrocellulose filters, is more powerful as it allows detailed analysis of viral mRNA from specific regions of the genome, using cloned DNA fragments. DNA probes may be of any length and may represent either or both DNA strands. The advantages of this technique in studying individual mRNA species from a heavily transcribed region of the genome are illustrated by examining the temporal transcription patterns of HSV-1 IE genes (Figure 11). The behaviour of four of the five major IE genes was determined by analysis of mRNA made under IE conditions and at 3 hr and 6 hr PI, using DNA probes containing U_S (EcoRI h) or the junctions of IR_L/TR_L with U_L (BamHI b and BamHI e). Only IE mRNAs were made under IE conditions, whereas at 3 hr and 6 hr PI increasingly complex transcription patterns were revealed. mRNA species corresponding in size to the IE mRNAs were present at 3 hr and 6 hr PI, suggesting that the IE genes continued to be expressed. To examine if this was the case, 6h Northern blot strips were reprobed with smaller fragments from the regions known to encode the IE mRNAs. Probing with BamHI n, x and z confirmed the presence at 6 hr of the 2 kb mRNAs thought to represent IE mRNAs-4 and -5 (Figure 11). Use of single-stranded DNA probes cloned in bacteriophage M13 showed that the 2 kb mRNAs

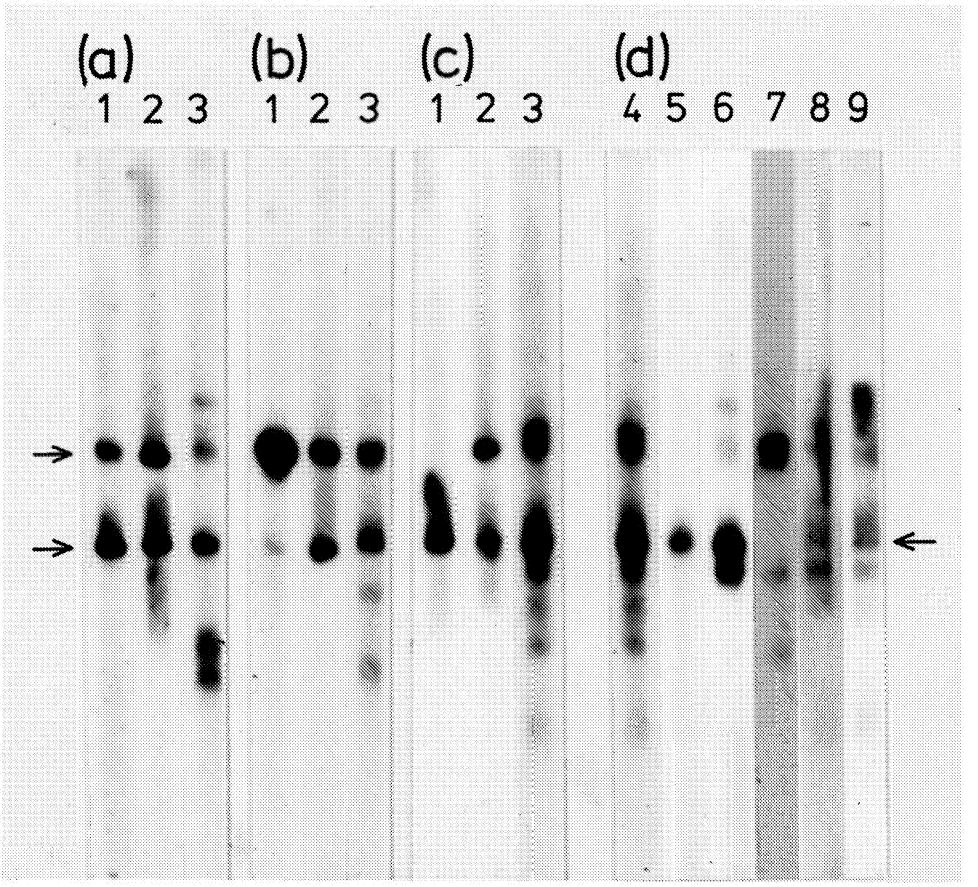


FIGURE 11. Northern blot analysis of HSV-1 mRNA. Tracks 1, 2 and 3 represent IE, 3 hr and 6 hr mRNA respectively. The DNA probes used were: panel (a) BamHI b; panel (b) BamHI e; panel (c) EcoRI h. Panel (d) shows 6 hr mRNA probed with the following cloned DNA fragments: EcoRI h (track 4); the left (track 5) and right (track 6) HindIII sub-fragments of BamHI n; BamHI j (track 7); BamHI z (track 8); BamHI x (track 9). The 3 kb (IE mRNA-1) and 1.75 kb (IE mRNAs-2, -4, -5) mRNAs are indicated.

present at 3 hr and 6 hr PI had the same orientations as the equivalent IE mRNAs.

Transcription patterns and the behaviour of individual mRNA species may be examined over large regions of a genome by Northern blot analysis. However, in order to map precisely mRNAs on the genome and to determine their structures it is necessary to use

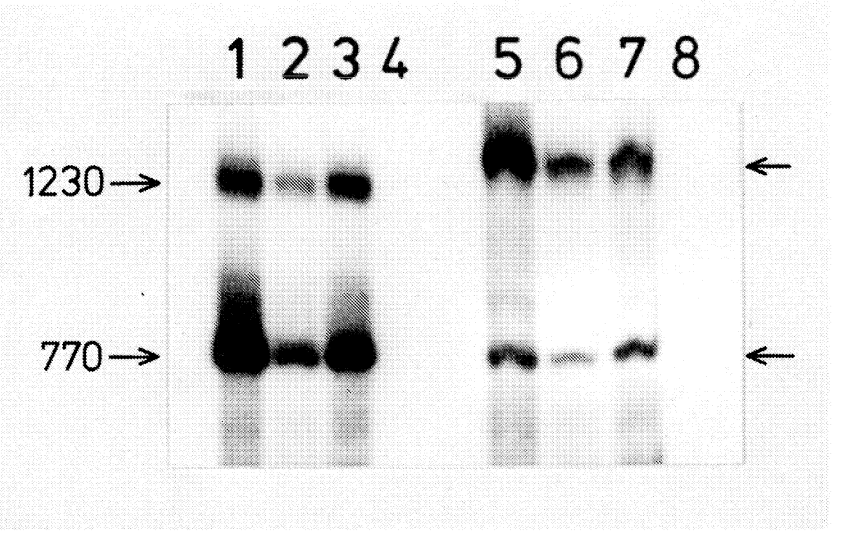


FIGURE 12. Temporal transcription pattern of IE gene 5. 5'-labelled BamHI x was hybridised to: 1&5, 15 µg of IE mRNA; 2&6, 15 µg of 3 hr mRNA; 3&7, 15 µg of 6 hr mRNA; 4&8, 20 µg of mock-infected mRNA. Samples 1-4 were nuclease S1 digested, samples 5-8 were exonuclease VII digested. The nuclease-resistant material was analysed on a 6% denaturing polyacrylamide gel.

the nuclease digestion techniques of Berk and Sharp (68). The termini and splice points of individual mRNAs may be located using DNA probes which have been end-labelled at specific sites. This is particular useful where different mRNAs are generated from the same region of the virus genome by differential splicing or by independent initiation and termination events which result in overlapping mRNAs. The temporal transcription pattern of IE gene 5 was examined by this method. Nuclease S1 and exonuclease VII digestion of hybrids formed between BamHI x 5'-labelled at the BamHI sites and mRNA made under IE conditions and at 3 hr and 6 hr PI resulted in the appearance of two bands on a denaturing gel (Figure 12). These represent the unspliced and spliced forms of IE mRNA-5 (70). The relative abundance of the two forms was constant in the different mRNA preparations, indicating that there was no alteration in splicing pattern during infection. The absolute concentration did change, however, and the increase between 3 hr and 6 hr PI indicates that synthesis of IE mRNA-5 was

maintained until 6 hr PI. The observation that IE polypeptides were translated *in vitro* from mRNA made at late times PI (66,69) confirms the continued presence of IE mRNA. This is in sharp contrast to the results of Watson *et al.* (71), who found that both IE mRNAs-4 and -5 were present in very low amounts in mRNA made at early or late times PI. It is probable that continued expression of these genes represents a variable aspect of virus infection and is dependent on the virus/cell system used.

REFERENCES

1. MATTHEWS, R.E.F. (1982) Intervirology 17, 1-199.
2. SHELDRICK, P. & BERTHELOT, N. (1974) Cold Spring Harbor Symp. quant. Biol. 39, 667-678.
3. WADSWORTH, S., JACOB, R.J. & ROIZMAN, B. (1975) J. Virol. 15, 1487-1497.
4. HAYWARD, G.S., JACOB, R.J., WADSWORTH, S.C. & ROIZMAN, B. (1975) Proc. natn. Acad. Sci. USA 72, 4243-4247.
5. DELIUS, H. & CLEMENTS, J.B. (1976) J. gen. Virol. 33, 125-133.
6. CLEMENTS, J.B., CORTINI, R. & WILKIE, N.M. (1976) J. gen. Virol. 30, 243-256.
7. WILKIE, N.M. & CORTINI, R. (1976) J. Virol. 20, 211-221.
8. WILKIE, N.M., CORTINI, R. & CLEMENTS, J.B. (1977) J. Antimicrobial Chemotherapy 3 (Suppl. A), 47-62.
9. SKARE, J. & SUMMERS, W.C. (1977) Virology 76, 581-595.
10. CORTINI, R. & WILKIE, N.M. (1978) J. gen. Virol. 39, 259-380.
11. ROIZMAN, B., JACOB, R.J., KNIPE, D., MORSE, L.S. & RUYECHAN, W.T. (1979) Cold Spring Harbor Symp. quant. Biol. 43, 809-826.
12. GRAFSTROM, R.H., ALWINE, J.C., STEINHART, W.L. & HILL, C.W. (1974). Cold Spring Harbor Symp. quant. Biol. 39, 679-681.
13. GRAFSTROM, R.H., ALWINE, J.C., STEINHART, W.L., HILL, C.W. & HYMAN, R.W. (1975) Virology 67, 144-157.
14. WADSWORTH, S., HAYWARD, G. & ROIZMAN, B. (1976) J. Virol. 17, 503-512.
15. WAGNER, M.J. & SUMMERS, W.C. (1978) J. Virol. 27, 374-387.
16. KUDLER, L. & HYMAN, R.W. (1979) Virology 92, 68-81.
17. DAVISON, A.J. & WILKIE, N.M. (1981) J. gen. Virol. 55, 315-331.
18. DUMAS, A.M., GEELEN, J.L.M.C., WESTSTRATE, M.W., WERTHEIM, P. & VAN DER NOORDAA, J. (1981) J. Virol. 39, 390-400.
19. ECKER, J.R. & HYMAN, R.W. (1982) Proc. natn. Acad. Sci. USA 79, 156-160.
20. STRAUS, S.E., OWENS, J., RUYECHAN, W.T., TAKIFF, H.E., CASEY, T.A., VANDE WOUDE, G.F. & HAY, J. (1982) Proc. natn. Acad. Sci. USA 79, 993-997.
21. STEVELY, W.S. (1977) J. Virol. 22, 232-234.
22. POWELL, D.J. (1979) PhD thesis, University of Glasgow.
23. BEN-PORAT, T., RIXON, F.J. & BLANKENSHIP, M.L. (1979). Virology 95, 285-294.

24. LONSDALE, D.M., BROWN, S.M., SUBAK-SHARPE, J.H., WARREN, K.G. & KOPROWSKI, H. (1979) J. gen. Virol. 43, 151-171.
25. TIMBURY, M.C. (1971) J. gen. Virol. 13, 373-376.
26. KAPLAN, A.S. & VATTER, A.E. (1959) Virology 7, 394-407.
27. BROWN, S.M., RITCHIE, D.A. & SUBAK-SHARPE, J.H. (1973) J. gen. Virol. 18, 329-346.
28. MARSDEN, H.S., CROMBIE, I.K. & SUBAK-SHARPE, J.H. (1976) J. gen. Virol. 31, 347-372.
29. WILKIE, N.M. (1973) J. gen. Virol. 21, 453-467.
30. BEN-PORAT, T., DEMARCHI, J.M. & KAPLAN, A.S. (1974) Virology 61, 29-37.
31. DUMAS, A.M., GEELEN, J.L.M.C., MARIS, W. & VAN DER NOORDAA, J. (1980) J. gen. Virol. 47, 233-235.
32. TWIGG, A.J. & SHERRATT, D.J. (1980) Nature, London 283, 216-218.
33. TANAKA, T. & WEISBLUM, B. (1975) J. Bacteriol. 121, 354-362.
34. RAO, R.N. & ROGERS, S.G. (1979) Gene 7, 79-82.
35. COHEN, S.N., CHANG, A.C.Y. & HSU, L. (1972) Proc. natn. Acad. Sci. USA 69, 2110-2114.
36. BOYER, H.W. & ROULLAND-DUSSOIX, D. (1969) J. molec. Biol. 41, 459-472.
37. GRUNSTEIN, M. & HOGNESS, D.S. (1975) Proc. natn. Acad. Sci. USA 72, 3961-3965.
38. HOLMES, D.S. & QUIGLEY, M. (1981) Anal. Biochem. 114, 193-197.
39. KOMANO, T. & SINSHEIMER, R.L. (1968) Biochim. biophys. Acta 155, 295-298.
40. CLEWELL, D.B. & HELINSKI, D.R. (1970) Biochemistry 9, 4428-4440.
41. WILKIE, N.M., CLEMENTS, J.B., BOLL, W., MANTEI,N., LONSDALE, D. & WEISSMAN, C. (1979) Nucleic Acids Research 7, 859-877.
42. OTSUKA, A. (1981) Gene 13, 339-346.
43. STRAUS, S.E., HAY, J., SMITH, H. & OWENS, J. (1983) J. gen. Virol. 64, 1031-1041.
44. PRESTON, V.G., DAVISON, A.J., MARSDEN, H.S., TIMBURY, M.C., SUBAK-SHARPE, J.H. & WILKIE, N.M. (1978) J. Virol. 28, 499-517.
45. DAVISON, A.J. & SCOTT, J.E. (1983) J. gen. Virol., in press.
46. PLUMMER, G. (1964) Brit. J. exp. Pathol. 45, 135-141.
47. TIMBURY, M.C. & SUBAK-SHARPE, J.H. (1973) J. gen. Virol. 18, 347-357.
48. HONESS, R.W., POWELL, K.L., ROBINSON, D.J., SIM, C. & WATSON, D.H. (1974) J. gen. Virol. 22, 159-169.
49. KIEFF, E.D., HOYER, B., BACHENHEIMER, S.L. & ROIZMAN, B. (1972). J. Virol. 9, 738-745.
50. WATSON, D.H., WILDY, P., HARVEY, B.A.M. & SHEDDEN, W.I.H. (1967) J. gen. Virol. 1, 139-141.
51. CAUNT, A.E. & SHAW, D.G. (1969) J. Hygiene, Cambridge 67, 343-352.
52. LUDWIG, H.O., BISWAL, N. & BENYESH-MELNICK, M. (1972) Virology 49, 95-101.
53. RAND, T.H. & BEN-PORAT, T. (1980) Intervirology 13, 48-53.
54. DAVISON, A.J. & WILKIE, N.M. (1983) J. gen. Virol., in press.
55. MURCHIE, M-J. & McGEOCH, D.J. (1982) J. gen. Virol. 62, 1-15.
56. YEO, J., KILLINGTON, R.A., WATSON, D.H. & POWELL, K.L. (1981) Virology 108, 256-266.

57. LITTLER, E., YEO, J., KILLINGTON, R.A., PURIFOY, D.J.M. & POWELL, K.L. (1981) J. gen. Virol. 56, 409-419.
58. MOCARSKI, E.S. & ROIZMAN, B. (1982) Cell 31, 89-97.
59. DAVISON, A.J. & WILKIE, N.M. (1983) J. gen. Virol. 64, 1-18.
60. JACOB, R., MORSE, L. & ROIZMAN, B. (1979) J. Virol. 29, 448-457.
61. MAXAM, A.M. & GILBERT, W. (1980) Methods in Enzymology 65, 499-560.
62. MOCARSKI, E.S. & ROIZMAN, B. (1981) Proc. natn. Acad. Sci. USA 78, 7047-7051.
63. LONSDALE, D.M., BROWN, S.M., LANG, J., SUBAK-SHARPE, J.H., KOPROWSKI, H. & WARREN, K.G. (1980) Ann. N. Y. Acad. Sci. 354, 291-308.
64. SMITH, G.P. (1976) Science 191, 528-535.
65. WATSON, R.J., PRESTON, C.M. & CLEMENTS, J.B. (1979) J. Virol. 31, 42-52.
66. ANDERSON, K.P., COSTA, R.H., HOLLAND, L.H. & WAGNER, E.K. (1980) J. Virol. 34, 9-27.
67. THOMAS, P.S. (1980) Proc. natn. Acad. Sci. USA 77, 5201-5205.
68. BERK, A.J. & SHARP, P.A. (1978) Proc. natn. Acad. Sci. USA 75, 1274-1278.
69. LEE. G.T.-Y., PARA, M.F. & SPEAR, P.G. (1982) J. Virol. 43, 41-49.
70. RIXON, F.J. & CLEMENTS, J.B. (1982) Nucleic Acids Research 10, 2241- 2255.
71. WATSON. R.J., COLBERG-POLEY, A.M., MARCUS-SEKURA, C.J., CARTER, B.J. & ENQUIST, L.W. (1983) Nucleic Acids Research 11, 1507-1522.

8

THE CLONING AND SEQUENCING OF SITES OF LINKAGE BETWEEN ADENOVIRUS DNA AND CELLULAR DNA: RECOMBINATION OF FOREIGN DNA WITH THE MAMMALIAN GENOME

WALTER DOERFLER, REINHOLD GAHLMANN, SILVIA STABEL, RENATE DEURING, MANFRED SCHULZ, URSULA LICHTENBERG, DIRK EICK, ROLF JESSBERGER, AND REINER LEISTEN
Institute of Genetics, University of Cologne, Cologne, Germany

INTRODUCTION

We have studied the mechanism of recombination between human adenovirus DNA and the mammalian genome in considerable detail. Human adenovirus type 2 (Ad2) and type 12 (Ad12) have been investigated. This research was conducted as a model, as it were, to elucidate the mechanism of insertion of foreign DNA into the mammalian genome. There is no doubt that practically any foreign DNA molecule can be taken up by mammalian cells and can be inserted as a whole or in part into the host genome. It is not clear yet how frequent these integration events are. Insertion could be transient in certain cases, and stable integration might be a relatively rare event. As often in molecular biology of eukaryotes, adenoviruses have served a very useful function in the investigation of reaction mechanisms (for recent reviews see 1-3). At least certain aspects of the mechanism of insertion of foreign (viral) DNA into the mammalian genome can be investigated by studying the modes of integration of adenovirus DNA. In previous work, we have used clonal lines of adenovirus-transformed cells or of adenovirus type 12-induced tumor cells to determine patterns of adenovirus DNA integration in mammalian cells and to gauge the gamut of possibilities for foreign DNA insertion (for reviews 4, 5). The clonality of the cell lines used facilitates detailed analyses of the sites of adenovirus DNA insertion by molecular cloning and determinations of their nucleotide sequences. This approach, however, forecloses the possibility to

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

study all the recombination events actually occurring, since transformed or tumor cells represent a highly selected cell population. This possibility could be realized, in principle though in practice only with difficulties, by choosing the logical way and by studying viral-host DNA recombinations in abortively (6, 7) or productively infected cells (8-10). In these systems many different recombination products are generated, and there are considerable difficulties in analyzing them in detail. So far, we have therefore resorted to a compromise solution and elucidated patterns of adenovirus DNA integration in some 70 different transformed cell lines, tumors or cell lines established from these tumors (4). Moreover, we have cloned nine and sequenced seven sites of junction between adenovirus and cell DNA derived from several different cell systems (5) including the junction from a symmetric recombinant (SYREC2) between Ad12 DNA and human cell DNA (10, 11).

For a number of reasons, we consider it important to investigate in detail the cellular mechanism involved in the uptake and fixation of foreign DNA in mammalian cells. Acquisition of foreign DNA sequences may have played a decisive role in evolution and continues to do so in inducing significant genetic change in eukaryotic cells. Moreover, the mechanism at least of viral oncogenesis involves foreign DNA insertion. Lastly, the application of genes to cells, one of the paramount goals in gene technology, particularly the insertion of foreign genes at specific sites in the mammalian genome, will require more detailed research on the mechanisms involved in the fixation of foreign DNA.

For the insertion of adenovirus DNA into host DNA homologous or heterologous recombination mechanisms could be relevant. The data accumulated so far would tend to favor a mechanism akin to heterologous recombination which can utilize patch homologies between segments of the viral genome and many different sequences in the host genome. The mechanism of insertion does not appear to be dependent on the occurrence of such patch homologies.

The main results presented in this overview can be summarized as follows. In the adenovirus-transformed cell lines, Ad12-in-

duced tumors or tumor lines established from them, Ad12 DNA was usually inserted intact or nearly intact (see below) and colinear with the virion DNA. Integrated Ad2 DNA, on the other hand, carried internal deletions. This apparent difference may reflect the permissivity of the hamster or mouse cell systems used. Hamster cells are nonpermissive for Ad12. At the junction sites investigated by sequence analyses, the adenovirus genomes had recombined exclusively at their termini with host DNA. Similar conclusions had been suggested by Southern blot analyses of the integration patterns of adenovirus DNA in about 70 different transformed lines or tumors (for review, see reference 4). Insertion of adenovirus DNA has been observed in unique or repetitive cellular DNA sequences. At least for some of the junctions, the data justify the conclusion that the mechanism of adenovirus DNA insertion is different from that of bacteriophage λ DNA in Escherichia coli and of retroviral DNA in vertebrate genomes. Insertion of adenovirus DNA is accompanied by deletions of between 2 and 174 nucleotides at the viral termini (5). In cell line HE5, not a single cellular nucleotide was deleted, duplicated or rearranged at the integration site. In another instance (tumor CBA-12-1-T), at least 1500 to 1600 cellular nucleotides were deleted. We have also investigated a symmetric recombinant (SYREC2) between Ad12 DNA (left terminal 2081 nucleotide pairs) and human cellular DNA. This recombinant is encapsidated into Ad12 virions. The occurrence of this recombinant constitutes proof for recombination between viral and cellular DNA in productively infected cells (8, 9).

SUMMARY OF EARLIER FINDINGS

The mode of insertion of the adenoviral genome is influenced by the permissivity of the virus-cell system. As pointed out previously, the Ad12 genome is usually inserted in the hamster or mouse genome without internal deletions (12-16), although in a few Ad12-induced tumors, deletions of parts of the Ad12 genome have been observed (17). Ad12 DNA cannot replicate in hamster cells (6, 12, 18, 19, 20) and in mouse cells (16). In contrast, the patterns of integration of the DNAs of Ad2 and adenovirus

type 5 (Ad5)in hamster cells, which are permissive or at least semipermissive for Ad2 and Ad5, reveal extensive deletions of the viral genome (21-23). It is therefore reasonable to postulate that Ad12 DNA can persist intact or nearly intact in hamster cells, because the nonpermissive disposition of hamster cells toward Ad12 does not allow Ad12 DNA to replicate, and thus cells containing the entire Ad12 genome are not selected against. In Ad2- or Ad5-infected hamster cells, however, complete free viral genomes will replicate and eventually destroy the host cells. Transformants with intact integrated Ad2 or Ad5 genomes are not likely to be found, but in most cases the persisting viral genomes carry deletions. Of course, there may exist additional, more complex reasons for the differences in integration patterns found in Ad2- and Ad12-transformed cells.

The patterns of integration of highly oncogenic simian adenovirus type 7 (SA7) in virus-induced hamster tumor cells are complicated and imply that deletions and amplifications of the SA7 genome have occurred (E. Timme, H. Soboll, R. Neumann and W. Doerfler, unpublished results; compare also reference 24).

In general, multiple copies of adenoviral DNA were inserted in the host genome, in one cell line usually at the same site or at a very limited number of sites. The number of viral genome equivalents per cell ranged from one to two to over 30, both in cells that were transformed in culture and in Ad12-induced tumor cells isolated from tumor-bearing hamsters or mice. From several lines of evidence it appeared plausible to argue that one or a limited number of genomes were integrated initially and that viral DNA sequences, probably in conjunction with abutting cellular sequences, were postintegrationally amplified. Adenoviral DNA was found to be inserted into unique or repetitive cellular DNA sequences in different cell lines or tumors (25-27, 42). In a few Ad12-transformed cell lines, terminal sequences of the integrated adenoviral genomes were disproportionately amplified (14). Multiple inserted adenoviral genomes were usually not arranged in true tandems, with one entire genome being followed by another (14, 16, 17). In the cases analyzed in this respect, it was found that individual genomes were separated by sequences not identical

to authentic viral DNA. These sequences could have constituted cellular or possibly rearranged viral DNA. Deletions and rearrangements at the right termini of most of the 20-22 Ad12 genomes integrated in cell line T637 have been demonstrated (28).

From the analyses of over 70 different adenovirus-transformed and tumor cell lines investigated in our laboratory and from the results of several other laboratories, no evidence emerged for the notion that adenoviral DNA would insert at identical or similar sites in the host genome. Similar conclusions were derived from studies on the insertion of SV40 or polyoma DNA (29-35). Absence of specificity was documented by the results of Southern blot analyses and/or of determinations of nucleotide sequences across the sites of junctions.

From several Ad12-transformed or tumor cell lines, morphological revertants were isolated which had arisen spontaneously and had lost part or all of the integrated adenoviral genome copies. Cell line T637, which carried 20-22 genome equivalents of Ad12 DNA, lost all of these copies or retained one-half or one genome equivalent or one genome and a fraction of a second per revertant cell. The revertants were different from the parent line (36-38). From an Ad12-induced hamster tumor line, morphological variants were isolated that had gradually lost all detectable traces of viral DNA, but preserved their oncogenic phenotype, the absence of viral DNA notwithstanding. It was therefore concluded that - at least in these variants - persistence of adenoviral DNA was not an essential precondition for expression of the oncogenic phenotype (17). It could not be ruled out that a viral DNA fragment of extremely short length (10-100 nucleotides) might have persisted in these cells. The occurrence of the morphological revertants and variants simultaneously proved that foreign DNA could not only be fixed in cellular DNA but could also be excised from it more or less completely.

By reassociation kinetics measurements (12) as well as by Southern (39) blot analyses (14), scattered fragments of viral DNA could be detected in an integrated form in Ad12-tranformed cell lines in addition to the intact or nearly intact Ad12 DNA molecules. Apparently, apart from the nearly intact integrated

viral DNA molecules, some cell lines also contained fragments of viral DNA which might have been inserted, independent of the intact viral DNA molecules. It is not known how stably these scattered fragments of viral DNA are inserted in the cellular genome.

Analyses of nucleotide sequences at several sites of integration of adenovirus DNA in hamster, mouse, or rat DNA revealed that terminal viral nucleotides were deleted at the sites of junction (25-27, 40-42, see also Table 1). In the Ad2-transformed hamster cell line HE5, cellular nucleotides were not deleted at the site of insertion (41). In the mouse tumor CBA-12-1-T, cellular DNA was deleted at the site of junction (42). A detailed discussion of the structure and the peculiarities of each site of junction analyzed will be presented below.

The mechanism of foreign (viral) DNA insertion in mammalian cells is not yet understood. It is conceivable that patchy homologies between viral and cellular DNA sequences might have played a role in some recombination events (25-27, 33, 42, 43). Such short homologies, particularly certain combinations of nucleotide patches, might stabilize the recombination complex. Moreover, patch homologies between adenoviral DNA and cellular DNA are abundant and might help to explain the apparent lack of specificity in insertion sites. Depending on the combinations of patch homologies selected, insertion of foreign DNA might be facilitated at a very large number of sites. At some sites, integration of foreign DNA might be fatal for the cell. When nonessential sites in unique DNA or in repetitive DNA are hit by the insertion event, the cell might survive and even gain selective advantages from the added genetic information. Depending on the site of insertion, viral DNA sequences may be linked to cellular enhancers or vice versa and cells with viral DNA inserted at such sites may exhibit the transformed phenotype due to more efficient expression of viral or cellular DNA.

It will also be interesting to elucidate in what conformation the adenoviral genome can recombine with host DNA. Circular viral DNA - perhaps not a covalently linked circle, but one stabilized by the terminal viral protein (44) - has sometimes been impli-

cated in that precursory function (21, 23, 44). Evidence for the presence of covalently closed circular viral DNA molecules in abortively or productively infected cells has been difficult to obtain (45, 46). Recently, evidence has been published for adenoviral molecules which are joined end to end (47), but it is not yet certain that these molecules actually represent circular structures.

The expression of integrated viral genomes in transformed or tumor cell lines will not be dealt with in this chapter. Suffice it to say that, in general, early viral genes are expressed and late viral genes are shut off (19, 48-51), although in most Ad12-transformed hamster lines the late regions of the Ad12 genome persist in a perfectly preserved form. In general, these late regions are extensively methylated (13, 52-54). In Ad12-induced rat tumor cell lines, some of the late segments of the integrated Ad12-genomes can be expressed (55).

CLONING AND SEQUENCING OF JUNCTION FRAGMENTS

The problem of viral DNA insertion, in particular with respect to the possible specificity of integration sites and the mechanism of integration, could be approached by cloning sites of junctions between viral and cellular DNA from a number of transformed and tumor cell lines and by determining the nucleotide sequences at these sites. In general, the following procedures were employed:

1. The sizes of the junction fragments were determined as off-size fragments, usually after cleavage of transformed cell DNA or tumor cell DNA with different restriction endonucleases, blotting and hybridization to ^{32}P-labeled, cloned adenovirus DNA fragments from the termini of viral DNA. Off-size fragments representing host-virus DNA junctions were identified by comparison with the cleavage patterns of authentic viral DNA. This approach has been described in detail (13, 14), and the integration maps of many different adenovirus-transformed and tumor cell lines have been published (for review see reference 4). All these cell or tumor lines had been obtained by transforming cells with Ad2 or Ad12 or

by inducing tumors in animals with Ad12. Cells analyzed in our laboratory so far were not transformed with viral DNA fragments.

2. DNA fragments with a length corresponding to the off-size fragments to be cloned were selected by gel electrophoresis or by zone sedimentation on sucrose density gradients. The latter method was usually preferred. Gradient fractions were characterized by subsequent gel electrophoresis. Adenovirus DNA cut with the same restriction endonuclease was used as size marker and off-size positions were determined in this way.

3. The selected size-class of EcoRI-cut DNA was ligated with the "arms" of λgtWES·λB DNA (56) or λCharon DNA (57). The arms of the λ vector DNAs, i.e., the terminal EcoRI fragments of the vector, were purified by zone sedimentation on sucrose density gradients. The ligated DNA was then packaged in vitro into phage heads (58, 59). Phage plaques containing the junction fragment to be cloned were identified by the method of Benton and Davis (60), using cloned terminal viral DNA fragments as hybridization probes. It proved advantageous to perform the initial cloning of junction sites in λ vectors, since with that vector system large enough numbers of plaques per petri dish could be produced and screened. Comparable colony numbers could not have been attained using plasmids, since, upon insertion of foreign DNA into plasmids, transformation efficiencies sometimes dropped precipitously. A comparable drop in packaging and plaquing efficiencies was not observed with phage λ DNA. Usually, the desired fragments were detected as clones in about 0.5×10^6 plaques, equivalent to about 40-50 petri dishes. In some instances, much higher numbers of plaques were screened without detecting a cloned junction fragment. With DNA from the Ad12-transformed cell line T637, several millions of plaques were screened without isolating a single junction fragment. Internal fragments of adenoviral DNA could be readily cloned from T637 DNA (U. Winterhoff and W. Doerfler, unpublished). It has been shown that in this cell line peculiar DNA structures at the sites of junction between Ad12 and cell DNA and between Ad12 DNA copies probably forestall the cloning of the junction sites (61).

4. Frequently, suitable subfragments of the λ junction clone, still containing the site of linkage, were subcloned in pBR322 (62), pUR222 (63), pUR250 (64), or in bacteriophage M13 DNA (65), and were then analyzed further.

5. The cloned insert was characterized by restriction analyses using restriction endonucleases that frequently cut adenovirus DNA. Junction fragments were again identified by Southern blotting and by comparison with the authentic restriction pattern of adenovirus DNA. These junction fragments assumed off-size positions relative to the known virion DNA fragments. In most instances, the off-size fragments or parts thereof were subcloned in a suitable vector either directly or using appropriate linkers when necessary.

6. Subsequently, a detailed restriction map of the subcloned junction fragment was established. It was also ascertained that the subcloned fragment still contained cellular DNA by hybridizing the clone back to cellular DNA from transformed or untransformed cells of the same species. Moreover, it could be demonstrated that cloning and subcloning of the junction fragments did not lead to deletions or rearrangements of the DNA insert in the clones we have analyzed. Thus, we have not obtained evidence for cloning artefacts.

7. Employing the Maxam and Gilbert sequencing technique (66, 67) and terminal labeling at the 5' or 3' ends of distinct fragments, the nucleotide sequence across the sites of junction was determined. In most instances, it was considered necessary to sequence several hundred nucleotides into the cellular segment, as it was conceivable that decisive recombinatorial signals might have been situated remote from the actual site of junction. More recently, the method of Sanger et al. (68) was also used to determine nucleotide sequences.

8. A special approach was designed to clone certain fragments that did not fit into available λ phage vectors or in order to avoid the use of nucleotide linkers (R. Gahlmann and W. Doerfler, unpublished results). The DNA from transformed or tumor cells was cut with the restriction endonucleases PstI, or HindIII, or BamHI. The off-size fragment to be cloned was then size-selected

by zone sedimentation on sucrose density gradients and ligated into the appropriate site of plasmid pBR322 DNA. The resealed plasmid carrying the insert was subsequently cleaved with EcoRI and ligated to the arms of phage λgtWES·λB DNA. On in vitro packaging of λ DNA, appropriate plaques containing the junction fragment were identified by the Benton-Davis technique using terminal virion DNA fragments as probes.

CELL LINES AND JUNCTION SITES ANALYZED

Adenovirus-transformed cell lines and adenovirus-induced tumors or cell lines established from them constitute a source of clonal recombinants between viral and cellular DNA. As much as it would be desirable to investigate such recombinants shortly after their inception, sites of insertion could not easily be studied at that stage, as insertions occur at many different sites. Thus, it was more realistic to concentrate on the study of clonal lines. Eventually, we will have to return to investigations on cells shortly after infection with virions or subviral particles or after transfection with viral DNA or with cloned DNA fragments. Such experiments have been initiated.

In Table 1, the sites of junction between adenovirus DNA and cell DNA that have been cloned and sequenced in our laboratory have been summarized. Different adenoviruses and cells from different species were used in these investigations. In symmetric recombinant (SYREC2) DNA, which was found to be encapsidated into adenovirions, linkage between Ad12 DNA and human KB cell DNA was shown by sequence analysis (11). SYREC DNA was generated in productively infected human cells (10), and its existence constituted proof for the occurrence of recombination between the adenoviral and host genomes in productively infected cells (8, 9).

The sequence data from all the transformed and tumor cell lines represented unequivocal proof for the covalent linkage between adenovirus-DNA and cellular DNA (6). The findings that adenovirus DNA in infected and transformed cells was associated with high molecular weight DNA (6, 7, 69) and that adenovirus DNA

Table 1. Sequenced sites of junction between adenoviral and cellular DNAs

No.	Cell line	Species	Adenovirus	Copy number	Terminus	Cellular sequence	Viral nucleotides deleted
1	CLAC3	hamster	Ad12	4 - 5	left	unique	45
2	CLAC1	hamster	Ad12	10 - 13	left	repetitive	174
3	HE5	hamster	Ad2	2 - 3	right	unique	8
4	HE5	hamster	Ad2	2 - 3	left	unique	10
5	SYREC - Ad12-KB cell recombinant			-	(left)	repetitive	-
6	CBA-12-1-T	mouse	Ad12	> 30	left	unique	9
7	T1111(2)	hamster	Ad12	10 - 11	left	unique	64

was linked to unique (25, 26, 42) or repetitive DNA sequences (27) argued for the chromosomal location of inserted viral DNA.

In situ DNA hybridization analyses have been performed in a CELO virus (avian adenovirus)-induced rat tumor cell line. The tumor cells carry about 33 % of the viral DNA molecule linked to cellular DNA. This viral-cellular sequence has been repeated 160 times and is distributed on only a few chromosomes per hypotetraploid tumor cell (70).

In the following section, the nucleotide sequences of individual integration sites determined in our laboratory will be presented and their analyses discussed. These sequence data did not reveal evidence for specific integration sites based on nucleotide sequence, in the sense that adenoviral DNA integrated at a few highly specific sequences in all cell lines. Certain regularities could, however, be observed; e. g., in all lines investigated in this respect, the viral DNA was linked via its termini to cellular DNA and viral nucleotides were deleted at the site of junction. Only in an exceptional case was Ad2 DNA linked to cell DNA via internal fragments (23). As pointed out previously, we consider it likely that the terminal protein of adenovirus DNA or terminal viral DNA sequences or both can play a role in integration. It has also to be considered that signals of possible specificity for the integration event might not reside in nucleotide sequence, but rather in the structure of chromatin or of protein-DNA complexes. At the present stage of technical developments, the possibilities to unravel such structural signals are very limited. Hence any statements with respect to specific sites of viral-cellular recombinations, or the lack thereof, will have to be phrased with caution at the present time. The concept of specificity of integration sites has to be viewed in the light of highly complex interactions between viral and cellular genomes.

NUCLEOTIDE SEQUENCES AT INDIVIDUAL SITES OF JUNCTION

Ad12-induced hamster tumor line CLAC3

The tumor cell line CLAC3 (14) carries four to five Ad12 genome equivalents per diploid genome. Using the λgtWES.λB vector

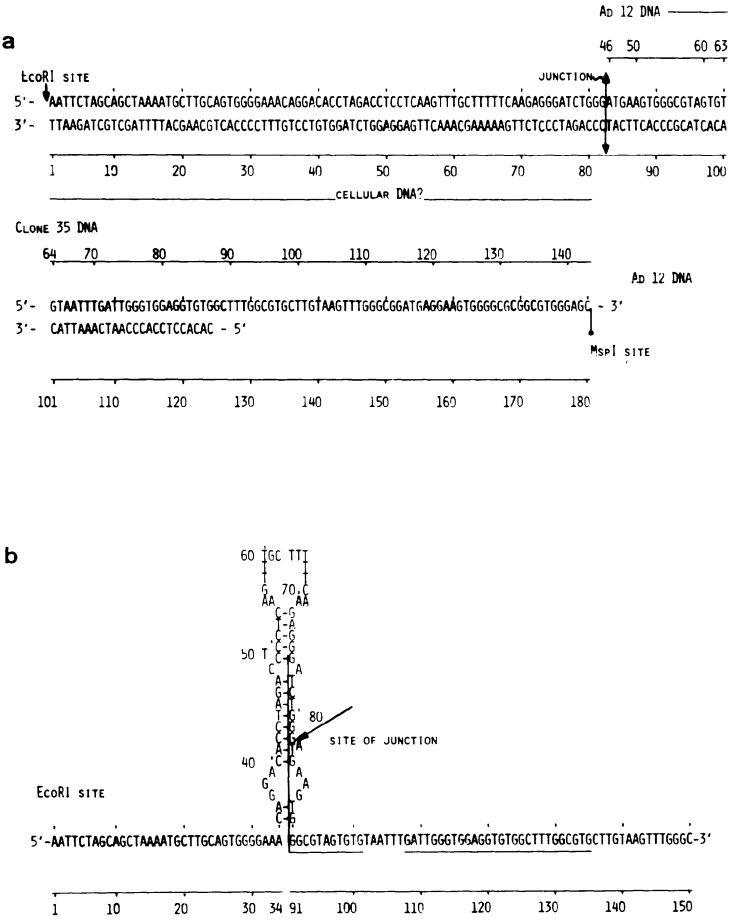


FIGURE 1. Junction site between Ad12 DNA and hamster cell DNA in cell line CLAC3 (25). a) Nucleotide sequence at site of linkage. b) Region of dyad symmetry and possible stemmed loop at site of junction. There is at present no proven way to ascertain the actual occurrence of such structures in vivo.

DNA as cloning vehicle, a left terminal site of junction between Ad12 DNA and hamster cell DNA was cloned and sequenced (25). The colinearly inserted Ad12 genome exhibited the authentic virion DNA sequence starting with base pair 46 (Fig. 1a). The first 82 base pairs of the cloned fragment were not viral and contained scattered homologies of octa- to undecanucleotide patches to sequences within the left terminal 2722 base pairs of Ad12 DNA. The phenomenon of patch homologies and its significance will be discussed below. No homologies were observed between the deleted string of the 45 left terminal nucleotides of viral DNA and the cellular sequence replacing them (cf. however, line CLAC1). At the site of junction, a stemmed loop could be constructed, based on extensive regions of dyad symmetry (Fig. 1b). Across the site of junction and close by in the viral DNA sequence, there were stretches of 27 base pairs that exhibited 70% homology (underlined sequences in Fig. 1b). The significance of these peculiarities for the recombination event is not understood.

Ad2-transformed hamster cell line HE5

The cell line HE5 (23, 71, 72) was generated by transformation of primary hamster embryo cells with UV-inactivated Ad2 (72). Each cell contained approximately two copies of Ad2 DNA, and the viral DNA exhibited a major internal deletion (23) extending from about 35 to 82 map units of Ad2 DNA (41). The junction sites of both termini of Ad2 DNA with hamster cell DNA, as well as the site of linkage of the two Ad2 DNA fragments at the internal joint, were cloned and sequenced (26, 41). The unoccupied site of cellular DNA from cell line HE5 corresponding to the insertion site was also cloned and sequenced (26). This site was unique or occurred only a few times per cell. The results indicated that there was an almost perfect insert of the Ad2 DNA molecule via the termini. Not a single cellular nucleotide pair was lost, altered, or repeated in the integration event, while eight and ten nucleotide pairs of Ad2 DNA were deleted at the right and left termini, respectively. Figure 2 presents a survey of nucleotide sequences at the site of insertion. The cellular DNA se-

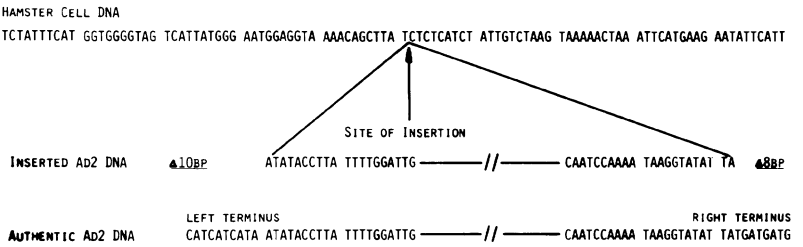


FIGURE 2. Insertion of Ad2 DNA into cellular DNA in cell line HE5 (41). Insertion without deletion, amplification or rearrangement of cellular nucleotides. Δ Indicates the deletion of 10 left terminal and 8 right terminal adenoviral nucleotides at the site of junction. The viral genome is schematically presented by its terminal sequences and an interrupted line designating the remainder of the genome with an internal deletion between map units 35 and 82.

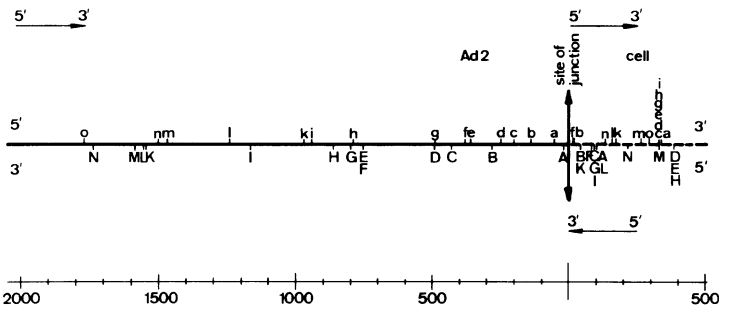


FIGURE 3. Maps of patches of homologies between the right terminus of Ad2 DNA and the abutting hamster cell DNA sequence in cell line HE5 (26). Scale at the bottom indicates number of nucleotides from junction site. Letters refer to individual nucleotide patches (not shown).

quence and the abutting viral sequences at either end exhibited short patch homologies.

Patch homologies ranging in length from dodeca- to octanucleotides were detected by computer analyses at sites also quite remote from the points of linkage. Patches exceeding octanucleotides in length detected at the right terminus have been mapped as shown in Fig. 3. The nucleotide sequence at the right arm of the Ad2 genome was compared with randomly selected sequences of 401 nucleotides in length from vertebrate or prokaryotic DNA. In all cases, similar numbers of patch homologies were observed, indicating that patch homologies of up to 12 nucleotides long were abundant between Ad2 DNA and randomly chosen DNA sequences (26). It is conceivable that such homologies play a role in the insertion event, perhaps stabilizing the recombination complex. The abundance of patch homologies, which were quite different in sequence in each case, was consistent with the apparent lack of specificity with respect to integration sites of adenovirus DNA

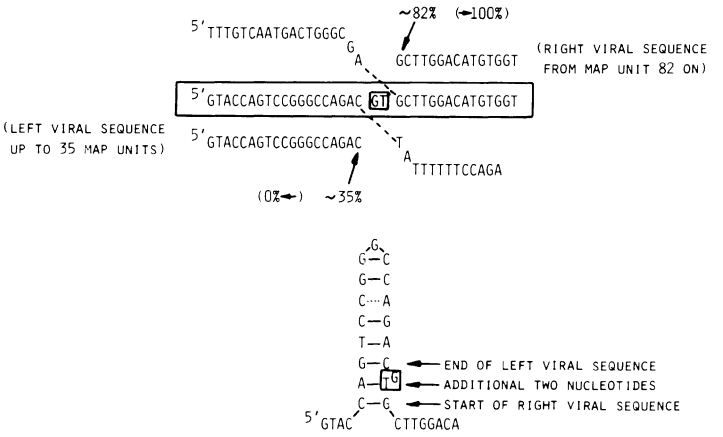


FIGURE 4. Internal viral junction site in cell line HE5 (41). Sequence at the junction site is boxed. The additional GT dinucleotide is also indicated. Sequences outside the box and connected by broken lines are the immediately adjacent viral sequences that have been deleted. A region of dyad symmetry is apparent at the site of linkage (bottom).

(4), since many different sites could have been chosen in this way. Patch homologies thus may help the recombination process, they do not seem to be essential. Recombination via patchy sequence homologies might be considered a mechanism in between homologous and heterologous recombination.

The internal junction between the right and left arms of Ad2 DNA remaining after the deletion was also cloned and sequenced (41) (Fig. 4). Comparison of the sequence data with those of authentic Ad2 DNA (73, 74) revealed that map position 35 from the left end was linked to map position 82 from the right end. A dinucleotide, GT, of unknown derivation was interspersed between the two arms of Ad2 DNA (Fig. 4). Intercalation of nucleotides between recombining molecules has also been described in somatic CV1 monkey cells, in which SV40 and pBR322 DNA molecules were rejoined (75).

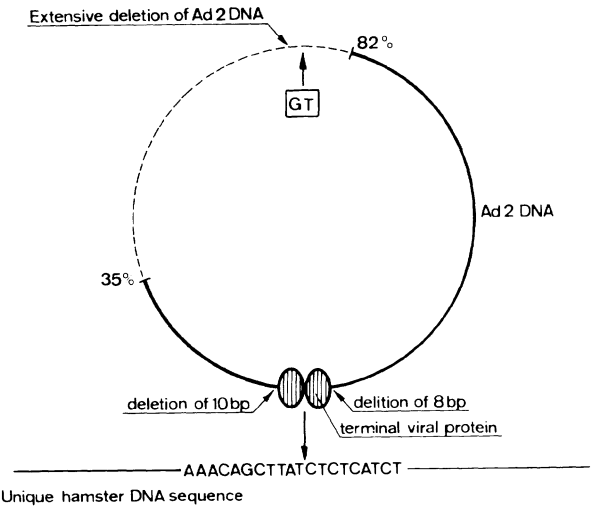


FIGURE 5. Hypothetical model of Ad2 DNA insertion (41). A circular intermediate has not been proven. In fact, a linear intermediate is equally likely. This model ascribes a guiding function in recombination to the terminal viral protein, or to terminal DNA sequences, or to both. Other models cannot be ruled out.

It is also interesting to note that at the site of junction between the two arms a short inverted repeat can give rise to a cruciform structure (Fig. 4).

The apparent mode of insertion of adenovirus DNA preserving the continuity of cellular DNA sequence in this cell line and involving linkage to cellular DNA by the termini of viral DNA suggests a model of foreign DNA insertion in which both termini of Ad2 DNA would have to interact with adjacent nucleotides in cellular DNA either simultaneously or in succession. One possible, but unproven intermediate structure of viral DNA participating in this event would be a circular adenovirus DNA molecule in which the termini were linked together noncovalently via the terminal viral protein (41). It is not clear at what stage the deletion was introduced into Ad2 DNA. Such a deletion could have occurred before, during, or after the integration event. In the transformation experiment that led to cell line HE5, Ad2 virions irradiated with ultraviolet light had been used. This procedure might predispose the genome to deletion events and might also stimulate recombination. A possible model of the integration event is presented in Fig. 5.

We have also investigated RNA with homologies to the cellular DNA sequence at the junction site (76). The hamster cell DNA sequence of about 400 nucleotides to the right of the integrated Ad2 DNA is homologous to low molecular weight RNAs (Fig. 6). The predominant size class is an RNA population of about 300 nucleotides which is present in about 20 copies per cell. Minority populations of about 150 and possibly 80 nucleotides are also detectable. This low molecular weight cellular RNA is not polyadenylated and is found predominantly in the cytoplasm. The nucleotide sequence of the DNA segment homologous to this RNA does not contain open reading frames in excess of a sequence theoretically encoding an 18 amino acid polypeptide. Thus, it is unlikely that the low molecular weight RNA can be translated. The nucleotide sequence does not exhibit any similarity to known low molecular weight RNAs of eukaryotic origins. The low molecular weight cellular RNA has been found expressed in HE5 cells and also in various organs of 15 - 16 week old animals and in embryo-

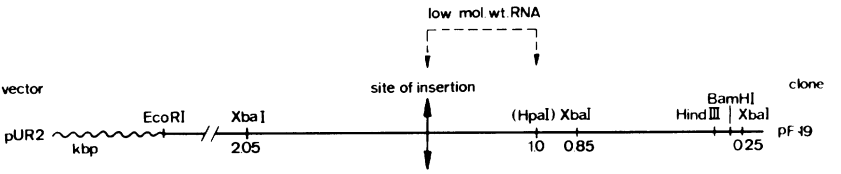


FIGURE 6. Restriction map of the site of insertion of Ad2 DNA in cell line HE5. Site of insertion has been designated by a double-headed arrow. The bracket indicates the region of homology to the low molecular weight, cellular RNA (76).

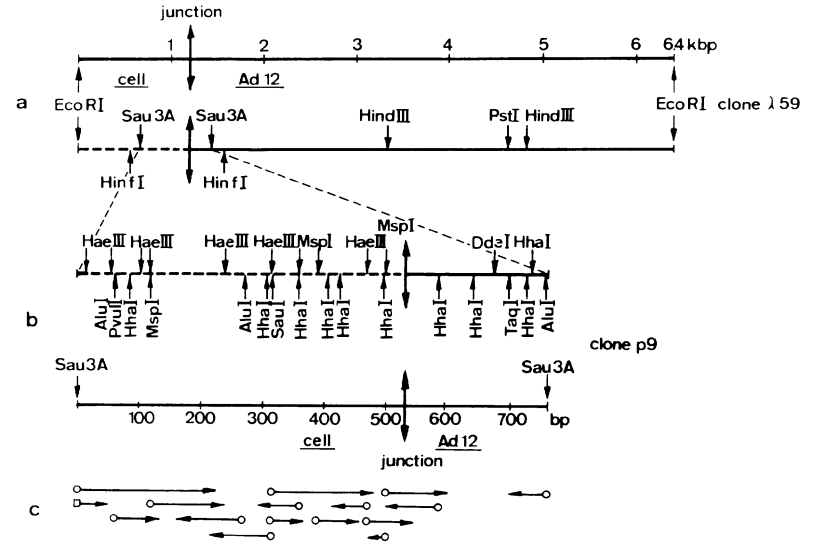


FIGURE 7. Restriction maps (a, b) and sequencing strategy (c) at site of junction between Ad12 DNA and repetitive hamster cell DNA in cell line CLAC1 (27). The site of junction is designated by a double-headed vertical arrow.

nal tissues, as well as in BHK21 cells and in the Ad12-transformed hamster cell line T637. The same or a similar RNA is also present in mouse cells. Homologous RNA sequences have not been detected in human, amphibian or insect cells. The biological function of this RNA is not known. Perhaps, foreign DNA can be more easily inserted at cellular sites that are actively transcribed. The RNA homologous to cellular DNA sequences adjacent to the site of insertion of Ad2 DNA could indicate transcriptional activity at that site (76). Transcriptional activity may constitute an important feature for sites of foreign DNA insertion.

Ad12-induced tumor cell line CLAC1

Cell line CLAC1 (14, 27) was established from an Ad12-induced hamster tumor, and each cell of this line contained 10-13 copies of colinearly integrated viral DNA, which seemed to be inserted at a limited number of sites. A left terminus of Ad12 DNA linked to cellular DNA was molecularly cloned (Fig. 7) and sequenced (Fig. 8). The first 174 nucleotides of Ad12 DNA at the site of linkage were deleted (Fig. 8b). Within 43 nucleotides of cellular DNA starting from the linkage point, there were one hepta-, one tri-, two tetra-, and one pentanucleotide which were identical and arranged in the same order as in the 174 deleted viral nucleotides and the cellular sequence replacing them (underlined sequences in Fig. 8b). In addition, there were patch homologies, twenty two octa-, twelve nonanucleotides and one decanucleotide between the left terminal 2320 bp of Ad12 DNA and the sequenced 529 cellular nucleotides at the site of junction. Thus, two types of patch homologies have been noted: One between the deleted terminal viral DNA sequence and the cellular sequence replacing it (27), another one between the persisting viral DNA and the adjacent cellular DNA sequence. In the cellular sequence, an internal undecanucleotide repeat can be detected (brackets in Fig. 8a). The sequence GCCC is repeated six times in succession (underlined sequence in Fig. 8a), and the nucleotide sequences GCC, GCCC, and GCCCC occur 25, 12, and two times respectively and comprise 25% of the entire cellular sequence of 529 base pairs.

The cellular DNA sequence corresponding to the fragment spanning the junction site was also cloned from BHK21 (B3) hamster cells and was determined by the Maxam-Gilbert technique. This cellular sequence was represented in the hamster genome several hundred times as determined by Southern blot analyses. Up to the linkage site with viral DNA, this cellular sequence was almost identical to the equivalent sequence from CLAC1 hamster cells (Fig. 9).

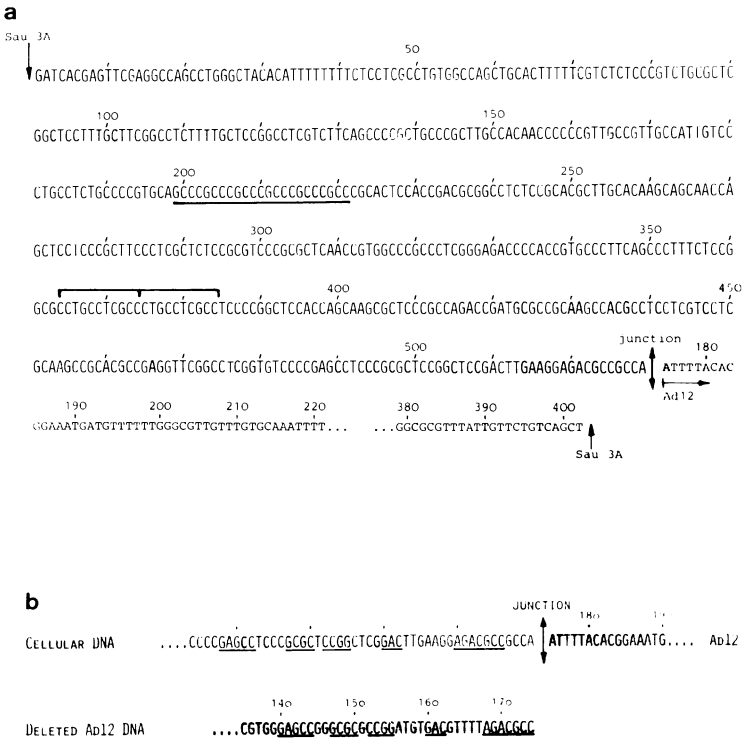


FIGURE 8. Nucleotide sequence at site of junction between the left terminus of Ad12 DNA and hamster cell DNA (27) in cell line CLAC1. a) Nucleotide sequence. b) Junction site with patch homologies between deleted viral DNA sequence and cellular sequence replacing it. Patch homologies are underlined. Ad12 sequence is shaded.

Based on the results gleaned from this analysis, a model can be suggested, in which the insertion of Ad12 DNA into cellular DNA in cell line CLAC1 was mediated by multiple, short sequence homologies (Fig. 10) in an invasive recombination event. Multiple sets of short patch homologies might be recognized as patterns in independent integration events. As is apparent from Fig. 10, this model also accounts for the loss of terminal viral DNA sequences in line CLAC1. Such losses have proved a general phenomenon in the insertion of adenoviral DNA. The model proposed was not intended to describe all aspects of the insertion mechanism. The model refers to the situation at the site of recombination at a certain moment in one particular recombination event that led to cell line CLAC1 and does not predict whether linear or circularized viral DNA molecules are intermediates in the reaction. As mentioned above, it is conceivable that circularized adenovirus DNA is required for recombination.

Another molecular feature of this model predicts (Fig. 11) that depending on the combination of patch homologies, which may happen to interact, different sites of recombination could be

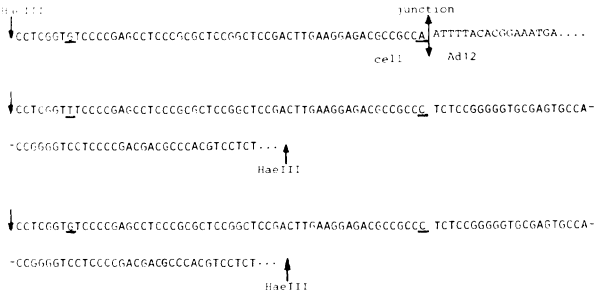


FIGURE 9. The recloned cellular sequence from nonvirus-transformed hamster cells is identical to the cellular junction sequence in cell line CLAC1 (27). Clone p9 represents the junction sequence from cell line CLAC1. Clones p7 and λ12 are different clones from BHK21(B3) hamster cells.

selected. Thus, a certain combination of patch homologies may be decisive and would also account for the large number of possible integration sites. It should be added, however, that distinct patches, like those at the junction site analyzed in cell line CLAC1, cannot always be found and that the insertion mechanism cannot be entirely dependent on but may be enhanced by this feature.

Linkage between Ad12 DNA and human KB cell DNA in SYREC2 DNA which is encapsidated into virus particles

We have previously described a symmetric recombinant SYREC between the left terminus of Ad12 DNA and human KB cell DNA (10). The human cell DNA contained, at least in part, repetitive sequences recurring several hundred times in cellular DNA. SYREC DNA was encapsidated into viral particles and exhibited the same or nearly the same length as authentic Ad12 DNA. It was likely that SYREC DNAs constituted a collection of similarly but not identically constructed molecules. Accordingly, the length of the left terminal Ad12 DNA segment involved in the SYREC structure might vary somewhat from recombinant to recombinant. In any event, the presence of the left terminal sequences was important, as it was demonstrated that the left-most 400 nucleotides and specifically a sequence between bp 290 and 390 of adenovirus DNA were essential in packaging viral DNA (77, 78). On denaturation and renaturation, the recombinant molecules were converted to molecules half the length of Ad12 DNA. Thus, SYREC DNA represented a symmetrically duplicated inverted repeat of the type ABCDD'C'B'A' with the left terminus of Ad12 DNA flanking the molecule on either end. The occurrence of SYREC molecules provided proof for the generation of virus-host DNA recombinants in productively infected cells (8, 9, 79). It was therefore required to confirm the recombinant structure of SYREC DNA by determining the nucleotide sequence across the point of fusion between Ad12 DNA and human KB cellular DNA. A 6.4 kb BamHI fragment was cloned from SYREC2 DNA into the plasmid pBR322. This clone might also serve as a tool to construct eukaryotic vectors based on the adenovirus replicon. The immediate junction site was

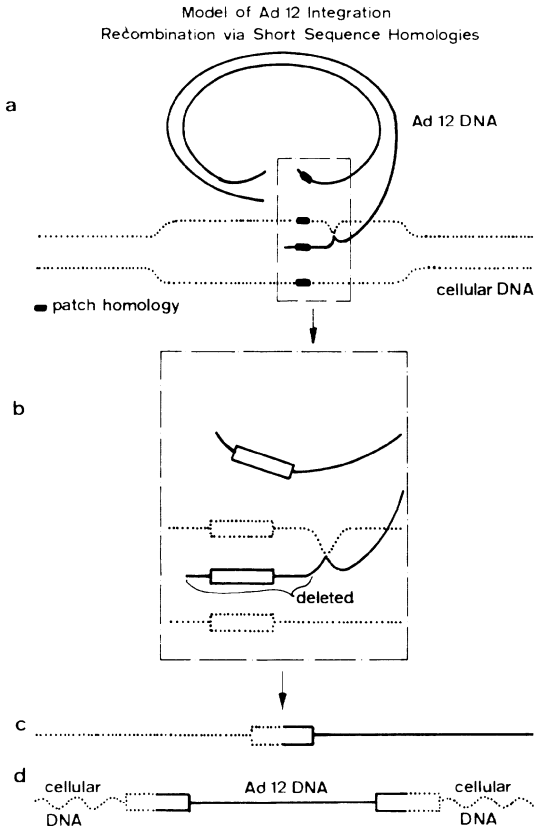


FIGURE 10. a) Hypothetical model of adenovirus DNA integration (27). The terminus of one strand of adenovirus DNA invades double-stranded cellular DNA. The intermediate in recombination could be circular or linear. Site of recombination could be determined by patterns of patch homologies. Recombination occurred outside the realm of these patches. b) The model explains the deletion of terminal nucleotides. It is, of course, possible that the circular structure forms only at the time of insertion. c) and d) Inserted viral DNA.

recloned as a 1.4 kb MspI fragment, again in the plasmid pBR322. The results in Fig. 12 present the nucleotide sequence across the joint between the left terminus of Ad12 DNA and KB cell DNA (Fig. 12a). A comparison with the nucleotide sequence at the left terminus of authentic Ad12 DNA (80-82) revealed that the SYREC2 DNA analyzed contained 2081 nucleotides from the left end of Ad12 DNA covalently linked to KB cellular DNA (11). The 6.4 kb BamHI fragment hybridized to Southern blots of KB DNA. Thus, the DNA

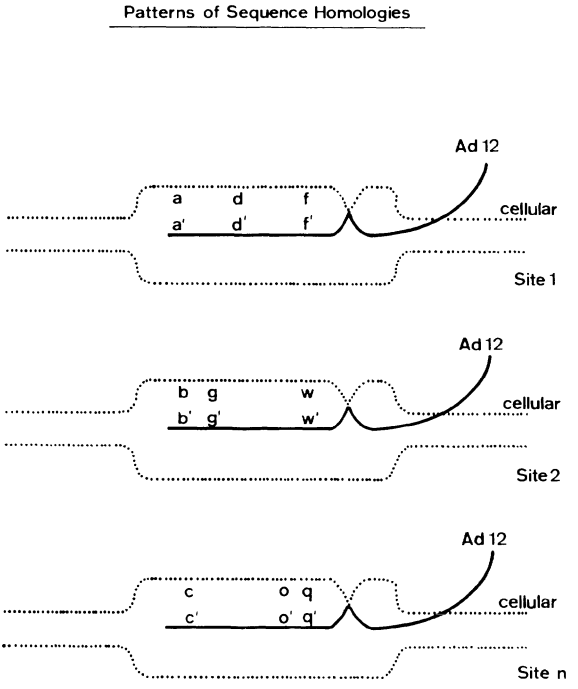


FIGURE 11. Patterns of patch homologies determine site of insertion (27); letters a, d, f; b, g, w; and c, o, q represent combinations of patch homologies.

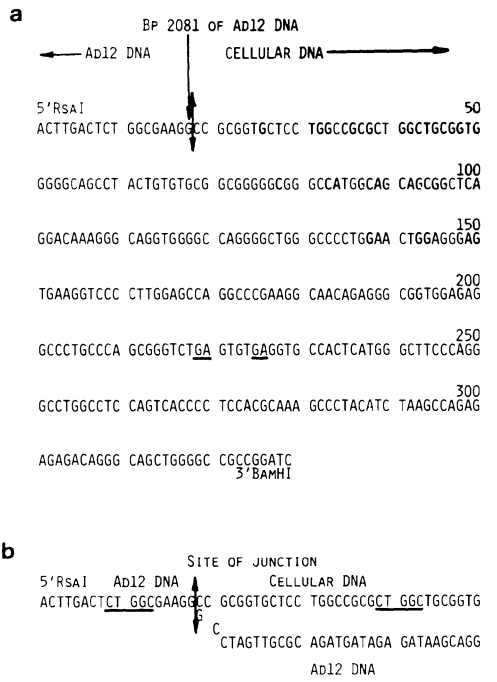


FIGURE 12. Site of junction between the left terminal 2081 nucleotides of Ad12 DNA and KB cellular DNA in SYREC2 DNA (10, 11). a) Nucleotide sequence. The Ad12 E1b protein starts at base pair 1846 of Ad12 DNA and continues in an open reading frame beyond the site of junction into cellular DNA for another 66 codons. Two termination codons TGA have been indicated. b) Comparison between the deleted viral nucleotide sequence and the cellular sequence replacing it. Sequence homologies are not apparent. CTGGC is a common pentanucleotide between the remaining viral and the adjoining cellular DNA sequences.

fragment adjacent to the 2081 nucleotides of Ad12 DNA was indeed cellular. Sequence comparisons between the KB cellular sequence and the deleted Ad12 sequence from nucleotide 2082 on to the right showed no obvious homologies that might have been instrumental in recombination (Fig. 12b).

It was also interesting to note that the SYREC1 DNA molecule isolated in 1980 had only about 700-1150 nucleotides of Ad12 DNA, again derived from the left terminus (10). The recombinant SYREC2 described here was isolated in 1977 and carried the first 2081 nucleotides from the left end of Ad12 DNA as determined by sequence analyses (11). These data suggest that SYREC DNA molecules may undergo alterations during continued passage, e.g., selective deletions of viral nucleotides. It is, however, also conceivable that at different times, different SYREC populations became the preponderant ones or that different recombinants were chosen for the study in different experiments.

Preliminary results indicate that the sequence immediately adjacent to the Ad12 DNA in SYREC2 recombinants (note open reading frame to the right of arrow in Fig. 12a) is transcribed into RNA in human cells (U. Freisem and W. Doerfler, unpublished results). It will be interesting to investigate whether this RNA is also translated and what function such a protein might have.

Linkage between the left terminus of Ad12 DNA and mouse cell DNA in the Ad12-induced tumor CBA-12-1-T

Cells of the Ad12-induced mouse tumor CBA-12-1-T contain > 30 copies of integrated viral DNA (42). Restriction analyses using the PstI restriction endonuclease revealed that there is one major site of linkage between the right and the left termini of Ad12 DNA and mouse cell DNA. The inserted viral DNA molecules are not arranged in true tandems, since terminally linked MspI fragments of authentic sizes have not been detectable (16). It is unknown at present how that many viral genomes have been accommodated. One model proposes that, initially, one viral DNA molecule is inserted into cellular DNA and that subsequently the integrated viral genome is amplified together with the flanking cellular DNA sequences (83).

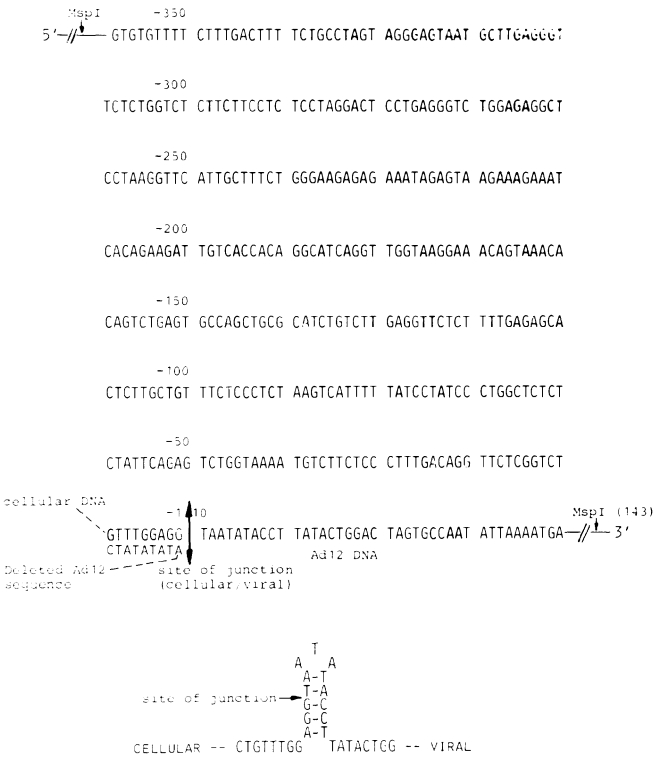


FIGURE 13. Nucleotide sequence at the left terminal site of junction between Ad12 and mouse cell DNA (42) in the tumor CBA-12-1-T.
The nucleotide sequence from the MspI site in cellular DNA to the left-most MspI site in Ad12 DNA (nucleotide 143 from the left terminus of virion DNA) was determined. Only part of the viral DNA sequence is shown. It was identical with the published sequence (80-82), except that the first 9 viral nucleotides were deleted. In the sequence shown, this deleted sequence was added for comparison underneath the cellular sequence replacing it. A short stemmed loop at the site of junction is also indicated.

We have cloned and sequenced the left terminal junction sequence between Ad12 DNA and mouse cell DNA (Fig. 13), as well as the preinsertion (unoccupied) site from normal CBA/J mouse cells (Fig. 14) (42). The preinsertion sequence is characterized by a transition from unique to more abundant to repetitive mouse DNA close to the site of viral DNA integration. At the site of insertion a cellular DNA fragment of at least 1500-1600 bp has been deleted. At the left site of linkage of viral DNA, nine nucleotides are missing. It has been pointed out previously that nucleotides 8 to 10 from the end of adenovirus DNA may represent a frequent site of recombination. This site is located just in front of a consensus sequence shared by all adenovirus DNAs, a

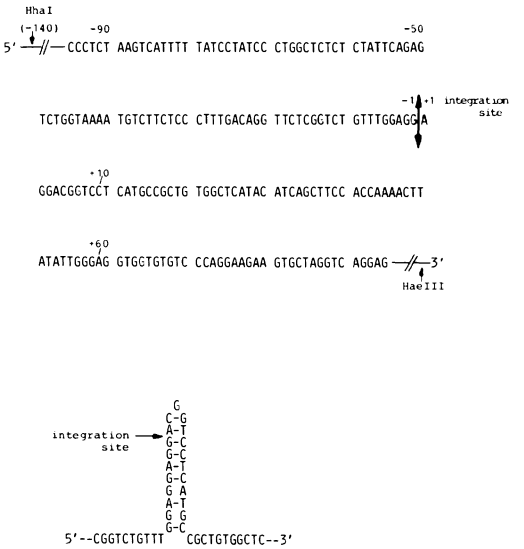


FIGURE 14. Nucleotide sequence at preinsertion site in mouse DNA (42). The site of integration of Ad12 DNA is designated by a double-headed arrow. At this site a sizeable palindromic sequence can be detected. The HhaI-HaeIII sequence was only partly determined.

sequence that has been shown to be essential in viral DNA replication (84-87). Recombination at this site has also been observed in adenovirus type 16 (77, 78), in a mutant of Ad2 (88), and between both termini of Ad2 DNA with hamster cell DNA in cell line HE5 (41). Recombination in the tumor CBA-12-1-T between nucleotide 10 of Ad12 DNA and mouse DNA (this report) represents a further example. Up to the site of linkage the cellular DNA sequence in CBA-12-1-T tumor DNA and the preinsertion sequence in CBA/J mouse cells were identical. The nucleotide sequence at the site of linkage (Fig. 13) and at the preinsertion site (Fig. 14) revealed palindromic stretches of 5 and 10 nucleotide pairs, respectively. Scattered patch homologies (8-10 nucleotide pairs long) were observed between adenovirus DNA and cellular DNA. A hypothetical model for DNA arrangements at the site of recombination is presented (Fig. 15).

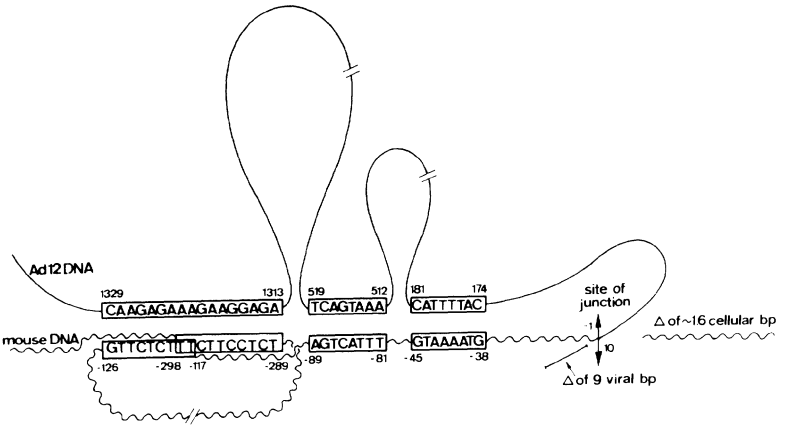


FIGURE 15. Hypothetical sequence arrangement in recombination complex in the tumor CBA-12-1-T (42). This scheme is based on sequence homologies at or close to the site of junction as determined by computer analysis.

Linkage between the left terminus of Ad12 DNA and hamster cell DNA in the Ad12-induced tumor T1111(2)

The tumor T1111(2) was induced by injecting Ad12 into newborn Syrian hamsters; each cell contained 10-11 viral genome equivalents (17). A left terminal off-size fragment was cloned in λgtWES·λB DNA, and a HhaI fragment carrying the site of junction was subcloned in pBR322. The off-size band proved interesting, as it disappeared upon subsequent passage of the T1111(2) cells in culture (17). The nucleotide sequence across the site of junction was determined (Fig. 16a) (U. Lichtenberg and W. Doerfler, manuscript in preparation). The data revealed that the terminal 64 nucleotides had been deleted at the site of junction. Thus, deletions of terminal viral nucleotides again proved a general principle in this type of recombination process. Unlike the findings in tumor cell line CLAC1, comparison of the deleted sequence of 64 nucleotides of Ad12 DNA with the hamster cellular sequence replacing it did not reveal patchy homologies (Fig. 16b). Thus, the insertion event was apparently not directed by a mechanism similar to the one proposed for cell line CLAC1 or, alternatively, the cellular DNA with homologies to viral DNA had been deleted during integration.

An interesting peculiarity of this junction site was uncovered by computer analysis of the nucleotide sequence (Fig. 16a). The left terminal nucleotide number 65 of Ad12 DNA was linked to cellular DNA of the unique type or of low abundance in hamster cell DNA. At 135 nucleotides to the left from this junction site, 64 nucleotides of Ad12 DNA from the left viral DNA terminus (nucleotides number 1297 to 1361) were inserted in the opposite direction into cellular DNA. It is unknown at what stage or in what way this insertion or partial duplication of viral sequences had occurred. This finding indicates that complicated rearrangements of sequences can occur occasionally before, during, or after the integration event.

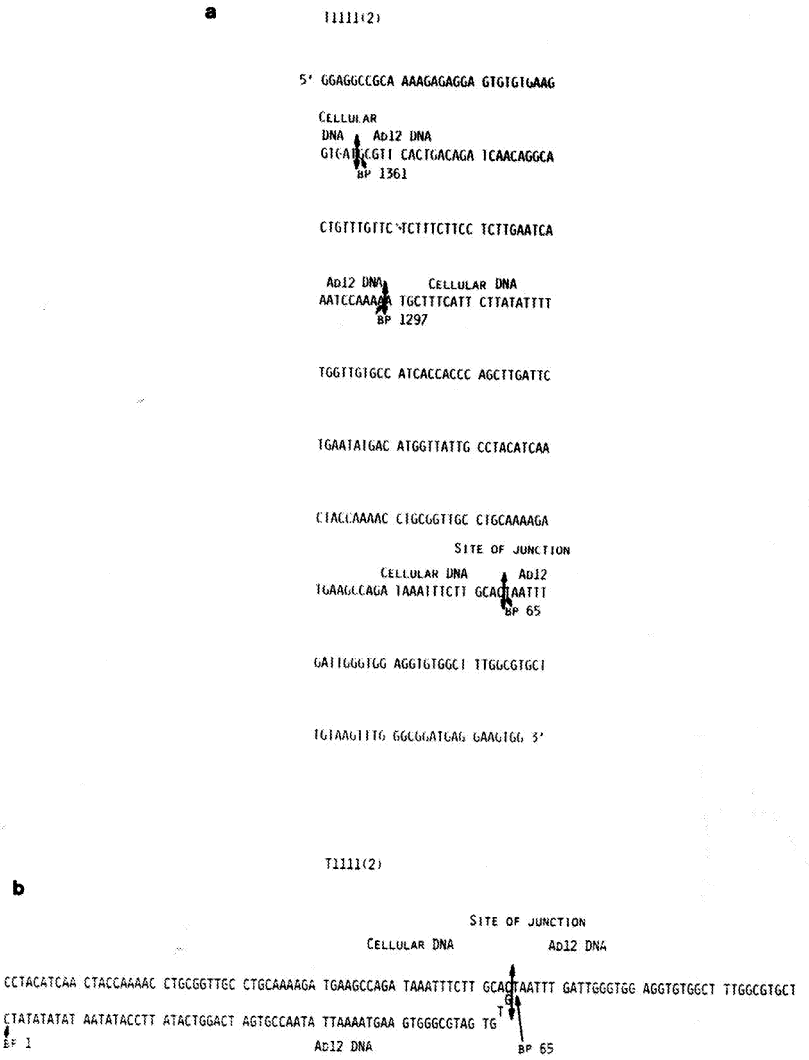


FIGURE 16. Site of junction between the terminal Ad12 DNA and hamster cell DNA in the tumor T1111(2) (U. Lichtenberg and W. Doerfler, manuscript in preparation). a) Nucleotide sequence. Note the inverted insert of viral nucleotides into cellular DNA. b) Sequence comparison between 64 deleted adenoviral nucleotides and the hamster cell sequence replacing them.

Left terminal junction site of Ad12 DNA in cell line HA12/7

The hamster cell line HA12/7 was derived from primary hamster embryo cells which were transformed in culture by Ad12. Each cell carried about three to four Ad12 DNA equivalents, and the integration patterns were described in detail (14). These patterns were characterized by a peculiar selective amplification of left terminal Ad12 DNA sequences (14). Moreover, there is evidence that the E1 region of integrated Ad12 DNA is expressed abundantly in cell line HA12/7 (50, 51). We have cloned and are sequencing the left and right terminal junction sites of viral DNA from this line, since it was conceivable that the terminal amplifications were relevant to explain the levels of expression of the E1 region. A PstI off-size fragment of about 4 kb (see Fig. 9c in reference 14) was cloned into pBR322, and the EcoRI-cut plasmid vector was subsequently cloned into λgtWES·λB DNA to facilitate detection of clones containing the junction fragment. Similarly,

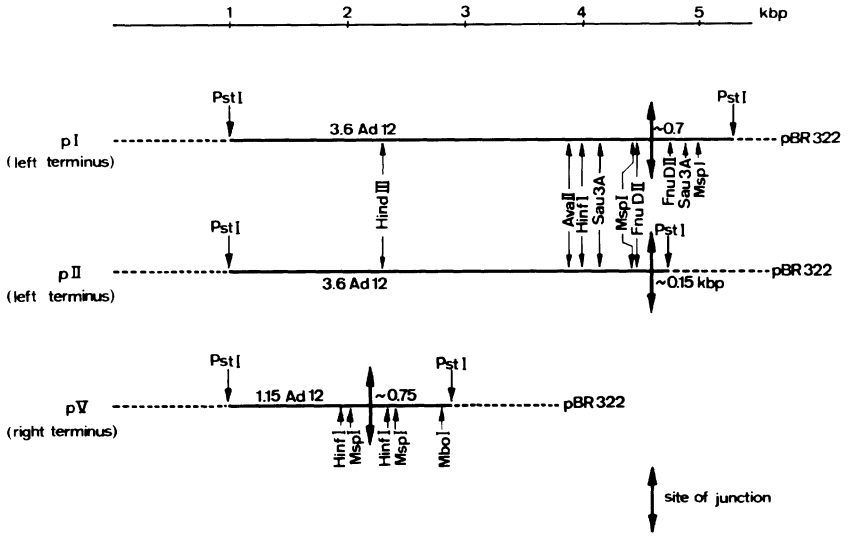


FIGURE 17. Restriction maps of the left and right junction sites from cell line HA12/7. (R. Jessberger, S. Stabel, and W. Doerfler, unpublished results).

the right terminal site of junction was cloned by the same procedure. Restriction maps of both sites are shown in Fig. 17 (R. Jessberger, S. Stabel, and W. Doerfler, unpublished results). Experiments are in progress to determine the nucleotide sequences of these sites.

Excision of amplified viral DNA at palindromic sequences from the Ad12-transformed hamster cell line T637

In the Ad12-transformed cell line T637, 20-22 copies of viral DNA are integrated into cellular DNA. Morphological revertants of cell line T637 (36) that had lost all but one or one-half of a viral DNA copy were considered useful tools in elucidating the mechanisms of insertion and excision of the Ad12 DNA molecules. It was shown that specific sites of linkage between viral genomes were lost in many of the morphological revertants. The same sites proved highly sensitive to endogenous nucleases, as shown by autoincubation of T637 nuclei. The failure of all attempts to clone the sites of junction in cell line T637 (see section 3, item 3) suggested that excision-prone palindromic sequences might exist at these sites and prompted experiments in which T637 DNA was denatured, briefly renatured, and subsequently digested with S1 nuclease. The fold-back structures generated in this way were of distinct sizes and could be localized to the termini of the integrated Ad12 genome. Evidence was adduced that identical DNA sequences at these termini generated fold-back structures and at the same time were highly sensitive to autodigestion by endogenous nucleases. One of these sites appeared to be less sensitive to nuclease digestion than the other one. The results of experiments in which T637 DNA was treated with endonuclease VII (89), an enzyme presumably recognizing Holliday structures (90) in DNA (91), strengthened the notion that palindromic DNA sequences existed at the sites of linkage of Ad12 genomes in cell line T637. It appeared likely that only part of these sequences assumed the cruciform configuration at a given time. Thus, there was evidence for special sequence arrangements at the sites of viral-viral or viral-cellular DNA junctions (61).

COMPUTER ANALYSES OF CELLULAR SEQUENCES AT JUNCTION SITES

The cellular nucleotide sequences linked to the left or right terminus of adenovirus DNA were computer analyzed for the presence of common features. The following parameters were included in the analyses.

Sequence homologies including patch homologies to Ad2 or Ad12 DNA

The data in Fig. 18 list sequence homologies found between cellular junction sequences as listed (v-c, viral-cellular sequence or c-v, cellular-viral sequence in direction of reading),

	HE5(r)		CLAC3		CLAC1		SYREC2		T1111(2)	
	v-c	c-v	v-c	c-v	v-c	c-v	v-c	c-v	v-c	c-v
PBR322 TOTAL SEQUENCE	10:1 9:1 8:1	11:1 10:1 8:10	8:4	9:1 8:3	12:1 10:2 9:9 8:25	11:1 10:3 9:11 8:25	10:2 9:7 8:19	10:1 9:2 8:24	9:2 8:7	9:4 8:7
AD12 LEFT TERMINAL 2320 NUCLEOTIDES	9:5 8:10	11:1 9:2 8:11	11:2 9:1	9:2 8:4	9:5 8:16	10:1 9:7 8:6	9:1 8:5	10:1 9:1 8:8	9:2 8:4	8:2
AD2 LEFT 38 %	12:1 10:3 9:10 8:40	9:13 8:36	10:1 9:3 8:12	9:3 8:8	12:1 11:2 10:8 9:41 8:152	11:1 10:7 9:24 8:102	14:1 11:2 10:6 9:15 8:61	11:3 10:3 9:6 8:64	11:1 10:1 9:2 8:25	9:2 8:17
AD2 EcoRI F, D, E, C FRAGMENTS	12:1 11:1 10:2 9:11 8:50	11:2 10:2 9:11 8:52	10:1 9:3 8:12	8:12	12:1 11:3 10:8 9:16 8:62	12:1 11:2 10:9 9:17 8:84	11:1 10:4 9:10 8:33	12:1 10:1 9:7 8:35	9:3 8:14	11:2 10:3 9:6 8:19

FIGURE 18. Computer-aided analyses of some of the cellular junction sequences. The nucleotide sequences of cellular junction sequences read in both directions (5'-3'), viral-cellular (v-c) or cellular-viral (c-v), were compared with the aid of a CDC76 computer, to the entire sequence of the E. coli plasmid pBR322 (92), to the left terminal 2320 nucleotides of Ad12 DNA (80, 81), and to the left and right terminal sequences of Ad2 DNA (73) as indicated. The designations used are the following: 8:2 two common octanucleotides, 10:1 one common decanucleotide, etc. Compare legend to Fig. 19 for relative length of cellular sequences.

and the left terminal HindIII-G fragment of Ad12 DNA (2320 bp) (80, 81), or the left terminal 38% of Ad2 DNA, or the sequence comprised by the right terminal EcoRI fragments F, D, E, and C of Ad2 DNA (73). As expected from previous analyses on the occurrence of patch homologies (26), such homologies are very abundant. They comprise sequence identities of up to 14 nucleotides. Homologies shorter than eight nucleotides have not been tabulated, as they occur very frequently. Obviously, the longer the viral and cellular sequences included in the analyses, the more abundant is the number of sequence homologies detected. Such sequence homologies could theoretically be used in directing recombination events. In some instances, they may actually have been used (Figs. 8b, 15). In other cases, no homologies have been found. Thus, patch homologies represent a very sizeable repertoire that can potentially become important in directing recombination events. A purely statistical treatment of sequence homologies is obviously not sufficient. As revealed by the abundance of sequence homologies, their occurrence is statistically predicted, and they could be useful as signals in recombination events.

Sequence homologies to the nucleotide sequence of the cloning vector pBR322

The entire nucleotide sequence of plasmid pBR322 was included in this comparison, mainly to rule out cloning artifacts in the sense that longer nucleotide sequences from the plasmid vector could have recombined with the cellular junction sequence. The data presented in Fig. 18 provide no evidence whatsoever for this possibility. Patch homologies occur at rates comparable to those found between cellular junction sequences and adenovirus DNA. It is concluded that cloning artifacts have not occurred with the junction sequences analyzed. This possibility had already been ruled out by cloning and sequencing the original unoccupied cellular sequence that had remained unaltered in cell lines HE5, CLAC1 and CBA-12-1-T (26, 27, 41, 42). It is also apparent from the data presented that patch homologies in cellular junction

sequences are about equally frequent to the sequence of the prokaryotic vector as to sequences of human adenoviruses. Thus, patch homologies appear to be a common phenomenon.

Comparison of cellular junction sequences

Lastly, some of the cellular junction sequences were compared with each other (Fig. 19). Again, the matrix reveals only short sequence homologies at frequencies that may not exceed values statistically expected. Moreover, there are no striking similarities or identities of nucleotide sequences when these patch homologies are compared with each other. Hence, a cellular nucleotide sequence common to all junction sites involving adenovirus DNA integration does not exist. Allowing for maximal flexi-

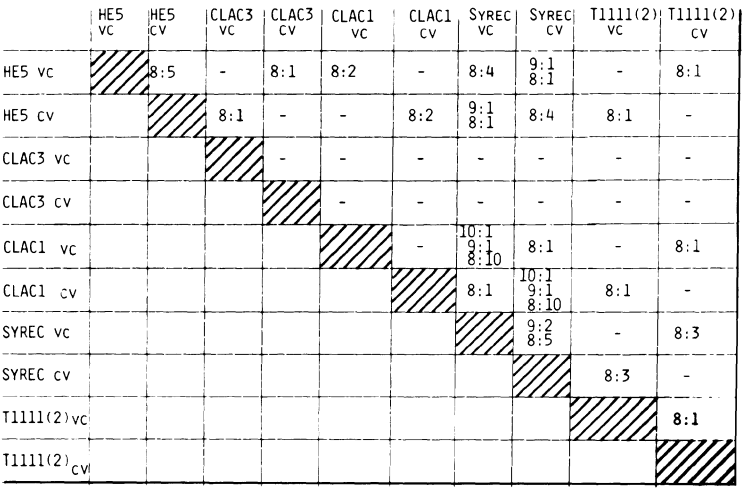


FIGURE 19. Computer-aided comparison of some of the cellular junction sequences among each other. The lengths in nucleotide pairs of the cellular DNA sequences are the following: HE5 401, CLAC3 82, CLAC1 529, SYREC 304, T1111(2) 170. Thus, strictly speaking, comparisons of frequencies of patch homologies have to take into account the relative nucleotide numbers of each cellular sequence.

bility of the recombination mechanism in eukaryotic cells, patch homologies as abundantly found may be utilized in stabilizing recombination complexes, but it cannot be claimed that such homologies are absolutely essential, since they are not always found.

MAIN CONCLUSIONS

We do not understand yet the mechanisms of recombination in molecular terms that explain the insertion of foreign (viral) DNA into the genome of mammalian cells. For a number of reasons, it will be very important to elucidate this mechanism. One may have to accept the possibility that more than one functional pathway for insertion exists.

In comparing results on the analyses of several different sites of linkage between adenoviral and cellular DNAs (for reviews, see references 4, 5), the notion emerges that recombination of adenovirus and cellular DNAs can lead to a considerable variety of phenomena with respect to the structure of the junction site. This variety suggests that recombination may not occur by a rigidly defined mechanism but seems to tolerate variability. Common features in this recombination event, on the other hand, involve recombination at the termini of adenovirus DNA and deletion of terminal viral nucleotides ranging between 2 and 174 bp in the instances examined so far. Nucleotide sequence specificities or duplications of cellular nucleotides at the sites of integration have not been observed. This finding does not preclude the possibility that in some instances (27, 42) patchy sequence homologies between viral and cellular DNA can be utilized to stabilize the recombination complex. Cellular DNA was preserved completely (41) or extensive deletions were generated at the integration site (42). At other sites of junctions, selective amplifications of terminal viral nucleotides (14), deletions and rearrangements (28) or palindromic sequence arrangements (61) have been described. It is impossible to decide at present whether this apparent variability in the structure of junction sites reflects peculiarities of the insertion mechanism or ensues rather as a consequence of postintegrational events.

From the data collected in this review, we shall try to deduce a few general features which may help to formulate new experimental approaches. In evaluating these general features, we are aware of the fact that we have analyzed highly selected, clonal lines of adenovirus-transformed cells or adenovirus-induced tumors. In these cells the sites of insertion may reflect in a complex way the mode of selecting for transformed cells with certain growth properties. Proximity of viral promoters and/or enhancers to certain cellular sequences and vice versa may bestow upon the carrier cell of such inserted genomes growth advantages that have led to the selection of these cells. Thus, the cells selected for analyses do not necessarily have to reflect the entire gamut of possible recombinations between viral and host DNA.

What are the most important general findings gleaned from extensive analyses of host-viral DNA junction sites?

1. The insertion sites have been studied in a series of transformed or tumor cell lines which were usually initiated by adenovirions, not by transfection with viral DNA. Exceptions have been noted (10, 93). The mode of introducing foreign DNA into the host cell may be of importance in guiding or selecting the mechanism of recombination. Usually adenovirus DNA has been linked to cellular DNA via the terminal viral sequences. This finding suggested a mechanism that somehow gave special significance to the termini of viral DNA. A circular intermediate, in particular one that was formed by the interaction of the terminal protein molecules, appeared conceivable. On the other hand, in some instances the recombination event had also involved internal viral DNA sequences (21, 23). Thus, one type of mechanism may not exclusively apply; there may be alternative ways of inserting foreign DNA.

2. It has been previously pointed out (83) that colinear insertion of the nearly intact genome of Ad12 and the insertion of partly deleted Ad2 genomes may be a consequence of the degree of permissivity of the interaction between virus and host cells. Ad12 infects hamster and mouse cells abortively, Ad2 replicates

in hamster cells. In some of the Ad12-induced hamster tumors analyzed, fragments of viral DNA were integrated. Moreover, cell lines carrying multiple copies of nearly intact Ad12 genomes contained, in addition, nonstoichiometric sets of fragments of the viral genome.

3. In considering a model that emphasizes the functional role of the viral protein in the recombination event, this function could be exerted in a circular, as well as a linear, intermediate. In fact, a circular structure may never be formed, but the two viral DNA termini may contact the same cellular site in immediate succession, the second terminus perhaps only after cellular DNA has been endonucleolytically cleaved. It is also conceivable that the viral termini have a special propensity to undergo recombination, because the terminal sequences are the sites of initiation of viral DNA replication and hence particularly prone to local denaturation and/or interactions with protein or DNA molecules.

4. In cell line HE5, it was striking that insertion of Ad2 DNA did not lead to deletion of a single cellular nucleotide at the site of linkage. In cell line HE5, the original cellular sequence was unaltered (41). In the tumor CBA-12-1-T cellular DNA was deleted (42). Again, several possibilities appear to be realized.

5. Repetitions of cellular nucleotides at the sites of linkage between adenovirus and cellular DNA were so far not observed. This result rendered the possibility of a transposon-like mechanism unlikely as reported for integration of the retroviral proviruses (94).

6. Deletions of viral nucleotides of varying lengths (comprising 2, 8, 10, 9, 45, 64, or 174 nucleotides) were seen at all junctions analyzed. An invasive model of recombination (cf. Fig. 10) would explain these microdeletions. In cell line HE5, hamster cell DNA was linked to nucleotide 11 on the left terminus and to nucleotide 9 on the right terminus of the integrated Ad2 genome (41). In the mouse tumor CBA-12-1-T, cellular DNA was connected to nucleotide 10 on the left terminus of Ad12 DNA (42). In two variants of Ad16, the viral DNA carried a reduplication of left terminal DNA sequences on the right terminus of the Ad16 genome.

In the DNA of both variants, the 8 right terminal nucleotides were deleted (78). It was therefore tempting to speculate that a hot spot of recombination might exist around nucleotides 8 to 10 at the adenovirus DNA termini. These nucleotides are located just in front of the essential consensus sequence at which viral DNA replication is initiated (84-87).

7. We have discussed the hypothetical role that patch homologies or combinations of patches could play in the recombination event, perhaps by stabilizing the recombination complex. We should like to emphasize, however, that such patch homologies have not always been found or are sometimes located remote from the site of junction (Fig. 15). Patches may not be essential in some integration events. One has to consider patch homologies of up to 14 nucleotides in length and certainly combinations of nucleotide patches involving even longer stretches (cf. Fig. 8b) as naturally occurring. These homologies could then be exploited by the mechanism of recombination. This concept is consistent with the large number of different recombination sites actually found. In the sequences analyzed, we have noted two types of patch homologies; one type existing between neighboring cellular and preserved viral DNA sequences, another one between the deleted viral nucleotides and the cellular sequence replacing them.

8. Another interesting phenomenon is the insertion of nucleotides of unknown origin at the internal junction of Ad2 DNA in cell line HE5 (cf. Fig. 4). Is this insertion haphazard or does it point to a complex recombination event involving other (cellular?) sequences?

9. Adenoviral DNA integration has been observed in unique or repetitive cellular sequences. It is always possible that repetitive cellular sequences were generated as a consequence of the insertion event or were due to another viral function. This reservation obviously does not hold for the data obtained with cell line CLAC1, since in that case integration had occurred into preexisting repetitive cellular DNA (27).

10. We have also studied the expression of cellular junction sequences in normal and adenovirus-transformed cells and have found a low molecular weight cytoplasmic, non-polyadenylated RNA homologous to cellular DNA adjacent to the right site of junction in cell line HE5 (76). The data obtained are compatible with but do not prove the hypothesis that adenovirus DNA can integrate at transcriptionally active sites of cellular DNA.

11. The mechanism of adenovirus DNA insertion is not like the integration of bacteriophage λ DNA into the bacterial chromosome at one highly specific site, since all cellular junction sequences are different. Moreover, the mechanism of adenovirus DNA insertion does not resemble that of retroviral proviruses or of transposons. There must be yet another mechanism or other mechanisms of recombination in mammalian cells.

Comparisons with the integration of bacteriophage λ DNA are interesting also in that λ DNA can integrate with low frequency at nonspecific sites. Several of these sites have been sequenced (95). The highly significant finding has been reported that at presumably nonspecific sites the internal hexanucleotide TTTATA of the highly specific "attachment site" GCTTTTTTATACTAA sequence (96) has been preserved. Thus, even part of this highly specific sequence can be recognized. A hexanucleotide is consequently sufficient to direct recombination mechanisms. This hexanucleotide is certainly a less conspicuous signal than some of the patch combinations we have observed at sites of adenovirus DNA integration. The data on sequence peculiarities at nonspecific sites of λ DNA insertion tend to caution one toward a purely statistical interpretation of patch homologies and their possible function. Highly specific proteins that have the capacity to recognize short sequence identities and utilize them could be involved in recombination.

12. In searching for signals to direct the recombination event, palindromic DNA structures may also play a role. Future research may have to be directed, in particular, toward specific cellular chromatin structures that may predispose for foreign DNA insertion.

13. The sequence data on the SYREC2 DNA molecule (cf. Fig. 12) at the site of junction between Ad12 DNA and human cell DNA (11) and on the Ad7 - human cell DNA recombinant (93) provide unequivocal proof for the occurrence of recombinants between adenovirus and cellular DNA in productive infection (8).

14. A comparison of many of the features of adenovirus DNA insertion with those derived from investigations on the SV40 (32, 33) or polyoma virus system (34, 35) reveal striking similarities. The recombination mechanisms of a mammalian cell - there may in fact be more than one exclusive way of inserting foreign DNA - may not be capable of differentiating among different foreign DNA molecules. This assumption would rather tend to belittle the role that the terminal adenovirus protein might have in the insertion process. Since viral-cellular recombinations with simian virus 40, polyoma virus, and adenovirus DNAs have many features in common, it is conceivable that the terminal adenovirus protein or the terminal DNA sequence has a guiding function in recombination and that the actual recombination is catalyzed by cellular functions. It is, however, remarkable that adenovirus DNA frequently recombines via its termini, whereas SV40 DNA or polyoma virus DNA recombine at random sites of the viral sequence.

ACKNOWLEDGMENTS

We thank Petra Böhm for editorial assistance. Research in the authors' laboratory was supported through SFB74-C1 of the Deutsche Forschungsgemeinschaft, and by grants from the Federal Ministry of Research and Development, Bonn-Bad Godesberg, Germany and the Ministry of Science and Research of the State Nordrhein-Westfalen, Düsseldorf, Germany.

REFERENCES

1. Doerfler, W. (Ed.) The Molecular Biology of Adenoviruses, vol. 1. Springer Verlag, Berlin, Heidelberg, New York, 1983.
2. Doerfler, W. (Ed.) The Molecular Biology of Adenoviruses, vol. 2. Springer Verlag, Berlin, Heidelberg, New York, 1984.
3. Doerfler, W. (Ed.) The Molecular Biology of Adenoviruses, vol. 3. Springer Verlag, Berlin, Heidelberg, New York, 1984.
4. Doerfler, W. Curr. Top. Microbiol. Immunol. 101:127-194, 1982.
5. Doerfler, W., Gahlmann, R., Stabel, S., Deuring, R., Lichtenberg, U., Schulz, M., Eick, D. and Leisten, R. Curr. Top Microbiol. Immunol. 109:193-228, 1983.
6. Doerfler, W. Proc. Natl. Acad. Sci. USA 60:636-643, 1968.
7. Doerfler, W. J. Virol. 6:652-666, 1970.
8. Burger, H. and Doerfler, W. J. Virol. 13:975-992, 1974.
9. Schick, J., Baczko, K., Fanning, E., Groneberg, J., Burger, H. and Doerfler, W. Proc. Natl. Acad. Sci. USA 73:1043-1047, 1976.
10. Deuring, R., Klotz, G. and Doerfler, W. Proc. Natl. Acad. Sci. USA 78:3142-3146, 1981.
11. Deuring, R. and Doerfler, W. Gene 26:283-289, 1983.
12. Fanning, E. and Doerfler, W. J. Virol. 20:373-383, 1976.
13. Sutter, D., Westphal, M. and Doerfler, W. Cell 14:569-585, 1978.
14. Stabel, S., Doerfler, W. and Friis, R.R. J. Virol. 36:22-40, 1980.
15. Kuhlmann, I. and Doerfler, W. Virology 118:169-180, 1982.
16. Starzinski-Powitz, A., Schulz, M., Esche, H., Mukai, N. and Doerfler, W. EMBO J. 1:493-497, 1982.
17. Kuhlmann, I., Achten, S., Rudolph, R. and Doerfler, W. EMBO J. 1:79-86, 1982.
18. Doerfler, W. Virology 38:587-606, 1969.
19. Ortin, J., Scheidtmann, K.-H., Greenberg, R., Westphal, M. and Doerfler, W. J. Virol. 20:355-372, 1976.
20. Esche, H., Schilling, R. and Doerfler, W. J. Virol. 30:21-31, 1979.
21. Sambrook, J., Botchan, M., Gallimore, P., Ozanne, B., Pettersson, U., Williams, J. and Sharp, P.A. Cold Spring Harbor Symp. Quant. Biol. 39:615-632, 1974.
22. Visser, L., van Maarschalkerweerd, M.W., Rozijn, T.H., Wassenaar, A.D.C., Reemst, A.M.C.B. and Sussenbach, J.S. Cold Spring Harbor Symp. Quant. Biol. 44:541-550, 1979.
23. Vardimon, L. and Doerfler, W. J. Mol. Biol. 147:227-246, 1981.
24. Tikchonenko, T.I. Curr. Top. Microbiol. Immunol. 110:169-189, 1984.
25. Deuring, R., Winterhoff, U., Tamanoi, F., Stabel, S., and Doerfler, W. Nature 293:81-84, 1981.
26. Gahlmann, R., Leisten, R., Vardimon, L. and Doerfler, W. EMBO J. 1:1101-1104, 1982.
27. Stabel, S. and Doerfler, W. Nucleic Acids Res. 10:8007-8023, 1982.
28. Eick, D. and Doerfler, W. J. Virol. 42:317-321, 1982.

29. Ketner, G. and Kelly, T.J., Jr. Proc. Natl. Acad. Sci. USA 73:1102-1106, 1976.
30. Botchan, M., Topp, W. and Sambrook, J. Cell 9:269-287, 1976.
31. Sambrook, J., Greene, R., Stringer, J., Mitchinson, T., Hu, S.-L. and Botchan, M. Cold Spring Harbor Symp. Quant. Biol. 44:569-584, 1980.
32. Stringer, J.R. J. Virol. 38:671-679, 1981.
33. Stringer, J.R. Nature 296:363-366, 1982.
34. Hayday, A., Ruley, H.E. and Fried, M. J. Virol. 44:67-77, 1982.
35. Mendelsohn, E., Baran, N., Neer, A. and Manor, H. J. Virol. 41:192-209, 1982.
36. Groneberg, J., Sutter, D., Soboll, H. and Doerfler, W. J. gen. Virol. 40:635-645, 1978.
37. Groneberg, J. and Doerfler, W. Int. J. Cancer 24:67-74, 1979.
38. Eick, D., Stabel, S. and Doerfler, W. J. Virol. 36:41-49, 1980.
39. Southern, E.M. J. Mol. Biol. 98:503-517, 1975.
40. Visser, L., Reemst, A.C.M.B., von Mansfeld, A.D.M. and Rozijn, T.H. Nucleic Acids Res. 10:2189-2198, 1982.
41. Gahlmann, R. and Doerfler, W. Nucleic Acids Res. 11: 7347-7361, 1983.
42. Schulz, M. and Doerfler, W. Nucleic Acids Res. 12: 4959-4976, 1984.
43. Gutai, M.W. and Nathans, D. J. Mol. Biol. 126:275-288, 1978.
44. Robinson, A.J., Younghusband, H.B. and Bellet, A.J.D. Virology 56:54-69, 1973.
45. Doerfler, W., Lundholm, U. and Hirsch-Kauffmann, M. J. Virol. 9:297-308, 1972.
46. Doerfler, W., Lundholm, U., Rensing, U. and Philipson, L. J. Virol. 12:793-807, 1973.
47. Ruben, S., Bacchetti, S. and Graham, F. Nature 301:172-174, 1983.
48. Sharp, P.A., Gallimore, P.H. and Flint, S.J. Cold Spring Harbor Symp. Quant. Biol. 39:457-474, 1974.
49. Sawada, Y. and Fujinaga, K. J. Virol. 36:639-651, 1980.
50. Schirm, S. and Doerfler, W. J. Virol. 39:694-702, 1981.
51. Esche, H. and Siegmann, B. J. gen. Virol. 60:99-113, 1982.
52. Sutter, D. and Doerfler, W. Cold Spring Harbor Symp. Quant. Biol. 44:565-568, 1979.
53. Sutter, D. and Doerfler, W. Proc. Natl. Acad. Sci. USA 77:253-256, 1980.
54. Kruczek, I. and Doerfler, W. EMBO J. 1:409-414, 1982.
55. Ibelgaufts, H., Doerfler, W., Scheidtmann, K.H. and Wechsler, W. J. Virol. 33:423-437, 1980.
56. Tiemeier, D., Enquist, L. and Leder, P. Nature 263:526-527, 1976.
57. Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.-A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. Science 196:161-169, 1977.
58. Hohn, B. and Murray, K. Proc. Natl. Acad. Sci. USA 74:3259-3263, 1977.

59. Hohn, B. Methods Enzymol. 68:299-309, 1979.
60. Benton, W.D. and Davis, R.W. Science 196:180-182, 1977.
61. Eick, D., Kemper, B. and Doerfler, W. EMBO J. 2:1981-1986, 1983.
62. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and Boyer, H.W. Gene 2:95-113, 1977.
63. Rüther, U., Koenen, M., Otto, K. and Müller-Hill, B. Nucleic Acids Res. 9:4087-4098, 1981.
64. Rüther, U. Nucleic Acids Res. 10:5765-5772, 1982.
65. Messing, J., Gronenborn, B., Müller-Hill, B. and Hofschneider, P. Proc. Natl. Acad. Sci. USA 74:3642-3646, 1977.
66. Maxam, A.M. and Gilbert, W. Proc. Natl. Acad. Sci. USA 74:560-564, 1977.
67. Maxam, A.M. and Gilbert, W. Methods Enzymol. 65:499-560, 1980.
68. Sanger, F., Nicklen, S. and Coulson, A.R. Proc. Natl. Acad. Sci. USA 74:5463-5467, 1977.
69. Groneberg, J., Chardonnet, Y. and Doerfler, W. Cell 10:101-111, 1977.
70. Yasue, H. and Ishibashi, M. Virology 116:99-115, 1982.
71. Johansson, K., Persson, H., Lewis, A.M., Pettersson, U., Tibbetts, C. and Philipson, L. J. Virol. 27:628-639, 1978.
72. Cook, J.L. and Lewis, A.M., Jr. Cancer Res. 39:1455-1461, 1979.
73. Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E, Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E. and Roberts, R.J. J. Biol. Chem. 257:13475-13491, 1982.
74. Roberts, R.J., Sciaky, D., Gelinas, R.E., Jiang, B.-D., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, K.E., O'Neill, K.E. and Gingeras, T.R. Cold Spring Harbor Symp. Quant. Biol. 47:1025-1037, 1982.
75. Wilson, J.H., Berget, P.B. and Pipas, J.M. Mol. Cell. Biol. 2:1258-1269, 1982.
76. Gahlmann, R. and Doerfler, W., submitted, 1984.
77. Hammarskjöld, M.-L. and Winberg, G. Cell 20:787-795, 1980.
78. Hammarskjöld, M.-L. and Winberg, G. Abstract Gulbenkian Workshop on the Molecular Biology of Adenoviruses, Sintra, Portugal, 1983.
79. Tyndall, C., Younghusband, H.B. and Bellet, A.J.D. J. Virol. 25:1-10, 1978.
80. Sugisaki, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I., Shimojo, H., Sawada, Y., Uemizu, Y., Uesugi, S. and Fujinaga, K. Cell 20:777-786, 1980.
81. Bos, J.L., Polder, L.J., Bernards, R., Schrier, P.I., van den Elsen, P.J., van der Eb, A.J. and van Ormondt, H. Cell 27:121-131, 1981.
82. Kimura, T., Sawada, Y., Shinawa, M., Shimizu, Y., Shiroki, K., Shimojo, H., Sugisaki, H., Takanami, M., Uemizu, Y. and Fujinaga, K. Nucleic Acids Res. 9:6571-6589, 1981.
83. Doerfler, W., Stabel, S., Ibelgaufts, H., Sutter, D., Neumann, R., Groneberg, J., Scheidtmann, K.H., Deuring, R. and Winterhoff, U. Cold Spring Harbor Symp. Quant. Biol. 44:551-564, 1979.

84. Steenberg, P.H., Maat, J., van Ormondt, H. and Sussenbach, J.S. Nucleic Acids Res. 4:4371-4389, 1977.
85. Arrand, J.R. and Roberts, R.J. J. Mol. Biol. 128:577-594, 1979.
86. Shinagawa, M. and Padmanabhan, R. Biochem. Biophys. Res. Comm. 87:671-678, 1979.
87. Tolun, A., Aleström, P. and Pettersson, U. Cell 17:705-713, 1979.
88. Brusca, J.S. and Chinnadurai, G. Virology 129:382-392, 1983.
89. Kemper, B. and Garabett, M. Eur. J. Biochem. 115:123-131, 1981.
90. Holliday, R. Genet. Res. 5:282-304, 1964.
91. Mizuuchi, K., Kemper, B., Hays, J. and Weisberg, R.A. Cell 29:357-365, 1982.
92. Sutcliffe, J.G. Cold Spring Harbor Symp. Quant. Biol. 43:77-90, 1978.
93. Shimizu, Y., Yoshida, K., Ren, C.-S., Fujinaga, K., Rajagopalan, S., and Chinnadurai, G. Nature 302:587-590, 1983.
94. Majors, J.E. and Varmus, H.E. Nature 289:253-258, 1981.
95. Ross, W. and Landy, A. Cell 33:261-272, 1983.
96. Landy, A. and Ross, W. Science 197:1147-1160, 1977.

9

CLONING OF PAPILLOMAVIRUS DNA

Lutz GISSMANN* and Elisabeth SCHWARZ

Institut für Virologie, Zentrum für Hygiene, Hermann-Herder-Str. 11, 7800 Freiburg, FRG

*Present adress: Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 6900 Heidelberg, FRG

INTRODUCTION

Papillomaviruses induce epithelial proliferation in man and in a number of animal species (for review see 1, 2). Replication of the virus particles occurs within those tumors, but is restricted to the upper layers of the epidermis (3, 4) indicating that only cells at a certain level of differentiation can support the virus multiplication. This might be the reason for the lack of an in vitro system to propagate the virus particles (for review see 5). Thus, investigation of papillomaviruses depends upon their isolation from biopsy material which contain - as a property of the virus type-different quantities of virus particles. Even those human papillomaviruses which are of special interest because their association with malignant tumors is suggested by epidemiological or histological studies (for review see 6) are poorly replicated in the respective lesions. Therefore the availability of molecularly cloned viral DNA is a prerequisite to investigate the physical organization as well as the biological behaviour of these viruses in more detail.

This article first describes different cloning strategies (Table 1) which were employed for papillomavirus genomes emphasizing the different types of human papillomaviruses (HPV). For details concerning the cloning technology per se the reader is referred to the excellent laboratory manual by Maniatis et al. (7). In the following chapters use of the cloned material in experiments on molecular characterization, transforming ability and detection of the

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

papillomavirus genomes in tumor biopsies will be reported.

Cloning of human papillomavirus DNA from virus particles

This approach can only be used with papillomaviruses which are produced in considerably high quantities within the different papillomas. The virus particles are purifed in CsCl equilibrium gradients (8) or by velocity centrifugation (9). Supercoiled DNA extracted from the virions is obtained after equilibrium centrifugation in a 4.5 M CsCl solution containing intercalating dyes (10). The papillomavirus DNA is then cleaved with restriction endonucleases to identify the most suitable enzyme for the insertion into a given bacterial plasmid. Single restriction enzyme recognition sites which are also present in the vector permit the cloning of the circular papillomavirus DNA in one piece and the proper recovery from the plasmid after DNA amplification. Cloning in different fragments must be performed if a suitable enzyme cleaves the viral DNA more than once (11) or if only partial cloning of the genome is permitted by the national safety guidelines (12).

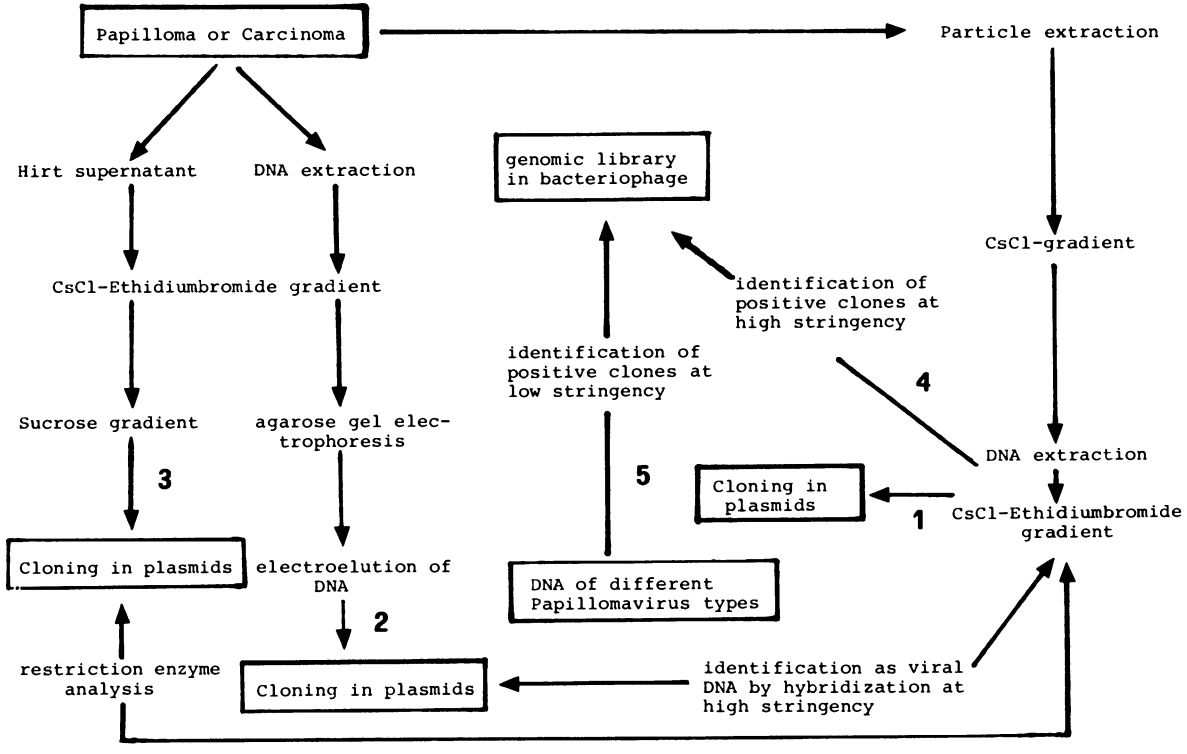
Whereas the availability of the complete genome is desired for biological studies (13, 14, 15), restriction enzyme analysis (12, 16, 17, 18, 19, 20), comparison with other papillomavirus types or for detection of this DNA in other tumors by hybridization (see below), fragments of the papillomavirus DNA are required in some instances. Subcloning into smaller pieces is the only way to obtain clean probes for the analysis of definite areas of the papillomavirus DNA. Subgenomic fragments were used to identify the transforming region of BPV (21), for mapping of BPV transcripts (22, 23) and for DNA sequencing by the M 13 dideoxy method (24, 25, 26).

Cloning of HPV genomes from cellular DNA

In contrast to the epidermal proliferations at the hands, feet, trunk or head which usually contain high particle concentrations, the papillomas located at other sites of the body (i.e. penile or vulval warts, flat intravaginal or cervical lesions and laryngeal papillomas) which are induced preferentially by two different virus types,

Table 1

Cloning strategies used for Papillomavirus DNAs



HPV 6 or HPV 11 (27), harbour these viruses only in minute quantities (28, 29). The replication of viral DNA, however, usually occurs at a higher extent (30). Thus approximately 100 ng of supercoiled DNA could be recovered from one genital wart (condylomata acuminatum) by centrifugation of total cellular DNA in CsCl-ethidium bromide equilibrium gradients (31). This DNA was further purified by gel electrophoresis and its viral origin was shown by hybridization with DNA which was extracted from CsCl gradient purified particles derived from the same genital wart (31). Cloning of the molecules in pBR322 (12) unequivocally proved this human papillomavirus to be a thus far unknown type.

A similar approach was used for cloning of HPV 8 (32) or of HPV 5 and HPV 9 DNA, respectively (18), whose restriction enzyme pattern has been shown before from virus particle derived material (33). The DNAs of HPV 5 and 9 were extracted directly from the clinical biopsies using a modified Hirt procedure (34) and further purified in a CsCl-ethidium bromide equilibrium gradient followed by centrifugation through a sucrose gradient.

HPV 11 genomes were first identified in a juvenile laryngeal papilloma as circular molecules of 8 kb in length by Southern blot analysis with ^{32}P-labelled HPV 6 DNA at conditions of reduced stringency (19). Since the total amount of this tumor DNA containing only 10 genome equivalents of viral DNA per cell was very limited, the theoretical yield of less than 1 ng HPV 11 DNA did not promise a successful cloning if one follows the same protocol as described before. Instead of this, a genomic library was constructed in a derivative of bacteriophage lambda (L 47) (35) after cleavage with Bam HI which linearized the HPV 11 DNA. HPV DNA positive recombinants were identified by the Benton and Davis plaque hybridization technique (36) using the cloned HPV 6 DNA as radioactive probe. HPV 11 DNA was then subcloned in pBR322 to have a more favourable insert-to-vector ratio in the recombinant DNA (37). Similarly, the DNAs of HPV 2 and 3 were cloned at their single Eco RI or Bam HI

sites, respectively, HPV 5 and HPV 10 DNA was obtained in two different Bam HI fragments (38). Viral DNA recovered from agarose gels after separation from the wart DNA by CsCl-ethidium bromide centrifugation served as probe for the identification of HPV positive plaques.

HPV 13, although containing three Bam HI sites exists in a single clone, obtained by partial cleavage of cellular DNA from a focal epithelial hyperplasia (Morbus Heck) in which this papillomavirus has been identified by Southern blot hybridization at conditions of low stringency with ^{32}P-labelled HPV 6 DNA (20).

The HPV DNAs described above were cloned either directly from virus particles or from purified DNA, respectively, or a well-characterized viral DNA was used to identify HPV positive recombinants in a shot-gun cloning experiment. Due to the sequence relationship of different papillomavirus DNAs the latter approach can also be followed to screen for papillomavirus sequences if no homologous DNA is available as probe. Using hybridization conditions of low stringency (39), HPV genomes were identified in clones from the genomic libraries of two different cervical carcinomas in bacteriophage lambda (40, 41). Hybridization at high stringency conditions proved these sequences to have very little if any homology with the other HPV genomes described thus far but revealed presence in a number of different genital tumors (40, 41). A more detailed analysis indicated that complete and defective viral genomes persist within those cancer biopsies as oligomeric circular molecules (42) or are integrated in the cellular DNA (43). This of course raises the question whether such sequences are in fact viral genomes. In the following the criteria are summarized which should be fulfilled to designate a cloned sequence which is not detected in virus particles as HPV DNA.

1. Homology with other HPV DNAs

This can be tested by hybridization of the material in question with the other HPV DNAs at conditions of different

stringency.

2. Homology with other HPV genomes at different parts of the molecule

In Southern blot hybridizations radioactively labelled HPV DNA should recognize different restriction enzyme fragments of the unknown sequence and vice versa. This posulate not being a platitude is substantiated by the recent cloning of HPV related cellular sequences between 2.5 and 15 kb in size from different cervical and laryngeal carcinomas and from the HeLa cell line (44). They consist of repetitive as well as unique cellular DNA harbouring a short stretch of homology to a 0.6 kb fragment covering the end of the late and part of the untranslated region (45) within the HPV 6 genome (46).

3. Colinearity with defined HPV genomes

Restriction enzyme mapping of the DNA to be tested and hybridization of the purified fragments with cleaved DNA of a well defined HPV type must show a linear relationship between both genomes. It is superfluous to mention that the final proof of a similar organization as for the other papillomaviruses comprising an early and late as well as untranslated regions cannot be undertaken until the complete nucleotide sequence is presented.

4. Presence of the DNA (at least in some tumors) as episomes of approximately 8 kb in length which is characteristic for papillomavirus DNA necessary to be encapsidated into infectious virus particles.

Analysis of the tumor DNA after CsCl-ethidium bromide equilibrium centrifugation or on a two-dimensional gel electrophoresis discriminating between circular and linear molecules (47) unequivocally shows the occurrence of circular papillomavirus DNAs of monomeric length.

As summarized for the DNAs cloned from different tumors (40, 41) and from the HeLa cell line of cervical and laryngeal origin (44), HPV 16 fulfills all the requirements

Table 2

Properties of HPV sequences isolated from human tumors

Designation	Origin	homology to non stringent	homology to stringent	colinear with	cellular DNA	8kb circle
HPV 16 7.2kb	cervical carcinoma	HPV 1,2,3,4,5, 6,8,9,10, 11,13,16	HPV 12,14, 15	HPV 6b "HPV" 18	-	+
"HPV" 18 7.8kb	cervical carcinoma	HPV 1-6, 8-17	HPV 2, 11, 13	HPV 6b HPV 16	-	-
HPV related 2.5, 15kb	cervical carcinoma (HPV 16 positive)	HPV 3,6,10,11, 16	-	only short piece of homology	+	-
HPV related 4, 4.7kb	cervical carcinoma ("HPV"18 positive)	HPV 6,1o,11,13, 16	-	not tested	+	n.t.
HPV related 4kb	laryngeal carcinoma	HPV 16 (others not tested)	n.t.	n.t.	+	n.t.
HPV related 6kb	KB Cell-Line (HeLa)	HPV 1,2,3,4,5, 6,8,9,11, 13,16,18	n.t.	n.t.	+	n.t.

Table 3

List of cloned Papillomavirus DNAs

Virus Type	Vector	Method*	Origin	Reference
HPV 1	pBR 322	1	common wart	16, 17
HPV 2	pBR 322	1	hand wart	17
HPV 3	pBR 322 L 47	3, 4	flat wart	38, 48
HPV 4	pBR 322	1	common wart	17
HPV 5	pBR 322 L 47	3, 4	epidermodysplasia verruciformis (e.v.)	18
HPV 6	pBR 322	2	genital wart	12
HPV 7	not yet cloned		butcher's warts	49, 50
HPV 8	pBR 322	2	e.v.	32
HPV 9	pBR 322	3	e.v.	18
HPV 10	pBR 322 L 47	1, 3, 4	e.v., flat wart	38, 48, 51
HPV 11	L 47 pBR 322	5	laryngeal papilloma	19
HPV 12	pBR 322	3	e.v.	48
HPV 13	L 47	5	Morbus Heck	20
HPV 14	pBR 322	3	e.v.	52
HPV 15	pBR 322	3	e.v.	52
HPV 16	L 47 pBR 322	5	cervical carcinoma	40
HPV 17	pBR 322	3	e.v.	52
HPV 18	L 47	5	cervical carcinoma	41
HPV 19-24	pBR 322	3	e.v.	52
BPV 1	pBR 322 pAT 153	1	cutaneous fibro-papillomas	53, 54
BPV 2	pAT 153	1	"	54
BPV 3	pAT 153 L 47	4	atypical bovine papillomas	55
BPV 4	pAT 153	1	alimentary tract papillomas	54
BPV 5	pAT 153	1	rice grain like papillomas	56
BPV 6	pAT 153	1	teat papillomas	57
CRPV	pBR 322	1	Cottontail rabbit papillomas	15, 46

* according to Table 1

listed above, another DNA which was not yet found in a packageable size was tentatively designated as "HPV 18". In contrast, the nature and function of the HPV related sequences, listed in table 2, remain unknown.

In table 3, the cloned DNAs which are available from human, bovine or cottontail rabbit papillomaviruses are summarized.

Cell transformation by molecularly cloned papillomavirus DNA

Up to now no tissue culture system is available for the propagation of papillomaviruses. However, BPV virions are capable of in vitro transformation of bovine, mouse and hamster cells (58, 59, 60, 61, 62, 63, 64). Furthermore, BPV DNA extracted from the virus particles can also transform rodent cells in vitro (65, 66). By Lowy et al. (21) it was demonstrated for the first time that transformation of NIH 3T3 and C127 mouse cells can also be achieved by molecularly cloned BPV 1 and BPV 2 DNA. The BPV 1 DNA was cloned in pBR322 either at the single Bam HI site or at the single Hind III site (53). After separation of the pBR322 sequences from the BPV 1 DNA by restriction endonuclease digestion, the DNAs were transfected into the mouse cells. Foci of transformed cells appeared within two weeks. Albeit with a lower efficiency, transformation was also successful by using a subgenomic Bam HI - Hind III fragment which covers 69% of the entire BPV 1 genome. The 31% counterpart fragment as well as any of the other subgenomic fragments tested failed to induce transformation. The transformed cell lines all grew in soft agar and induced tumors in athymic mice.

In C127 mouse cells transformed either by BPV 1 virions, by the linearized BPV 1 DNA or by the 69% subgenomic fragment (BPV 1_{69T}), the BPV 1 DNA was present as a circular extrachromosomal molecule in about 10-200 copies per cell (13). No evidence for integration of BPV 1 sequences into the host genome was obtained indicating that in the case of BPV cell transformation is mediated by the nonintegrated

viral DNA.

To obtain a high transformation rate it was absolutely necessary to separate the BPV sequences from the pBR322 moiety prior to transfection. Transfection of the intact circular hybrid DNA resulted in a reduction of the transformation efficiency to approximately 1% of that of the linear full-length BPV 1 DNA (67). The inhibitory effect of the covalently linked pBR322 DNA was even more drastic when using the BPV 1 $_{69T}$ fragment for transformation (67, 68). This observation and the reproducibly lower transformation efficiency of the 69% fragment suggested that BPV1 $_{69T}$ does not possess the full transformation capacity and that some sequences are located within the 31% nontransforming fragment which play a supporting role in the transformation process. Attempts to replace pBR322 by other bacterial vector sequences which will not inhibit the transformation acticity of BPV 1 DNA will be discussed later in connection with the use of BPV 1 DNA as a eukaryotic vector. The localization of the transforming sequences in the BPV 1 genome will be discussed in the next chapter.

Campo and Spandidos (14) reported the transformation of NIH 3T3 cells by the molecularly cloned DNA of BPV type 4. This is particularly interesting because virus particles of BPV 4 are not able to transform mouse cells (57) in contrast to BPV 1 or BPV 2 virions. The authors used the pBR322 derivative pAT153 as bacterial vector and found that the separation of the BPV DNA from the pAT153 DNA only slightly improved the transformation efficiency.

In addition to C127 and NIH 3T3 cells, molecularly cloned BPV DNA is also able to transform various other rodent cells in vitro, such as Fisher rat embryo cells (69), FR3T3 rat fibroblasts (66) and primary hamster embryo cells (57).

Recently, it has been shown for the cottontail rabbit papilloma virus CRPV that the virus particles as well as the DNA molecularly cloned in pBR322 can transform NIH 3T3 and C127 mouse cells in vitro (15). In contrast to BPV 1,

the covalently linked pBR 322 sequences did not hamper transformation. The CRPV DNA persists in the transformed cells as a multicopy episome.

All attempts to productively infect or to transform tissue culture cells with human papillomavirus particles have failed so far. Cultured human epidermal keratinocytes which resemble in part the natural target tissue of HPV and undergo differentiation to some extent could be infected with HPV 1 and the viral DNA persisted as a replicating extrachromosomal episome with 50-200 copies per cell (70). However, there was no detectable synthesis of virion proteins and no viral particles could be isolated from the infected cultures. HPV-specific RNA could be detected in the nucleus, but only in very small amounts in the cytoplasm suggesting that some step(s) in the production of stable cytoplasmic HPV mRNA cannot proceed properly in the cultured keratinocytes.

Cotransfection of C127 cells with the BPV1 $_{69T}$ fragment and HPV 11 DNA linearized at the Bam HI site resulted in transformed cells which contained both the BPV 1 and HPV 11 sequences covalently joined in a single molecule of about 13 kb (71). Furthermore, BPV1 $_{69T}$ was used to introduce into C127 cells HPV 11 or HPV 16 DNA which were ligated to the BPV1 $_{69T}$ fragment prior to transfection. The transformation efficiency of these recombinant DNAs was higher than that of BPV1 $_{69T}$ alone (71).

The morphological transformation of C127 mouse cells was reported by the use of molecularly cloned HPV 5 DNA (72).

DNA sequence analysis and genome organization of papillomaviruses

By using molecularly cloned DNA, the complete nucleotide sequences of two human papillomaviruses HPV 1a (11, 25) and HPV 6b (45) and of two animal papillomaviruses BPV 1 (24, 73) and CRPV (74) have been determined. The DNA sequence analysis and the comparison of the different papillomavirus sequences together with data on the transformation

and transcription of particularly BPV 1 make it possible to get insights into the structural and functional organization of the papillomaviruses. The BPV 1, HPV 1a and HPV 6b genomes consist of 7944, 7814 and 7902 base pairs, respectively. Based on characteristic partial homologies, the BPV 1 and HPV 1a sequences have been aligned (25) and they have been subjected to a detailed comparative analysis (75). The HPV 6b sequence can be matched with the two other sequences as well (45). The determination of the open reading frames in the two DNA strands of each papillomavirus genome showed clearly that the three papillomaviruses have a very similar genome organization which, however, is quite distinct from that of the taxonomically related polyomaviruses: all of the major open reading frames are located on only one DNA strand, indicating that viral transcription is unidirectional over the entire genome. An alignment of the HPV 6b, HPV 1a and BPV 1 genomes emphasizing the similar arrangement of the coding regions is depicted in Fig. 1.

By in vitro transformation of mouse cells, the BPV 1 genome was functionally dissected into a 69% transforming (BPV 1_{69T}) and a 31% non-transforming (BPV 1_{31NT}) Bam HI - Hind III subgenomic fragment (21). BPV 1 specific transcripts detected in transformed cells map all within the 69% segment (23) and the BPV 1_{69T} DNA is maintained in the transformed cells as an extrachromosomal multicopy plasmid (13), indicating that the BPV 1_{69T} fragment contains the genetic elements for transformation and autonomous replication. Transcripts hybridizing to the 31% fragment are found only in the peripheral parts of bovine warts where virus replication takes place (76), suggesting that the late functions are encoded in this part of the genome. In agreement with this assumption, the 31% segment is found to be covered completely by two large open reading frames (Fig. 1). They have been designated L1 and L2, because they are thought to code for the structural (late) proteins of the virus. In the 69% segment two larger open reading frames E1 and E2 and several smaller ones E3 to E8 are located (Fig. 1).

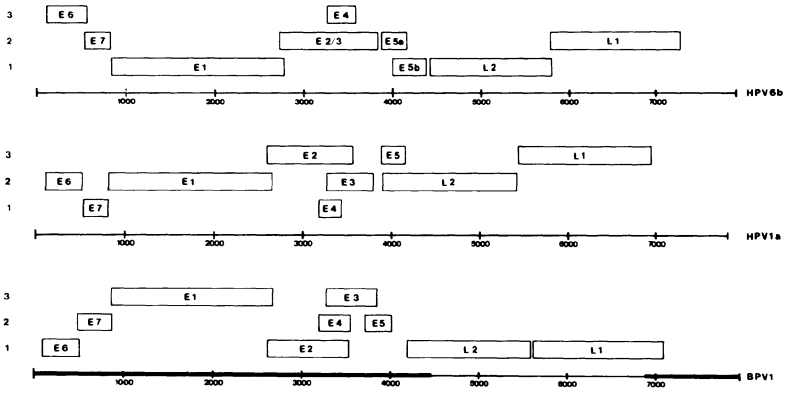


FIGURE 1. Alignment of the genomes of HPV 6b, HPV 1a and BPV 1. For numbering of nucleotide positions the genomes were linearized at a conserved sequence containing a Hpa I recognition site in HPV 1a and BPV 1. The 69% transforming fragment of HPV 1 DNA framed by a Bam HI and Hind III site at positions 4451 and 6959, respectively, is marked by a heavy bar. With the exception of E5, the open reading frames are indicated starting at the position of their respective potential ATG translational start codon. Reading frame E8 which is nonhomologous between the three genomes is omitted. Reading frames E5 of HPV 1a and BPV 1 are also nonhomologous (75).

The sequence comparison in the E and L coding regions of BPV 1 and HPV 1a revealed a distinct pattern of clustered sequence homologies (75). A very similar pattern of homology is observed when HPV 6b is compared with either BPV 1 or HPV 1, showing that blocks of homologous amino acid sequences are often conserved within all three papillomavirus sequences.

The most prominent ones are found at the beginning of L2, throughout L1, in the C-terminal half of E1 and in the N-terminal part of E2. Differences in the nucleotide sequences within these conserved regions often concern the third codon positions and therefore do not alter the amino acid sequences, a fact which argues strongly that these open reading frames represent genuine genes or exons. The separate coding regions E2 and E3 in HPV 1 and BPV 1 are joined in the HPV 6 genome to give one continuous open reading frame E2/3. This characteristic difference to HPV 1 and BPV 1 is also found in the CRPV sequence (74). Reading frames E6 and to a minor extent E7 exhibit as a common structural feature a homologous spacing of the tetrapeptide Cys-X-X-Cys. Mapping data of mRNA transcripts from BPV 1 transformed cells (23) and from HPV 1-infected COS monkey cells (77) indicate strongly that the various open reading frames in the E region are transcribed into mRNA and actually code for viral peptides.

The region between the end of L1 and the beginning of E6 seems to be noncoding in all three papillomavirus genomes. It extends for 943 bp and 982 bp in BPV 1 and HPV 1a, respectively (75) and has a smaller size of 715 bp in HPV 6b. Characteristic structural features common to the three sequences are: an A-T rich region, polyadenylation signals AATAAA, directly repeated sequences and promoter-like sequences which are located in front of the E6 reading frame. They may constitute the promoter for transcription of the E region genes. The noncoding region of HPV 1 has been shown to contain a sequence which supports autonomous replication indicating that it is part of the origin of replication (78). By constructing specific deletion mutants of the

cloned BPV 1 DNA and assaying their ability to transform mouse cells, it has been shown that the BPV1 $_{69T}$ fragment contains two discontinuous segments which are both required for transformation (79, 80). One of them is located upstream of the E region and can be functionally replaced by either a retroviral long terminal repeat of the SV40 early promoter suggesting that it contains a transcriptional regulatory element. The other segment is bounded by the Bam HI site and is presumed to contain the transforming gene(s). The same authors could demonstrate that deletions affecting the E1 open reading frame resulted in an integration of the BPV 1 sequences indicating that the E1 gene product is required for extrachromosomal replication of BPV 1 DNA. A transcriptional enhancer element has been localized in the BPV 1 genome at the extreme Bam HI end of the BPV1 $_{69T}$ fragment (81). In addition, transcriptional control sequences have been identified in the BPV 1 genome by a functional assay in which subgenomic Hind III- Hae III fragments of BPV 1 DNA were tested for their ability to replace the promoter of the HSV 1 tk gene thus leading to a conversion of LA tk^- cells to the tk^+ phenotype upon transfection with the recombinant plasmids (82). Four fragments were found to be able to induce the expression of the tk gene, one of them which spans the Bam HI site, in an orientation independent, i.e. enhancer-like manner. This result is compatible with the localization of a BPV 1 enhancer element by Lusky et al.(81). Two other fragments showed an orientation dependent, i.e. promoter like induction of tk gene expression. Their localization within the BPV 1 genome proximal to the E region is in agreement with the localization of a promoter element by Nakabayashi et al. (79) and Sarver et al. (80).

BPV 1 as a eukaryotic cloning vector

The ability of molecularly cloned BPV 1 DNA to transform mouse cells (21) and to establish itself in the transformed cells as a multicopy circular extrachromosomal plasmid (13) was exploited for the development of BPV 1 - derived

eukaryotic cloning vectors. In a first series of experiments, a recombinant plasmid was constructed composed of the BPV1 $_{69T}$ fragment, pBR322 DNA and a 1.62 kb fragment containing the entire rat preproinsulin gene I (rI_1) together with regulatory signals for transcription (83). The pBR322 sequences were separated by restriction endonuclease digestion from the BPV1 $_{69T}$ - rI_1 moiety before transfection of C127 mouse cells and selection of morphologically transformed cells. All transformed cells tested contained multiple copies of the BPV1 $_{69}^{T}$ - rI_1 recombinant DNA present exclusively as circular nonintegrated molecules. The preproinsulin gene was correctly transcribed and spliced and the processed gene product proinsulin was demonstrated to be synthesized in substantial amounts and to be secreted into the medium.

Constitutive expression of foreign eukaryotic genes introduced into mouse cells by using the BPV vector system has been demonstrated for the human ß-globin gene (68) and for the hepatitis B virus surface antigen gene (84, Wang et al., cited in 85). HBsAg particles were secreted in high amounts into the medium.

Regulated expression of genes inserted into BPV 1 - derived vectors has been demonstrated for the human ß-interferon gene (86, 87) and for the human growth hormone gene (88). The interferon gene was inducible by inactivated Newcastle disease virus or polyriboinosinic acid-polyribocytidylic acid. The growth hormone gene was linked to the promoter region of the mouse metallothionein gene. Its expression could be induced by cadmium and large quantities of human growth hormone were secreted into the medium (88). The BPV-metallothionein hybrid should be useful in general as a eukaryotic expression vector if the human growth hormone gene is replaced by other gene sequences which are thereby placed under the control of the metallothionein promoter for efficient and regulatable expression.

In order to get independent of the BPV-induced transformation as the sole selective marker, the Escherichia

coli gpt gene coding for xanthine-guanine phosphoribosyl-transferase was joined to the BPV replicon (89). Cells that express the gpt gene can be selected in HAT medium containing xanthine and mycophenolic acid (90). After transfection of the hybrid DNA into C127 cells, cells were selected either for a BPV-induced transformed phenotype or for a gpt-induced ability to grow in selective media. Many of the cell lines obtained expressed both phenotypic markers. The hybrid DNA was present as circular extrachromosomal molecules but was found to be extensively rearranged in many cases. In tk^- cells transfected with BPV recombinant DNA containing the HSV tk gene and selected for the tk^+ phenotype, nearly all of the BPV tk plasmid DNA was found to be in an integrated state (91, 92).

In a further step towards a broad versatility, the BPV-vector system was modified in such a way that it can be used to shuttle genes between mammalian cells and bacteria. For this purpose it was necessary to combine the BPV 1 segment with bacterial vector sequences that allow replication and selection in bacteria. As has been mentioned already, the capacity to transform mouse cells, however, is drastically reduced when the entire BPV 1 DNA or the 69% transforming fragment is covalently linked to pBR322 DNA (21). For this reason, it was necessary to separate the BPV 1 segment from the bacterial plasmid sequences prior to transfection into mouse cells. This prerequisite for efficient transformation, however, eliminated the possibility for subsequent recovery of the hybrid plasmid DNA in bacteria. The pBR322 vector sequences were replaced by certain deletion derivatives of pBR322 which lack the sequence that inhibits the replication of SV40 - pBR322 recombinant DNA in monkey cells (93). Transfection of mouse cells with an intact recombinant plasmid composed of the entire BPV 1 DNA covalently linked to the pBR322 deletion derivative pML2d resulted in a transformation efficiency as high as with BPV 1 DNA separated from the pML2d DNA (67). The transformation capability of the $BPV1_{69T}$ fragment, however, was severely

inhibited by the covalent linkage even of the pML2d DNA and dropped to 1/100th as compared to the digested DNA (67). As an unexpected finding, the block to transformation of mouse cells could be relieved by the insertion of a human ß-globin gene fragment into a plasmid consisting of BPV1 $_{69T}$ DNA and the pML2 analogue pBRd (68). The human ß-globin gene fragment seems to contain sequences which can substitute for the 31% nontransforming BPV 1 fragment in stimulating transformation. The recombinant DNAs were present in the transformed mouse cells as extrachromosomal multicopy episomes and no major rearrangements could be detected (67, 68). Furthermore, plasmids indistinguishable from the input DNA were recovered by subsequent transformation of bacteria with low molecular weight DNA from the transformed mouse cell lines. In experiments reported by Binétruy et al. (66), however, an efficient transformation of rat and mouse fibroblasts with interact BPV 1 - pML2 DNA was not obtained unless the recombinant DNA was transfected into the cells by polyethylene glycol-induced fusion with bacterial protoplasts. The extrachromosomal plasmids in the transformed cell lines exhibited an oligomeric structure and, if not rearranged, could be recovered in Escherichia coli after cleavage with restriction enzymes and circularization of the monomeric molecules.

In conclusion, it has been shown that molecularly cloned BPV 1 DNA or the BPV1 $_{69T}$ fragment can act efficiently as eukaryotic cloning vectors. Recombinant plasmids consisting of BPV 1 and certain deletion derivatives of pBR322 have in addition the ability to shuttle gene between mammalian and bacterial cells. Efficient and faithful expression of a variety of eukaryotic genes has been obtained by using the BPV 1 vector system. Transfected cells can be selected due to their BPV-induced transformed phenotype. Addition of a second selectable marker makes it possible to use the BPV vector system even with cells not susceptible to BPV-induced transformation. The foreign DNA sequences are amplified in the transformed cells as part of the extrachromosomally replicating BPV episome. Because no integration

occurs, the sequence environment of the foreign genes remains unchanged upon transfection into the cells.

Detection of papillomavirus DNA in tumors

Applying hybridization conditions of varying stringency in Southern blot analysis, cloned papillomavirus DNA may be used as probe for detection of identical or related sequences in benign papillomas as well as in malignant tumors.

HPV 6 or 11 DNA were shown to be present in the majority of genital warts (19, 27) and laryngeal papillomas (19, 94, 95, 96) as well as in clinical suspicious tissue of the larynx obtained from patients with a history of laryngeal papillomas (95). Unclassified HPV DNA was also found in laryngeal papillomas by hybridization at low stringency (97).

This method can be used for typing of any papillomavirus DNA in biopsy materials if the different HPV DNAs are available. This of course requires various hybridization experiments with the individual ^{32}P-labelled probes. Alternatively, labelling of total cellular DNA extracted from the clinical material to be analyzed and hybridization to the cloned HPV DNA which has been blotted onto nitrocellulose filters allows the identification of the papillomavirus type in the respective tissue(Fig. 2). As shown from reconstitution experiments by mixing different quantities of HPV DNA to normal cellular DNA prior to labelling, this procedure permits the detection of ten genome equivalents per cell (98).

As an association between papillomavirus infection and development of malignant tumors has been discussed for many years (6), the availability of cloned papillomavirus DNA is necessary to screen those materials for the presence of viral sequences (Table 4).

Acknowledgements:

We are grateful to Dr. H. zur Hausen for critical reading and Miss S. Näger for typing this manuscript.

We thank our colleagues M. Dürst, H. Ikenberg, A. Stremlau, A. Kleinheinz, K. Dartmann and M. Boshart for stimulating

FIGURE 2. Detection of HPV 6 DNA in a genital wart.

1 µg of total cellular DNA was radioactively labelled by incorporation of ^{32}P-TTP and blot hybridized to HPV 3 (cloned in lambda), HPV 4 (cloned in pBR322), HPV 5 (cloned in lambda), HPV 6 and 8 (cloned in two fragments in pBR 322) and HPV 11 (cloned in a pBR 322 derivative). Vector DNA was not cleaved off except for HPV 11.
PM2DNA was included as marker.
Positive reaction with HPV 11 DNA is due to the close sequence relationship between HPV 6 and 11 (19).

Courtesy of Dr.H.Ikenberg

Table 4 Occurrence of HPV DNA in malignant human tumors

Papillomavirus	Tumor	Frequency	Reference
HPV 3	e.v.	low	52
HPV 5	e.v.	high	33,52,99, 100
HPV 6	verrucous carcinoma (Buschke Löwenstein-tumor)	high	101,102, 103
HPV 8	e.v.	high	52
HPV 10	genital tumor	low	51
HPV 11	genital tumor	low	27
HPV 14	e.v.	low	52
HPV 16	genital tumor	high	40,104
HPV 18	genital tumor	medium	41

discussions and Drs. G. Orth, D. Kremsdorf, M. S. Campo, P. Howley, W. F. H. Jarrett, M. Lusky, D. Lowy and I. Giri for communicating unpublished data.

Original work described in this article was supported by the Deutsche Forschungsgemeinschaft (SFB 31: Tumorentstehung und -Entwicklung)

REFERENCES

1. Lancaster WD, Olson C, 1982.
   Microbiol. Rev., 46, 191-207.
2. Pfister H, 1983.
   Rev. Physiol. Biochem. Pharmacol.
   Springer Verlag, in press.
3. Almeida JD, Howatson AF, Williams MG, 1962.
   J. Invest. Dermatol., 38, 337-345.
4. Grussendorf EI, zur Hausen H, 1979
   Arch. Derm. Res., 264, 55-63.
5. Rowson KEK, Mahy BWJ, 1967.
   Bacteriol. Rev., 31, 110-131.
6. zur Hausen H, 1977.
   Curr. Top. Microbiol. Immunol., 78, 1-30.
7. Maniatis T, Fritsch EF, Sambrook J, 1982.
   Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. Gissmann L, zur Hausen H, 1976.
   Proc. Natl. Acad. Sci. USA, 73, 1310-1313.
9. Favre M, Breitburd F, Croissant O, Orth G, 1975.
   J. Virol., 15, 1239-1247.
10. Bauer W, Vinograd J, 1968.
    J. Mol. Biol., 33, 141-171.
11. Clad A, Gissmann L, Meier B, Freese UK, Schwarz E, 1982.
    Virology, 118, 254-259.
12. de Villiers EM, Gissmann L, zur Hausen H, 1981.
    J. Virol., 40, 932-935.
13. Law MF, Lowy DR, Dvoretzky I, Howley PM, 1981.
    Proc. Natl. Acad. Sci. USA, 78, 2727-2731.
14. Campo MS, Spandidos DA, 1983.
    J. Gen. Virol., 64, 549-557.
15. Watts SL, Ostrow RS, Phelps WC, Prince JT, Faras AJ, 1983.
    Virology, 125, 127-138.
16. Danos O, Katinka M, Yaniv M, 1980.
    Eur. J. Biochem. 109, 457-461.
17. Heilman CA, Law MF, Israel MA, Howley PM, 1980.
    J. Virol., 36, 395-407.
18. Kremsdorf D, Jablonska S, Favre M, Orth G, 1982.
    J. Virol., 43, 436-447.
19. Gissmann L, Diehl V, Schultz-Coulon HJ, zur Hausen H, 1982.
    J. Virol., 44, 393-400.
20. Pfister H, Hettich I, Runne U, Gissmann L, Chilf GN., 1983.
    J. Virol., in press.
21. Lowy DR, Dvoretzky I, Shober R, Law MF, Engel L., Howley, PM, 1980.
    Nature, 287, 72-74.
22. Freese UK, Schulte P, Pfister H, 1982.
    Virology, 117, 257-261.
23. Heilman CA, Engel L, Lowy DR, Howley PM, 1982.
    Virology 119, 22-34.
24. Chen EY, Howley PM, Levinson AD, Seeburg PH, 1982.
    Nature, 299, 529-534.
25. Danos O, Katinka M, Yaniv M, 1982.
    EMBO J., 1, 231-236.

26. Dartmann K, Schwarz E, Gissmann L, zur Hausen H, unpublished data.
27. Gissmann L, Wolnik L, Ikenberg H, Koldovsky U, Schnürch HG, zur Hausen H, 1983. Proc. Natl. Acad. Sci. USA, 80, 560-563.
28. Dunn AEG, Olgilvie MM, 1968. J. Ultrastruct. Res.,22, 110-131.
29. Boyle WF, Riggs JL, Oshiro JS, Lennette EH, 1973. Laryngoscope,83, 1102-1108.
30. Grussendorf-Conen EI, Gissmann L, Hölters J, 1983. J. Invest. Dermatol., in press.
31. Gissmann L, zur Hausen H, 1980. Int. J. Cancer, 25, 605-609.
32. Pfister H, Nürnberger F, Gissmann L, zur Hausen H, 1981. Int. J. Cancer, 27, 645-650.
33. Orth G, Favre M, Breitburd F, Croissant O, Jablonska S, Obalek S, Jarzabek-Chorzelska M, Rzesa G, 1980. In: Viruses in naturally occurring cancers, (Eds. Essex M, Todaro G, zur Hausen H), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 259-282.
34. Orth G, Breitburd F, Favre M, Croissant O, 1977. In: Origins of human cancer, (Eds. Hiatt HH, Watson JD, Winsten JA), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 1043-1068.
35. Loenen WAM, Brammar W, 1980. Gene, 20, 249-259.
36. Benton WD, Davis RW, 1977. Science,196, 180-182.
37. Gissmann L, Takada K, unpublished data.
38. Pfister H, Gassenmeier A, Gissmann L, unpublished data.
39. Law MF, Lancaster WD, Howley PM, 1979. J. Virol.,32, 199-207.
40. Dürst M, Gissmann L, Ikenberg H, zur Hausen H, 1983. Proc. Natl. Acad. Sci. USA, 80, 3812-3815.
41. Boshart M, Gissmann L, zur Hausen H, manuscript in preparation.
42. Dürst M, Boshart M, Gissmann L, zur Hausen H, manuscript in preparation.
43. Gissmann L, Dürst M, Boshart M, unpublished data.
44. Dürst M, Boshart M, Ikenberg H, Schwarz E, Gissmann L, unpublished data.
45. Schwarz E, Dürst M, Demankowski C, Lattermann O, Zech R, Wolfsperger E, zur Hausen H, submitted for publication.
46. Dürst M, Gissmann L, unpublished data.
47. Wettstein FO, Stevens JG, 1982. Proc. Natl. Acad. Sci. USA, 79, 790-794.
48. Kremsdorf et al., submitted for publication.

49. Orth G, Jablonska S, Favre M, Croissant O, Obalek S, Jarzabek-Chorzelska M, Jibard N, 1981. J. Invest. Dermatol., 76, 97-102.
50. Ostrow RS, Krzyzek R, Pass F, Faras AJ, 1981. Virology, 108, 21-27.
51. Green M, Brackman KH, Sanders PR, Loewenstein PM, Freel JH, Eisinger M, Switlyk SA, 1982. Proc. Natl. Acad. Sci. USA, 79, 4437-4441.
52. Orth G, et al. unpublished data.
53. Howley PM, Law MF, Heilman C, Engel L, Alonso MC, Israel MA, Lowy DR, Lancaster WD, 1980. In: Viruses in naturally occurring cancers, (Eds. Essex M, Todaro G, zur Hausen H) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp 233-247.
54. Campo MS, Coggins LW, 1982. J. Gen. Virol., 63, 255-264.
55. Coggins LW et al. submitted for publication.
56. Campo MS, unpublished data.
57. Jarrett WFH et al. unpublished data.
58. Black PH, Hartley JW, Rowe WP, Huebner RJ, 1963. Nature, 199, 1016-1018.
59. Thomas M, Boiron M, Tanzer J, Levy JP, Bernard J, 1964. Nature, 202, 709-710.
60. Boiron M, Levy JP, Thomas M, Friedman JC, Bernhard J, 1964. Nature, 201, 423-424.
61. Meischke HRC, 1979. J. Gen. Virol., 43, 471-487.
62. Dvoretzky I, Shober R, Chattopadhy SK, Lowy DR, 1980. Virology, 103, 369-375.
63. Morgan DM, Meinke W, 1980. Curr. Microbiol., 3, 247-251.
64. Moar MH, Campo MS, Laird H, Jarrett WFH, 1981. Nature, 293, 749-751.
65. Boiron M, Thomas M, Chenialle PH, 1965. Virology, 52, 150-153.
66. Binétruy B, Meneguzzi G, Breathnach R, Cuzin F, 1982. EMBO J., 1, 621-628.
67. Sarver N, Byrne JC, Howley PM, 1982. Proc. Natl. Acad. Sci. USA, 79, 7147-7151.
68. Di Maio D, Treisman R, Maniatis T, 1982. Proc. Natl. Acad. Sci. USA, 79, 4030-4034.
69. Graham GM, Babiss LE, Byrne J, Howley PM, Fisher PB, pers. comm.
70. La Porta RF, Taichman LB, 1982. Proc. Natl. Acad. Sci. USA, 79, 3393-3397.
71. Gissmann L, unpublished data.
72. Watts S, Phelps W, Ostrow R, Faras AJ, pers. comm.
73. Ahola H, Stenlund A, Moreno-Lopez J, Pettersson U, 1983. Nucleic Acids Res., 11, 2639-2650.

74. Giri I, Danos O, Yaniv M,
pers. comm.
75. Danos O, Engel LW, Chen EY, Yaniv M, Howley PM, 1983.
J. Virol., 46, 557-566.
76. Amtmann E, Sauer G, 1982.
J. Virol., 43, 59-66.
77. Chow LT, Broker TR, 1983.
In: Gene transfer and cancer, (Eds. Pearson ML, Sternberg NL),in press.
78. Lusky M, Botchan M,
pers. comm.
79. Nakabayaski Y, Chattopadhya SK, Lowy DR,
pers. comm.
80. Sarver N, Yang YC, Byrne JC, Howley PM,
pers. comm.
81. Lusky M, Berg L, Weiher H, Botchan M,
submitted for publication.
82. Campo MS, Spandidos DA, Lang J, Wilkie NM, 1983.
Nature, 303, 77-80.
83. Sarver N, Gruss P, Law MF, Khoury G, Howley PM, 1981.
Mol. Cell. Biol., 1, 486-496.
84. Stenlund A, Lamy D, Moreno-Lopez J, Ahola H, Pettersson U, Tiollais P, 1983.
EMBO J., 2, 669-673.
85. Stratowa C, Doehmer J, Wang Y, Hofschneider PH, 1982.
EMBO J., 1, 1573-1578.
86. Zinn K, Mellon P, Ptashne M, Maniatis T, 1982.
Proc. Natl. Acad. Sci. USA, 79, 4897-4901.
87. Mitrani-Rosenbaum S, Maroteaux L, Mory Y, Revel M, Howley PM, 1983.
Mol. Cell. Biol., 3, 233-240.
88. Pavlakis GN, Hamer DH, 1983.
Proc. Natl. Acad. Sci. USA, 80, 397-401.
89. Law MF, Sarver N, Howley PM, 1982.
In: Viral vectors (Ed. Gluzmann Y), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 79-85.
90. Mulligan RC, Berg P, 1981.
Proc. Natl. Acad. Sci., USA, 78, 2072-2076.
91. Lusky M, Berg L., Botchan M, 1982.
In: Viral vectors (Ed. Gluzman Y), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 99-107.
92. Sekiguchi T, Nishimoto T, Kai R, Sekiguchi M, 1983.
Gene,21, 267-272.
93. Lusky M, Botchan M, 1981.
Nature, 293, 79-81.
94. Mounts P, Shah KV, Kashima H, 1982.
Proc. Natl. Acad. Sci. USA, 79, 5425-5429.
95. Steinberg B, Topp WC, Scheider PS, Abramson AL, 1983.
New Engl. J. Med.,308, 1261-1264.
96. Ikenberg H, Gissmann L,
unpublished data.
97. Lancaster WD, Jenson AB, 1981.
Intervirology, 15, 204-212.

98. Gissmann L, Ikenberg H,
unpublished data.
99. Ostrow RS, Bender M, Niimura M, Seki T, Kawashima M,
Pass F, Faras AJ, 1982.
Proc. Natl. Acad. Sci. USA, 79, 1634-1638.
100. Pfister H, Gassenmeier A, Nürnberger F, Stüttgen G,
1983.
Cancer Res., 43, 1436-1441.
101. Gissmann L, de Villiers EM, zur Hausen H, 1982.
Int. J. Cancer, 29, 143-146.
102. Zachow KR, Ostrow RS, Bender M, Watts S, Okagaki T,
Pass F, Faras AJ, 1982.
Nature, 300, 771-773.
103. Boshart M, Schwarz E, Gissmann L, Gross G, zur Hausen H,
unpublished data.
104. Ikenberg H, Gissmann L, Gross G, Grussendorf-Conen EI,
zur Hausen H,
submitted for publication.

CLONING OF RETROVIRUS DNA

10

CLONING OF BOVINE LEUKEMIA VIRUS PROVIRAL INFORMATION

J. DESCHAMPS[1], A. BURNY[1,2], and R. KETTMANN[1,2]

[1]Department of Molecular Biology, University of Brussels
[2]Faculty of Agriculture, Gembloux, Belgium

Bovine leukemia virus (BLV) is an exogenous retrovirus which induces lymphoid proliferation (persistent lymphocytosis and/or lymphosarcoma)in cattle (1). Understanding the phenomenon of cell transformation at the molecular level will depend on the availability of specific probes. Among these, a cloned provirus is a mandatory requirement to investigate the presence of a cellular oncogene in the BLV genome and to determine the complete nucleotide sequence of a BLV provirus. For this reason, a BLV provirus having SacI restriction sites only in the long terminal repeat (LTR) regions was chosen for the molecular cloning procedure described below.

To prepare 8.3 kb eukaryotic DNA fragments, the internal SacI fragments of λgtWESλB were replaced by DNA inserts (in the size range of 0-11.38 kb) bearing SacI cohesive ends (2). For this purpose, 1 mg of high molecular weight DNA from tumor tissue containing BLV proviral sequences was digested to completion with SacI. It should be noted that the clonal cell population contained a unique BLV provirus having SacI sites only in the LTR of the proviral genome. The DNA fragments were separated on a 1% low-melting point agarose gel (Bethesda Research Laboratories, Inc.); λ-HindIII DNA was used as a molecular weight marker. The region of the gel containing fragments of about 8.3 kb (which is the size of the internal viral SacI fragment) was phenol-extracted as described by the manufacturer. After ethanol precipitation, the DNA was purified on an Elutip column (Schleicher and Schüll) as described by the manufacturer. About 40 µg of 8.3 kb fragments were recovered.

To prepare λgtWESλB arms, the phage was grown and purified essentially as described by Maniatis et al. (3). For preparation of phage DNA, purified phage was dialyzed against TNE (10 mM Tris-HCl, pH 8.3; 150 mM NaCl; 1 mM EDTA) and incubated for 15 min in the presence of ribonuclease A (20 µg/ml) and for 30 min in the presence of 200 µg/ml of proteinase K and 0.5% SDS. After phenol extraction, the DNA was ethanol precipitated. After centrifugation, the DNA was dissolved in TE (10 mM Tris-HCl, pH 7.5;

1 mM EDTA) buffer. To prepare the end fragments of λgtWESλB, 200 μg of phage DNA were digested to completion by SacI (1000 units of enzyme for 3 hr at 37°C) and then heated for 10 min at 68°C. After addition of $MgCl_2$ to a final concentration of 3 mM, the cohesive ends were annealed by incubation for 1 hr at 42°C. The annealed material was subsequently centrifuged on two 10-40% linear sucrose gradients (1 M NaCl; 20 mM Tris-HCl, pH 8.0; 10 mM EDTA) at 27,000 rpm for 22 hr at 15°C and 0.5 ml fractions were collected. Fractions were analyzed on a 0.5% agarose gel and those containing the λgtWESλB annealed arms were pooled.

Ligation of the 8.3 kb eukaryotic DNA fragments to λgtWESλB DNA arms was performed in a reaction mixture containing 10 μg of vector arms DNA, 2.5 μg of 8.3 kb eukaryotic insert DNA, 60 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT and 2 units of T4 DNA ligase (Bethesda Research Laboratories, Inc.). Incubation was performed at 22°C for 1 hr. Ligation efficiency was checked by agarose gel analysis of one aliquot of the ligated material. Another aliquot was packaged into phage and the number of plaque-forming units (pfu) was determined. In this case, about 300,000 pfu/μg of DNA were obtained.

In vitro packaging of DNA into phage particles was done according to Maniatis et al. (4). In vitro packaging extracts were prepared by freezing and thawing lysates and sonication of extracts. Packaging efficiency was 3×10^7 pfu/μg with intact λ phage DNA (Bethesda Research Laboratories, Inc.). Packaging was performed in 12 reactions each containing 1 μg of 8.3 kb recombinant DNA according to Blattner et al. (5). The reactions were subsequently pooled and titrated on bacterial cells (300,000 pfu/μg of phage DNA).

To screen the recombinant phages, 800,000 recombinant phages were plated on 2×10^{10} stationary phase bacterial cells on 22 x 22 cm disks. To prevent the top agar from adhering to the nitrocellulose filter when it was lifted from the plate, plates were dried for several hours and 0.7% agarose was used as the top agar. The plates containing about 100,000 recombinant molecules were incubated at 37°C for 16 hr. Plates were refrigerated for 3 hr before Biodyne A membranes (Pall Process Filtration, Ltd.) were applied for 3 min and then treated as described by the manufacturer. Filters were hybridized with a ^{32}P-labeled BLV rep probe as previously described (6). After autoradiography for 16 hr, 8 plaques containing BLV information were identified and subsequently purified and amplified (4).

The restriction map of one of these viral inserts is represented in Fig. 1.

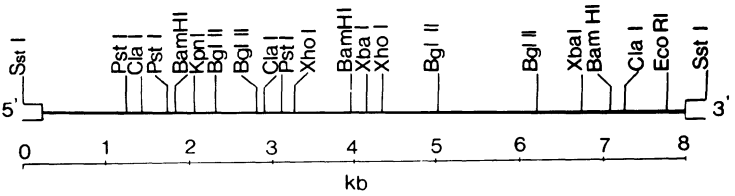


FIGURE 1. Restriction map of the cloned BLV provirus.

Use of the cloned bovine leukemia virus DNA allowed us, for the first time, to definitely rule out the existence of any cellular sequences in the BLV genome. It thus appears that BLV does not contain a cellular oncogene (6). The use of monoclonal antibodies allowed the identification of three important antigenic sites on the surface of the BLV envelope gp51 glycoprotein. These sites referred to as F, G and H were shown to be involved in the biological activities of the virus (infectivity, virus-neutralizing activity, complement-dependent cytotoxicity) (7). It also appeared that the neutralizing antibody-inducing sites of gp51 are subjected to antigenic variation among BLV isolates of the same or different geographical origins (7). It will be interesting to investigate the correlation of gp51 antigenic changes with variations in nucleotide sequence when the exact location of the coding region of the env gene in BLV proviral DNA is known. The mechanism of leukemogenesis by BLV is unknown. In tumors, the provirus is integrated at many sites and is in a repressed state. One hypothesis regarding cell transformation is that a virus-coded protein plays a key role in the initial transforming event but is not required for the maintenance of the tumor state. The availability of a cloned provirus should help investigating the coding capacity of the px region of the BLV genome, the only region the function of which is not yet known.

REFERENCES

1. Burny,A. 1984. In:Mechanisms of Viral Leukaemogenesis (Eds. J.M. Goldman and O. Jarrett), Churchill Livingstone, London (in press).
2. Leder, P., Tiemeier, D. and Enquist, L. 1977. Science 196:175-177.
3. Maniatis, T., Hardison, R.D., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. 1978. Cell 15:687-701.
4. Maniatis, T., Fritsch, E.F. and Sambrook, J. 1982. Molecular Cloning: A Laboratory Mannual. Cold Spring Harbor Laboratory, U.S.A.

5. Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Sheldon, E.L. and Smithies, O. 1977. Science 196:161-169.
6. Deschamps, J., Kettmann, R. and Burny, A. 1981. J. Virol. 40:605-609.
7. Bruck, C., Portetelle, D., Burny, A. and Zavada, J. 1982. Virology 122:353-362.

11

CLONING OF ENDOGENOUS 'RETROVIRUS-LIKE' GENES: THE MURINE VL30 FAMILY.

ELI KESHET, AHUVA ITIN , GALIT ROTMAN
Department of Virology, The Hebrew University-Hadassah Medical School, Jerusalem, Israel.

SUMMARY

Uninfected cells contain DNA sequences related to the genetic information of infectious retroviruses. In addition to these so-called 'endogenous proviruses', certain vertebrate cells contain gene families that are distinguished by several 'retrovirus-like' properties but are, nevertheless, unrelated to retroviruses (i.e., there is no nucleic acid homology with known retrovirus). The possible genetic origin of 'retrovirus-like' gene families and their potential biological significance are discussed. These issues were experimentally approached in our laboratory with a particular endogenous 'retrovirus-like' family, namely, the murine VL30 family. VL30 DNA elements are ubiquitous in murine cells where multiple copies are dispersed throughout the genome. The 30S RNA that is constitutively transcribed from these genes can be efficiently packaged in C-type virions and is capable of subsequent transmission to other cells via pseudovirion infection.

The article reviews our recent studies with cloned VL30 DNA units, highlighting the following observations: distribution patterns of VL30 elements in different species of the genus Mus and within a given species are polymorphic; there is a hundred-fold variation in the reiteration frequency of VL30 DNA among different species; the LTRs possessed by VL30 DNA units are bounded by inverted repeats ending with (5')TG...CA(3'). The nucleotide sequences at the junctions with mouse flanking DNA are suggestive of duplication of four nucleotides at the integration target and of an 'integration from within' mechanism. These structural features are also shared by proviruses and transposable elements. Adjacent to the 5' LTR is an 18-base sequence complementary to $tRNA^{pro}$ and adjacent to the 3' LTR a purine-rich tract ending with AATG. These sequences resemble the putative primer binding sites for (-) and (+)

strands, respectively, of proviral DNA. The nature of sequence heterogeneity within the VL30 family and heterogeneity of VL30-associated LTRs are described. Evidence is provided for the participation of VL30 DNA in recombinations with endogenous proviruses and for the recombinant nature of certain VL30-associated LTRs.

INTRODUCTION: ENDOGENOUS 'RETROVIRUS-LIKE' GENE FAMILIES .

A unique feature of retroviruses that distinguishes them from all other animal viruses, is the presence in the chromosomes of normal, uninfected cells, of sequences closely related to those of infectious retroviruses. These genetic elements, designated 'endogenous proviruses' are inherited as stable Mendelian genes, are evolutionarily unstable, resembling structurally normal proviruses, and vary greatly in expression (for a recent review, see ref. 1).

The variability in expression accounts for the fact that only a small fraction of the endogenous retrovirus-related DNA that can be detected physically in normal cell DNA, can also be detected genetically (i.e. loci associated with spontaneous expression or induction of a complete virus). Moreover, only a fraction of the retrovirus-related DNA in normal cells is engaged in the expression of identifiable retrovirus-related RNA and proteins. Consequently, the best parameter employed, in order to define the overall amount of genetic information present in the normal cell that is retrovirus-related, is the nucleic acid sequence homology with that particular virus.

The existence of an infectious counterpart for endogenous retroviruses formed the basis for the hypothesis that endogenous retroviruses were established in the genome through infections of germinal tissues by infection-competent retroviruses. Alternatively, it has been hypothesized that endogenous proviruses have evolved from normal cellular constituents, presumably transposable genetic elements (2).

Other types of genetic elements that naturally reside in normal cells are 'retrovirus-like' elements. These are genetic units that do not share nucleotide sequences with known retroviruses (exogenous or endogenous), but, nevertheless, possess 'retrovirus-like' properties. A major problem arises in attempting to define and expose the cellular repertoire of gene families that resemble proviruses but, yet, do not cross-hybridize with proviruses . Provided that 'retrovirus-like'

elements are incapable of independent cell-to-cell transmission,the first question that arises is: which retroviral features should be considered both necessary and sufficient to define a genetic unit as 'retrovirus-like'. Many retroviral features are also shared by other cellular elements, most notably, the long terminally repeated sequences (LTR) possessed by retroviruses and eukaryotic transposable elements, which have several structural and functional features in common. These include short, terminal, inverted repeats, the loss upon integration of the distal nucleotides of the LTR ("integration from within") and duplication of a short cellular sequence at the site of union with both proviruses and transposable elements,(for review, see ref. 3 and 4) . Moreover, features that until recently were considered unique to retroviruses were also found in bona fide transposable elements. These include the existence of a sequence complementary to tRNA adjacent to the LTR (that presumably serves as a primer binding site for reverse transcriptase) (5), the assembly of the LTR from sequences present at both ends of the terminally redundant RNA (6) (i.e. a general structure analogous to the retroviral U3-R-U5 arrangement of the LTR), and the encapsidation of a transposon RNA in a 'virus-like' capsid associated with a reverse transcriptase-like activity (7). The issue of whether these similarities reflect a diverging rather than converging evolutionary relationship (8) is still controversial, although the recent discovery of a significant nucleic acid homology (60%) between the LTR's of a copia-like transposable element and avian leukosis virus strongly suggests a divergent evolution (9).

In principle, it is conceivable that the cell contains multiple gene families displaying a continuum of 'retrovirus-like' characteristics. This may reflect either a progressive degenerative pathway of once-competent retroviruses that have been inserted into the germ cells, or, alternatively, stages in the sequential acquisition by a cellular transposable element of retrovirus-like properties through complex recombinations (8). It is thus not entirely clear how to define the members of a 'retrovirus-like' group.

A second problem in unveiling endogenous 'retrovirus-like' gene families stems from the lack of cross reactivity with known retroviruses, i.e., the lack of molecular probe to aid the screening of genomic libraries. It is possible therefore, that the repertoire of endogenous

'retrovirus-like' families has not yet been fully appreciated.

Indeed, the fashion through which the two known murine 'retrovirus-like' families were initially found attests to this possibility. The intracisternal A-particle (IAP) family was identified through the formation of intracellular particles resembling retroviral nucleocapsids. The amounts of these particles in certain mouse tumor cell lines allowed their purification for the initial biochemical characterization (10) and subsequent cloning (11, 12). The so-called VL30 gene family was identified through the capacity of the 30S RNA that is constitutively transcribed from these genes to be efficiently packaged in C-type virions and subsequently released from the cells (13-15). It seems that VL30 RNA possesses an efficient cis-acting packaging signal recognized by C-type virion components. This particular sequence is apparently responsible for the 'contamination' of mouse-grown retroviral virions, by large amounts (over 50% in certain cells) of VL30 RNA. It is likely that additional 'retrovirus-like' families that do not possess these readily identifiable phenotypes of IAP and VL30 have as yet escaped detection.

It has been shown that exogenously introduced retroviral LTRs may affect the expression of cellular genes at or in the vicinity of their integration site. This may occur by either direct insertion inactivation (16) or by positively regulating adjacent genes. The latter can be achieved by both promoter-insertion mechanisms (17) or by enhancing the expression of a nearby gene that is driven by its own promoter (18). Endogenous proviruses are apparently capable of similar effects. For example, the dilute coat color mutation in DBA(2) mice cosegregates with a single endogenous murine leukemia virus-related provirus, and reversion of the mutation is accompanied by loss by excision of the provirus (19). Very recent evidence indicates that 'retrovirus-like' IAP elements can exert transposable element-like effects. Insertion of IAP DNA into mouse immunoglobulin light chain DNA rendered the gene functionally defective (20), whereas integration of an IAP DNA within the c-mos gene led to increased expression of the gene to levels sufficient to induce cellular transformation (21).

Two major factors have to be taken into account when considering the significance of endogenous 'retrovirus-like' elements as potential insertion mutagens. Firstly, what are the nature and frequencies of the ele-

ment-mediated DNA rearrangements? Rearrangements may include both transpositions and recombination with other cellular sequences leading to creation of novel genomic linkages, and specifically, to juxtapositioning of the LTR of the element adjacent to new cellular genes. Secondly, what is the capacity of the LTR of VL30 elements to promote expression of downstream genes or to affect expression of neighboring genes by other mechanisms. It should be emphasized that such a putative role for 'virus-like' genetic units is a nonprogrammed one and is unrelated to the inherited (if any) role of 'virus-like' genes.

In this article we describe recent work from our laboratory with the murine VL30 'retrovirus-like' gene family. The main issues addressed are: the structural organization of the element respective to the genomes of proviruses and transposable elements, and, in particular, the sequence arrangement and sequence diversity of the VL30-associated LTRs, and in addition, the participation of VL30 elements in DNA rearrangements, particularly recombinations.

THE 'RETROVIRUS-LIKE' VL30 FAMILY

Experimental approach

30S RNA was obtained from purified virions released by Balb/c mouse cells chronically infected with MuLV. Virion preparations were composed of roughly equimolar amounts of MuLV RNA and VL30 RNA encapsidated in MuLV virions. Complementary DNA synthesized from this RNA was used as the hybridization probe for screening mouse genomic library. The mouse library was constructed from Balb/c embryonic DNA, partially digested with EcoRI and ligated with Charon 4A lambda DNA vector. Colonies that gave positive signals in an *in situ* plaque hybridization assay were selected and the VL30 DNA clones were unambiguously identified as those recombinant phages detecting 30S RNA in northern blots prepared from polyadenylated RNA derived from uninfected Balb/c cells (22).

Clones containing complete VL30 DNA units were identified through heteroduplexing independently cloned VL30 DNAs with each other (22). Only clones containing the complete 5.2-5.4 kb long element and mouse flanking sequences at both sides were selected for further studies.

Physical maps of restriction enzyme cleavage sites were obtained for some representative VL30 clones (23). The approximate locations of the LTRs were initially determined by heteroduplexing (24) followed by fine

physical mapping with restriction endonucleases and detection of symmetrical cuts, and finally by sequencing.

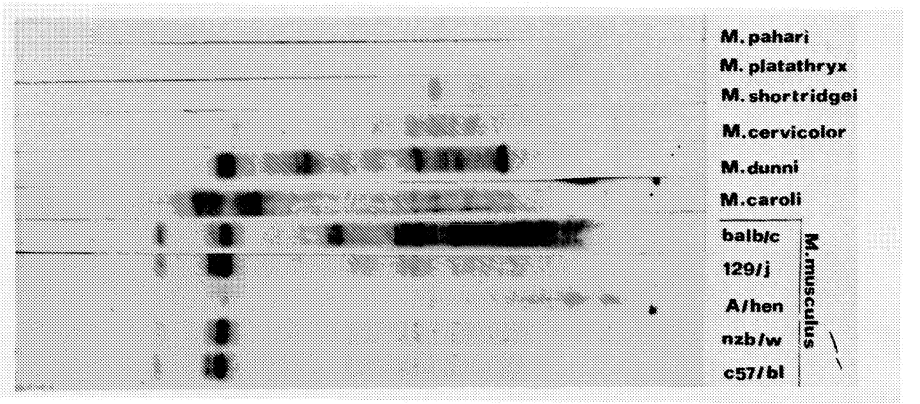
To aid analysis of VL30 sequences that are present in molecular contexts other than 'standard' VL30 units, a battery of subgenomic VL30 probes was obtained by subcloning relatively short VL30 segments in either plasmid or M13 phage vectors. Colonies with incomplete representation of VL30 sequences were identified by screening the genomic library with multiple subgenomic VL30 probes. The VL30 LTR unit was also subcloned in the πvx miniplasmid (25) and homologous recombination was used to facilitate isolation of VL30 LTR-containing genomic DNA fragments. This system was exploited in order to identify genomic linkages with defined non-VL30 sequences for which specific hybridization probes were available.

Occurrence reiteration and polymorphic distribution of VL30 elements

VL30 units are ubiquitous in all murine cells. Fig. 1 shows a pattern of VL30-containing restriction fragments in DNAs derived from different species of the genus Mus and from different strains of M. musculus.

Despite the diverse geographic origins of the species examined, they all contained multiple copies of VL30 DNA. Thus, VL30 genetic information must have been established in the genus Mus prior to speciation. Similar results were also obtained by Courtney et al. (26). In this respect, VL30 genetic information differs from known families of murine endogenous proviruses shown to be relatively recent acquisitions by the germ cells. There is, however, a great variation in the number of VL30 copies in different species, ranging from only a few VL30 units in certain Asian species (e.g. M. phari) and up to over one hundred copies in other species (e.g. all strains of M. musculus) (Fig. 1). Direct enumeration by hybridization kinetic analysis has confirmed that M. musculus contains over one hundred VL30 (23). Thus, VL30 DNA is about one order of magnitude more abundant than MuLV-related or mouse mammary tumor virus-related endogenous proviruses. Interestingly, the reiteration frequency of IAP, the other known murine 'retrovirus-like' family, also varies greatly among different species (e.g. only 25 copies per haploid genome of M. cervicolor but about 1000 copies in M. musculus)(27). Since an infectious counterpart of VL30 has not been encountered and considering its reiteration frequency, it is likely that the variation in VL30 copy

Fig. 1. Detection of VL30 sequences in DNAs of different species of the genus Mus. 15µg DNA extracted from the liver of each of the indicated Mus species or from the indicated strain of M. Musculus was digested with EcoRI, electrophoresed through 0.7% agarose gel, blotted, hybridized with a VL30-specific probe and autoradiographed. The hybridization probe was a subclone of VL30 DNA that is devoid of DNA sequences flanking the cloned VL30 unit, labeled by in vitro nick translation.



number reflects intracellular amplifications rather than recurrent infections of germinal tissues.

VL30 DNA is grossly species specific. Under relaxed hybridization conditions, however, VL30 DNA hybridized with other vertebrate DNA, including human DNA. The cross-hybridization between mouse VL30 and rat genomic DNA was further analyzed. All cross-reactivity could be accounted for by the cross-reactivity between the respective VL30 elements (28). Rat VL30 constitutes a 'retrovirus-like' gene family with properties similar to murine VL30. Noteworthy is the capacity of the 30S transcripts of these 'genomes' to be rescued by viral infection and their subsequent transmissibility to other cells (29). The cross-hybridization between mouse and rat VL30 units also suggests a genetic relatedness between the two. It is likely that similar genetic elements are present in the genomes of

other vertebrate species.

VL30 units are dispersed throughout the mouse genome. Our initial heteroduplexing data have already suggested that VL30 units are nontandemly arranged and are flanked by unrelated DNA sequences. Using blot hybridization analysis with DNAs derived from a panel of mouse-hamster somatic hybrid cell lines, we have found that VL30 units are distributed among different mouse chromosomes (23).

Blot hybridization of DNAs derived from different inbred strains of M. musculus as well as from feral mice has shown polymorphism in restriction enzyme digestion patterns. Since polymorphic fragments of a size larger than the VL30 unit size were also detected, these data were interpreted as reflecting polymorphic size distribution of VL30 among different strains of M. musculus (in addition to reflecting microheterogeneity within VL30 units). Courtney et al. (26), using a probe derived from sequences flanking a particular VL30 unit, have recently shown that a cellular locus that is occupied by VL30 in one strain is not occupied by VL30 DNA in another strain.

Both the apparent amplifications of VL30 DNA and its polymorphis distribution suggest that new cellular loci can be potentially occupied by VL30 units and that VL30 can be potentially juxtaposed next to different cellular genes.

Structural organization of VL30 genomes

Members of the VL30 family are heterogeneous in respect to both size and sequence arrangement (see following section). The size of a standard VL30 DNA unit was estimated as 5.2-5.4 kb on the basis of heteroduplexing (22) and blot hybridization experiments (23). This estimate roughly corresponds to the amount of genetic information that is required to encode 30S RNA. It is likely, therefore, that like full-length RNA of retroviruses, 30S RNA is transcribed from a continuous stretch of DNA. Subgenomic VL30 RNAs could not be detected (14). Analysis of the methylation patterns of VL30 DNA revealed that while some VL30 DNA units are strongly methylated, other copies of VL30 DNA are not (23). This observation supports the notion that only a fraction of VL30 elements are transcriptionally active. Proteins encoded by VL30 genes have not yet been detected although VL30 RNA was found to be associated with polyribosomes (30).

We have sequenced large segments of randomly selected VL30 clones. Some noteworthy features are schematically shown in Fig. 2, and summarized as follows: The VL30 LTRs are bounded by inverted repeats ending in (5')TG...CA(3'). Computer-aided comparison revealed no significant sequence homology between the LTRs of VL30 and MuLV proviruses. At the site of union with flanking DNA there is a duplication of a tetranucleotide sequence of mouse DNA. At the inner junction of the 5' and 3' LTRs are the complementary dinucleotide sequences TT and AA, respectively . Thus, the structure shown in Fig. 2 can be interpreted as the result of an 'integration from within'. That is, the loss upon integration of the dinucleotides AA and TT from the distal 5' and 3' ends, respectively, of the LTR.

Immediately adjacent to the 5' LTR is an 18 bp sequence that is complementary to the 3' end of $tRNA^{pro}$. This finding suggests that , as with mammalian retroviruses, $tRNA^{pro}$ serves as a primer for synthesis of VL30 (-) strand DNA and that synthesis of VL30 DNA is initiated close to the 5' end of the RNA molecule. Adjacent to the 3' LTR is a purine-rich sequence (17 out of 21 nucleotides are A or G) ending in AATG. A purine-rich tract ending in AATG resides in the same location in retrovirus proviruses and is believed to span the site of binding of the primer for (+) strand DNA synthesis (4). It is likely therefore that this site is also utilized as (+) strand primer binding site during reverse transcription of 30S RNA.

It is not known whether VL30 LTRs share the general structure U3-R-U5 with proviruses. Moreover, it has not been reported that 30S RNA is terminally redundant. We have previously shown, however, that short polyA-containing fragments of VL30 RNA do hybridize with VL30 DNA fragments residing at both termini of the VL30 unit (24). When viewed in conjunction with the observation that VL30 DNA synthesis is initiated no further than several hundred base pairs away from the 5' end of the RNA, these data suggest that indeed the VL30 LTR too is assembled from nucleotide sequences residing at both 3' and 5' ends of the RNA.

It thus appears that the structure of the LTR, the nature of the junctions with mouse-flanking sequences, and the existence of putative primer binding sites for the synthesis of both (-) and (+) strands of DNA, disclose a strong resemblence to retroviruses in both the integration mechanism and the mode of reverse transcription. It should be pointed out,

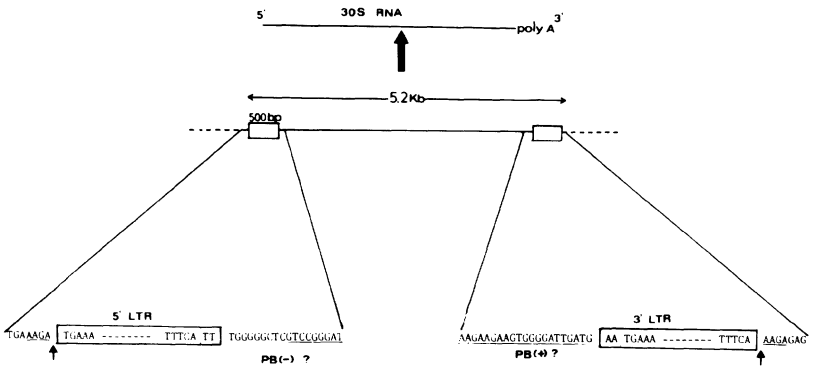


Fig. 2. Genome organization of VL30 DNA. The organization of a standard VL30 DNA unit is schematically shown. The junctions of the LTRs with both internal VL30 sequences (solid line) and mouse flanking sequences (broken line) were determined as described in Experimental Approach. The transcriptional orientation of VL30 DNA in respect to the 30S RNA was determined by the use of short radiolabeled, polyadenylated subsets of 30S RNA as hybridization probes (24) and challenge of blots of defined subgenomic VL30 DNA fragments (data not shown). The nucleotide sequences at both the inner and outer LTR junctions are indicated and some interesting features of these sequences are specified in the text.

however, that all these features were also recently found to be shared by eukaryotic transposable elements. Thus, the attempt to place VL30 on an evolutionary scale (that is, as evolutionary intermediates linking transposable elements and retroviruses or as degenerative descendants of proviruses) seems an almost impossible task.

An important 'retrovirus-like' feature is a packaging signal acting in cis for the encapsidation of the genomic RNA molecule. Experiments are underway to identify the location and nature of the putative packaging signal of VL30 in respect to those of retroviruses (31,32).

Heterogeneity and evidence for VL30-mediated recombinations

The multiple copies of VL30 units present in the mouse genome are not identical. Oligonucleotide fingerprinting analyses have shown that 30S RNA molecules yield non-stoichiometric amounts of oligonucleotides (14), suggesting heterogeneity of VL30 RNA. Several levels of heterogeneity of VL30 DNA units exist. We have analyzed heteroduplexes between independently cloned VL30 DNAs and used the heteroduplexing data as a major

criterion in assessing sequence relatedness among VL30 elements (Table I).

Some combinations of cloned VL30 units gave an uninterrupted heteroduplex that was roughly 5.2-5.4 kb long (class I in Table I), indicating a high degree of sequence homology. Yet, when the physical maps of class I VL30 units were compared, many differences were detected (23). This sequence divergence presumably reflects scattered base changes throughout the VL30 units.

A second level of heterogeneity disclosed by heteroduplexing was the presence of nonhomologous segments, up to several hundred bp long that interrupt the shared VL30 sequences. These nonhomologous regions were visible as single-stranded 'bubbles' interrupting duplexed regions (class II in Table I). The nature of these presumptive non-VL30 substitutions is not known. In other VL30 units, even larger substitutions with non-VL30 genetic information were observed. In certain clones the majority of VL30 sequences were replaced by foreign DNA (class III in Table I). These structures presumably reflect VL30-mediated recombinations leading to capture of other cellular DNAs by VL30 elements.

A well-known example of VL30-mediated recombinations are the genomes of the sarcomagenic retroviruses - the Harvey (Ha) and Kirsten (Ki) strains of mouse sarcoma virus (MSV). These viruses were isolated through passage of Moloney-MuLV in rats and selection for newly transforming viruses. Subsequent analysis showed that the oncogenic potential was acquired concomitantly with the acquisition of two unrelated sets of rat sequences (33, 34). It is intriguing that in the two independent isolates of rat-derived sarcoma viruses, VL30 information was transduced onto the viral genome although the selection applied was exclusively for capture of the oncogenes.

Since the generation of the apparent tripartite Ha-MSV and Ki-MSV genomes involved recombination between MuLV and VL30, we screened a mouse embryonic gene library for 'footprints' of recombination between VL30 and MuLV-related genetic information. Several clones that are consistent with this structure were indeed isolated (including the class III clone whose heteroduplex is shown in Table I). Blot hybridization analysis with a panel of defined subgenomic VL30 probes indicated that this clone (designated VM) contained VL30-related LTRs and additional limited subsets of VL30 information adjacent to both LTRs. The centrally located VL30 sequences, however, were replaced by non-VL30 sequences.The non-VL30

Table I. Heterogeneity of VL30 DNA.

CLASS	HETERODUPLEX WITH A STANDARD VL30 UNIT	DESCRIPTION	MuLV-RELATED DNA
I		AN UNINTERRUPTED 5.2 KB LONG DUPLEX	NO
II		DUPLEX IS INTERRUPTED BY SHORT SEGMENTS OF DNA NOT SHARED WITH STANDARD VL30	NO
III		CLONE CONTAINS VL30 LTRS CENTRALLY LOCATED VL30 SEQUENCES WERE REPLACED BY MuLV RELATED DNA	gag AND pol determinants. No MuLV-related LTRs

A standard VL30 DNA unit cloned in a charon 4A lambda vector was heteroduplexed with the indicated VL30 DNAs cloned in the same vector, and in the same orientation. Tracings of typical heteroduplexes are shown. J_V and J_f denote the vector-insert junctions and the flanking DNA-VL30 DNA junctions, respectively. Hybridization with MuLV was examined by the use of a cloned, circularily permuted MuLV provirus that contains all viral sequences.

sequences were identified as MuLV-related gag and pol determinants. Blot hybridization with specific subgenomic MuLV probes indicated that the MuLV genome is only partially represented in VM DNA (e.g. there are no MuLV-related LTRs) and that the VL30 and MuLV-related sequences are arranged in a parallel 5'- 3' transcriptional orientation (35). It is likely that this structure arose as a result of recombination between VL30 and MuLV-related endogenous proviruses, possibly through formation of heterodimeric RNA, and subsequent stable deposition of the recombinant DNA in the mouse genome. Interestingly, this structure is reciprocal in nature to the recombination that created Ha-MSV and Ki-MSV where VL30 information was captured between MuLV LTRs.

It might be speculated that the specific recombination described above reflects a more general tendency of VL30 to recombine with cellular genes and that due to an even more frequent recombination with retroviruses,

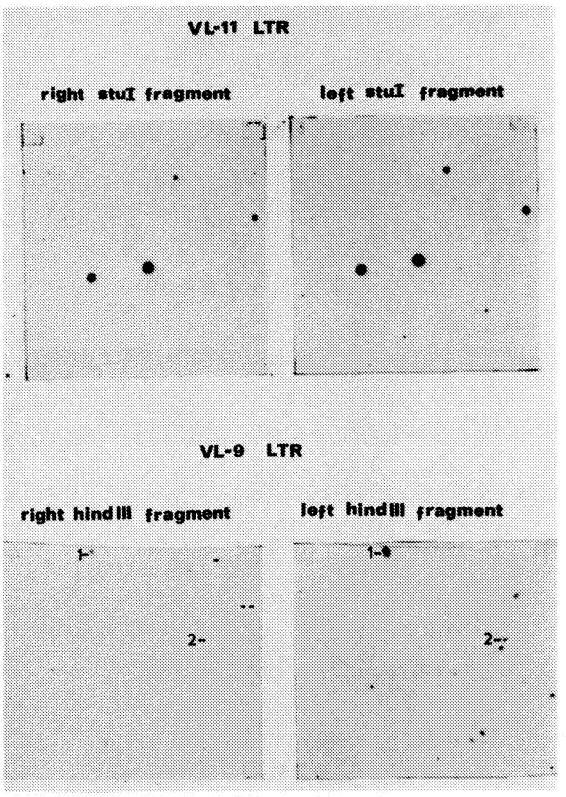


Fig. 3. Hybridization of mouse library phages with probes corresponding to different parts of VL30 LTR. Phages from the genomic mouse library were plated at a density of 5000 pfu/plate and phage imprints were obtained from each plate on two duplicate nitrocellulose filters. A restriction endonuclease that cleaves once within the LTR (HindIII in the case of LTR11 and StuI in the case of LTR9) was used to generate LTR probes that correspond to the two complementary parts of the same LTR. These LTR halves were hybridized with nitrocellulose filters obtained from the same plate. (In each case, probes were devoid of VL30 internal sequences). Autoradiograms were superimposed in order to identify clones that are reactive with only part of the LTR.

VL30 elements may act as mediators in facilitating the transduction of cellular genes into the genomes of retroviruses (34).

Differences in restriction enzyme recognition sequences among VL30 units also spanned the LTR regions, suggesting heterogeneity of VL30-associated LTRs. The nature of LTR heterogeneity was further studied by

direct DNA sequencing of cloned LTRs. Computer-aided comparison of the nucleotide sequences of two randomly selected LTRs showed an almost identical DNA sequence followed by DNA sequences that are completely unrelated to each other (A. Itin and E. Keshet, to be published). This result suggested to us that the point of sequence divergence within the LTR is the site of recombination with nucleotide sequences that are not part of a 'standard' LTR.

In order to examine this possibility, specific probes were obtained that correspond to the different domains of the putative recombinant LTR. These probes were used to challenge multiple nitrocellulose imprints obtained from the same plating of the genomic library phages. Most of the clones detected by one probe were indeed not detected by the other probe (Fig. 3). This result indicated that the two sets of sequences which are physically linked in the particular LTR unit studied are unlinked in other loci of the mouse genome and suggested that the LTR was assembled from two remote genomic subsets, presumably through recombination.

In principle, it is possible that recombination between LTR components might lead to the generation of LTR units with altered regulatory capacities. This possibility is currently being examined by ligating different VL30 LTR components next to an assayable gene in the appropriate shuttle vector, introducing the recombinant plasmids into mouse cells by transfection and assaying for promoter and enhancer activities.

REFERENCES

1. Coffin, J.M. in: The molecular biology of tumor viruses, part III, RNA tumor viruses (Eds. R.A. Weiss, N. Teich, H.E. Varmus and J.M. Coffin). Cold Spring Harbor Laboratory, 1982, pp. 1109-1203.
2. Temin, H.M. Persp. Biol. Med. 14: 11-26, 1970.
3. Temin, H.M. Cell 21: 599-600, 1980.
4. Varmus, H.E. Science 216: 812-820, 1982.
5. Eibel, H., J. Gafner, A. Stotz and P. Philippsen. Cold Spring Harbor Symp. Quant. Biol. 45: 609-617, 1980.
6. Elder, R.T., E.Y. Loh and R.W. Davis. Proc. Natl. Acad. Sci. USA 80: 2432-2436, 1983.
7. Shiba, T. and K. Saigo. Nature (London) 302:119-124, 1983.
8. Shimotohno, K., S. Mizutani and H.M. Temin. Nature (London) 285: 550-554, 1980.

9. Kugimiya, W., H. Ikenga and K. Saigo. Proc. Natl. Acad. Sci. USA 80: 3193-3197, 1983.

10. Lueders, K.K. and E.L. Kuff. Cell 12: 963-972, 1977.

11. Lueders, K.K. and E.L. Kuff. Proc. Natl. Acad. Sci. USA 77: 3571-3575, 1980.

12. Ono, M., M.D. Cole, A.T. White and R.C.C. Huang. Cell 21: 465-473, 1980.

13. Hawk, R.S., D.H. Troxler, D. Lowy, P.H. Duesberg and E.M. Scolnick. J. Virol. 25: 115-123, 1978.

14. Besmer, P., U. Olshevsky, D. Baltimore and H. Fan. J. Virol. 29: 1168-1176, 1979.

15. Sherwin, S.A., U.R. Rapp, R.E. Benveniste, A. Sen and G.I. Todaro. J. Virol. 26: 257-264, 1978.

16. Varmus, H.E., N. Quintrell and S. Ortiz. Cell 25: 23-36, 1981.

17. Hayward, W.S., B.E. Neel and S.M. Astrin. Nature (London) 290: 475-480, 1980.

18. Payne, G.S., S.A. Courtneidge, L.B. Crittenden, A.M. Fadly, J.M. Bishop and H.E. Varmus . Cell 23: 311-322, 1981.

19. Jenkins, N.A., N.G. Copeland, B.A. Taylor and B.K. Lee. Nature (London) 293: 370-374, 1981.

20. Kuff, E.L., A. Feenstra, K.K. Lueders, L. Smith, R. Hawley, N. Hozumi and M. Shulman. Proc. Natl. Acad. Sci. USA 80: 1992-1996, 1983.

21. Rehavi, G., D. Givol and E. Canaani. Nature (London) 300: 607-610, 1982.

22. Keshet, E., Y. Shaul, J. Kaminchik and H. Aviv. Cell 20: 431-439, 1980.

23. Keshet, E. and A. Itin. J. Virol. 43: 50-58, 1982.

24. Keshet, E. and Y. Shaul. Nature (London) 289: 83-85, 1981.

25. Maniatis, T., E.F. Fritsch and J. Sambrook. In: Molecular cloning (a laboratory manual). Cold Spring Harbor Laboratory, pp. 353-362, 1982.

26. Courtney, M.G., P.K. Elder, D. Steffen and M.J. Getz. J. Virol. 43: 511-518, 1982.

27. Kuff, G.L., K.K. Lueders and E.M. Scolnick. J. Virol. 28: 66-74, 1978.

28. Itin, A., G. Rotman and E. Keshet. Virology, (in press).

29. Scolnick, E.M., W.C. Vass, R.S. Hawk and P.H. Duesberg. J. Virol. 29: 964-972, 1979.

30. Scolnick, E.M., Williams, D., Maryak, J., Vass, W., Goldberg, R.J. and W.P. Parks. J. Virol. 20: 570-582, 1976.

31. Watanabe, S. and H.M. Temin. Proc. Natl. Acad. Sci. USA 79: 5986-5990, 1982.

32. Mann, R., R.C. Mulligan and D. Baltimore. Cell 33: 153-159, 1983.

33. Chien, Y-H., M. Lai, T.Y. Shih, I.M. Verma, E.M. Scolnick, R. Roy-Burman and N. Davidson. J. Virol. 31: 752-760, 1979.

34. Ellis, R.W., D. Defeo, J.M. Maryak, H.A. Young, T.Y. Shih, E.H. Chang, D.R. Lowy and E.M. Scolnick. J. Virol. 36: 408-420, 1980.

35. A. Itin and E. Keshet. J. Virol., (in press).

CLONING OF THE DNA FORM OF AN RNA VIRUS GENOME

12

POLIOVIRUS cDNA CLONED IN BACTERIAL PLASMIDS

MARC GIRARD, MICHEL DREANO, HELENA KOPECKA, DANIELE BENICHOU, and SYLVIE VAN DER WERF
Unit of Molecular Virology, Department of Virology, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15 France

1. INTRODUCTION

The genome of poliovirus is a 7440 nucleotide long, infectious, single-stranded RNA molecule of "plus" strand polarity (1). Its 5' end is covalently attached to a small MW protein termed VPg (for a review see reference 2), whereas its 3' end is polyadenylated (Fig.1). Most of the progress in our understanding of the organization of the viral genome comes from the recent determination of the complete nucleotide sequence of the RNAs of both the Mahoney strain (wild type) and the attenuated Sabin strain (Lsc2ab strain) of type 1 virus (3-5).

Analysis of the RNA nucleotide sequence has revealed the existence of a long open reading frame which extends from an AUG at nucleotide 743 to a double stop codon (UAGUAA) at nucleotide 7370 (Fig.1). This open reading frame can code for a polyprotein of ca. 247,000 daltons, NCVP00, which is cleaved *in statu nascendi*, most probably by host cell proteinases (6-8), to generate 3 primary precursor proteins which map on the genome in the order: NCVP1a, NCVP3b, and NCVP1b (Fig.1). Each of these proteins is in turn processed by a series of proteolytic cleavages to generate all the known poliovirus proteins.

The exact location on the RNA of the coding sequence for each of the proteins, which was approximately known from pactamycin mapping experiments (8,9), was definitely established by comparing the experimentally determined amino acid sequence of the proteins with that deduced from the nucleotide sequence in the open reading frame of the genome (3,11,12). Thus, the four poliovirus capsid polypeptides VP4, VP2, VP3 and VP1 were shown to map in that order on the genome and to result from

cleavage of the structural protein NCVP1a (10,13). During the initial stages of virion morphogenesis, VP4 and VP2 remain covalently linked together under the form of their precursor polypeptide, VP0. The procapsids, which are the proteinic precursor of the virions, are made of 60 copies each of VP0, VP1 and VP3. Cleavage of VP0 into VP2 and VP4 takes place at the last step in morphogenesis, after a progeny viral RNA molecule has entered the preformed procapsid (14-16).

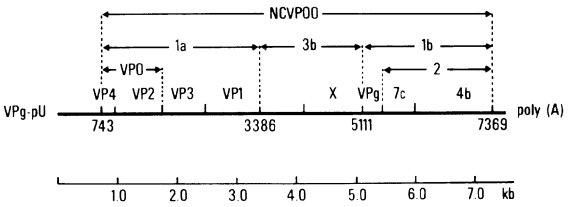


FIGURE 1. Organization of the poliovirus genome and protein processing pathway. The viral RNA is figured by a heavy line. Its primary translation product, NCVP00, is cleaved as indicated into three main precursor proteins which are themselves cleaved further to generate the different viral polypeptides (see text). The numbers on the RNA genome represent the nucleotide number at the exact sites of initiation and termination of translation, and at those at which the initial proteolytic cleavage of NCVP00 occurs. Nucleotide numbering is after Dorner et al. (12).

The other two primary cleavage products of NCVP00, NCVP3b and NCVP1b, are similarly processed by endopeptidic cleavages. NCVP3b is the precursor to non-structural viral proteins of unknown functions, such as NCVPX (17), whereas cleavage of NCVP1b ultimately generates VPg, the 5' terminal protein, NCVP7c, a specific viral proteinase (18), and NCVP4b, the RNA

replicase (19-23) which can readily be detected by the poly(A)-dependent poly(U)-polymerase activity it exhibits *in vitro* in the presence of oligo(U).

The cleavage of viral proteins occurs at specific sites characterized by the sequence Gln-Gly (3,24),due to the action of NCVP7c (18).The rule that viral proteins begin by Gly and their amino terminus and terminate by Gln at their carboxy end suffers however a few exceptions: for example,both the cleavage between VP1 and NCVP3b and that between NCVP6a and NCVP6b occur at a Tyr-Gln site.The former is probably mediated by a cellular proteinase, but it is not known whether this is also the case for the latter. Another known exception is the cleavage between VP4 and VP2,which occurs at an Asn-Ser site.This cleavage is however linked to virion maturation, and it is tempting to think that it could therefore be mediated by yet another unidentified proteinase.

In spite of this overall apparent simplicity,many processes which control the expression and/or replication of the viral genome are still a matter of speculation.For example it is still unclear whether precursor polypeptides can function *in vivo* before they are cleaved,perhaps fulfilling other functions than those of their cleavage products.There is actually evidence that this might be the case.Baron and Baltimore (25) have shown that anti-VPg antibodies immunoprecipitate pre-VPg molecules that are bound to nascent RNA molecules:Pre-VPg, but not VPg,could therefore be involved in the initiation of RNA synthesis.VPg,on the other hand,is covalently bound to the 5' end of all viral "+" and "-" RNA molecules,where it could play a different albeit still unknown function.

Also,the function of many of the viral polypeptides remains to be elucidated.For example,nothing is known of the role of NCVPX,which is present in abundance in cytoplasmic extracts from infected cells.Conversely,we totally ignore the nature of the proteins responsible for the inhibition of the host cell RNA and protein syntheses and for the pathogenicity of the virus

Recombinant DNA technology permits the synthesis and

cloning in Escherichia coli of the cDNA copy of a viral RNA genome.Were the expression of a cloned poliovirus cDNA copy to occur with efficiency after transfection of permissive cells, the application of site-directed mutagenesis to the cloned cDNA could offer an unique opportunity for the functional and genetic analysis of the viral genome.Expression of the cloned cDNA sequences in E.coli or in animal cells could also open the way to the development of new anti-poliovirus vaccines composed of isolated capsid polypeptides or of selected parts of the polypeptides,which would offer the advantage of being completely devoid of pathogenicity.For these reasons,we decided to undertake the molecular cloning of the genome of poliovirus type 1.Part of these results have been published (26). A similar work has been reported by Racaniello and Baltimore (4,27). We have now also expressed in E.coli the cDNA sequence coding for the capsid polypeptide VP1 (28,29).

2. MOLECULAR CLONING OF THE GENOME OF POLIOVIRUS TYPE 1

The Mahoney strain of poliovirus type 1 was grown in HeLa cells in suspension cultures and viral RNA was extracted as described (30).The strategy we followed for the cloning of the cDNA is summarized in Fig.2. In brief,AMV* reverse transcriptase was used in the presence of oligo(dT) as a primer to synthesize a cDNA copy of the viral RNA starting from the poly(A) tail of the genome.The resulting RNA-cDNA hybrid molecules were trimmed of excedentary RNA sequences by treatment with RNases A and T1, tailed with poly(dC),and directly inserted into the poly(dG) tailed PstI site of pBR322.Recombinant plasmids were used to transform E.coli bacteria and transformants were selected on the basis of tetracycline resistance and ampicillin sensitivity.

The use of RNA-DNA hybrid molecules for the transformation of bacteria has been reported previously (34,35).We chose to use this procedure because of its ease, and in order to favor the recovery of plasmids carrying the 5' end of the viral genome.

A total of about 800 transformed colonies was obtained starting from approximately 300 ng of poliovirus RNA-cDNA

*avian myeloblastis virus

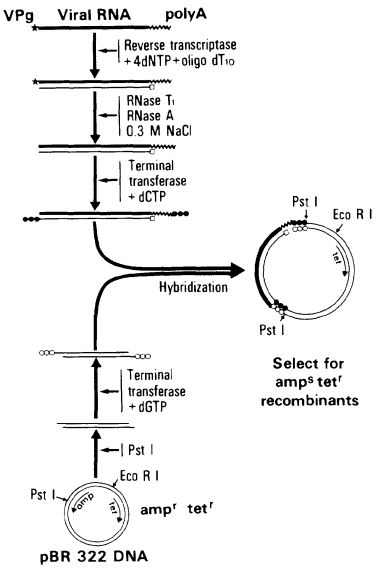


FIGURE 2. Strategy followed for the molecular cloning of the poliovirus genome. Viral RNA (100 µg/ml) was incubated with 200 units of AMV reverse transcriptase (a generous gift of J.W.Beard) for 2 hrs at 37°C in 50 mM Tris-HCl pH 8.3, 8 mM $MgCl_2$, 0.4 mM DTT, 80 mMKCl, 4 mM Na pyrophosphate, 0.4 mM each of dATP, dGTP, dCTP and dTTP and 10 µg/ml of oligo(dT)10. The reaction was stopped with 10 mM EDTA. Resulting RNA-cDNA hybrid molecules were trimmed for 30 min at 37°C with 25 µg/ml RNase A and 0.05 u/ml RNase T1 in the presence of 0.3 M NaCl. The reaction was stopped with 0.2% SDS and the material was extracted with phenol and with chloroform:isoamylalcohol (24:1) and precipitated with ethanol. The trimmed RNA-cDNA hybrid molecules were purified by centrifugation through a 13-50% sucrose gradient in 10 mM Tris-HCl pH 7.5, 0.1 M NaCl, 1 mM EDTA, 0.5% SDS. Fractions containing full length molecules were pooled and precipitated with 2 vol ethanol. The hybrids were then dC-tailed by incubating for 1 hr at 30°C with 10 units of terminal nucleotidyl transferase in a 20 µl reaction containing 100 mM K-cacodylate pH 7.0, 2 mM MnCl2, 0.1 mM DTT and 20 mM (3H)-labeled dCTP. The reaction was stopped with 10 mM EDTA and 0.2% SDS and chilled on ice. The SDS precipitate was discarded by centrifugation at 10,000 rpm for 10 min. Ten µg of PstI-digested pBR322 was similarly dG-tailed for 45 min at 30°C using 50 units of terminal transferase and 2.5 mM dGTP. The dG-tailed vector was treated as above and 450 ng were annealed with 300 ng of the dC-tailed poliovirus cDNA-RNA hybrid molecules in 50 µl of 20 mM Tris-HCl, pH 7.4, 0.3 M NaCl, 1 mM EDTA. Hybridization was for 10 min at 65°C, followed by 1 hr at 60°C, 1 hr at 50°C and progressive cooling to room temperature overnight. Competent E.coli 1106 (803 $r_k^- m_k^-$) (32) was transformed and tetracycline-resistant colonies were selected (33). An initial screen was performed for sensitivity to ampicillin, after which recombinant plasmids were selected by colony filter hybridization. Symbols: ∿∿: poly(A); ✱ VPg; □ oligo(dT); ●●● : poly(dC); ○○○ : poly(dG); thin lines: DNA strands; heavy line : viral RNA strand.

hybrid molecules.The reason for this low efficiency probably lies,at least in part, in the choice of the RNases used to treat the RNA-DNA hybrid molecules,as,in more recent experiments where the same procedure was followed except that the hybrids were trimmed with RNase T2 (BRL,20 units/ml),a frequency of 20-60 transformants per ng of vector was observed (unpublished results and G.Barry, personal communication). Transformation with recombinant plasmids containing DNA-RNA inserts has been reported to yield cDNA clones with extensive deletions (35).We have not however observed this type of deletion.

Recombinant plasmids were first screened by *in situ* hybridization (36,37) using as a probe ^{32}P-end-labeled,short,single-stranded cDNA molecules that were synthesized by limited reverse transcription of oligo(dT)-primed viral RNA.Such a probe should be specific for the 3' end of the genome.Only about 10% of the clones showed positive hybridization with the probe,suggesting that most of the clones did not contain the 3' end of the genome.As a confirmation,we were unable, using labeled poly(dA) as a hybribization probe,to detect recombinant plasmids carrying the 3' terminal poly(A) sequence. The sequence of the insert extending the closest to the 3'end of the genome stopped about 50-100 nucleotides from the poly(A)tract (41).These results were most surprising in view of the fact that synthesis of the cDNA had been initiated at the 3'-terminal poly(A) of the viral RNA using oligo(dT) as a primer.

We interpret the absence of the real 3' end of the genome in this first series of recombinant plasmids as a consequence of the fact that using RNases T1 and A,excedentary 3'poly(A) sequence had not been removed from the RNA-DNA hybrid molecules. Indeed,we have later obtained recombinant plasmids containing the 3' end and part of the poly(A) tract of the genome when the RNA-cDNA hybrid molecules were treated with RNase T2 prior to tailing and insertion into the vector (see plasmid H3 in Fig.3 as an example).

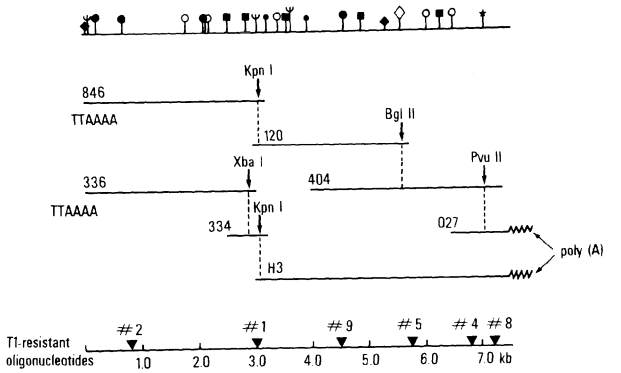


FIGURE 3. Positions of some of the cloned cDNA inserts relative to the map of the poliovirus genome and strategy followed for the construction of full-length cDNA inserts. The figure shows a limited restriction enzyme map of poliovirus as experimentally determined from the map of each of the cloned cDNA inserts (26,41) and as deduced from the nucleotide sequence of the genome (3,4). The inserts from plasmids 846,120,404 and 027 were used to construct a full-length cDNA insert by successive in vitro recombinations, as indicated. This generated plasmid pPV1-1022 (see text). The inserts from plasmids 336,334 and H3 were similarly used to construct a full-length cDNA insert as shown. This generated plasmid pPV1-1515.

The sequence of the cDNA insert from plasmids 846 and 336 begins by TTAAAA. The sequence of the cDNA insert in plasmids 027 and H3 ends by a poly(A) stretch which is 19 and 15 nucleotides long, respectively. The viral RNase T1-resistant oligonucleotides that were used as hybridization probes for the screening of the recombinant plasmids by colony filter hybridization (see text) are represented by the black triangles at the bottom of the figure. Symbols of the restriction enzymes cleavage sites: ◆ BglI; Ψ :KpnI; ● :BamH1; ○ :PstI; □ :XbaI; ● :BclI; ◇ :BglII; o :HindIII; * :PvuII.

As none of the plasmids contained the 3' end of the genome, we decided to clone a double-stranded cDNA copy of the RNA in order to obtain a cloned 3'end. AMV reverse transcriptase was used in the presence of oligo(dT) for the synthesis of the 1st cDNA strand, whereas E.coli DNA polymerase I (Klenow fragment)

was used for that of the 2nd strand. The cDNA was treated with nuclease S1,(dC)-tailed,and cloned into the (dG)-tailed PstI site of pBR322, after which recombinant plasmids were screened by colony filter hybridization using labeled poly(dA) as a probe. A series of plasmids containing the 3' end of the poliovirus genome and part of the poly(A) stretch was thus selected (see for example plasmid 027 in Fig.3).

To screen the other recombinant plasmids,we used poliovirus-specific RNase T1-resistant oligonucleotides as hybridization probes (Fig.4).Most of these oligonucleotides had been mapped approximately on the viral RNA at the time of this study (38-40).Oligonucleotides number 5,4 and 8 were used as probes specific for the 3' part of the genome,oligonucleotides number 1 and 9 for the middle of the genome, and oligonucleotide number 2 for the 5' part (Fig.3).Hybridization was carried out at 42°C in 2XSSC as described (26).This hybridization procedure worked well with all the RNase T1-resistant oligonucleotides tested,provided they were at least 22 nucleotides long. No attempt was made at determining the proper hybridization conditions for probes of smaller size.

None of the poliovirus recombinant plasmids studied hybridized to all the T1 oligonucleotides tested,but only to one, or to a few oligonucleotides located in close vicinity on the poliovirus RNA map (26 and Fig.4).This suggested that none of the cloned inserts represented the entire poliovirus sequence. This was confirmed by analyzing the size of the different plasmids.The longest of the poliovirus inserts that we ever obtained is about 4.4 kb long (clone H3,Fig.3).In the first cloning experiment (26),the longest insert obtained was 3.2 kb long (clone pPV1-846,Fig.3).It was therefore obvious that the cloned inserts corresponded to scattered fragments of the genome.

The reason why the cloned inserts corresponded to a scatter of overlapping fragments of the genome (Fig.3) and not to the totality of the genome in one piece,is not yet clear.A possible explanation is that single-strand nicks,which acted as sites for elongation by terminal deoxynucleotidyl transferase,were introduced into the RNA-DNA hybrid molecules during reverse

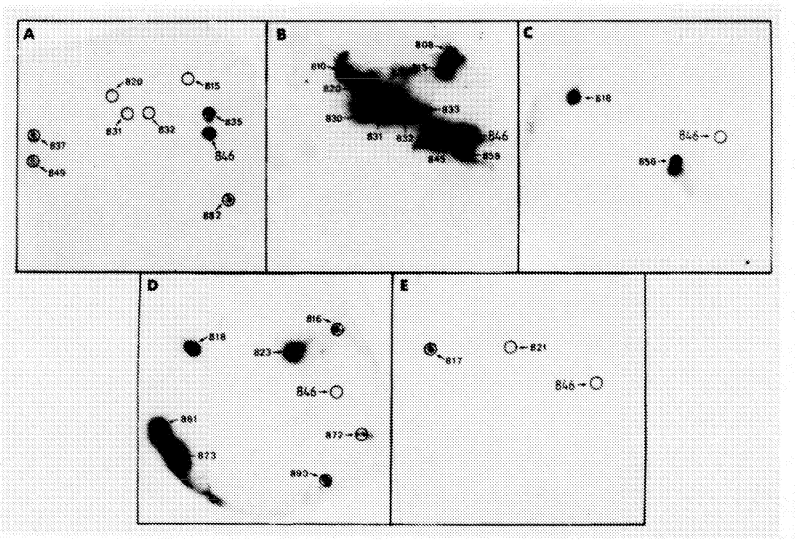


FIGURE 4. Colony filter hybridization of a series of bacterial clones containing poliovirus recombinant plasmids with RNase T1-resistant oligonucleotides. RNase T1-resistant oligonucleotides from poliovirus RNA were separated by 2D-electrophoresis, eluted, and separately 5' end labeled using 250 μCi (γ^{32}P)-ATP and 4 units polynucleotide kinase. The DNA from a series of bacterial colonies was transferred onto nitrocellulose filter replicas (36) and hybridized at 42°C with labeled T1 oligonucleotides number 1,2,4,8 and 9 (panels A-E respectively). The location of the oligonucleotides on the genome map is indicated in Fig.3. Hybridization was for 70 hr at 42°C, after which the filters were processed for autoradiography after extensive washing as described (26).

transcription of the RNA or during RNase treatment of the hybrid molecules, or at the time of tailing of the molecules. As a scatter of cloned inserts has also been observed when double-stranded poliovirus cDNA molecules were cloned (4) or when cloning the double-stranded cDNA molecules of foot-and-mouth disease virus (44,45, and G. Barry, personal communication), fragmentation of the cDNA at the time of tailing stands as the most probable explanation.

Restriction enzyme mapping of each plasmid (41) and analysis under the electron microscope of the heteroduplex molecules formed by annealing of the DNA of each two plasmids

together (42) were used to orient the cDNA inserts with respect to each other and to the poliovirus map (Fig.3).This analysis was rendered easier by the fact that several inserts were found to overlap.Nucleotide sequencing of the cDNA inserts (43) revealed that at least 2 of them, those of plasmids pPV1-846 and -336,began by the sequence TTAAAACA... (26) which is that of the real 5' end of the viral genome (3,4).Thus, the 5' end of the viral RNA had indeed been successfully cloned in one step by the procedure we had chosen.

From the results obtained using these different methods, it was clear that the set of cloned inserts covered the totality of the viral genome under the form of overlapping cDNA fragments extending from nucleotide 1 at the 5' end of the genome to the poly(A) stretch at the 3' end.We could therefore reconstruct with virtual certainty the restriction map of the viral genome from the individual restriction map of each of the different plasmids (26,41).The map obtained (fig.3) was compared to, and found to be consistent with the maps derived from the nucleotide sequence. As an illustration, table 1 lists the positions of 45 restriction enzyme cleavage sites. We found that restriction enzymes which did not cleave the poliovirus cDNA were EcoRI, AvaI, HpaI, PvuI, SalI,SmaI, XhoI and XmaIII.

The absence of an AvaI site was rather surprizing,as this site can be found at nucleotide 2980 in the sequence reported by Racaniello and Baltimore (4) for the Mahoney strain of poliovirus type 1 and is also predicted,at the same position,from the sequence of the Sabin strain (5).We have sequenced several of our plasmids and found that the AvaI recognition sequence CCCGAG at position 2980 is replaced by the sequence CCCGAC (data not shown). The same sequence has been reported by Kitamura et al. for the RNA of the Mahoney strain (4).It is therefore quite likely that these discrepancies reflect a one-base substitution between clones of the Mahoney strain which were derived from the same original virus stock but which were propagated independently in different laboratories.

Table 1. Location of some of the restriction endonuclease recognition sequences in the poliovirus genome.

Enzyme	Number of sites	and location [a]
BglII	1	(5601)
AccI	1	(6218)
NruI	1	(1172)
PvuII [b]	1	(7053)
BclI	2	(3217 3963)
BglI	2	(35 5318)
BstEII	2	(3235 3925
EcoRV	2	(5805 6024)
HindIII	2	(6056 6516)
KpnI	3	(66 3064 3660)
PstI	3	(1809 2243 3417)
BalI	5	(627 3447 3682 4908 6282)
BamHI [c]	5	(220 670 2099 2129 4600)
HaeII	5	(2467 2966 4842 5617 5823)
HindII	5	(3913 5192 5240 6770 6990)
XbaI	5	(2546 2861 3581 4886 6304)

a. The coordinates indicated are those of the first base in the sequence of the restriction site as determined from the nucleotide sequence (3,4).Numbering is after Dorner et al. (12).Most of the cleavage sites indicated have been verified by gel electrophoresis of restriction enzyme digests of the recombinant poliovirus plasmids.

b. There is only one PvuII site contrary to earlier reports (26,41).

c. The existence of the two BamHI sites at position 2099 and 2129 was not detected by restriction enzyme analysis but predicted from the nucleotide sequence.

3. CONSTRUCTION OF FULL-LENGTH cDNA INSERTS

To reconstruct a full-length poliovirus cDNA molecule, plasmids carrying appropriate overlapping cDNA fragments were recombined *in vitro*, taking advantage of the presence of unique restriction sites in the regions of overlaps. As shown in Fig.3, a 5700 nucleotides long poliovirus cDNA insert was constructed in this way by recombining *in vitro* plasmids pPV1-846 and -120 at the KpnI site (poliovirus nucleotide 3064) both plasmids shared in their region of cDNA overlap. The construction is detailed in Fig. 5.

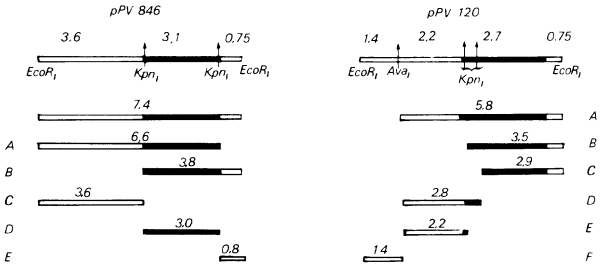


FIGURE 5. Details of the construction of plasmid pPV1-958. The DNA of plasmid 846 was digested with EcoRI and partially with KpnI and the 5 resulting fragments (A-E, left side of the figure) were separated by agarose gel electrophoresis. The DNA of plasmid 120 was treated similarly, in addition to which it was further restricted with AvaI, generating 6 fragments (A-F) as indicated to the right of the figure. The 6.6 kb EcoRI-KpnI fragment (A) from plasmid 846 and the 3.5 kb KpnI-EcoRI fragment (B) from plasmid 120 were electroeluted from the gel and ligated using T4 DNA ligase. This reconstructed a new recombinant plasmid, pPV1-958, containing the first 5750 poliovirus nucleotides inserted at its PstI site (Fig.3).

The resulting recombinant plasmid, pPV1-958, contained the entire poliovirus sequence from the 5' terminal nucleotide to approximately nucleotide 5750, i.e. the entire sequence coding for the structural protein NCVP1a, and for the non-structural protein NCVP3b (28). The insert in plasmid 958 was next elongated towards the 3' end of the genome by *in vitro* recombination with plasmid 404 at the poliovirus BglII site. The cDNA of the resulting plasmid was spliced, in turn, at the PvuII site, to the cDNA insert of plasmid 027 (Fig.3). In this way, a plasmid containing a poliovirus cDNA copy of genomic length was generated and designated pPV1-1022.

As illustrated in Fig.3, another recombinant plasmid containing a full-length cDNA copy,pPV1-1515, was constructed using the same methodology but starting from 3 altogether different cDNA fragments of the genome,carried by plasmids 336, 334, and H3 respectively.

Colinearity of the reconstructed full-length cDNA inserts with viral RNA was ascertained by electron microscope analysis of the R-loops formed between the EcoRI restricted plasmid DNAs and viral RNA extracted from purified virions.Annealing of the molecules was done at elevated temperature and in high formamide concentration,i.e. under conditions which favor DNA-RNA over DNA-DNA hybridization (46).

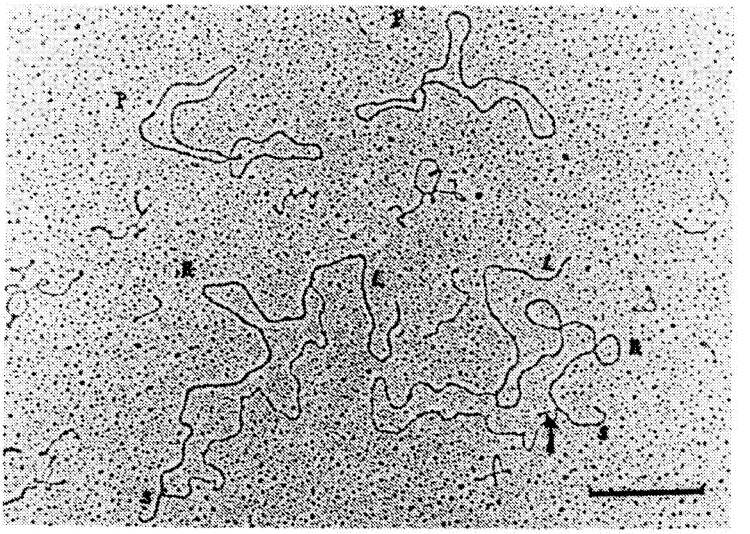


FIGURE 6.R-loops between the DNA of pPV1-1515 and viral RNA. 0.1 µg of EcoRI digested plasmid DNA was mixed with 0.1 µg of purified viral RNA.The mixture was annealed in 15 µl 0.1 M Tricine-NaOH,pH 8.0,70% formamide,0.5 M NaCl,10 mM EDTA,for 3 hr at 52°C,spread,and examined as previously described (42). Symbols: L:EcoRI-PstI long arm of pBR322; s: PstI-EcoRI small arm of pBR322; R: R-loops;P:molecules of bacteriophage PM2 DNA used as internal size standards.Relative segment lengths were determined by comparison with that of the PM2 molecules (9.725 kb). The arrow refers to the hybridization tail of the 3' end of the RNA. The bar is 0.5µm.

The length of the R-loops examined varied from 7.0 to 7.4 kb (Fig.6). Each R-loop was flanked by 2 double-stranded arms of unequal length,corresponding to the long and short EcoRI-PstI arms of pBR322,respectively. In addition, a tail of non-hybridized RNA was clearly detected,in many molecules,at the extremity of the R-loop closest to the small arm of the vector (arrow in Fig.5). There are several reasons to believe that such a tail resulted from a branch migration phenomenon initiated at the 3' terminal poly(A) stretch of the viral RNA.

Firstly,a similar tail has been observed with all the preparations of plasmids containing a full-length poliovirus cDNA insert that we examined,including those of plasmid pVR-106 (27),kindly provided by Racaniello and Baltimore.Secondly,the tail was no longer visible when RNA-DNA hybrid molecules formed by hybridizing viral RNA and fully denatured (single-stranded) plasmid DNA were examined (results not shown).Thirdly,determination of the nucleotide sequence at the 3' end of the full-length poliovirus inserts in plasmids 1022 and 1515 has failed to reveal the presence of any insertion, inversion or deletions that could account for a lack of hybridization of the plasmids DNAs with the 3' end of the viral RNA. It is therefore highly likely that the free 3' end tails of the RNA detected in the electron micrographs of R-loops are an artifact resulting from local reannealing of the two strands of the plasmid DNA.

The biological properties of the full-length cDNA copies have been investigated by transfecting human (HeLa) or monkey (CV1 and VERO) cells in culture with the DNA of the plasmids, using the calcium phosphate technique of Graham and van der Eb (47) as recently modified by Chu and Sharp (48).Two types of assays were performed to measure the infectivity of the recombinant plasmids.In the first type of assay,cell cultures were transfected with plasmid DNA while still in suspension (48),after which they were seeded in 35 mm dishes under growth medium supplemented with serum, and incubated at 37°C.After 4 hr incubation,the cells were shocked with glycerol (49), washed, and incubated again at 37°C under liquid medium.Total cpe was then monitored.In the second type of assay,transfected

cells which had been seeded under liquid medium,incubated for 4 hr at 37°C,and shocked with glycerol,were overlaid with soft agar and incubated at 37°C. Plaques were counted after 2 days incubation at 37°C.

For reasons which are not yet clear,only pPV1-1515 has been found infectious so far,and not pPV1-1022.We have compared the infectivity of the former with that of plasmid pVR-106 constructed by Racaniello and Baltimore (27). The specific infectivity of pVR-106 was 20-50-fold higher than that of pPV1-1515 (table 2),but viral RNA was still several hundred-fold more infectious than pPVR-106 (on a molar basis) when measured in a parallel assay (results not shown).Under liquid medium,cpe was rarely observed before 4-5 days after transfection with 5 µg of pPV1-1515 whereas it was observed within 3 days in the case of transfection with 2 µg of pVR-106.Seroneutralisation assays were performed to ascertain that both cpe were indeed due to the production of type 1 poliovirus. Titrations of virus after freeze-thawing of the transfected cell cultures with positive cpe usually yielded a value of 2×10^8 pfu per 10^6 transfected cells.

Among the several factors studied which were found to influence the efficiency of the transfection assays,the nature of the cell line used for transfection turned out to be most critical:VERO cells were definitely and reproducibly more sensitive than CV1 cells or HeLa cells.They were also easier to handle,and resisted well to the glycerol shock.The use of polyethylene glycol instead of glycerol to shock the cells (50) was difficult to control and did not seem to improve the efficiency of the assay.Attempts at replacing transfection with calcium by fusion of the cells with bacterial protoplasts (51) were not met with success.The most efficient method in our hands was that of the calcium phosphate precipitate as applied to cells in suspension,followed by a glycerol shock 4 h later (48).Under these conditions,a specific infectivity of 2-13 pfu/ µg of DNA was consistently observed with the DNA of pVR-106 (table 2).

Table 2. Specific activity of plasmids containing a full-length poliovirus cDNA insert.

Plasmid	Specific infectivity[a] (pfu/μg DNA)	
	Experiment 1	Experiment 2
PVR-106	13.0	7.0
pPV1-1515	0.2	<0.2
pR15-56A[b]	<0.2	<0.2
pR15-56B[c]	3.0	12.0
pR15-70A[d]	N.D.	9.0
pR15-70B[e]	N.D.	<0.2

a. The DNA of the indicated plasmids was transfected onto freshly trypsinized VERO cells in suspension using from 1 to 5 μg of DNA per 10^6 cells (47). The cells were then seeded into 6 cm plastic 35 mm dishes under warm medium containing 6.25 mM CaCl2 and 0.05 X Hepes saline buffer (2 X Hepes saline is 10 g/l Hepes, pH 7.05, 16 g/l NaCl, 0.74 g/l KCl, 0.25 g/l Na2HP04.2H20, 2 g/l dextrose) and transferred to a CO_2 incubator at 37° C. The medium was carefully withdrawn 4 hr later and the cell monolayers were overlaid with 5 ml of 20% glycerol in 0.1X Hepes saline. One and half min later, the cells were washed,overlaid with 0.9% agar, and incubated at 37°C. Plaques were counted 2-3 days later.

b. Recombinant plasmid pR15-56A contains the first 5602 poliovirus nucleotides from pVR-106 (up to the BlgII site) and the last 1838 nucleotides of the genome and the poly(A) stretch from pPV1-1515.

c. Plasmid pR15-56B contains the 1st 5602 nucleotides from pPV1-1515 and the 3' end of PVR-106.

d. Plasmid pR15-70A contains the first 7056 poliovirus nucleotides from pVR-106 (up to the PvuII site) and the last 384 nucleotides and the poly(A) stretch from pPV1-1515.

e. Plasmid pR15-70B contains the 1st 7056 nucleotides from pPV1-1515 and the last 384 nucleotides and the poly(A) from pVR-106.

In an attempt to understand what were the reasons for the lower infectivity of pPV1-1515,hybrid plasmids were constructed by recombining the 3' part of pVR-106 with the 5' part of pPV1-1515, and vice versa, and the infectivity of the resulting chimeric plasmids was tested by plaque assay (table 2). Substituting the 1838 3'-terminal nucleotides from pVR-106 to

those of 1515 increased the infectivity of the latter up to that of pVR-106, showing that most of the reason for the lower infectivity of 1515 lied in the sequence of the 3' end of the cDNA insert.

Table 2 shows that replacement of the BglII-PvuII fragment from pPV1-1515 (poliovirus nucleotides 5602-7056) by the corresponding fragment from pVR-106 should be sufficient to confer substantial infectivity to the plasmid, whereas substitution of the 5602-7056 fragment from pVR-106 by the corresponding fragment from pPV1-1515 should abolish the infectivity of the plasmid. It is most probable that the reason for the very low infectivity of pPV1-1515 is to be found in the nucleotide sequence of the insert between poliovirus nucleotides 5602 and 7056. Determination of this sequence is presently being undertaken.

Determination of the nucleotide sequence of the region of the cDNA from the PvuII site (poliovirus nucleotide 7053) to the 3' border of the insert, failed to reveal differences in the poliovirus coding sequences of pVR-106 and pPV1-1515. The only difference which was observed between both plasmids was that of the length of the poly(A) stretch (table 3).

Another important difference between both plasmids was observed at the 5' end of the inserts: the poly(dG) stretch of pPV1-1515 was found to be 4 times shorter than that of pVR-106 (table 3). The rest of the 5' sequence up to the PstI site at nucleotide 1809 appeared to be the same in both plasmids (results not shown). The influence of the poly(dG) stretch on the expression of the cDNA might be of importance as GC rich sequences have been found to be important for the function of several eukaryotic promoters (54,55).

However, recombinant pR15-56B, which carries the first two thirds of pPV1-1515, with the 3' end of pVR-106, was approximately as infectious as pVR-106 itself on a molar basis. It is therefore probable that the 5' end sequences play little role in the specific infectivity of the plasmids.

Table 3. Number of residues within repetitive sequence elements at the 5' and 3' flanks of the poliovirus cDNA insert in plasmids pVR-106 and pPV1-1515.

Plasmid	Number of residues [a]		
	Poly(dG)	Poly(dA)	Poly(dC)
pPV1-1515	15	15	17
PVR-106	50-60	37	17

a. The number of G residues preceding the 5' end of the poliovirus insert, that of A residues in the 3' terminal poly(A) of the insert, and that of C residues following the poly(A) stretch, were determined by sequencing the DNA of pPV1-1515 using the dideoxy-chain termination method (52), after sub-cloning of the appropriate DNA fragments in phage M13mp8 and/or M13mp9 (53). The data relative to pVR-106 are from reference 27. The presence of the flanking poly(dG) and poly(dC) elements results from the tailing of the DNA by terminal deoxynucleotidyl transferase at the time of cloning (see legend to Fig.2).

4. CONCLUSION

Cloning of the poliovirus genome under the form of RNA-cDNA hybrid molecules into E.coli plasmid pBR322 resulted in the production of a family of recombinant plasmids containing scattered overlapping fragments of viral cDNA corresponding to the totality of the viral genome cloned in multiple pieces. In vitro recombination between appropriate plasmids allowed to generate new recombinant plasmids containing genomic length cDNA inserts. One at least of these plasmids, pPV1-1515, was infectious, i.e. it was able to regenerate infectious poliovirions when transfected into susceptible primate cells. The infectivity of pPV1-1515 was however extremely low, as measured by direct plaque assay. We compared its infectivity to that of recombinant plasmid pVR-106, which was kindly made available to us by Racaniello and Baltimore. pVR-106 contains a full-length poliovirus cDNA insert and is infectious (27). Comparison of pVR106 and pPV1-1515 and study of recombinants

constructed between both plasmids suggests that the reason for the very low infectivity of pPV1-1515 lies in the nucleotide sequence of the insert between nucleotides 5602 and 7656. Nucleotide sequencing of this region of the plasmid DNA will be of interest to determine the exact nature of the defect. Surprizingly,although pPV1-1515 was not able or barely able to generate plaques when the transfected cells were overlaid with agar immediately after transfection,it was perfectly able after a few days incubation to generate cpe in cell monolayers kept under liquid medium, and the presence of type 1 poliovirus was readily demonstrated in the cell lysate by plaque assay. The basis for such phenomenon remains unclear.

It will be of interest to investigate, at the molecular level, how the 1st molecule of infectious self-replicating viral RNA is generated in a successfully transfected cell.It is not known for example whether the incoming plasmid must integrate into the cell DNA prior to transcription,or whether it is directly transcribed by the host cell RNA polymerase after it has entered the cell nucleus.Introducing eukaryotic promoters upstream from the poliovirus cDNA insert into the recombinant plasmid DNA will help clarify this point. It would also be of interest to unravel the details of the transcription mechanism and of the possible post-transcriptional processing mechanisms which allow to generate from the cDNA insert in the plasmid a molecule of viral RNA beginning with the correct viral 5' end sequence and terminating with a normal poly(A) tail at the 3' end. It has been shown that the poly(A) stretch is required for the infectivity of poliovirus RNA (56). It will be of interest to determine whether the length of the poly(A) in the viral RNA molecules generated upon transfection of the cells is that of the cDNA insert in the plasmid used for transfection, or has been elongated up to the size of the standard poly(A) stretch normally found in wild type poliovirus. Experiments aimed at answering these questions are in progress.

ACKNOWLEDGEMENTS

We thank Françoise Suarez and Odile Fichot for efficient technical help and Charles Dauguet for help with the electron microscope. We gratefully acknowledge the help of Akiko Kitamura for the preparation of the RNase T1-resistant oligonucleotides and that of François Bregégère for the initial cloning experiments. This work was helped by NATO Research grant n° 75/80, by C.N.R.S. grant N°A1-5037 and ERA N°1030, and by INSERM grant N°124010.

REFERENCES

1. Baltimore,D. Bacteriol.Rev.35:235-241,1971.
2. Kitamura,N., Adler,C., and Wimmer,E. Ann.N.Y.Acad.Sci. 345:183-201,1980.
3. Kitamura,N., Semler,B.L., Rothberg,P.G., Larsen,G.R., Adler,C.J., Dorner,A.J., Emini,E.A., Hanecak,R., Lee,J.J., van der Werf,S.,Anderson,C.W. and Wimmer,E. Nature (London) 291:547-553,1981.
4. Racaniello,V. and Baltimore,D. Proc.Natl.Acad.Sci. U.S.A. 78: 4887-4891, 1981.
5. Nomoto,A., Omata,T.,Toyoda,H.,Kuge,S.,Horie,H.,Kataoka,Y., Genba,Y., Nakano,Y. and Imura,N. Proc.Natl.Acad.Sci. U.S.A. 79:5793-5797,1982.
6. Jacobson,M.F. and Baltimore,D. Proc.Natl.Acad.Sci. U.S.A. 61:77-84,1968 .
7. Summers,D.G. and Maizel,J.V. Proc.Natl.Acad.Sci. U.S.A. 59:966-971,1968.
8. Rueckert,R.R., Matthews,T.J., Kew,O.M., Pallansch,M., McLean,C., and Omilianowski,D. In: The Molecular Biology of Picornaviruses,(Ed.Perez-Bercoff,R.),Plenum, New York pp.113-125, 1979.
9. Taber,R., Rekosh,D. and Baltimore,D. J.Virol. 8:395-401, 1971.
10. Larsen,G.R.,Anderson,C.W., Dorner,A.J.,Semler,B.L., and Wimmer,E. J.Virol. 41: 340-344,1982.
11. Semler,B.L., Anderson,C.W., Kitamura,N., Rothberg,P.G., Wishart,W.L., and Wimmer,E. Proc.Natl.Acad.Sci. U.S.A. 78: 3464-3468, 1981.
12. Dorner,A.J., Dorner,L.F., Larsen,G.R., Wimmer,E., and Anderson,C.W. J.Virol. 42: 1017-1028, 1982.
13. Rekosh,D. J.Virol. 9:479-487, 1972.
14. Jacobson,M.F.,and Baltimore,D. J.Mol.Biol. 33: 369-378, 1968 b.
15. Putnak,J.R., and Philipps,B.A. Microb.Rev.,45:287-315,1981.
16. Rueckert,R.R. In: Comprehensive Virology, (Eds. Fraenkel-Conrat,H. and Wagner,R.R.),Plenum, New York, vol.6, pp.131-213, 1976.

17. Semler,B.L., Hanecak,R., Anderson,C.W. and Wimmer,E. Virology. 114:589-594,1981.
18. Hanecak,R., Semler,B.L.,Anderson,C.W. and Wimmer,E. Proc. Natl.Acad.Sci. U.S.A. 79: 3973-3977,1982.
19. Van Dyke,T.A.,and Flanegan,J.B. J.Virol. 35:732-740,1980.
20. Baron,M.H., and Baltimore,D. J.Biol.Chem. 257:12359-12366, 1982.
21. Baron,M.H., and Baltimore,D. J.Virol. 43: 969-978, 1982.
22. Flanegan,J.B., and Baltimore,D. J.Virol. 29: 352-360,1979.
23. Etchison,D. and Ehrenfeld,E. Virology 107:135-143, 1980.
24. Emini,E.A., Elzinge,M., and Wimmer,E. J.Virol. 42:194-199, 1982.
25. Baron,M.H., and Baltimore,D. Cell 28: 395-404, 1982.
26. Van der Werf,S., Brégégère,F., Kopecka,H., Kitamura,N., Rothberg,P.G.,Kourilsky,P.,Wimmer,E. and Girard,M. Proc.Natl.Acad.Sci. U.S.A. 78: 5983-5987, 1981.
27. Racaniello,V., and Baltimore,D. Science 214:916-919,1981.
28. Van der Werf,S.,Dreano,M., Bruneau,P., Kopecka,H., and Girard,M. Gene 23:85-93, 1983.
29 Van der Werf,S.,Wychowski,C.,Bruneau,P., Blondel,B., Crainic,R., Horodniceanu,F. and Girard,M. Proc.Natl.Acad. Sci. U.S.A. in press, 1983.
30. Lee,Y.F., Kitamura,N., Nomoto,A. and Wimmer,E. J.Gen. Virol. 44:311-322, 1979.
31. Deng,G. and Wu,R. Nucleic Acids Res. 9:4173-4188, 1981.
32. Murray,N.E.,Brammer,W.J. and Murray,K. Mol.Gen.Genet. 150:53-61, 1976.
33. Villa-Komaroff,L., Efstratiatis,A.,Broome,S., Lomedico,P., Tizard,R., Naber,S.P., Chick,W.L. and Gilbert,W. Proc.Natl.Acad.Sci. U.S.A. 75: 3727-3731, 1978.
34. Wood,K.O., and Lee,J.C. Nucleic Acids Res. 3:1961-1971, 1976.
35. Zain,S.,Sambrook,J.,Roberts,J.J., Keller,W.,Fried,M. and Dunn,A. Cell 16:851-861, 1979.
36. Grunstein,M., and Hogness,D. Proc.Natl.Acad.Sci.U.S.A. 72:3961-3965, 1975.
37. Cami,B., and Kourilsky,P. Nucleic Acids Res. 5:2381-2390, 1978.
38. Lee,Y.F., Kitamura,N., Nomoto,A. and Wimmer,E. J.Gen. Virol. 44:311-322,1979.
39. Nomoto,A.,Jacobson,A.,Lee,Y.F., Dunn,J.J. and Wimmer,E. J.Mol.Biol. 128:179-196, 1979.
40. Stewart,M.L.,Crouch,R.J., and Maizel,J.V. Virology 104: 375-397, 1980.
41. Kopecka,H.,Van der Werf,S., Brégégère,F.,Kitamura,N., Dreano,M.,Rothberg,P.G., Wimmer,E., Kourilsky,P. and Girard,M. Develp.biol.Standard. 50: 301-309, 1982.
42. Girard,M.,Bénichou,D., van der Werf,S., and Kopecka,H. Annales Virol.(Institut Pasteur) 132:271-285, 1981.
43. Maxam,A.M. and Gilbert,W. Methods Enzymol. 65:499-560, 1980.
44. Kupper,H.,Keller,W., Kurz,C., Forss,S.,Schaller,H., Franze,R.,Strohmaier,K.,Marquandt,O., Zaslavsky,V.G. and Hofschneider,P.H.Nature (London) 289:555-559, 1981.

45. Bothroyd,J.C.,Highfield,P.E.,Cross,G.A.M.,Rowlands,D.J., Lowe,P.A.,Brown,F., and Harris,T.J.R. Nature (London) 290:800-802, 1981.
46. Thomas,M.,White,R.L., and Davis,R.W. Proc.Natl. Acad.Sci. U.S.A. 73:2294-2298, 1976.
47. Graham,F.L., and van der Eb, A.J. Virology 52:456-467, 1973.
48. Chu,G., and Sharp,P. Gene 13:197-202, 1981.
49. Frost,E., and Williams,J. Virology 91:39-50, 1978.
50. Shen,Y.M.,Hiaschkorn,R.R.,Mercer,W.E.,Surmacz,E., Tsutsui,Y., Soprano,K.J., and Baserga,R.M. Molec.Cell. Biol. 2:1145-1154, 1982.
51. Rassoulzadegan,M., Binetruy,B. and Cuzin,F. Nature 295: 257-259,1982.
52. Sanger,F., Nicklen,S., and Coulson,A.R. Proc.Natl.Acad. Sci. U.S.A. 74:5463-5467, 1977.
53. Messing,J. Third Cleveland Symp.on Macromolecules; Recombinant DNA,(Ed.A.Walton),Elsevier, Amsterdam,pp. 143-153, 1981.
54. Everett,R.D.,Baty,D., and Chambon,P. Nucleic Acids Res. 11:2447-2464, 1983.
55. McKnight,S.L.,and Kingsbury,R. Science 217:316-324,1982.
56. Spector,D.H. and Baltimore,D. Proc.Natl.Acad.Sci. U.S.A. 71:2983-2987, 1974.

CLONING OF PLANT VIRUSES AND VIROID GENES

13

CLONING AND MANIPULATING CAULIFLOWER MOSAIC VIRUS

L.K. DIXON and T. HOHN, FRIEDRICH MIESCHER INSTITUTE, BASEL

I. INTRODUCTION

Most of the several hundred groups of plant viruses identified have an RNA genome, only the geminiviruses and caulimoviruses have DNA genomes. Geminivirus particles are composed of two shells fused together probably at pentameric corners. They contain single-stranded circular DNA, 2-3 Kb in length. As a group, they infect a wide range of mono and dicotolydonous host plants. Two different DNA molecules are necessary for infection by whitefly transmissible bean golden mosaic virus (BGMV; 1), tomato golden mosaic virus (TGMV; 2,3) and casava latent virus (CLV; 4). In contrast to these bipartite viruses some of the leafhopper transmissible strains such as chloris striate virus may be monopartite and require only one DNA molecule for infectivity (5). Double-stranded replicative intermediates have been isolated (2,6) and used for molecular cloning.

In addition to their interest as model systems for studying virus-plant cell interactions, both caulimoviruses and geminiviruses have been suggested as potential vectors to introduce foreign genes into plant cells.

This review focuses on cauliflower mosaic virus (CaMV), the most studied member of the caulimovirus group. The first section describes the viral life cycle (see 7,8,9,10, for a detailed account of knowledge before 1982). Experiments involving the cloning and manipulation in vitro of the viral DNA are then described. Finally, progress towards understanding the viral life cycle and developing the virus as a useful vector are discussed and outlooks for the future assessed.

II. CAULIFLOWER MOSAIC VIRUS LIFE CYCLE

A. Physical and Phytopathological Properties

CaMV infects mainly Crucifereae, although some strains can also infect Datura, a member of Solanaceae (10). CaMV is transmitted in nature by aphids and perhaps by leaf to leaf contact. Non-aphid transmissible

strains exist. These can still be transmitted by aphids that have previously fed on plants infected with a transmissable strain which suggests that the transmissable strain produces a substance during infection which allows transmission of the defective strain (11,12).

Virus infection causes chlorotic lesions giving leaves a mosaic appearance. Once established, the infection spreads systemically through the tissues of the infected plant causing vein clearing, wrinkling and stunting of new leaves.

When these leaves are examined by electron microscopy, characteristic electron-dense masses of irregular size and shape are found in the cytoplasm of infected cells (Fig.1). Two types of these "inclusion bodies" have been described. One type is vacuolated and contains virus particles within these vacuoles and embedded in the matrix material. The other type is non-vacuolated and is devoid of virus particles (10). Purified inclusion bodies may be solubilised releasing virus particles. Neutron diffraction studies reveal that the DNA is layered against the inside surface of the capsid shell with the centre of the particle free of DNA and protein (13).

Inclusion bodies of different CaMV strains differ in appearance and stability (14,15). In addition, the growth temperature of the host plant effects inclusion body structure (16).

B. CaMV DNA

CaMV DNA is circular, double stranded and 8000 base pairs long. The complete nucleotide sequences have been determined for strains S (17), CM1841 (18) and DH (19) and are available by code CAMVG2, CAMVG1 and CAMVDH from the EMBL sequence library. Additional sequencing and mapping data on a variety of strains of CaMV and other caulimoviruses exist (20, 21,22,23,24,25,26,27). Some features of these sequences are shown in Table 1 and Fig.2. The sequence reveals six large (I-VI) and two small (VII and VIII) open reading frames (ORF) all encoded on the minus strand of the DNA (Fig. 2). These are closely spaced, generally alternate in reading frame and overlap little. An exception is ORF VIII which is located within ORF IV but in a different reading frame. There is one large (700 bp) intergenic region.

CaMV DNA isolated from virus particles is relaxed (Fig. 3, reviewed: ref. 7). Melting the double-stranded DNA yields 3 linear single-stranded

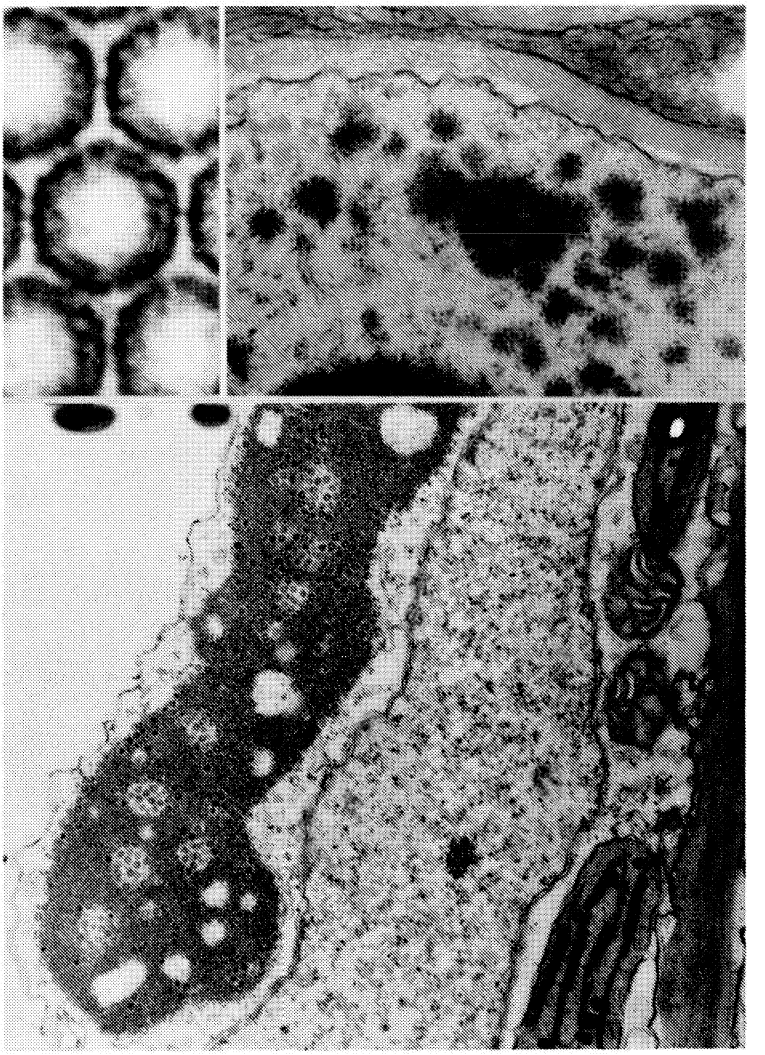


Fig.1
Electronmicroscopic observation of cauliflower mosaic virus. Upper left, virus particles (optically filtered image from a paracrystalline packing); upper right, inclusion bodies, type with none or few virus particles. Lower, inclusion bodies, type with vacuoles and virus particles. The photographs were kindly provided by J. Ménissier, M. Würtz and M. Gibaud.

Table 1. Some important sites in CaMV DNA (see also Fig. 2).

CaMV(S)DNA		Some important sites	
1 - 14		**T**GGT ...	Met tRNA homology, (-)PBS(1)
14		**A**TGA	Start ORFVII
130		*GGT**G**ACC*	*BstEII site*
302		**T**AA	End ORFVII
365		**A**TGG	Start ORFI
733		*GAC**C**ATGCT*	*TthI site*
1346		**T**AA	End ORFI
1350		**A**TGA	Start ORFII
1634	TTAAGAGTGGGGG**G**GT		Start plus strand, (+)PBS(3)
1646		*CTC**G**AG*	*XhoI site*
1827		**T**AA	End ORFII
1831		**A**TGG	Start ORFIII
2044		*CGT**A**CC*	*KpnI site*
2202		**A**TGG	Start ORFIV
2218		**T**AG	End ORFIII
3098		*GTT**A**AC*	*HpaI site*
3634	or	**A**TGG	} Start ORFV
3679		**A**TGA	
3669		**T**GA	End ORFIV
4219	TTCAGAGGGGAGG**A**GG		Start plus strand, (+)PBS(2)
4839		*GTC**G**AC*	*SalI site*
5390		*CTG**C**AG*	*PstI site*
5663		**T**TAATT	} "CAT" Boxes
5669		**C**TAATT	
5671		**T**AA	End ORFV
5732		**T**ATTTAAA	"TATA" Box
5764		AAT**C**AGACC	Start small transcript
5777		**A**TGG	Start ORFVI
5828		*GTG**C**AC*	*SacI site*
7337		**T**GA	End ORFVI
7405		**T**ATATAA	"TATA" Box
7436		AGG**A**CACGC	Start large transcript
7599		**A**ATAAA	Polyadenylation site
7616		TTC**C**TAAAA	End transcripts
7720		*GAC**G**CC*	*HgiDI site*
8024		AAA**T**	Start minus strand (reverse to the numbers)

Sequences at the sites are given in the second column, bold letters refer to the gase in the sequence numbered in the first column. Signals for DNA, RNA and protein synthesis are given, as well as some prominent restriction sites.

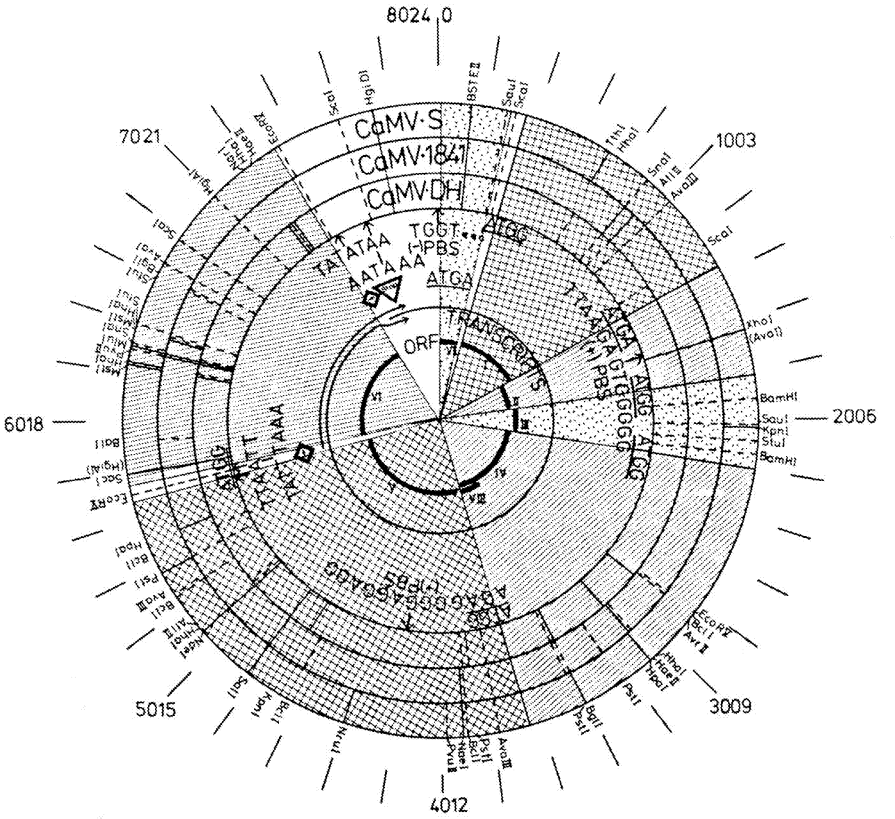


Fig.2
Physical map of the three sequenced CaMV strains, CaMV S (17), CaMV 1841 (18) and CaMV DH (19). Location of restriction sites occurring just once (——) or only 2 or 3 times (---) are indicated. Prominent restriction sites missing are PvuI, SphI and SmaI. Numbering is given for CaMV S, the other strains are 8031 and 8016 bp long. Signal sequences are indicated on the map and their exact locations are shown in Table 1. (-)PBS, primer binding site for minus DNA strand synthesis (site of single strand overlap in packaged (-)DNA); (+)PBS, primer binding site for plus DNA strand synthesis (sites of single strand overlaps in packaged (+)DNA). Start (◁▷) and polyadenylation (▽ ATGA ▽) signals for transcription. Thin curved arrows indicate major transcripts, segments and bold arrows symbolize the open reading frames in common to all three strains.

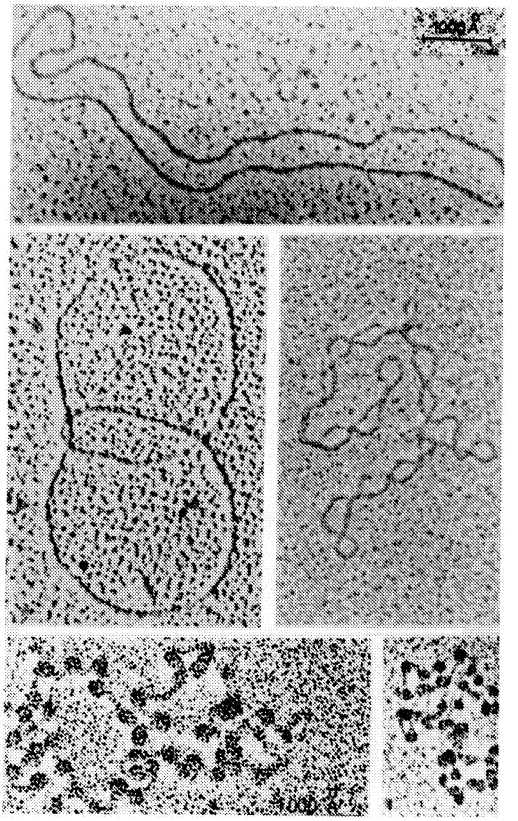


Fig.3
Cauliflower mosaic virus DNA. Relaxed circular (upper panel) and relaxed knotted (middle left panel) forms as packaged in virus particles, supercoiled form (middle right panel) and minichromosomes (lower panels) as isolated from infected nuclei. Electron micrographs kindly provided by J. Ménissier, Strasbourg.

DNA species, a full length minus strand and two fragments of plus strand, 1/3 and 2/3 full length. This indicates that there are three discrete interruptions in the DNA, one in the minus strand the others in the plus strand (20,25,28). These interruptions are also sensitive to S1 nuclease treatment and analysis has revealed that the interruptions are sites of single strand sequence overlap (17,29,30). The degree of overlap varies within the population of DNA molecules but is predominantly 8 bases in the minus strand for Δ1 and 18 and 15 bases in the plus strand for Δ2 and

Δ3. The position of the 5' end is always fixed (29,32). The 14 nucleotides at the 5' end of CaMV DNA, immediately adjacent to and downstream from Δ1 are complementary to the 3' end of a plant met tRNA (30,31; see chapter II-D).

C. CaMV transcription

Two major transcripts of CaMV have been detected (32,33,34). These are polyadenylated, capped (35) and share the same 3' terminus (36,37,38). One of these (19S) is transcribed from ORF VI (39,40,41) whilst the other (35S) is a full length transcript of the viral genome with a terminally redundant sequence overlap of 180b. The terminal repeat may result from overreading the stop signal during the first round of transcription.

Minor transcripts have also been detected (38,42,43), one encompasses the entire genome starting and terminating at Δ1, whilst another (8S) extends from the 35S promoter to Δ1 (38).

Preparations of isolated nuclei incorporate radioactive precursors into virus-specific RNA which indicates that the nucleus is the site of CaMV transcription (35). Transcription is inhibited by low concentrations of α-amanitin suggesting that the CaMV genome is transcribed by RNA polymerase II (35). The template for transcription is supercoiled DNA, lacking the single strand overlaps typical of DNA isolated from viral particles (44,45) and bound to histones (46) (Fig.3). Recently, various types of subgenomic supercoiled DNA have been observed in infected nuclei. Whether these could act as template for specific transcripts has to be studied (46a).

D. CaMV DNA Replication

Conflicting reports exist on the site of CaMV replication. The accumulation of tritiated thymidine in inclusion bodies was interpreted as an indication that these virus-specific organelles are the site of DNA replication (47,48). However, Ansa et al (49) suggest that it replicates in the nucleus since isolated nuclei incorporate some radioactivity into CaMV DNA.

Recently, a model has been proposed (30,31,50,51,52) (Fig.4) which resolves these seemingly conflicting reports and other unexplained features of the viral DNA. According to this model CaMV DNA enters the nucleus during infection (Fig 4, Step 1), is repaired to produce super-

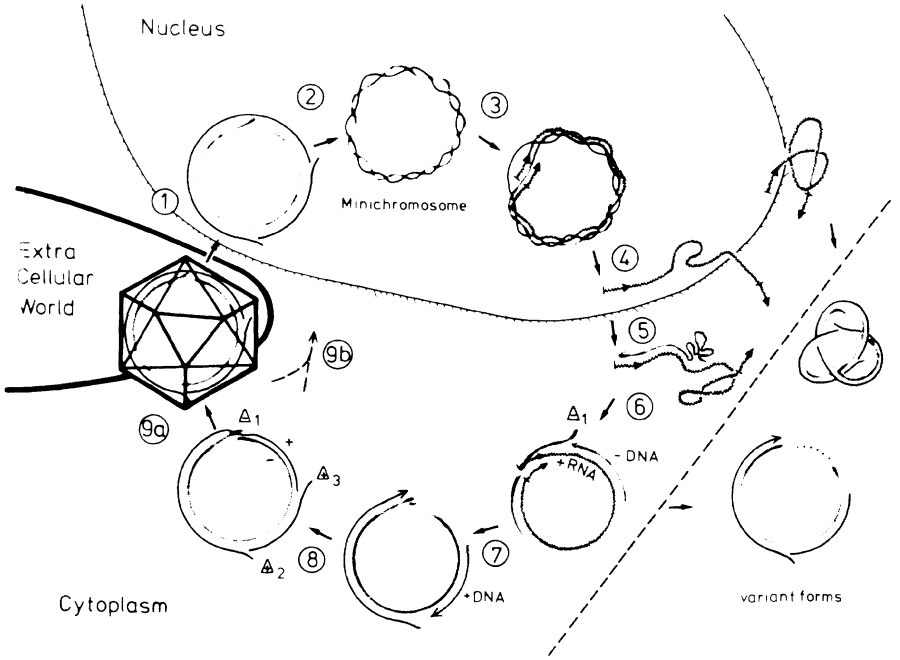


Fig.4
Model of CaMV genome replication. The numbers refer to the individual steps proposed. Black lines, plus DNA; grey lines, minus DNA; zigzag lines, RNA. ⚠ , △, S1 nuclease sensitive sites of plus and minus DNA strands (primer binding sites on RNA). From Pfeiffer and Hohn (31).

coiled DNA, and then forms minichromosomes (Fig 4, Step 2). This repair process may account for the low level of incorporation of radioactive nucleotides into CaMV-specific DNA within the nucleus. Transcription of minichromosomes by RNA polymerase II (Fig 4, Step 3) provides a means for transporting the genomic information from the nucleus to the cytoplasm (Fig 4, Step 4), where it can either be translated or replicated by reverse transcription. It is proposed that tRNA primer binds to the 35S transcript at the tRNA complementary site adjacent to Δ1 (see chapter I-C). Reverse transcription starts from this primer and copies the RNA into minus-strand DNA (Fig 4, Step 5). A template switch occurs at the end of the transcript, the nascent DNA strand switches to the 3' end of the original or another 35S RNA molecule. Reverse transcription then continues, finally passing the original minus primer binding site displacing the primer and some nucleotides at the 5' end of the viral DNA. Thereby the DNA minus-strand terminal repeat of 8 nucleotides is generated (Fig 4, Step 6).

Plus strand DNA synthesis is proposed to begin at primer binding sites located adjacent to and upstream from Δ2 and Δ3. The proposed primer binding sites show remarkable homology with plus primer binding sites of retroviruses (51). Synthesis proceeds from Δ3 to Δ2 and from Δ2 to Δ1 and after another template switch back to Δ3. Once again some nucleotides are displaced at the 5' end of the plus strand DNA forming the plus-strand terminal repeats at Δ2 and Δ3 (Fig 4, Steps 7 and 8). Double-stranded DNA thus formed with the typical single strand interruptions of CaMV DNA may either be packaged into viral particles (Fig 4, Step 9a) or re-enter the nucleus (Fig 4, Step 9b).

Evidence for this model comes from studies which demonstrate that CaMV uses two different templates for the synthesis of its RNA and DNA (31). A preparation of subcellular particles from CaMV-infected cells consisting of nuclei, inclusion bodies and a few starch grains, when leached in hypertonic buffer, released nucleoprotein complexes which could be fractionated by sucrose gradient centrifugation and assayed for endogenous RNA and DNA synthesis. CaMV-specific RNA was synthesised by faster sedimenting complexes containing CaMV minichromosomes. CaMV-specific DNA was synthesised by slower sedimenting complexes containing virus-specific DNA as well as RNA. DNA synthesis was 50% inhibited by RNAase treatment and 40% inhibited by actinomycin D, suggesting that the synthesis is performed

on both RNA and DNA templates as proposed in the reverse transcription model (31,53). Volovitch et al (54) showed that this activity can also copy poly C with an oligo dG primer and that it differs from the classical γ-like DNA polymerases. The activity is present in isolated inclusion bodies (55). It migrates in polyacrylamide activity gels as a 75 Kd protein (55a).

Isolating DNA molecules from virus-infected cells provides further evidence. Some DNA molecules may be intermediates that accumulate during the replication process. For example, single-stranded DNA molecules 725 bases long extending from Δ1 to the 35S promoter have been isolated (30, 56,56a), that have covalently attached ribonucleotides and are complementary to the plus-strand DNA. These may be nascent minus-strand DNA molecules with attached remnants of the RNA primer. Accumulation of these replication intermediates may be caused by a delay in the template switch at the terminus of the 35S transcript. Other possible intermediate structures formed during the replication process have been reported (50,56).

Pietrzak and Hohn (unpublished) examined which sequences are necessary for formation of single-strand interruption Δ3. Shortened fragments containing preΔ3 were inserted into a deletion mutant lacking this region. Although all inserts were stable, the single-strand interruption was only formed when sequences upstream of preΔ3 were inserted in native orientation. This result agrees with the prediction from the reverse transcription model for viral replication that a primer binding site in native orientation is required for Δ3 formation.

E. CaMV Translation and Protein Products

The CaMV genome has 6 to 8 possible open reading frames (ORF) (see B.). To date polypeptides and functions have been assigned or proposed to few of these. Al Ani et al. (57) presented evidence that the viral capsid comprises one major polypeptide of molecular weight 42000 daltons from which the polypeptides observed in earlier studies (58,59,61,62) were derived either by degradation or dimerization. The capsid protein is lysine rich (60). Examination of the DNA sequence suggests that ORF IV could code for such a protein. Plasmids containing ORF IV produce capsid protein in bacteria providing final proof that ORF IV codes for the viral capsid protein (62). Translation of ORF IV would produce a 55000-dalton polypeptide containing acidic amino acids flanking a lysine-rich 42000

dalton central region. The acidic amino acids may be removed by proteolytic processing yielding the mature 42000-dalton capsid protein (17,63).

In addition to viral capsid protein, purified solubilized inclusion bodies contain a major virus-encoded polypeptide of molecular weight 62000 or 66000 daltons. Translation in vitro of the 19S RNA transcribed from ORF VI yields a 66000 dalton product (34,39,41) that is immunoprecipitated by antisera prepared against purified inclusion body protein (40), a final proof that ORF VI codes for inclusion body protein.

An 18000-dalton polypeptide which co-purifies with virus inclusion bodies may be the protein product of ORF II and is apparently involved in transmission of the virus by aphids. Evidence comes from comparison of wild type virus with virus mutated in ORF II (64,21,65,15,66).

ORF V shares homology with retrovirus reverse transcriptase genes indicating that it may also code for a reverse transcriptase (67,54). The products and their functions of the other ORFs remain unknown and it is also still unclear by what means the CaMV genome is translated. As discussed above ORF VI has its own mRNA which can be translated in vitro in eukaryotic systems. Individual mRNAs for the other ORFs have not been detected and attempts to translate the 35S RNA in vitro have, as yet, not been successful . Sieg and Gronenborn (68) suggested that translation may be from a polycistronic mRNA and that the ribosome moves directly from the termination codon of one reading frame to the initiation codon of the next. They term this the "relay race model" and provide evidence (see section III-E) for such a mechanism of translation for ORFs I, II and III.

III. GENETIC ENGINEERING OF CaMV

The strategy to develop CaMV into a useful DNA vector follows the basic scheme used for viruses of the other kingdoms:

1) The viral DNA is cloned in bacterial vector to allow its rapid replication and facilitate purification and in vitro manipulations.
2) The cloned DNA is used to reinfect the original host.
3) Conditions are determined that allow introduction, maintenance and expression of payload DNA.
4) Transformation systems for single cells (protoplasts) are developed that allow additional regions of its genome to be dispensed with, for example, regions that might be required for cell penetration and cell to cell movement may be deleted in single-cell systems,

viral genome to a non-pathogenic replicating vector (replicon) or to a vector that integrates into the host chromosome (integron). Both types of vector have a potential use; the replicon vector may be used to study the expression in high copy number of genes in a defined environment, whilst the integron vector may be preferred for longer term experiments involving the regeneration of intact plants and transmission of inserted DNA to progeny plants.

5) The vector is optimized for expression of payload DNA using either the viral or other promoters.

Techniques for cloning CaMV DNA and reintroduction into plants have been well established, work on the other stages is still in progress.

A) Cloning of CaMV DNA in bacteria

The various unique restriction sites of CaMV DNA (Fig 2) offer possibilities for cloning the complete viral genome into bacterial vectors (69,70,71,72,73,74). In most cases either moderate copy-number plasmids such as pBR322 or high copy number pBR322 derivatives such as pBR328 (75), pAT153 (76) or pUC7 (77) have been used. The preferred cloning site is Sal I since insertions at this site in pBR322 can easily be screened, it is unique in most CaMV strains, and cloning the viral DNA at this rather than other sites produces more stable DNA hybrids (see below).

Although cloning of viral DNA initially removes variability from the population of viral DNA molecules (78) a new variation can be created during the growth of the bacterium harbouring the hybrid plasmid. DNA sequences useless or toxic to either the plasmid replicon or the bacterial host may be eliminated and altered by selective pressure. Alterations within the viral DNA result from cloning and mutant progeny of faithful clones can quickly outgrow their parent (8,74). CaMV strains cloned in the PstI site of pBR322 frequently acquire deletions within ORF IV (capsid protein gene).Clones in the BamHI site acquire insertion elements in the same region, whilst clones in the SalI site are rarely affected (8). The high incidence of alterations within the capsid protein gene indicates that this region is expressed in bacteria by the production of a toxic product. Daubert et al. (62) showed that CaMV capsid protein antigen accumulates in bacteria containing CaMV hybrid plasmids. Strong binding of this protein to DNA might interfere with DNA function and therefore be toxic.

CaMV clones are observed to be unstable when stored as bacterial colonies in stabs or agar plates and as a practical consequence preservation of collections as either DNA or bacterial suspensions in 50% glycerol at -70°C is recommended.

B. Reintroduction of original and cloned CaMV DNA into plants

Whilst CaMV is naturally transmitted by aphids it is also infectious when applied mechanically to the plant leaf (70,71). In this case the aphid transmission factor and its corresponding gene (ORF II) are not required.

Naked and encapsidated DNA are equally infectious and it is inconsequential whether the DNA is linear, relaxed circular or supercoiled (70). Specific infectivity ranges between 1-10 plants per µg DNA. Moreover DNA cloned in a bacterial vector, amplified as a hybrid molecule in bacteria and re-excised at the cloning site does not lose its specific infectivity.(70,71). CaMV DNA molecules remaining attached to the vector are not infectious.

C. Interactions between CaMV DNA molecules of an inoculum

Recircularization of linear CaMV DNA demonstrates that plant cells are capable of performing ligation. Lebeurier et al. (80) showed that plant cells can also ligate CaMV DNA inoculated as two separately cloned restriction fragments to yield a fully infectious virus. When a plant cell is inoculated separately with CaMV defective genomes the virus remains uninfective. If, however, a pair of such genomes cloned at an homologous restriction site are excised and inoculated the plant cells recombine the strands without any defects, thereby regenerating infectivity of the virus (81; Fig.5). This phenomenon is explained by end-to-end ligation followed by a single cross over event which resolves the mixed dimers. A similar system has recently been reported in a polyoma virus/animal cell system (82). Double cross over events are suggested when infective virus is recovered from plant cells inoculated with pairs of defective CaMV genomes, but this time, cloned in different restriction sites (80,89; Fig.5). Similarly overlapping cloned CaMV DNA fragments can also be resolved, by plant cells, to wild-type infectious virus (80; Fig.5)

Tandemly arranged cloned CaMV DNA can infect host plants without prior excision from the bacterial plasmid (80,84; Fig.5). Thus the virus

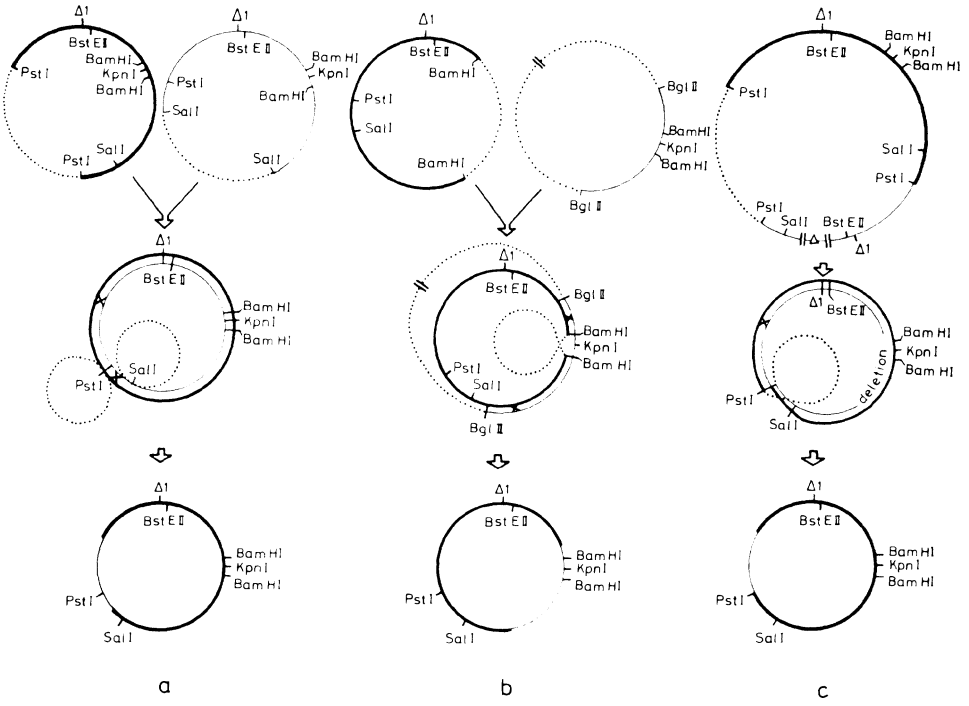


Fig.5
Suggested scheme for recombination events (crossover). (a) Recombination between two full-length CaMV DNAs. (b) Recombination between two overlapping CaMV DNA fragments. (c) Recombination of tandemly dimerized CaMV DNA. One line is a double-stranded DNA. ....., Plasmid vector sequences; ——, recombination partner 1; ——, recombination partner 2.
Δ1, position of gap 1 (see ref.2) corresponding to origin point of the CaMV DNA map. Δ , Deletion in CaMV DNA genome (c). From Lebeurier et al. (80).

genome can be transferred directly from the bacterium to the eucaryotic cell as is the case in SV 40 or polyoma/animal virus systems (85,86).

The replication mechanism involving a single-strand RNA intermediate suggests recombination mechanisms alternative to crossover events, namely recombination by template switch of the nascent DNA strand (31,51, Fig.6, Table 2). Types of interactions between CaMV DNA molecules observed and possible mechanisms of recombination on a crossover or on a replicative basis are summarized in Table 2.

Table 2. Recombination of inoculating CaMV DNA

Type of Inoculum of CaMV DNA	Possible Mechanism of Resolution	Ref.
linear	single ligation	70
two restriction fragments comprising total genome	double ligation	80
two mutant linears with homologous ends	ligation and crossover or template switch of nascent (-)DNA	81
tandemly arranged cloned DNA not excised	crossover or reverse transcription from complete transcript	80,84
pairs of CaMV DNA, cloned at different sites, excised or not	double crossover or template switch of nascent (-)DNA	80,83
pairs of overlapping CaMV DNA fragments, not excised	double crossover	80

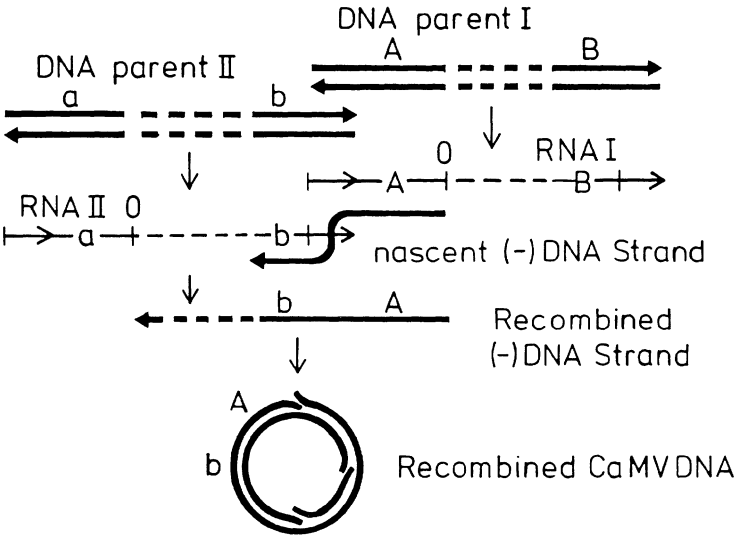


Fig.6
Model for CaMV DNA replicative recombination (interstrand switch of the nascent minus DNA strand; see also Table 2).

D. In vitro mutagenesis of CaMV

1. "Linker" insertions

Short DNA sequences (8 to 30 bp) carrying restriction enzyme sites ("linkers") have been inserted into cloned CaMV DNA (65,87,88,81). Insertion of a short DNA sequence approximately conserves the genome length and thus should not interfere with packaging of the viral DNA into particles. Introduction of a specific restriction enzyme site allows the position of the insert to be easily mapped and if the site is unique, additional manipulations, such as insertions or deletions, are facilitated. In these studies either CaMV DNA molecules linearized at a specific restriction site or random collections of linear cloned CaMV DNA molecules, each of which had received one cut with a restriction endonuclease by limited digestion in the presence or absence of ethidium bromide was used. In the presence of ethidium bromide supercoiled DNA is cleaved whilst digestion of linear or nicked circular DNA is inhibited (89). "Linkers", either commercially available or from plasmids bearing multiple restriction enzyme site linkers were ligated to the collection of full length linear cloned CaMV DNA molecules and the position of the inserted "linker" mapped after trans-

formation of E coli cells.

The functional effect of each mutation was analysed by releasing CaMV DNA from the plasmid vector DNA by restriction at the cloning site, inoculation of plants and observation of symptoms. Mutants could be classified into those which caused normal symptoms after the usual incubation period (normal), those which caused normal symptoms after an extended incubation period (delayed), those which caused milder symptoms (mild) and those which caused no symptoms.

Insertions into ORFs I, III and V were lethal whether they caused a frameshift mutation or not, indicating that the products of these ORFs are essential for viral infectivity. Insertions at the 5' termini of ORFs IV and VI were also lethal, but viable mutants were obtained when insertions were closer to the 3' termini of these ORFs. Thus, although the products of ORF IV (capsid protein) and ORF VI (inclusion body protein) are essential for viral infectivity, minor modifications may be introduced near their C termini without destroying infectivity. Two mutants with in-frame insertions near the 3' terminus of ORF VI caused milder symptoms than the wild type virus indicating that the inclusion body protein is involved in determining the severity of symptoms.

A 10 bp insertion near the 3' terminus of ORF IV, enabling a novel stop codon close to the original, but in different reading phase, to be read, caused a delay in the appearance of symptoms. When viral DNA, isolated after the first cycle of infection, was reinoculated on plants, symptoms appeared within the usual period. Sequencing of the viral DNA isolated after the first and second cycles of infection showed that the viral DNA population selectively lost a single base pair at the site of the insertion. The original reading frame was thus restored (90).

Insertions into ORFs II and VII did not destroy viral infectivity. However, development of symptoms was delayed if insertions into ORF II caused a frameshift mutation (69,87). No delay in the development of symptoms was observed if the insertion into ORF II was in frame or if a second insertion was introduced into a delayed mutant restoring the original reading frame . These results show that the products of ORFs VII and II are not essential for viral infectivity and also indicate the importance of maintaining the reading phase of ORF II.

Some insertions within the large intergenomic region were viable (e.g. at position 7671) whilst others were lethal (e.g. at position 7991,

42 bp downstream from Δ1) (87).

2. Deletion Mutants

Isolation of a naturally occurring non-aphid transmissible viral strain (CM 4184) with a deletion of 421 bp within ORF II (21), indicated that ORF II is not essential for infectivity but may be required for aphid transmission of the virus. Sequencing of this isolate revealed that the deletion occurred between short homologoues sequences (21). Viable deletion mutants of ORF II have also been constructed in vitro (68,15,91,64,66). Aphid-transmissible strains were rendered non-aphid transmissible by deletions or linker insertions (see above) within ORF II (15,64,65,66). Inclusion bodies extracted from plants infected with a deletion mutant lacked an 18,000 molecular weight polypeptide which was present in wild type virus infected cells (15,64). Deletions within ORF II of strain S altered the appearance, as observed by electron microscopy, of viral inclusion bodies. Those from cells infected with the deleted strain were less dense and viral particles free in the cytoplasm could be observed. Immunological analysis confirmed this result (15). These results indicate that ORF II codes for a product which is required for aphid transmission of the virus and which is an 18000 dalton polypeptide associated with and important for the structure of viral inclusion bodies. However, non-aphid transmissible strains of virus exist which have no major deletions in ORF II (64). It remains to be determined if these have point mutations which render the ORF II product non-functional.

Viable mutants with deletions in ORF VII have been constructed. These extend from sequences adjacent to the tRNA homologous region beyond the stop codon of ORF VII. As would be predicted from the reverse transcription model, deletions extending into the region of tRNA homoloy (the minus strand primer binding site) are lethal (87).

An in-inframe deletion of 221 bp within ORF III was lethal confirming that the product of ORF III is essential for infectivity (70).

3. Stability of Mutant DNA

Certain mutants albeit in non-essential regions of CaMV DNA also interfere with virus virulence. Consequently symptoms in the plant are weak and appear late but in later stages of infection return to normal

severity; progeny virus show normal virulence (65,87). Comparing the sequences of the progeny genome to that of the original inoculum revealed that inserts are either totally or partially removed (68), that interrupted ORFs are restored by frameshift mutations (90) and point mutations are reverted (92). This indicates a high mutation rate of CaMV DNA molecules within the plant cell in combination with selection pressure for a restored or supressed molecule. Inserted sequences interfering with virus functions are especially unstable if flanked by repeated sequences. Obviously the deletion frequency is dependent on the length of the recombination target, since a DNA fragment inserted into ORF II was rapidly excised in plant cells if bounded by long repeated sequences, but was maintained if the viral genome was bonded by shorter ones. (91,93). Clearly, to improve the chances of inserted DNA being stably maintained in the viral genome, the creation of repeated sequences should be avoided.

Further examples and the reasons for some types of interference with virulence namely interference with translation are given in the following section.

E) Strategies to Construct a Plant Vector

1) Transforming Virus (Fig.7A)

To construct a DNA vector based on CaMV, "payload" DNA must be inserted into the viral genome. This insert should not affect the viral infectivity and, to be of any value, must not only be stably maintained but also expressed when introduced into plant cells. Certain requirements on this "transforming virus" approach must be met in order to achieve this goal.

a) Systemic Infection Functions

As already discussed (section III-D), ORFs I, III, IV, V and VI produce proteins necessary for systemic viral infection and should remain intact in any "transforming virus" CaMV vector.

b) Packaging Limits

The amount of DNA that can be packaged into the shell of icosahedral viruses is strictly limited. Gronenborn *et al* (88) inserted different-sized fragments of DNA into the unique XhoI site in ORF II of strain CM1841. A fragment of 254 bp could be accommodated but one of 499 bp could

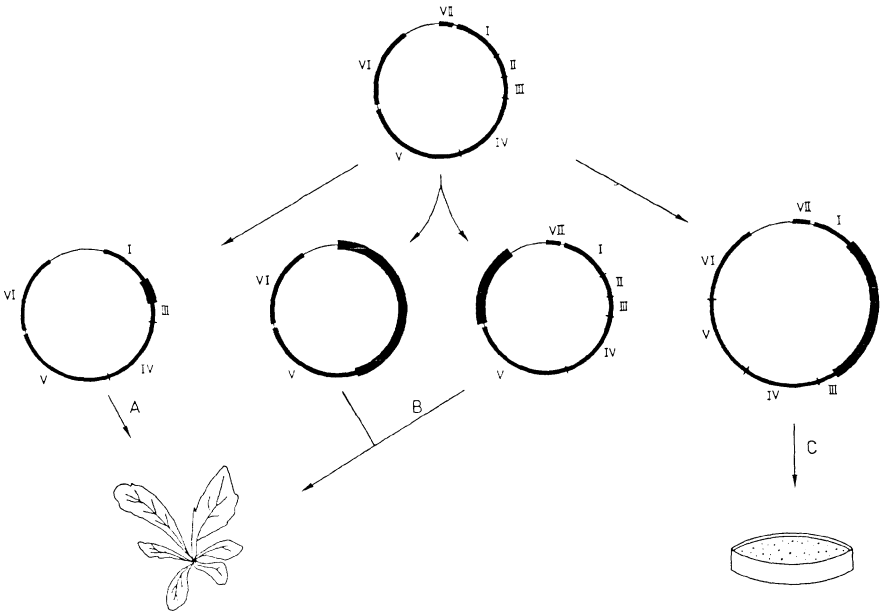


Fig.7
Strategies to construct and use a plant vector. A) Replacement of non-essential regions by gene of interest (transforming virus). B) Complementation of two mutant viruses with DNA replacements in various regions (helper approach). C) Infection of protoplasts using super long hybrid virus DNA. Bold lines virus open reading frames; double bold lines payload DNA.

not, indicating a size limitation for the insertion of foreign DNA into CaMV.

Additional space can be provided by deleting unnecessary parts of the virus genome, for example ORF II (see section III-D-2) and ORF VII (87,94) and thereby space is provided for a "payload" totalling 800-1000 bp. Exploration to discover dispensable DNA stretches in the intergenomic region is in progress and might lead to deletion of a further 100 bp.

c) Conservation of Replication functions

In addition to any gene product required for the genome replication, replication signal sequences must also be preserved. According to the reverse transcription model, these signals are the promoter for the large transcript, the attenuated transcription stop signal and primer binding sites for minus and plus strand DNA synthesis. It is expected that DNA inserts containing additional transcription stop sequences would interfere with production of the large transcript and therefore interfere with the replication cycle. Additional promoters, on the other hand, might be permitted. Systematic studies to confirm these limitations are under way.

d) No Intereference with Translation

Close packing of the ORFs might be an essential feature for translation. A 351 bp fragment, derived from the galactosidase gene of E.coli, was inserted in sense and antisense directions into the XhoI site in ORF II of CM1841 (68).The fate of the inoculated DNA was analysed by recloning the viral DNA and sequencing the region surrounding the insert. Inserted DNA in the sense orientation was rapidly excised from more than 95% of the viral DNA population. Sequencing revealed that the excision took place at a 12 bp region of homology between a site near the 3' end of the inserted DNA and a site at the 5' end of ORF II of the viral DNA.

When the insert was in opposite orientation translation of the modified ORF II would stop at nonsense codons occurring in the insert. In this case infection symptoms developed only after an eclipse period of 2 to 3 months. During this eclipse period the progeny of the hybrid CaMV had all acquired different deletions, sharing one feature in common: they only remained stable when translation of region II could proceed to the start of ORF III uninterrupted by stop codons. Based on these and other results obtained from "linker" insertions and in vivo-constructed deletions of CaMV (see section III-D), Sieg and Gronenborn (68) suggested that CaMV encodes a polycistronic mRNA. The ribosome after passing the terminator codon of an ORF, does not immediately dissociate from the mRNA but can initiate synthesis of a new polypeptide chain, provided an AUG codon occurs in the close vicinity of the previous stop codon. The same ribosome may thus be passed from one ORF to the next (relay race model), the smaller the intergenic space between the reading frames the more efficient the linked translation would be. ORF II can be deleted to the ATG initiator

codon of ORF II arguing against the presence of ribosome binding sites preceding each ORF (68).

Daubert et al. (65) have also reported difficulties with insertions of larger DNA fragments into ORF II. Inserts of 65 and 265 bp within ORF II were not infective. Since these insertions alter the reading frame these results can once again be explained by the relay race model of translation. The validity of the relay race model for other regions of the CaMV genome has still to be shown.

Taking into account all these limitations hybrid CaMV molecules are being constructed (91) in which excised ORF II and VII are replaced by exactly tailored coding regions of bacterial selective markers such as the aminoglycosid phosphotransferase conferring kanamycin resistance from Tn5 (95), the methotrexate - insensitive dihydrofolate reductase from R67 (96,97) and chloramphenicol acetyltransferase, conferring chloramphenicol resistance from plasmid pBR325 (98,99).

2. Helper Approach (Fig. 7B)

Since space available for "payload" DNA on a single CaMV DNA molecule is small, one strategy is to divide the necessary CaMV genes into two DNA molecules and thus create a multipartite or helper system. Many plant viruses (e.g. gemini viruses; see Introduction) have their genomes naturally distributed on two or several separately packaged nucleic acid molecules and therefore the construction of an artificial bipartite CaMV systems seems plausible. Signal sequences must be present on both molecules but a helper system would still provide space for 6000 additional base-pairs.

Preliminary experiments were performed using pairs of different lethal CaMV mutants as inoculum. Infectivity could be restored but analysis of progeny virus revealed that this infectivity was due to recombination of defective mutants rather than complementation (80,81). Recombination could theoretically occur by crossover events or by a template switch of the nascent DNA minus strands during reverse transcription (Figs.5,6; Table 2). Since both DNA molecules of a helper system must share similar signal sequences and these can serve as recombination targets development of a helper system is very difficult. Final solutions may make use of overlapping deletions, or pairs of only distantly related viruses with little DNA homology or of genomes with a scrambled order of genes.

3. Infection of protoplasts (Fig. 7C)

Many of the CaMV functions might only be required for the systemic spreading of the disease through the plant and not for infection of a single cell. For example, packaging of the virus genome might be unnecessary for replication of the virus within the plant cell. If so the capsid protein gene (ORF IV) could be deleted and the size limits imposed for packaging of the virus genome in particles be ignored. Experiments are under way in several laboratories to develop efficient plant protoplast transformation systems. Transformation with naked DNA (100,101,102), fusion of protoplasts with bacterial spheroplasts containing CaMV hybrid plasmids (103) or infection of plant protoplasts with agrobacterium containing the CaMV information within the T stretch of their ti plasmid are possible routes.

F. CONCLUSION

Cauliflower mosaic virus is studied for its interesting life cycle, as a tool to learn about molecular events in the plant cell and as a potential vector for the introduction of foreign DNA into plant cells. Progress in any of these directions also advances knowledge in the others. Thus the reverse transcription model for viral replication defines the CaMV replicator which would be needed in a vector; attempts to introduce heterologous DNA sequences into the ORF II region led to the relay race model for translation; the separation of different stages of the CaMV life cycle between nucleus and cytoplasm indicates methods of information exchange between these cellular components. Both reverse transcription and relay race translation might have parallels within the healthy plant cell.

At this stage a CaMV plant vector would best be suited to accumulate and test certain gene products in the plant. CaMV spreads quickly to a large proportion of the plant cells where it is present at high copy number and vigorously expressed. These properties could allow study of the effect of certain gene products in the plant cell as well as harvest of high quantities of these products from infected plants. Stable transformation of plants and production of modified seeds, on the other hand, might presently be more easily achieved by use of the integrative ti plasmid vector.

ACKNOWLEDGMENTS

We thank Chris Rowland for extensive editing of this manuscript, Barbara Hohn, Pierre Pfeiffer, Genevieve Lebeurier, Normand Brisson, Maciej Pietrzak, Bruno Gronenborn, Jurek Paszkowski and Isabelle Koenig for many discussions, Josianne Ménissier for providing picture material and many colleagues for contributing unpublished information as indicated.

REFERENCES

1. Haber S., M. Ikegami, N.B. Bajet, and R.M. Goodman. Evidence for a divided genome in bean golden mosaic virus, a geminivirus. Nature 289 (1980) 324-326
2. Bisaro D.M., W.D.O. Hamilton, R.H.A. Coutts, and K.W. Buck. Molecular cloning and characterization of the two DNA components of tomato golden mosaic virus. Nucleic Acids Res. 10 (1982) 4913-4922
3. Hamilton W.D.O., D.M. Bisaro, and K.W. Buck. Identification of novel DNA forms in tomato golden mosaic virus infected tissue. Evidence for a two component viral genome. Nucleic Acids Res. 10 (1982) 4901-4912
4. Stanley J. and M.R. Gay. Nucleotide sequence of cassava latent virus DNA. Nature 301 (1983) 260-262
5. Marriott T.W. and R.H. Symons, pers. comm.; see Robertson et al.,(eds) "Plant infectious agents", Cold Spring Harbor, 1983
6. Stanley J. Infectivity of the cloned geminivirus genome requires sequences from both DNAs. Nature 305 (1983) 643-645
7. Hohn B. and T. Hohn. Cauliflower mosaic virus: a potential vector for plant genetic engineering. Mol. Biol. Plant Tumors (1982) 549-560, Academic Press, New York
8. Hohn T., K. Richards and G. Lebeurier. Curr. Top. Microbiol. Immunol. 96 (1982) 193-236
9. Howell S.H. Plant molecular vehicles: potential vectors for introducing foreign genes into plants. Ann. Rev. Plant Physiol. 33 (1982) 609-650
10. Shepherd R.J. DNA plant viruses. Ann. Rev. Plant Physiol. 30 (1979) 405-423
11. Lung M.C.J. and T.P. Pirone. Studies on the reason for differential transmissibility of cauliflower mosaic virus isolates by aphids. Phytopathology 63 (1973) 910-914
12. Lung M.C.J. and T.P. Pirone. Acquisition factor required for aphid transmission of purified cauliflower mosaic virus. Virology 60 (1974) 260-264
13. Chauvin C., B. Jacrot, G. Lebeurier and L. Hirth. The structure of cauliflower mosaic virus. A neutron diffraction study. Virology 96 (1979) 640-664
14. Xiong C., E. Balazs, G. Lebeurier, C. Hindelang, M. Stoeckel and A. Porte. Comparative cytology of two isolates of cauliflower mosaic virus. J. Gen. Virol. 61 (1982) 75-81
15. Givord L., C. Xiong, M. Giband, I. Koenig, T. Hohn, G. Lebeurier and L. Hirth. A second cauliflower mosaic virus gene product influences the structure of the viral inclusion body. EMBO J. (1984) in press
16. Yamaoka N. T. Morita, I. Furusawa and M. Yamamoto. Effect of temperature on the multiplication of cauliflower mosaic virus. J. gen. Virol. 61 (1982) 283-287
17. Franck A., H. Guilley, G. Jonard, K. Richards, and L. Hirth. Nucleotide sequence of cauliflower mosaic virus DNA. Cell 21 (1980) 285-294

18. Gardner R.C., A.J. Howarth, P. Hahn, M. Brown-Leudi, R.J. Shepherd and J. Messing. The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. Nucleic Acids Res. 9 (1981) 2871-2888
19. Balazs E., H. Guilley, G. Jonard and K. Richards. Nucleotide sequence of DNA from an altered-virulence isolate D/H of the cauliflower mosaic virus. Gene 19 (1982) 239-249
20. Hohn T., B. Hohn, A. Lesot, and G. Lebeurier. Restriction map of native and cloned cauliflower mosaic virus DNA. Gene 11 (1980) 21-31
21. Howarth A.J., R.C. Gardner, J. Messing and R.J. Shepherd. Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. Virology 112 (1981) 678-685
22. Volovitch M., N. Modjtahedi, Y. Chouikh and P. Yot. Conserved and non-conserved restriction sites are interspersed in the DNA of cauliflower mosaic virus. Biochem. Int. 3 (1981) 225-232
23. Volovitch M.D., G. Drugeon, J.P. Dumas, A.L. Haenni and P. Yot. A restriction map of cauliflower mosaic virus DNA strain PV-147. Mapping of the cleavage sites of HhaI, SacI, AvaI, PvuII, PstI, XbaI, EcoRI, BgII, HincII, HpaII and HindII plus III. Eur. J. Biochem. 100 (1979) 245-256
24. Hull R. and S.H. Howell. Structure of the cauliflower mosaic virus genome. Part 2: Variation in DNA structure and sequence between isolates. Virology 86 (1978) 482-493
25. Hull R. Structure of the cauliflower mosaic virus genome. Part 3: Restriction endonuclease mapping of 33 isolates. Virology 100 (1980) 76-90
26. Richins R.D. and R.J. Shepherd. Physical maps of the genomes of dahlia mosaic virus and mirabilis mosaic virus - two members of the caulimovirus group. Virology 124 (1983) 208-214
27. Hull R and J. Donson. Physical mapping of the DNAs of carnation etched ring and figwort mosaic virus. J. gen. Virol. 60 (1982) 125-134
28. Volovitch M., G. Drugeon, P. Yot. Studies on the single-stranded discontinuities of the cauliflower mosaic virus genome. Nucleic Acid Res 5 (1978) 2913-2925
29. Richards K.E., H. Guilley and G. Jonard. Further characterization of the discontinuities in cauliflower mosaic virus DNA. FEBS Lett. 134 (1981) 67-70
30. Guilley H. K.E. Richards and G. Jonard. Observations concerning the discontinuous DNAs of cauliflower mosaic virus. Embo J. 2 (1983) 277-282
31. Pfeiffer P and T. Hohn. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. Cell 33 (1983) 781-789
32. Hull R., S.N. Covey, J. Stanley and J.W. Davies. The polarity of the cauliflower mosaic virus genome. Nucleic Acid Res. 7 (1979) 669-677
33. Al Ani R., P. Pfeiffer, O. Whitechurch, A. Lesot, G. Lebeurier and L. Hirth. A virus specified protein produced upon infection by cauliflower mosaic virus. Ann. Virol. (Inst. Pasteur) 131 (1980) 33-53
34. Odell J.T. and S.H. Howell. The identification mapping and characterization of messenger RNA for p-66 a cauliflower mosaic virus coded protein. Virology 102 (1980) 349-359
35. Guilfoyle T.J. Transcription of the cauliflower mosaic virus genome in isolated nuclei from turnip leaves. Virology 107 (1980) 71-80
36. Covey S.N., G.P. Lomonossoff and R. Hull. Characterization of cauliflower mosaic virus DNA sequences which encode major polyadenylated

transcripts. Nucleic Acids Res. 9 (1981) 6735-6747
37. Dudley R.K., J.T. Odell and S.H. Howell. Structure and 5'-termini of the large and 19 S RNA transcripts encoded by the cauliflower mosaic virus genome. Virology 117 (1982) 19-28
38. Guilley H., R.K. Dudley, G. Jonard, E. Balazs, and K.E. Richards. Transcription of cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. Cell 30 (1982) 763-773
39. Covey S.N. and R. Hull. Transcription of cauliflower mosaic virus DNA. Detection of transcripts, properties, and location of the gene encoding the virus inclusion body protein. Virology 111 (1981) 463-474
40. Xiong C., S. Muller, G. Lebeurier and L. Hirth. Identification by immunoprecipitation of cauliflower mosaic virus in vitro major translation product with a specific serum against viroplasm protein. Embo J. 1 (1982) 971-976
41. Odell J.T., R.K. Dudley, and S.H. Howell. Structure of the 19 S RNA transcript encoded by the cauliflower mosaic virus genome. Virology 111 (1981) 377-385
42. Condit C., T.J. Hagen, T.D. McKnight and R.B. Meagher. Character ization and preliminary mapping of cauliflower mosaic virus transcripts. Gene 25 (1983) 101-108
43. Condit C. and R.B. Meagher. Multiple, discrete 35S transcripts of cauliflower mosaic virus J. Mol. Appl. Gen. 2 (1983) 301-314
44. Ménissier J., G. Lebeurier and L. Hirth. Free cauliflower mosaic virus supercoiled DNA in infected plants. Virology 117 (1982) 322-328
45. Olszewski N., G. Hagen and T.J. Guilfoyle. A transcriptionally active, covalently closed minichromosome of cauliflower mosaic virus DNA isolated from infected turnip leaves. Cell 29 (1982) 395-402
46. Ménissier J., G. de Murcia, G. Lebeurier and L. Hirth. Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA: knotted encapsidated DNA and nuclear minichromosome. EMBO J. 2 (1983) 1067-71
46a. Olszewski N., T. Guilfoyle. Nuclei purified from cauliflower mosaic virus-infected turnip leaves contain subgenomic, covalently closed circular cauliflower mosaic virus DNAs. Nucleic Acids Res. 11 (1983) 8961-9014
47. Kamei T., M. Rubio-Huertos and C. Matsui. Thymidine-^{3}H uptake by X-bodies associated with cauliflower mosaic virus infection. Virology 37 (1969) 506-508
48. Favali M.A., M. Bassi and G.G. Conti. A quantitative autoradiographic study of intracellular sites for replication of cauliflower mosaic virus. Virology 53 (1973) 115-119
49. Ansa O.A., J.W. Bowyer and R.J. Shepherd. Evidence for replication of cauliflower mosaic virus DNA in plant nuclei. Virology 121 (1982) 147-156
50. Hull R. and S. Covey. Characterization of cauliflower mosaic virus DNA forms isolated from infected turnip leaves. Nucleic Acids Res. 11 (1983) 1881-1895
51. Hohn T., M. Pietrzak, L. Dixon, I. Koenig, J. Penswick, B. Hohn and P. Pfeiffer. Involvement of reverse transcription in cauliflower mosaic virus replication. Plant Molec. Biol. R.B. Goldberg (ed.) Alan R. Liss; New York (1983) p. 147-155.
52. Hull R. and S.N. Covey. Does cauliflower mosaic virus replicate by reverse transcription? TIBS (1983) 119-121.//Replication of cauliflower mosaic virus DNA. Sci.Prog.Oxf. 68 (1983) 403-422.
53. Pfeiffer P., P. Laquel and T. Hohn. Search for the enzyme responsible for the reverse transcription step in cauliflower mosaic virus replication. In V. Hübscher and S. Spadari (eds.): Proteins involved in

DNA replication, Plenum Press, London, in press
54. Volovitch M., N. Modjtahedi, P. Yot and G. Brun. RNA-dependent DNA polymerase activity in cauliflower mosaic virus infected plant leaves. EMBO J. (1984) 309-314.
55. Modjtahedi N., M. Volovitch, L. Sossountzov, Y. Habricot, J.M. Bonne ville and P. Yot. Cauliflower mosaic virus-induced viroplasms support viral DNA synthesis in a cell-free system. Virology (1984), in press
55a Pfeiffer, P., P. Laquel and T. Hohn. Cauliflower mosaic virus replication complexes: characterization of the associated enzymes and of the polarity of the DNA synthesized in vitro. Plant Mol. Biol. 3 (1984) in press56.
56. Covey S.N., D. Turner and G. Mulder. A small DNA molecule containing covalently-lined ribonucleotides originates from the large intergenic region of the cauliflower mosaic virus genome. Nucleic Acids Res. 11 (1983) 251-264
56a Turner, D.S. and Covey, S.N. A putative primer for the replication of cauliflower mosaic virus by reverse transcription is virion-associated. FEBS letters 165 (1984) 285-289.
57. Al Ani R., P. Pfeiffer and G. Lebeurier. The structure of cauliflower mosaic virus. Part 2: identity and location of the viral polypeptides. Virology 93 (1979) 188-197
58. Tezuka, N. and T. Taniguchi. Effects of some chemicals on the multiplication of cauliflower mosaic virus. Ann. Phytopath. Soc. Japan 39, (1973) 259-265
59. Kelley D.C., V. Cooper and D.G.A. Walkey. Cauliflower mosaic virus structural proteins. Microbios 10 (1974) 239-245
60. Brunt A.A., R.J. Barton, J.H. Tremaine and R. Stace-Smith. The composition of cauliflower mosaic virus protein. J. Gen. Virol 27 (1979) 640-641
61. Hull, R., R.J. Shepherd and R.D. Harvey. Cauliflower mosaic virus; an improved purification procedure and some properties of the virus particles. J. Gen. Virol 31 (1976) 93-100
62. Daubert S., R. Richins, R.J. Shepherd, R.C. Gardner. Mapping of the coat protein gene of cauliflower mosaic virus by its expression in a prokaryotic system. Virology 122 (1982) 444-449
63. Hahn P. and R.J. Shepherd. Evidence for a 58-kilodalton polypeptide as precursor of the coat protein of cauliflower mosaic virus. Virology 116 (1982) 480-488
64. Woolston, C.J., S.N. Covey, J.R. Penswick and J.W. Davies. Aphid transmission and a polypeptide are specified by a defined region of the cauliflower mosaic virus genome. Gene 23 (1983) 15-23
65. Daubert S., R.J. Shepherd and R.C. Gardner. Insertional mutagenesis of the cauliflower mosaic virus genome. Gene 25 (1983) 201-208
66. Armour S.L., U. Melcher, T.P. Pirone, D.J. Lyttle and R.C. Essenberg. Helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. Virology 129 (1983) 25-30
67. Toh H., H. Hayashida and T. Miyata. Sequence homology between retro viral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature 305 (1983) 827-829
68. Sieg K. and B. Gronenborn. Evidence for polycistronic messenger RNA encoded by cauliflower mosaic virus. In prep.
69. Szeto W.W., D.H. Hamer, P.S. Carlson and C.A. Thomas, Jr. Cloning of cauliflower mosaic virus DNA in Escherichia-coli. Science 196 (1977) 210-212
70. Leubeurier G., L. Hirth, T. Hohn and B. Hohn. Infectivities of native and cloned cauliflower mosaic virus DNA. Gene 12 (1980) 139-146
71. Howell S.H., L.L. Walker and R.K. Dudley. Cloned cauliflower mosaic virus DNA infects turnips brassica-rapa. Science 208 (1980) 1265-1267

72. Meagher R.B., R.J. Shepherd and H.W. Boyer. The structure of cauliflower mosaic virus. Part 1: restriction endonuclease map of cauliflower mosaic virus DNA. Virology 80 (1977) 362-375
73. Howarth A.J., R.C. Gardner, J. Messing and R.J. Shepherd. Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. Virology 112 (1981) 678-685
74. Delseny M. and R. Hull. Isolation and characterization of faithful and altered clones of the genomes of cauliflower mosaic virus isolates Cabb B-JI, CM4-184, and Bari I. Plasmid 9 (1983) 31-41
75. Covarrubias L. Cervantes, A. Covarrubias, X. Soberon, I. Vichido, A. Blanco, Y.M. Kupersztoch-Portnoy and F. Bolivar. Construction and characterization of new cloning vehicles. V. Mobilization and coding properties of pBR322 and several deletion derivatives including pBR327 and pBR328. Gene 13 (1981) 25-35
76. Twigg A.J. and D. Sherratt. Trans-complementable copy number mutants of plasmid col El. Nature 283 (1980) 216-218
77. Vieira J. and J. Messing. The pUC plasmids, a M13mp7-derived system for insertion mutagenesis and sequencing with synthetic primers. Gene 19 (1982) 259-268
78. Hull R. and J.W. Davies. Genetic engineering with plant viruses and their potentials as vectors. Advances in Virus Research 28 (1983) 1-33
79. Hull R. and R.J. Shepherd. The structure of cauliflower mosaic virus genome. Virology 79 (1977) 216-230
80. Lebeurier G. L. Hirth, B. Hohn and T. Hohn. In vivo recombination of cauliflower mosaic virus DNA. Proc. Natl. Acad. Sci. USA 79 (1982) 2932-2936
81. Howell S.H., L.L. Walker, and R.M. Walden. Rescue of in vitro generated mutants of cloned cauliflower mosaic virus genome in infected plants. Nature 293 (1981) 483-486
82. Pomerantz B.J., M. Naujokas and J.A. Hassell. Homologous recombination between transfected DNAs. Mol. Cell. Biol. 3 (1983) 1680-1685
83. Walden R.M. and S.H. Howell. Intergenomic recombination events among pairs of defective cauliflower mosaic virus genomes. J. Mol. Appl. Genet. 1 (1982) 447-456
84. Walden R.M. and S.H. Howell. Uncut recombinant plasmids bearing nested cauliflower mosaic virus genomes infect plants by intragenomic recombination. Plant Mol. Biol. 2 (1983) 927-931
85 Schaffner W. Direct transfer of cloned genes from bacteria to mammalian cells. Proc Natl Acad Sci USA 77 (1980) 2163-2167
86. Fried M., B. Klein, K. Murray, P. Greenaway, J. Tooze, W. Boll and C. Weissmann. Infectivity in mouse fibroblasts of polyoma DNA integrated into plasmid pBR322 or lambdoid phage DNA. Nature 279 (1979) 811-816
87. Dixon L.K., I. Koenig and T. Hohn. Mutagenesis of cauliflower mosaic virus. Gene 25 (1983) 189-199
88. Gronenborn B., R.C. Gardner, S. Schaefer and R.J. Shepherd. Propagation of foreign DNA in plants using cauliflower mosaic virus as vector. Nature 294 (1981) 773-776
89. Shortle D. and D. Nathans. Local mutagenesis: A method for generating viral mutants with base substitutions in pre-selected regions of the viral genome. Proc. Natl. Acad. Sci. USA 75 (1978) 2170-2174
90. Koenig I. and T. Hohn, unpublished results
91. Brisson N., J. Paszkowski, J. Penswick, B. Gronenborn, I. Potrykus and T. Hohn. Expression of a bacterial gene in plants using cauliflower mosaic virus as vector. Nature, in press.

92. Hohn, T., B. Hohn and G. Lebeurier, unpublished results.
93. Pietrzak, M., unpublished results.
94. Dixon L. and T. Hohn. The presumptive start of relay race translation in cauliflower mosaic virus. In prep.
95. Davies D. and D.I. Smith. Plasmid-determined resistance to antimicrobial agents. Ann. Rev. Microbiol. 32 (1978) 469-518
96. Fling M.E. and L.P. Elwell. Protein expression in Escherichia coli minicells containing recombinant plasmids specifying trimethoprim-resistant dihydrofolate reductases. J. Bacteriol. 141 (1980) 779-785
97. O'Hare K. C. Benoist and B. Breathnach. Transformation of mouse fibroblasts to methotrexate resistance by a recombinant plasmid expressing a prokaryotic dihydrofolate reductase. Science 209 (1980) 1414-1421
98. Gorman C.M., Moffat, L.F. and B.H. Howard. Recombinant Genomes which express chloramphenicl acetyltransferase in mammalian cells. Mol. Cell. Biol. 2 (1982) 1044-1051
99. Herrera-Estrella L., A. Depicker, M. Van Montagu and J. Schell. Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. Nature 303 (1983 209-213
100 Maule A.J. Infection of protoplasts from several Brassica species with cauliflower mosaic virus following inoculation using polyethylene glycol. In prep.
101 Furusawa I., N. Yamaoka, T. Okuno, M. Yamamoto, M. Kohno and H. Kunoh. Infection of turnip brassica-rapa cultivar perviribis protoplasts with cauliflower mosaic virus. J. Gen. Virol. 48 (1980) 431-436
102 Howell S.A. and R. Hull. Replication of cauliflower mosaic virus and transcription of its genome in turnip leaf protoplasts. Virology 86 (1978) 468-481
103 Tanaka N., M. Ikegami, T. Hohn, C. Matsui and I. Watanabe. E. coli spheroplast-mediated transfer of cloned cauliflower mosaic virus DNA into plant protoplasts. MGG (1984) in press.

14

CLONING OF PLANT-VIRUS GENOMES OTHER THAN THAT OF CAULIFLOWER MOSAIC VIRUS

Léon HIRTH
Laboratoire de Virologie, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue Descartes, 67000 Strasbourg, France

INTRODUCTION

The cloning of plant-virus genomes in bacteria has for several years been fairly widely used to solve problems at the molecular level, notably for determining the nucleotide sequences of DNA and RNA viruses. There are many reasons for wanting to know the complete sequence of a plant-virus genome. The most important is probably to determine the primary structure of the proteins (other than the coat protein) involved, in ways that are at present unknown, in replication of the nucleic acid, generation of symptoms, transmission by insects, and so on.

In addition, a knowledge of the structure and the function of the various genes of a virus may be important in view of the possibility of using viruses as gene vectors in higher plants. Although the Ti plasmid can already be used to introduce some genes into plants (1), its use is limited to dicotyledons, and Ti is an integrated vector. Viruses could be used as non-integrated gene vectors in both monocotyledonous and dicotyledonous plants.

A knowledge of sequences allows sequence comparisons, which can shed some light on yet another matter, the function of genes. For example, the sequence analogy between genes coding for reverse transcriptase of duck hepatitis virus, avian leukemia virus (2), and the open reading frame V (ORF V) of cauliflower mosaic virus (CaMV) raises questions about the origin and significance of reverse transcriptase and of the relationships between retroviruses and the cell genome.

One of the most interesting prospects is the possibility, just mentioned, of using plant viruses as gene vectors. Dixon and Hohn, elsewhere in this volume (3) discuss in detail the introduction of foreign genes by means of a DNA segment of cauliflower mosaic virus (a plant DNA virus) and at present several laboratories are mounting projects to use RNA viruses as gene vectors. Such studies would require

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

at least that experiments be performed like those reported by Racaniello and Baltimore (4) with poliovirus cDNA, involving the transcription of the corresponding viral double-stranded cDNA matrix and the translation of the transcript.

By introducing a cDNA copy of a plant-virus RNA into a bacterial plasmid, one could, by means of hybridization, detect small quantities of viral RNA, and thus a collection of cDNA clones might be used to detect and identify viruses, especially those which are weak antigens. Viroids and virusoids will be easily detectable by DNA hybridization -- a subject raised in Dr. Owen's chapter.

In this chapter I describe briefly the main results that have been obtained from the cloning of plant-virus RNA and DNA genomes in bacteria. The customary techniques for introducing foreign DNA into bacterial plasmid viruses are used in this work, often without modification ; these techniques are described in detail elsewhere (5). I mention only in passing details of the introduction of the genome or DNA copy of the genome into bacterial plasmids or bacterial viruses.

I. Geminiviruses

Geminiviruses consist of two joined viral capsids containing two different single-stranded DNA molecules that have different coding capacities. These two molecules are located in the plant nucleus. They were discovered in 1976 by Matyis et al. (6) in beans grown in tropical regions of the Americas, and the first one was named bean golden mosaic virus (BGMV). Two main properties characterize this virus. One is the presence of paired particles in viral preparations: it seems (7) that the paired particles consist of two incomplete icosahedral particles (T = 1), each missing a pentamer. The second characteristic is that each half-particle contains a circular single-stranded DNA molecule.

Other geminiviruses have been discovered [for a review see Goodman (8)], and viral dsDNA molecules of some of them have been cloned.

In the case of tomato golden mosaic virus (TGMV), Cloning of viral DNA has been very useful for the determination and isolation of various species of viral DNA. The clone was obtained from a viral dsDNA isolated from infected plants (9,10). Indeed it seems that the replication of DNA of this virus takes place through a double-stranded DNA form. Digestion of viral dsDNA by Eco RI restriction enzyme allows DNA molecules to be inserted into the plasmid pAT 13 replicating in Escherichia coli HB 101.

Transformation and isolation of transformants containing the recombinant plasmid are performed according to the customary methods. The result of the digestion by Bam HI of many clones suggests the existence of two types of inserts. Thus the viral dsDNA is made up of two types of molecules, A and B, which have different restriction maps and thus different sequences (Fig. 1).

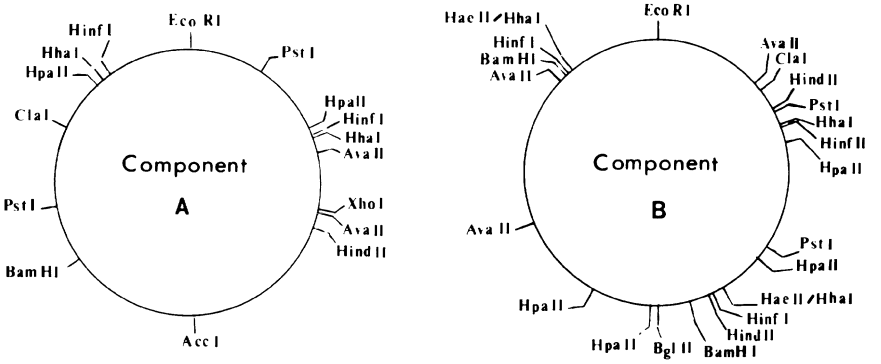


Figure 1. Restriction maps of cloned TGMV components.
It is clear that components A and B have different compositions and likely code for different proteins.

Molecular cloning in E. coli made it possible to determine the restriction maps of both the A and the B components of TGMV (11). From the two types of clones obtained it has been demonstrated that component A (in its double-stranded form) is 75 base-pairs longer than B. Recent data (10) have shown that there is some sequence homology between the two components but that they contain two different types of genetic information. Cloning of TGMV component A in pBH 404 preserves S1 nuclease-sensitive sites present in isolated viral double-stranded DNA ; this suggests the existence of stable cruciform structures in non-cloned supercoiled molecules of TGMV DNA. A and B clones may be used as DNA probes to detect various forms of viral DNA. Thus single-stranded DNA species have been detected that run faster than viral single-stranded DNA on agarose gel . The origin and role of such DNA molecules are unknown. Recently, hybridization experiments have shown that a probe corresponding to clone A has some homology with the DNA of cassava latent virus (CLV),

a geminivirus that infects cassava plants. The homology seems to be restricted to four relatively short regions of DNA probe A and CLV component 1. No homologies have been observed between probe A and component 2 of CLV.

A similar type of research has been performed by Marriott and Symons (12) with chlorosis striate mosaic virus (CSMV) isolated by Francki et al. (13). This virus has been sequenced by means of restriction fragments cloned in phage M13 mpq[1]. So far only four segments have been sequenced, and no overlaps between the four clones have been obtained; thus the complete sequence is not known. On the other hand, restriction maps of single-stranded DNA and of double-stranded DNA both digested by the same enzymes (MSPE, Fnu DII, Hae III, and Eco RI) give rise to the same results, which are consistent with a size of 2500 residues. These results also indicate that only one molecule of DNA is present in a viral preparation of CSMV; this contrasts with the structure of CLV, BGMV, and TGMV, which contain two DNA components. It is worthy to note that the complete sequence of CLV is in good agreement with those results (14).

Recent analysis of tobacco yellow dwarf virus (TYDV) has also shown the presence of only one species of single-stranded DNA, also of about 2500 residues (12). Thus two types of geminivirus exist in nature: one (CLV type) contains two species of DNA which code for several putative proteins using reading frames of both the plus and the minus DNA strands; the other (CSMV type) contains only one species of DNA, which codes for a small number of peptides. The significance of this difference in coding capacity may lie in the fact that transmission is very different in the two groups, being by insects in one and mechanical means in the other.

---

1. M13 mpq is a helical single-stranded DNA phage that infects E. coli without destroying it. The double-stranded replicative form can be used to introduce viral ds-DNA, and M13 phage particles with single-stranded DNA containing single-stranded viral insert DNA are extruded into the surrounding medium by the bacterium: by primer elongation (from universal primer) the complementary foreign DNA is synthesized and can be sequenced by the addition of labelled ddN (ATC and G) to the medium (for details see 5).

From the preceding results it is clear that cloning of viral DNA in bacteria has been very useful in determining the structure of the viral components and also their sequence. In the case of DNA viruses (for CaMV, see Dixon and Hohn's chapter) cloning is also useful because it provides relatively large quantities of intact or fragmented DNA for rapid sequencing and sequence comparison. Also, cloning will be useful for amplifying artificial mutants obtained by changing base pairs or by deletions allowing putative gene functions to be identified. Furthermore, in the case of the low-yielding geminiviruses, cloning allows the production of enough viral DNA to serve as a probe for the detection of viruses in plants. Table 1 shows the various geminiviruses that have been cloned and sequenced.

II. RNA viruses

Sequencing of viral RNA is also very useful. The nucleotide sequences reveal new functions or properties of the virus. However, in many cases, direct RNA sequencing procedures are difficult or impossible to carry out, especially when the viral RNA is very long. In the case of bacteriophage R17 for example (15) the complete sequence was obtained by the customary methods of RNA sequencing, but much time was spent on this determination. Recently the use of DNA recombinant technology has allowed the complete sequencing of several RNA plant viruses belonging to various groups, both single-component and multi-component. I describe the general rules of the techniques used in a limited number of examples.

A) Tobacco Mosaic Virus (TMV)

In the case of TMV, I would like to describe the strategy of sequencing more than the classical technology itself.

TMV RNA has no poly-A-terminal sequence and its sequence has been obtained (16) by synthesizing four to seven oligonucleotides that hybridize randomly to many regions of TMV RNA. These oligonucleotides acted as a primer and were elongated by reverse transcription. Overlaps of the sequences of the various clones were determined by computer. As some terminal regions do not give rise to cDNA samples, specific primers of 13 to 17 oligonucleotides were synthesized, corresponding to the complements of the 5' and 3' nucleotide sequences already determined. In many cases the RNA/DNA complex was digested by RNase H and the second

Table 1. Viruses for which the complete sequence has been established.

DNA viruses

Virus	Strain	Nucleotide number (base pairs)
Double-strand		
Cauliflower mosaic virus	Strasbourg strain (Cabb-S)	8024 (52)
	CM1841	8031 (53)
	DH strain	8016 (54)
Single-strand		
Geminiviruses		
Cassava latent virus		
DNA 1		2770 (14)
DNA 2		2720 (14)

RNA viruses

Monopartite genome		
Tobacco mosaic virus	vulgare	6937 (16)
Multipartite genome		
Brome mosaic virus	vulgare	
RNA 1		3234 (55)
RNA 2		2865 (55)
RNA 3		2117 (56)
Alfalfa mosaic virus		
RNA 1	strain 425	3644 (57)
RNA 2	"	3533 (58)
RNA 3	"	2037 (59)
RNA 3S	Strasbourg strain	2045 (Pinck et al., to be published)
Cowpea mosaic virus	vulgare	
RNA B		5889 (60)
RNA M		3425 (19)
Satellites:		
S. Tomato black ring virus	Scottish strain	1375 (17)
(n) S. Cucumber mosaic virus	Carna 5 (1)*	335 (22)
S. Tobacco necrosis virus		1240 (47)

*. Sequence of (1) Carna 5 has been determined and shows only small differences with n Carna 5 (61).

strand was synthesized by the addition of oligonucleotide primers and use of Klenow DNA polymerase in the presence of the four deoxyribonucleotide triphosphates.

The double-stranded DNA was digested with various restriction enzymes and the resulting fragments were inserted into the replicative form of phage M13 by ligation with Bam HI linkers. The various clones were detected by plaque hybridization to ^{32}P-labelled TMV cDNA clones or labelled TMV RNA oligonucleotides. Overlapping clones have been determined by restriction digest.

Another possibility is to tail the 3' end of the RNA by means of poly A polymerase and then to use techniques similar to those applied for polyadenylated RNAs (5).

B) Tomato black ring satellite (17)

In the case of the satellite, which is polyadenylated, a mixture of it and oligo-dT is transcribed with reverse transcriptase. Incubation of the resulting mixture with RNase H and the Klenow fragment of E. coli polymerase I allows the synthesis of a double-stranded viral DNA copy (18). This DNA is digested with S1 nuclease and the resulting double-stranded DNA is tailed by the addition of a dC residue. The tailed ds-DNA is inserted into the Pst I site of a dG tailed pBR322, and colonies are isolated by the usual means of ampicillin sensitivity and tetracycline resistance (5). The clones are amplified and plasmid DNA is isolated, the foreign DNA is cut with Pst I and separated by electrophoresis on agarose gel, and the fragments are sequenced. Northern hybridization with nick-translated hybrid-plasmid DNA has established the origin of the clones (RNA -1, -2, or -3 fragments). In general a complete cDNA copy of the satellite viral RNA is not obtained and often a polynucleotide chain of varying length is missing at the 5' end and it is necessary to use the method of primer extension : in this case a single-stranded DNA fragment complementary to the 5' end of the sequence part of the RNA is hybridized with the viral RNA and complementary DNA is obtained by primer extension . This DNA can be rendered double stranded and cloned in the bacterial plasmid (in any case the RNA/DNA hybrid can be introduced directly in the plasmid). Sequencing allows the detection of overlapping clones which permit the complete sequence to be established. Very often the correctness of the 5'-terminal sequence is verified by direct RNA sequencing methods.

C) Other viruses

Several RNA plant viruses have been sequenced in this way (see Table 1 which summarizes the characteristics of those that have been completely sequenced). Some viruses have their genetic information in a single RNA molecule, but other viruses have the information in several RNA molecules ; for example cowpea mosaic virus has two so-called genomic RNA molecules, whose complete sequences are known (19,20). In other cases, in addition to the genomic RNAs there are one or more subgenomic RNAs involved in the expression of all the genetic information (as with alfalfa mosaic virus, bromegrass mosaic virus, and cucumber mosaic virus).

Finally, a number of RNA viruses may have one or more satellite RNAs (21), whose complete sequences are known in some cases (22,23). A full knowledge of the sequences of RNA viruses, made possible only by the use of recombinant DNA technology, may lead to various developments, of which I will mention only the main ones.

III. General results from the complete sequences of some RNA viruses

A) The role of proteins

To understand the role of proteins (other than coat proteins) in the multiplication and spread of viruses, one must detect and locate these proteins in infected plants. Obviously, the sequence of a protein can be deduced from that of its gene (in the absence of splicing), and so one could envisage synthesizing peptides corresponding to the hydrophilic regions of this protein, which would lead in turn to production of specific antibodies (24) that react with the whole protein. This approach, recently applied to CaMV, has demonstrated that a putative gene (open reading frame III) does in fact code for a protein that had not been detected by polyacrylamide-gel electrophoresis techniques (25). Some Japanese authors (26) used the sequence of a strain of tobacco mosaic virus (TMV), established by cloning in M13, to construct peptides corresponding to the 30 K protein, to obtain the corresponding antibodies, and then to detect the protein in vivo. This protein was easily observed in the translation products of a 1500-nucleotide subgenomic RNA from the 3' end of TMV RNA.

B) Detection of additional mRNAs

Availability of the complete RNA sequence of TMV indicated that there may also be other, internal initiation regions. Since it is known (27) that messengers can be translated only starting from a free 5'-OH end, a search has been made for other messengers obtained by transcription of a negative strand, starting at the 3' end of the viral RNA and stopping at one of the potential internal initiation sequences. Zelcer et al. (29) reported that TMV contains 3 to 6 viral mRNAs, in addition to those corresponding to its three major proteins (28).

Although the double strands of these mRNAs have been discovered, their origin has remained unclear, because no one has yet been able to sequence them. Several groups (30,31,32,33) used the products of cloning of various fragments of cDNA, derived by copying of pieces from the 3' end of TMV RNA, to identify the additional mRNAs found in preparations of TMV and in RNAs from infected cells (31). Although these authors concluded that there are other messengers, and hence proteins, in addition to those usually described, Zaitlin et al. (34), also using TMV cDNA clones, recently questioned the existence of these substances (RNA and additional protein). They suggested that the substances are artefacts and, in particular, that in the case of experiments carried out with RNA extracted from infected plants, the presence of products of cellular origin provokes artefacts of hybridization. Zaitlin and his co-workers argued that in fact all these observations are artefactual. The question therefore remains open, in particular with regard to the formation of the RNAs coding for 30 K, 29 K, and 23 K proteins, which according to Hunter et al. (31) have a common C-terminal end but different N-terminal ends, corresponding to internal AUG sequences preceded by a "Butler box" (30). Such a superimposition of C-terminal ends has also been suggested (29) for longer mRNAs whose N-termini lie at internal AUG sequences within the viral RNA sequence.

Thus, the use of cDNA copies in the cases of single-component and multi-component RNA viruses has revealed details of genome organization that could not have been observed without the insertion of cDNA fragments of these viruses in bacterial plasmids and their concomitant amplification.

The type of research done with TMV has also been done with other viruses and has given similar results (32) — that is, a much greater number of replicative forms than the single form that would be expected

if 5' ends were the only starting points for translation into the potential proteins encoded by the virus and the corresponding subgenomic RNAs. These results are illustrated by the recent data obtained by Morris (35) indicating that the number of replicative forms is always greater than would have been expected. Dawson and Dodds (32) reported that the replicative forms which correspond to the classic genomic and subgenomic RNAs of TMV and cowpea chlorotic mottle virus became labelled much more rapidly than the additional replicative forms derived from the halt in synthesis of the minus strand at an internal initiation site. The mechanisms of synthesis of the two types of replicative forms are therefore different ; this observation should make it possible to eliminate the hypothesis that the double-stranded RNAs are artefactual. The contradictory results concerning this last question leave open the formulation of hypotheses and design of experiments to confirm the validity of the results obtained.

C) Comparison of sequences

By revealing the complete RNA sequences of plant viruses, the use of recombinant DNA methods has made it possible to compare the sequences from plant viruses with RNA sequences of very diverse origins (Table 1). For this the use of computers is of course essential.

These comparisons have produced results that are sometimes strange. Already sequence analogies between the putative protein encoded by CaMV gene 5 and the reverse transcriptases of Rous sarcoma virus and hepatitis B virus pose the problem of the analogy between some enzymes that participate in the replication of viral DNA in animals and plants (2).

Recently, Franssen et al. (36) showed that there are sequence analogies between the RNAs of cowpea mosaic virus and of the picornaviruses of animals. Once again, these analogies were observed only by means of recombinant DNA techniques, which provided the complete sequence of the two genomic RNAs that make up all the RNA information of cowpea mosaic virus (19,20). The RNA sequences of poliovirus (37,38) and of foot-and-mouth-disease virus (39,40) were established using the same techniques. The main homology is between the 87 K polypeptide of component B of cowpea mosaic virus and the region of poliovirus RNA that codes for the polymerase. In short, the replication systems of the two

viruses show analogies. Similarities have also been found between the 28 K protein of CPMV and a region of poliovirus to which (protease) functions have been attributed in maturation of the initial polyprotein.

Another sequence analogy has also been discovered between the 58 K polypeptide derived from cleavage of the 200 K polyprotein encoded in the BRNA of CPMV and the P2-x region of poliovirus RNA ; so far it has not been linked with any specific function common to the two viruses. Franssen et al. (36) reported homologies between some proteins encoded by alfalfa mosaic virus (AlMV), bromegrass mosaic virus (BMV), TMV, and the ns72 polypeptide of Sindbis virus (Haselhoff et al., in the press). As the number of complete nucleic acid sequences of animal and plant viruses increases, it will increasingly be possible to compare them, to speculate on the origins of RNA viruses, and perhaps to establish relationships between the various groups of animal and plant viruses. So far no link between the RNA or DNA viruses of prokaryotes and those of eukaryotes has been reported.

Recently, another type of sequence analogy has been discovered, between the open reading frame of intron 4 of the gene coding for cytochrome B of yeast and a particular region of the subgenomic RNA coding for the 30 K protein of TMV (41). Other sequence analogies have also been discovered between another strain of TMV (GTMV, cowpea strain), and another intron of cytochrome B. The degree of homology is shown in Table 2.

Table 2. Degree of homology between two strains of virus and two types of proteins encoded by introns of the cytochrome B gene of yeast [according to Zimmern (41)].

	CcTMVp30 %	TMVp30 %	b14 %
TMVp30	27		
b14	15	23	
aI5B	15	17	25

TMVp30 is the 30 kd protein expressed by a TMV subgenomic RNA.
CcTMVp30 is the 30 kd protein of the cowpea strain of TMV.
b14 is the fourth intron of the apocytochrome b gene of yeast (41).
aI5B is the gene for subunit 1 of cytochrome 2 oxidase of yeast (42).

The significance of these homologies is not clear, but they may indicate that RNA viruses originated when introns of various origins originally became independent entities and then linked up with a system for the replication of RNA (an RNA replicon). One would then have a system independent of DNA for its multiplication (41). Such RNA replicons seem to exist — the viroids (see Owens, this book).

IV. Perspectives

Several developments from the cloning of cDNA copies of viral RNA in bacterial plasmids can be foreseen.

A) Structure of viral RNAs

Viral RNA sequencing will clearly continue for some time and the catalogue of sequences known will grow in future years, undoubtedly revealing new homologies, which will enable the origin and evolution of viruses to be discussed.

B) Expression of cDNA copies

In addition, a double-stranded DNA corresponding to a defined viral cDNA and linked to a promoter sequence of prokaryotic or eukaryotic origin might be transcribed, thus giving rise to viral messengers able to replicate themselves. This would make it possible to alter at will (by means of additions or substitutions) the genetic content of the cDNA copy and to obtain gene vectors. This type of approach is based on the results obtained with poliovirus. Racaniello and Baltimore (4) have constructed a vector by introducing a complete double-stranded cDNA copy of poliovirus RNA into the Pst site of plasmid pBR322. This vector, amplified in bacteria, was inoculated into mammalian cell (HeLa) cultures and gave rise to infectious poliovirus.

Interpretation of these results suggests that they may be quite generally applicable and that plant viruses that have major analogies with poliovirus, for example, could replicate starting from a double-stranded cDNA inserted into a suitable vector. Such viruses exist, for example cowpea mosaic virus and tomato black ring virus. These two (multi-component) viruses (43) have their RNA 5' ends attached to a protein, and their genetic information is expressed in the form of a single protein, which then is cleaved into the various structural and nonstructural proteins isolated from infected cells.

On the basis of the sequence homologies described above, therefore, one could foresee experiments, like those performed with the picornaviruses, to study the expression, in sensitive plants or the corresponding protoplasts, of cDNA copies of the viral RNAs necessary for infection. One could then imagine using directed mutagenesis on cDNA copies and obtaining mutants that would make it possible to identify the function of the non-structural proteins. Also, such cDNA copies could be used as gene vectors.

In this latter case, satellite RNAs might be suitable material (46). The case of tomato black ring virus is an example (44,45). This virus can be accompanied by a satellite, which may or may not be present at the same time as the two genomic RNAs. It codes for a non-essential 48 K protein, of unknown function. One could envisage the insertion, between the noncoding regions of the cDNAs of this satellite, of a gene that might change the properties of the plant cell; the RNA transcribed from the cDNA matrix would be able to replicate thanks to the presence of the two genomic RNAs. It is thus possible to imagine the use of such a satellite as a gene vector.

The same approach has been suggested with the satellite of tobacco necrosis virus (47,48). In this case, however, the satellite RNA contains information for the coat protein of the satellite. This is a different situation from the satellite of TBRV, which is encapsidated by the coat protein whose information is carried by RNA 2. Other viruses also have one or more satellites accompanying their genomic RNA(s). This is the case, for example, with cucumber mosaic virus, which, in addition to the three genomic RNAs and one subgenomic RNA, has a satellite RNA whose sequence and function vary according to the type of family to which it belongs (49). Many other viruses have satellites, which have been more or less well identified ; a detailed list was drawn up by Murant and Mayo (21). These satellites could also be cloned and attempts to use them as gene vectors could be undertaken, as in the cases discussed above.

C) In vitro synthesis by cloning of viral proteins

Another prospect for the near future is the production of proteins corresponding to various genes of RNA or DNA viruses and the study of their properties and functions. It is well known (50) that for many genes in animals or humans the insertion of a foreign gene in the beginning of

a bacterial gene (such as ß-galactosidase) leads to the synthesis of a fusion protein by the bacteria. At present, attempts have begun to obtain fusion proteins with some genes of CaMV inserted in bacterial plasmids.

Thus, in my laboratory, attempts are being made to use a bacterial strain (JM103) containing a plasmid (derived from pUC9) to obtain synthesis of the product of the open reading frame V of CaMV suspected to code for a reverse transcriptase whose isolation, purification, and function have so far only been the subject of promising speculation. The same type of approach has been used by Hohn's group in Basel and is discussed in the article by Dixon and Hohn in this volume. Promising results have also been obtained by Fier's group in Ghent with the satellite of tobacco necrosis virus.

D) Diagnosis of virus diseases in plants

Another important point that may follow from cloning of the genome of DNA or RNA viruses is the use of these clones as identification probes. Introduced into a suitable plasmid, they could be amplified, labelled by nick translation, and used for the detection of the presence of a RNA in a plant or batch of plants by hybridization. This system could be particularly advantageous when the virus is a poor antigen, when it is a naked RNA (see Owens' chapter on viroids), or when the subunits encoded by a genome take part in the encapsidation of some heterologous genetic information [as in the cases of satellites, and of encapsidated viroids or virusoids (51,62)].

Whatever the future may hold, the technique of cloning of genetic information of RNA viruses is bound to develop substantially. In addition to all the roles just discussed, it is likely that directed mutagenesis starting from double-stranded DNA or from cDNA copies of RNA viruses will develop substantially, as a means of obtaining protective mutants. Likewise, the introduction of parts of a virus (for example the coat protein of a defined strain) into an appropriate vector might come to replace or complement the systems used hitherto for protection against viral infections.

REFERENCES

1. Caplan, A., Herrerra-Estrella, L., Inzé, D., Van Haute, E., Van Montagu, M., Schell, J. and Zambryski, P. Science 22:815-820, 1983.
2. Toh, H., Hayashida, H. and Miyata, T. Nature 305:827-829, 1983.
3. Dixon, L.K. and Hohn, T. This volume, 1984.
4. Racaniello, V.R. and Baltimore, D. Science 214:916-918, 1981.

5. Maniatis, T., Fritsch, E.F. and Sambrook, J. A cloning manual. Cold Spring Harbor Symposium, 1982.
6. Matyis, J.C., Silva, D.M. and Oliveira, A.M. Summa Phytopathol. 1: 267-275, 1975.
7. Hatta, T. and Francki, R.I.B. Virology 92:428-435, 1979.
8. Goodman, R.M. J. Gen. Virol. 54:9-21, 1981.
9. Hamilton, W.D.O., Bisaro, D.M. and Buck, K.W. Nucleic Acids Res. 10: 4901-4911, 1982.
10. Bisaro, D.M., Hamilton, W.D.O., Coutts, R.H.A. and Buck, K.W. Nucleic Acids Res. 10:4913-4922, 1982.
11. Buck, K.W., Coutts, R.H.A., Bisaro, D.M., Hamilton, W.D.O., Stein, V.E. and Sunter, G. Current Communications in Molecular Biology, Plant Infectious Agents, Cold Spring Harbor Symposium, 54-58, 1983.
12. Marriott, T.W. and Symons, R.H. Current Communications in Molecular Biology, Plant Infectious Agents, Cold Spring Harbor Symposium, 59-62, 1983.
13. Francki, R.I.B., Hatta, T., Boccardo, G. and Randles, J.W. Virology 101:233-241, 1980.
14. Stanley, J. and Gay, M.R. Nature 301:260, 1983.
15. Fiers, W. In: Comprehensive Virology (Ed. H. Fraenkel-Conrat), Plenum Press, New York, 1979, pp. 69-204.
16. Goelet, P., Lomonossoff, G., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. Proc. Natl. Acad. Sci. USA 79:5818-5822, 1982.
17. Meyer, M., Hemmer, O. and Fritsch, C. J. Gen. Virol. Submitted.
18. Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, Th., Rajewsky, K., and Baltimore, D. Cell 24:625-637, 1981.
19. Van Wezenbeek, P., Verver, J., Harmsen, J., Vos, P. and Van Kammen, A. The EMBO J. 2:941-946, 1983.
20. Lomonossoff, G.P. and Shanks, M. The EMBO J. 2:2253-2258, 1983.
21. Murant, A.F. and Mayo, M.A. Annu. Rev. Phytopathol. 20:49-70, 1982.
22. Richards, K.E., Jonard, G., Jacquemond, M. and Lot, H. Virology 89:395-408, 1978.
23. Collmer, C.W., Tousignant, M.E. and Kaper, J.M. Virology 127:230-234, 1983.
24. Lerner, R.A., Sutcliffe, J.G. and Shinnick, T.M. Cell 23:309-310, 1981.
25. Xiong, C., Lebeurier, G. and Hirth, L. Proc. Natl. Acad. Sci. USA, submitted.
26. Ooshika, I., Watanabe, Y., Meshi, T., Okada, Y., Igano, K., Inouye, K. and Yoshida, N. Virology 132:71-78, 1984.
27. Kozak, M. Nucleic Acids Res. 9:5233-5252, 1981.
28. Hirth, L. and Richards, K.E. Adv. Virus Res. 26:145-199, 1981.
29. Zelcer, A., Weaber, K.F., Balázs, E. and Zaitlin, M. Virology 113:417-427, 1981.
30. Goelet, P. and Karn, J. J. Mol. Biol. 154:541-550, 1982.
31. Hunter, T., Jackson, R. and Zimmern, D. Nucleic Acids Res. 11:801-821, 1983.
32. Dawson, W.O. and Dodds, J.A. Biochem. Biophys. Res. Commun. 107:1230-1235, 1982.
33. Dawson, W.O. Current Communications in Molecular Biology, Plant Infectious Agents, Cold Spring Harbor Symposium, 84-87, 1983.
34. Zaitlin, M., Palukaitis, P., Garcia-Arenal, F. and Sulzinski, M.A. Current Communications in Molecular Biology, Plant Infectious Agents, Cold Spring Harbor Symposium, 69-72, 1983.
35. Morris, T.J. Current Communications in Molecular Biology, Plant Infectious Agents, Cold Spring Harbor Symposium, 80-83, 1983.

36. Franssen, U., Leunissen, J., Goldbach, R., Lomonossoff, G. and Zimmern, D. The EMBO J. in press.
37. Racaniello, V.R. and Baltimore, D. Proc. Natl. Acad. Sci. USA 78: 4887-4891, 1981.
38. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecak, R., Lee, J.J., Van der Werf, S., Anderson, C.W. and Wimmer, E. Nature (London) 291:547-553, 1981.
39. Bodhroyd, J.C., Harris, T.J.R., Rowlands, D.J. and Love, P.A. Gene 17:153-161, 1982.
40. Robertson, B.H., Morgan, D.O., Moore, D.M., Grubman, M.J., Card, J., Fischer, T., Weddell, G., Dowbenko, D. and Yansura, D. Virology 126: 614-623, 1983.
41. Zimmern, D. J. Mol. Biol. 171:345-352, 1983.
42. De la Salle, H., Jacq, C. and Slonimski, P.P. Cell 28:221-232, 1982.
43. Hensgens, L.A.M., Bonen, A.M., de Haan, L., Van der Horst, G., van der Veen, R., van Ommen, G.J.B. and Grivell, L.A. Cell 32:379-389, 1983.
44. Hirth, L. In: Nucleic Acids and Proteins in Plants II (Eds. B. Parthier and P. Boulter), Springer-Verlag Vol. 14B, Berlin, 1982, pp. 302-336.
45. Murant, A.F., Mayo, M.A., Harrison, B.D. and Goold, R.A. J. Gen. Virol. 19:275-278, 1973.
46. Murant, A.F. and Taylor, M. J. Gen. Virol. 41: 53-61, 1978.
47. Van Emmelo, J., Devos, R., Ysebaert, M. and Fiers, W. J. Mol. Biol. 143:259-271, 1980.
48. Ysebaert, M., Van Emmelo, J. and Fiers, W. J. Mol. Biol. 143: 273-287, 1980.
49. Kaper, J.M. Plant Molecular Biology Reporter 1:2:49-54, 1983.
50. Harris, T.J.R. Genetic Engeneering 4:128-175, 1983.
51. Randles, J.W., Davies, C., Hatta, T., Gould, A.R. and Francki, R.I.B. Virology 108:111-122, 1981.
52. Franck, A., Guilley, H., Jonard, G., Richards, K. and Hirth, L. Cell 21:285-294, 1980.
53. Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Luedi, M., Shepherd, R.J. and Messing, J. Nucleic Acids Res. 9:2871-2888, 1981.
54. Balàzs, E., Guilley, H., Jonard, G. and Richards, K. Gene 19:239-249, 1982.
55. Ahlquist, P., Dasgupta, R. and Kaesberg, P. J. Mol. Biol. 172:369-383, 1984.
56. Ahlquist, P., Luckow, V. and Kaesberg, P. J. Mol. Biol. 153:23-38, 1981.
57. Cornelissen, B.J.C., Brederode, F.Th., Moormann, R.J.M. and Bol, J. Nucleic Acids Res. 11:1253-1265, 1983.
58. Cornelissen, B.J.C., Brederode, F.Th., Veenemann, G.H., Van Boom, J.H. and Bol, J.F. Nucleic Acids Res. 11:3019-3025, 1983.
59. Barker, R.F., Jarvis, N.P., Thompson, D.V., Loesch-Fries, L.S. and Hall, T.C. Nucleic Acids Res. 11:2881-2891, 1983.
60. Lomonossoff, G.P. and Shanks, M. The EMBO J. 2:2253-2258, 1983.
61. Collmer, C.W., Tousignant, M.E. and Kaper, J.M. Current Communications in Molecular Biology, Plant Infectious Agents, Cold Spring Harbor Symposium, 516-170, 1983.
62. Garger, S.J., Turpen, T., Carrington, J.C., Morris, T.J., Jordan, R.L., Dodds, J.A. and Grill, L.K. Plant Molec. Biol. Reptr. 1:21-25, 1983.

15

CLONING OF VIROID cDNA

ROBERT A. OWENS AND MICHAEL C. KIEFER*
Plant Virology Laboratory, U.S. Department of Agriculture, Beltsville, Maryland, USA and *Department of Botany, University of Maryland, College Park, Maryland, USA

A variety of strategies have been used to clone partial-length viroid cDNAs synthesized on single-stranded RNA templates. Major factors determining cloning strategy include the amount of sequence information available for a particular viroid and the anticipated use(s) for its cDNA. In vitro ligation of cDNA restriction fragments into the multiple cloning site of a pUC plasmid vector is an efficient strategy for several viroids. Full-length viroid cDNAs have been constructed from overlapping partial-length clones or cloned cDNA restriction fragments. Although not yet attempted, synthesis of viroid cDNA using denatured double-stranded viroid RNA templates seems ideally suited to cases where techniques using single-stranded templates are difficult to apply.

INTRODUCTION

Viroids, the smallest autonomously replicating pathogenic agents known, cause transmissible diseases in several economically important crop plants. Since the first viroid, potato spindle tuber viroid (PSTV), was characterized in 1971 (1), 10 additional viroid diseases have been reported. All known viroids are small, unencapsidated, and covalently closed circular RNA molecules containing 246-371 nucleotides arranged in a characteristic rod-like secondary structure. Because viroids appear to lack mRNA activity, they may be considered minimal genetic and biological systems. Comprehensive reviews of the biological and structural properties of these unusual pathogens are available (2, 3).

Comparison of the complete nucleotide sequences of 7 viroids suggest that presently known viroids can be divided into three groups (4): the first contains PSTV and several other viroids sharing extensive (60-70%) sequence homologies, a second group [coconut cadang-cadang viroid (CCCV)] shares only the strongly conserved central region with members of the PSTV

Y. Becker (ed.), RECOMBINANT DNA RESEARCH AND VIRUS.

group, and a third group [avocado sunblotch viroid (ASBV)] exhibits only scattered sequence homologies with PSTV.

Viroid cDNAs produced by recombinant DNA techniques have already been used in several areas of viroid research: nucleotide sequence determination (5-7), routine diagnosis of viroid diseases by nucleic acid hybridization (8, 9), and studies of viroid replication mechanisms requiring strand-specific hybridization probes (10-12). The recent demonstration that recombinant plasmids containing tandem dimers of full-length ds PSTV cDNA are infectious (13) has defined yet another application for cloned viroid cDNAs. Comparative sequence analyses of naturally occurring PSTV strains have identified regions of the genome involved in symptom production (reviewed in 3). In vitro mutagenesis of infectious cloned cDNAs can be used to systematically investigate the phenotypic effects of sequence changes in these and other regions of a viroid genome.

In this chapter we will describe the strategies which we have used to construct full-length PSTV and tomato apical stunt viroid (TASV) cDNAs using single-stranded RNA templates. Because different viroids require somewhat different cloning strategies, the advantages and limitations of alternative strategies will be discussed. Our discussion will conclude with a brief consideration of the apparent advantages of cloning viroid cDNA synthesized on denatured double-stranded (ds) RNA templates (14-16). Although this technique has not yet been applied to the cloning of viroid cDNA, it seems ideally suited to cases where techniques using single-stranded viroid RNA as template for cDNA synthesis are difficult to apply.

EXPERIMENTAL PROCEDURES

Construction of full-length PSTV and TASV cDNA clones involved six separate steps: determination of at least a partial viroid sequence by direct RNA sequencing techniques (17), nuclease S1 cleavage and polyadenylylation of the purified viroid RNA template (18), in vitro synthesis of ds viroid cDNA, insertion into a plasmid vector by either oligo(dC)-oligo(dG) tailing or ligation of doubly digested ds cDNA restriction fragments, determination of the nucleotide sequence of the cloned cDNA inserts, and assembly of the full-length ds cDNA by ligation of selected cDNA restriction fragments and recloning. Some of the specific methods detailed below have been adapted from ones found in "Molecular Cloning--A Laboratory Manual"

(19), a compilation of more general procedures that is an excellent source for tested protocols.

Direct RNA sequence determination

The complete nucleotide sequence of a PSTV strain inducing symptoms of intermediate severity (20) was available before we began to synthesize and clone PSTV cDNA. Provisional RNA sequences for TASV and tomato "planta macho" viroid (TPMV) were determined as described (17). Overlapping linear viroid fragments obtained by partial ribonuclease digestion of purified viroid RNA were labelled with ^{32}P by incubation with [γ-^{32}P]ATP plus T4 polynucleotide kinase and sequenced by partial enzymatic cleavage methods. Enzymes for RNA sequencing are available in kit form from P-L Biochemicals.

The partial enzymatic cleavage method yields sequence data rapidly but suffers from two limitations (5): underrepresentation of certain regions of the viroid sequence in the partial RNase digests and occasional difficulties in distinguishing C and U residues. Nevertheless, we obtained sufficient data to predict the distribution of restriction enzyme sites in double-stranded TASV cDNA and choose the most efficient cloning strategy.

Cleavage with S1 nuclease and polyadenylylation

Electrophoretically purified circular viroids were partially digested with S1 nuclease (Sigma) for 10 min at 37°C to produce primarily full-length linear molecules. The 50 µl reaction mixture (pH 4.5) contained 30 mM sodium acetate, 53 mM acetic acid, 300 mM NaCl, 1.8 mM $ZnCl_2$, 100 µg/ml viroid RNA, and 3000 units/ml S1 nuclease. Digestion was terminated by adding 2.5 µl of 2 M Tris base, and the RNA was extracted with phenol-chloroform, washed with ether, and recovered by ethanol precipitation. S1-cleaved viroids were polyadenylylated by incubation for 1 hr at 37°C in a 50 µl reaction mixture containing 50 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 240 mM NaCl, 0.1 mM ATP, 40 units/ml placental ribonuclease inhibitor (RNasin) (Biotec), 100 µg/ml viroid RNA (heated at 90°C for 2 min and quenched at 0°C before addition to the reaction mixture), and 60 units/ml *E. coli* poly(A) polymerase (Bethesda Research Labs). Polyadenylylated viroids were extracted with phenol/chloroform, washed with ether, and recovered by ethanol precipitation.

Synthesis of double-stranded viroid cDNA

Viroid cDNA was synthesized by incubation for 2 hr at 42°C in a 50 µl reaction mixture containing 66 mM Tris-HCl (pH 8.3), 50 mM KCl, 8 mM $MgCl_2$, 1 mM dithiothreitol, 0.015% Triton X-100, 1 mM each of dATP, dGTP, and TTP,

0.1 mM [α-^{32}P]dCTP (4 μCi/ml), 16.6 μg/ml each of $p(dT)_8dC$, $p(dT)_8dG$, and $p(dT)_8dA$ primers (P-L Biochemicals), 500 units/ml RNasin, 40 μg/ml polyadenylylated viroid (heated and quenched as above), and 640 units/ml reverse transcriptase (Life Sciences). cDNA synthesis was terminated and the RNA template removed by addition of 2 μl 500 mM EDTA (pH 8.1) and 25 μl 150 mM NaOH and incubation for 1 hr at 65°C. After neutralization with 2.5 μl 1.5 M acetic acid the cDNA was extracted with phenol-chloroform, chromatographed on Sephadex G-50, and recovered by ethanol precipitation.

Second-strand DNA synthesis was carried out in two stages: first, incubation for 90 min at 37°C in a 100 μl reaction mixture containing 100 mM HEPES (pH 6.9), 10 mM $MgCl_2$, 2.5 mM DTT, 70 mM KCl, 1 mM each of dATP, dGTP, dCTP, and TTP, 0.5 μg/ml viroid cDNA (heated and quenched as above), and 25 units/ml E. coli DNA polymerase I (large fragment, New England Biolabs). Second, incubation for 1 hr at 37°C in a 50 μl reaction containing 66 mM Tris-HCl (pH 8.3), 50 mM KCl, 8 mM MgCl, 1 mM dithiothreitol, 0.015% Triton X-100, 1mM each of dATP, dGTP, and TTP, 0.1 mM dCTP, 1-2 μg/ml ds cDNA, and 640 units/ml reverse transcriptase. DNA synthesis was terminated after each incubation by addition of 2 μl 500 mM EDTA (pH 8.1), and the ds cDNA was recovered by ethanol precipitation after phenol-chloroform extraction and chromatography on Sephadex G-50.

Construction of chimeric pUC9 plasmids

Double-stranded viroid cDNAs were cloned by two different techniques: S1-digested TPMV cDNA was inserted into the pUC9 vector PstI site via oligo(dC)-oligo(dG) tailing (19), and BamHl-HindIII restriction fragments of TASV ds cDNA were inserted into doubly-digested vector via in vitro ligation (21).

Prior to addition of oligo(dC) tails ds TPMV cDNA was incubated for 1 hr at 37°C with S1 nuclease (200 units/ml in the buffer used for the initial partial S1 digestion of viroid RNA) to remove non-base-paired regions. The S1 nuclease digestion was terminated by phenol-chloroform extraction, and ds cDNA eluting in the void volume during Sephadex G-100 chromatography was recovered by ethanol precipitation. Oligo(dG) and oligo(dC) tails were added to PstI-cleaved pUC9 and S1-treated ds TPMV cDNA respectively by incubation with deoxynucleotidyl transferase (P-L Biochemicals) (19). The tailed products were combined (50:1, wt/wt), annealed (19), and used to transform E. coli strain JM83 (21).

Double-stranded TASV cDNA was digested with BamHI plus HindIII to release two fragments that together contain the entire sequence of TASV. After phenol-chloroform extraction and ethanol precipitation the resulting ds cDNA fragments were fractionated by electrophoresis in a 5% acrylamide gel containing Tris-borate-EDTA buffer (19). Fragments of the appropriate sizes were excised from the gel, eluted, and recovered by ethanol precipitation. Each fragment (containing one BamHI and one HindIII terminus) was combined with BamHI/HindIII-digested pUC9 vector (1:100, wt/wt), incubated at 16°C for 20 hrs with T4 DNA ligase (P-L Biochemicals), and used to transform E. coli strain JM83 (21).

Transformants containing ds cDNA inserts were identified by screening for ampicillin-resistant colonies that appeared white in the presence of 5-bromo-4-chloro-3-indolyl-B-D-galactoside (21). Chimeric plasmids were isolated from JM83 by a rapid isolation procedure (22), and their ds viroid cDNA inserts were sized and mapped by single or multiple restriction endonuclease digestions followed by gel electrophoresis.

DNA sequence determination

Chimeric pUC9 plasmids were 3'-labeled at their unique HindIII or EcoRI sites by incubation of the cleaved DNAs with E. coli DNA polymerase I (large fragment) plus [α-^{32}P] dATP (400 Ci/mmole) (23). Labeled viroid inserts were excised by a second digestion with either EcoRI or HindIII, purified by gel electrophoresis, and sequenced by the base-specific chemical cleavage method (24).

Comparison of viroid RNA and cDNA sequences

Because at least one CEV isolate exhibits obvious molecular heterogeneity (25), sequence comparisons must be performed to establish the nucleotide sequence of a viroid via cloned cDNAs. Two types of comparisons can be performed: DNA sequences of independent cDNA clones representing the same region of the viroid genome can be compared or the DNA sequence of non-overlapping cDNA clones may be compared with the RNA sequence determined by direct methods. By these criteria neither TPMV nor TASV exhibits detectable molecular heterogeneity.

DISCUSSION

Restriction analyses of representative TPMV and TASV cDNA clones obtained by the procedures described in the EXPERIMENTAL PROCEDURES section are shown in Figure 1. Regions of the respective viroid sequences contained

in selected ds cDNA clones are shown in Figure 2. These figures illustrate some of the problems encountered in cloning ds viroid cDNAs by oligo(dC)-oligo(dG) tailing and the advantages of cloning selected doubly-digested ds cDNA restriction fragments.

Cloning via oligo(dC)-oligo(dG) tailing

Comparison of the mobilities of the ds TPMV cDNA inserts released by PstI digestion of recombinant pUC9 plasmids with appropriate size markers suggests that most clones contain no more than one half of the 360 nucleotide TPMV genome (Figures 1A and 2). Determination of the nucleotide sequences of several of the larger TPMV cDNA clones confirmed this inference. Construction of a full-length ds TPMV cDNA clone from our collection of partial-length clones (Figure 2) would be a complex undertaking, involving several cycles of fragment purification, ligation, and recloning. Construction of a full-length ds PSTV cDNA clone from similar overlapping clones proved considerably simpler (13). Figure 2 shows that contiguous AvaII-HaeIII fragments could be obtained from only two overlapping PSTV cDNA clones.

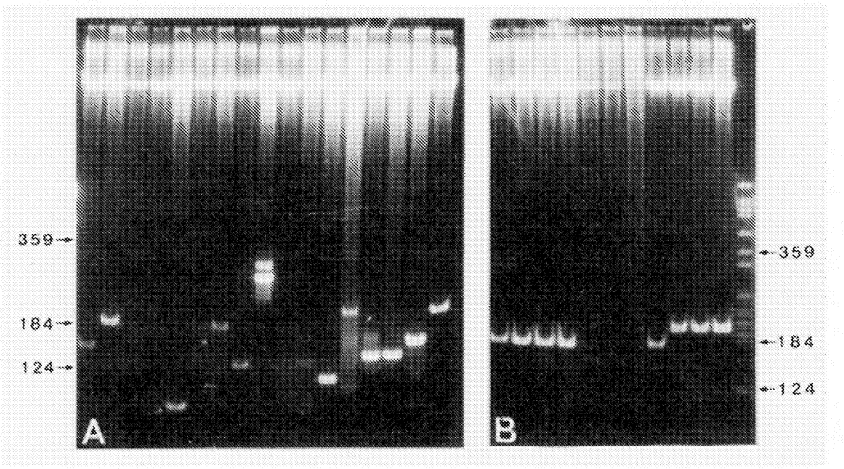


FIGURE 1. Sizes of TPMV- and TASV-specific inserts present in pUC9 recombinant plasmids. a) TPMV cDNAs, isolated via an oligo(dC)-oligo(dG) tailing procedure, were released by PstI digestion. b) TASV cDNAs, isolated via in vitro ligation of electrophoretically purified cDNA restriction fragments, were released by digestion with BamHI plus HindIII. Marker fragments were generated by HaeIII cleavage of a recombinant pBR322 plasmid containing a tandem dimer of PSTV cDNA inserted in the BamHI site. Arrows indicate the positions of fragments containing 359, 184, and 124 base-pairs.

A second difficulty associated with cloning viroid cDNAs via oligo(dC)-oligo(dG) tailing is the significant proportion of chimeric plasmids that release multiple inserts after digestion with PstI. These multiple inserts may be generated by successive deletions occurring within an "unstable" cDNA insert during plasmid replication. Examples of "unstable" ds TPMV cDNA clones may be seen in Figure 1A, and similar difficulties were encountered during the earlier cloning of ds PSTV cDNA via oligo(dC)-oligo(dG) tailing (Figure 2 in reference 9). The cause(s) of this phenomenon is unknown, but transfer of clones into the recA-deficient strain HB101 does not enhance their stability.

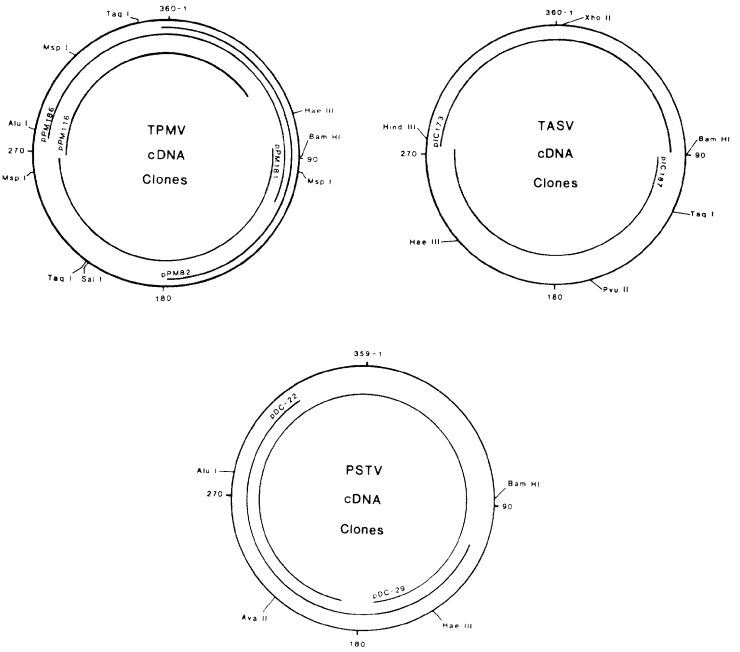


FIGURE 2. Mapping of partial-length viroid cDNA clones. The outer circles represent the complete nucleotide sequences of TPMV, TASV, or PSTV; positions of several restriction sites in each ds viroid cDNA are shown. Inner concentric curves indicate the portions of the complete genomes present in individual partial-length clones.

Cloning of cDNA restriction fragments

In the case of TASV we were able to circumvent both of these difficulties by inserting ds TASV cDNA restriction fragments into the multiple cloning site of plasmid pUC9. Preliminary direct RNA sequence determinations had shown that TASV contains single BamHI and HindIII recognition sites that divide the molecule almost exactly in half. Digestion of ds TASV cDNA released two discrete fragments of the expected sizes (173 and 187 bp), and sequence analysis of the cloned ds cDNAs fragments has shown that they do indeed contain the complete 360 nucleotide sequence of TASV.

Insertion of cDNA restriction fragments into the multiple cloning site of a pUC plasmid vector by *in vitro* ligation (21) is an efficient method for obtaining specific partial-length viroid cDNAs. Recombinant DNAs containing the individual BamHI-HindIII fragments are stable (Figure 1B), and construction of full-length TASV cDNA clones having BamHI termini was straight forward. Data presented in Table 1 demonstrates that this approach can be used to construct a number of full-length viroid cDNAs.

TABLE 1. Potential Cloning Sites in ds Viroid cDNAs

Enzyme	Viroid						
	CEV[a] (371)[c]	CSV[b] (354 or 356)	PSTV (359)	TPMV (360)	TASV (360)	CCCV (246)	ASBV (247)
BamHI	89[d]	84 or 86	87	89	90	54	
SalI				214			
HindIII	286	240 or 242			277		
		276 or 278					
AvaI	94	89 or 91	92	94	95	59	
	370	353 or 355	284	359	359		
XhoII	3	84	87	3	3	54	
	89			89	90		
PstI	40				40	33	

[a] citrus exocortis viroid
[b] chrysanthemum stunt viroid (2 isolates)
[c] number of nucleotides in each viroid
[d] position of first nucleotide in recognition sequence

The entire genomes of the first five viroids in Table 1, members of the PSTV group, can be cloned as pairs of restriction fragments. Fragments with BamHI and HindIII termini are suitable for CEV and TASV, while fragments with BamHI and SalI termini can be used for TPMV. AvaI fragments are the first choice for PSTV and CSV and can be used for any member of the PSTV group. Because AvaI has multiple recognition sites with the general sequence CPyCGPuG, its use in this type of cloning scheme may cause the introduction of point mutations in the AvaI sites of the cloned cDNA. The same problem may be encountered when XhoII fragments of ds cDNA are cloned after insertion into the pUC BamHI site.

Unfortunately the distribution of restriction sites within viroid cDNA is not always compatible with the use of the pUC multiple cloning site. Useful sites in CCCV are rather tightly clustered, while ASBV contains none of the most useful sites. Alternative approaches, including the use of specific synthetic oligodeoxynucleotide primers and denatured double-stranded RNA as template for cDNA synthesis, will be required to construct full-length cDNAs in these cases.

Synthetic primers for cDNA synthesis

Rohde et al. (26, 27) have demonstrated that a synthetic pentadecadeoxynucleotide can be used to specifically prime the synthesis of full-length PSTV and CEV cDNAs by reverse transcriptase. Maximum yields of PSTV cDNA, 8% of the template used, were 3-4 fold higher than that routinely obtained with polyadenylylated viroid RNA plus oligo(dT) primer. The particular primer used, d(TTCTTTTTTCTTTTC), is complementary to a polypurine sequence found in all five members of the PSTV viroid group and has been used to construct a full-length PSTV cDNA clone (7).

Although specific oligodeoxynucleotides prime efficient synthesis of full-length viroid cDNA, the double-stranded cDNAs subsequently recovered by cloning will not be full-length. The short hairpin loop formed at the 3'-terminus of single-stranded cDNA and used to prime the synthesis of the second DNA strand must be removed before the ds cDNA can be cloned. Full-length clones must, therefore, be constructed by in vitro ligation of contiguous fragments prepared from overlapping partial-length clones (7, 13).

Rearrangements occurring during synthesis of ds cDNA (28) allowed van Wezenbeek et al. (7) to isolate the necessary overlapping clones from PSTV cDNA synthesized using a single synthetic primer. Reverse transcriptase

will occasionally proceed through the primer region after synthesis of a full-length viroid cDNA and displace the full-length, primer-extended cDNA. The use of two synthetic oligonucleotides to prime cDNA synthesis at different points on the viroid RNA template provides a more efficient solution.

For example, the yield of BamHI-HindIII fragments of ds TASV cDNA (Figure 2B) could be greatly increased by using synthetic oligonucleotides complementary to regions of TASV containing the BamHI or HindIII recognition sites to prime TASV cDNA synthesis. No other alterations in the TASV cDNA synthesis and cloning protocol described above would be necessary. This approach will become feasible as the ability to synthesize specific oligonucleotide primers becomes more widespread.

cDNA synthesis on denatured double-stranded viroid RNA templates

The various possible strategies for the construction of full-length ds viroid cDNA clones discussed above share one common feature--the template for cDNA synthesis is single-stranded viroid RNA. An alternative approach involving cDNA synthesis from a denatured ds RNA template seems ideally suited to the one-step synthesis of full-length viroid cDNAs.

At least three groups have recently described procedures for the molecular cloning of double-stranded RNA virus genomes (14-16). The ds RNA templates had to be denatured and either poly(A) or poly(C) tails added before oligo(dT)- or oligo(dG)-primed cDNA synthesis could be catalyzed by reverse transcriptase. After the resulting mixture of cDNAs (including full and partial copies of both possible polarities) was annealed, the resulting DNA-DNA duplexes were filled in by the action of E. coli DNA polymerase I (large fragment) to complete any partial copies.

Reannealing complementary cDNAs prevents the loss of terminal sequences that must occur when the 3'-hairpin is used to prime second strand synthesis on single-stranded cDNA templates. Full-length ds cDNA clones of individual reovirus and rotavirus genome segments have been recovered after standard oligo(dC)-oligo(dG) tailing procedures were used to insert the ds cDNA into the PstI site of pBR322. If appropriate synthetic primers for cDNA synthesis are available, this methodology seems ideally suited to the one-step cloning of full-length ds viroid cDNAs.

Double-stranded viroid RNA is more difficult to purify than reovirus or rotavirus genomic RNAs, but quantities of ds PSTV RNA adequate for molecular cloning can be isolated from preparations of total cellular

RNA by either gel electrophoresis (29) or chromatography on CF-11 cellulose (11, 30). These ds RNAs, thought to be essential intermediates in PSTV replication (11, 30), contain circular and linear forms of PSTV, multimeric linear forms of the complementary strand (11, 30), and possibly multimeric forms of PSTV as well (31). Similar dsRNAs seem to be involved in the replication of other viroids (12).

Primers for cDNA synthesis on denatured ds viroid RNA templates can be chosen so that the 5'-portions of the complementary cDNAs will generate the recognition sequence of an appropriate restriction enzyme after self-annealing and filling in the 3'-termini by incubation with E. coli DNA polymerase I (large fragment). Restriction digestion of the resulting ds cDNA would release a full-length viroid cDNA fragment suitable for either direct cloning via in vitro ligation to linearized vector (see preceding section) or addition of synthetic linkers prior to ligation and cloning.

Although isolation of full-length viroid cDNA clones using this approach has not yet been reported, it would seem to be the most generally applicable of all the approaches discussed. The three prerequisites--determination of sufficient preliminary RNA sequence information for a particular viroid to allow appropriate primer sequences to be selected, ability to synthesize the two oligodeoxynucleotide primers required, and isolation of ds viroid RNA--are now within the capabilities of several laboratories engaged in viroid research.

REFERENCES

1. Diener, T.O. Virology 45, 411-428, 1971.
2. Diener, T.O. Viroids and Viroid Diseases. Wiley and Sons, New York, 1979.
3. Sänger, H.L. In: Encyclopedia of Plant Physiology New Series, 14B (Eds. B. Parthier and D. Boulter), Springer-Verlag, Berlin, 1982, pp. 368-454.
4. Haseloff, J., Mohamed, N.A., and Symons, R.H. Nature 299, 316-321, 1982.
5. Symons, R.H. Nuc. Acids Res. 9, 6527-6537, 1981.
6. Visvader, J.E., Gould, A.R., Bruening, G.E., and Symons, R.H. FEBS Letters 137, 288-292, 1982.
7. van Wezenbeek, P., Vos, P., van Boom, J., and van Kammen, A. Nuc. Acids Res. 10, 7947-4957, 1982.
8. Owens, R.A. and Diener, T.O. Science 213, 670-672, 1981.
9. Murphy, P.H., Barker, J.M., Gould, A.R., and Symons, R.H. Abstracts of Twelfth Int. Cong. Biochem. (Perth, Australia), p. 170, 1982.
10. Owens, R.A. and Cress, D.E. Proc. Natl. Acad. Sci. USA 77, 5302-5306, 1980.

11. Owens, R.A. and Diener, T.O. Proc. Natl. Acad. Sci. USA 79, 113-117, 1982.
12. Bruening, G.E., Gould, A.R., Murphy, P.J., and Symons, R.H. FEBS Letters 148, 71-78, 1982.
13. Owens, R.A., Kiefer, M.C., and Cress, D.E. In: Plant Infectious Agents-Viruses, Viroids, Virusoids, and Satellites (Eds. H.D. Robertson, S.H. Howell, M. Zaitlin, and R.L. Malmberg, Cold Spring Harbor Laboratory, New York, 1983 (in press).
14. Cashdollar, L.W., Esparza, J., Hudson, G.R., Camelo, R., Lee, P.W.K., and Joklik, W.K. Proc. Natl. Acad. Sci. USA 79, 7644-7648, 1982.
15. Imai, M., Richardson, M.A., Ikegami, N., Shatkin, A.J., and Furuichi, Y. Proc. Natl. Acad. Sci. USA 80. 373-377. 1983.
16. Both, G.W., Bellamy, A.R., Street, J.E., and Siegman, L.J. Nuc. Acids Res. 10, 7075-7088, 1982.
17. Haseloff, J. and Symons, R.H. Nuc. Acids Res. 9, 2741-2752, 1981.
18. Palukaitis, P. and Symons, R.H. FEBS Letters 92, 268-272, 1978.
19. Maniatis, T., Fritsch, E.F., and Sambrook, J. Molecular Cloning--A Laboratory Manual. Cold Spring Harbor Laboratory, New York, 1982.
20. Gross, H.J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H., and Sänger, H.L. Nature 273, 203-208, 1978.
21. Vieira, J. and Messing, J. Gene 19, 259-268, 1982.
22. Davis, R.W., Thomas, W., Cameron, J., St. John, T.P., Scherer, S., and Pagett, R.A. In: Methods Enzymol. 65 (Ed. R. Wu), Academic Press, N.Y., 1980, pp. 406-407.
23. Rüther, V., Koenen, M., Otto, K., and Müller-Hill, B. Nuc. Acids Res. 9, 4087-4098, 1981.
24. Maxam, A.M. and Gilbert, W. In: Methods Enzymol. 65 (Ed. R. Wu), Academic Press, N.Y., 1980, pp. 499-560.
25. Gross, H.J., Krupp, G., Domdey, H., Raba, M., Jank, P., Lossow, C., Alberty, H., Ramm, K., and Sänger, H.L. Eur. J. Biochem. 121, 249-257, 1982.
26. Rhode, W., Schnolzer, M., Rackwitz, H-R., Haas, B., Sellger, H., and Sänger, H.L. Eur. J. Biochem 118, 151-157, 1981.
27. Rhode, W., Schnolzer, M., and Sänger, H.L. FEBS Letters 130, 208-212, 1981.
28. Fagan, J.B., Pastan, I., and deCrombrugghe, B. Nuc. Acids Res. 8, 3055-3064, 1980.
29. Hadidi, A., Hashimoto, J., and Diener, T.O. Ann. Virol. (Inst. Pasteur) 133E, 15-31, 1982.
30. Branch, A.D., Robertson, H.D., and Dickson, E. Proc. Natl. Acad. Sci. USA 78, 6381-6385, 1981.
31. Branch, A.D., Willis, K.K., Odell, J.T., and Robertson, H.D. Abstracts of Fourth Int. Conf. Comp. Virol. (Banff, Canada), p. 196, 1982.

EXPRESSION OF VIRAL GENES IN BACTERIA

16

EXPRESSION OF CLONED GENES UNDER PHAGE λ CONTROL

AMOS B. OPPENHEIM, NURIT KATZIR*, SIMI KOBY AND HILLA LOCKER-GILADI
Department of Molecular Genetics, The Hebrew University-Hadassah Medical School, Jerusalem, and *Unit of Immunology and Microbiology, University Center for Health Sciences, Ben Gurion University of the Negev, Beer Sheba, Israel

SUMMARY

A system for the synthesis of specific gene products in *E. coli* is described. Genes cloned in phage lambda were shown to direct extensive synthesis on infection of a *nus* A host. Expression levels were higher than those reached with available plasmid expression vectors. The elevated level of gene expression is due in part to gene amplification. Of special interest is the fact that protein synthesis is confined to the cloned gene. Other lambda proteins do not accumulate under these conditions.

INTRODUCTION

Bacteriophage lambda (λ) can propagate by two exclusive modes: the lytic and the lysogenic (1). The lytic cycle is programmed by temporal gene expression. Initially, transcription initiated at the P_L and P_R promoters terminates at transcription termination signals t_R1 and t_L1 (see Fig. 1). Suppression of termination by the N protein permits the transcripts to be elongated into the early replication and recombination

genes. Following the expression of the Q gene product, the activator of late functions, rapid synthesis of late morphogenetic genes takes place. Finally, mature virions are produced and released from the lysing cell.

The lysogenic pathway occurs under conditions that favor the formation of the active cII gene product. The cII protein acts as a positive regulatory element that stimulates initiation of transcription essential for the synthesis of the cI repressor and integrase, the int gene product. The cI repressor represses P_L and P_R and prevents further expression of the early phage genes. The integrase directs the integration of the phage genome at a unique site in the bacterial chromosome. Induction of phage gene expression is achieved by the inactivation of the cI repressor protein. This is conveniently done by transferring a culture carrying a thermolabile repressor from 30° to 42° C.

The N gene product modifies the host RNA polymerase (2). This modification involves the participation of auxiliary host functions. One such function is coded by the nusA gene. E. coli mutated in nusA is unable to support λ growth (3). In such a host mutant, the λ N gene product is incapable of preventing transcription termination. Such bacterial strains are unable to support λ growth at elevated temperatures. It has been shown that the mutated nusA protein is unable to interact with the N protein (4).

Expression of a gene in E. coli is influenced by the level of transcription and translation of the genetic information. High level of expression is primarily dependent upon the presence of an effective promoter, an efficient ribosome binding site and a high number of gene copies.

In this work we describe the use of a novel system for obtaining high level expression of cloned genes. We find that phage λ carrying E. coli genes replicate extensively following infection of a nusA host. The infecting phage exclusively directs the synthesis of the cloned gene product without significant accumulation of any of the fifty-odd λ gene products. Our results indicate that the presence of the nusA mutation combined with phage DNA replication probably contribute to the accumulation of one specific gene product.

MATERIALS AND METHODS

Bacterial and phage strains

Strains used in this work are described in Table 1.

Infections

Cells were grown in TBMM (1% Bacto tryptone, 0.5% NaCl, 10 mM $MgSO_4$, 0.2% maltose, 0.1% yeast extract), to a density of $1X10^8$ cells/ml at 30^o, concentrated in TM (10 mM Tris, 10 mM $MgSO_4$, pH 7.4) and infected at a multiplicity of 3. The infected cells

TABLE 1. LIST OF BACTERIA AND BACTERIOPHAGE

BACTERIAL STRAINS	
Strain	Features
K37	*E. coli* W3102 *str*^{R}*gal*$^{-}$
K95	*E. coli* W3102 *str*^{R}*gal*$^{-}$*nusA1*
NS377	*E. coli nusA1 rif*^{R}-2
K554	*E. coli nusC rif*^{R}

PHAGE STRAINS		
Strain	No.	Features
AO 190		λ cI857P*lac*5
AO 844		λ cI857*gal*8
AO 936		λ cI857*cam*
AO 968		λ cI857*Nsus*7*Nsus*53P*lac*5
AO 759		λ cI857*trplac*
AO 758		λ cI857*Nsus*7*Nsus*53*trplac*
AO 956		λ cI857*Psus*902P*lac*5

were incubated at 40°C (zero time) for a few hours as stated. Incubations at 42°C or 43°C did not affect the results.

β-galactosidase assay

The assay for β-galactosidase was done as described by Miller (5). One unit is defined as the amount of enzyme hydrolyzing 1 mole of o-nitrophenyl-D-galactoside (ONPG) in 1 min at 30°C.

Chloramphenicol acetyltransferase (CAT) assay

The spectrophotometric assay for chloramphenicol acetyltransferase was done as follows: Samples (0.5 ml) were lysed by the addition of 2 drops of chloroform and 1 drop of 0.1% sodium dodecyl sulfate (SDS). The tubes were then vortexed for 10 seconds (5). The colorimetric assay medium for chloromphenicol acetyltransferase, modified from Shaw and Brodsky (6) contains: 10 mM Tris-HCl pH 7.8; 1 mM 5,5' dithiobis-2-nitrobenzoic acid (DTNB); 0.3 mM acetyl-CoA; 1 mM chloramphenicol; and 5-50 µl cell extract. The reactions were done at room temperature and the increase in absorbancy (412 mm) was followed. Enzyme units were determined in nmol DTNB reduced per min per $O.D._{650}$ of the samples.

Electrophoresis

1 ml samples were centrifuged for 10 min in an Eppendorf microcentrifuge. The precipitate was resuspended in sample buffer (0.2 M Tris pH 6.8, 8% SDS, 2 M β-mercaptoetnanol, 10% (v/v) glycerol, 0.025% bromophenol blue). Samples were placed in a boiling water

bath for 3 min and were spun to remove the insoluble fraction. Electrophoresis was on SDS-polyacrylamide gradient gels ranging from 10% to 26%.

RESULTS AND DISCUSSION

It has been previously observed (7, and A. Oppenheim, unpublished observation) that under N^- conditions, *trp* phages that carry the *trp* operon of *E. coli* permit extensive synthesis of the *trp* enzymes. Under these conditions phage gene expression was limited because of the absence of the *N* gene product. The cloned gene does not require the N protein for expression. An *E. coli* mutant carrying the *nusA* gene product is essential for λ *N* gene function. Thus infection of a *nusA* host mutant should mimic infections by λN^-. However, it was shown that some *N* function is expressed by λ on infection of a *nusA* host (3). In the following experiments we tested various aspects involved in the expression of genes cloned in phage λ following infection of *nus* bacterial mutants.

We have used several λ transducing phages as model systems. In the transducing strains abnormal excision of the prophage from the bacterial chromosome leads to the generation of viable phages in which dispensible functions are replaced by *E. coli* genes (8,9). We utilized phages carrying the *lac* operon, λ P*lac*5; and λ *gal*8, carrying the *gal* operon. In these

transducing phages the bacterial genes are inserted in the b region (Fig. 1). λ cam harbors the Tn9 transposon which confers resistance to the drug chloramphenicol, also located in the b region. This transposon is made of the gene coding for the enzyme chloromphenicol acetyltransferase bracketed by two IS1 (insertion elements) (10,11).

Infection of the nusA1 strain K95 with λPlac5, λ cam, and λ gal8 leads to extensive expression of the cloned genes. Figure 2 shows the kinetics of β-galactosidase synthesis following infection with λ Plac5. In contrast to the infection of the nusA host, very limited synthesis was observed on infection of the nus^+ permissive host. The presence of an active N function in infection of the nus^+ host permits infected cells to follow a productive lytic cycle. The reason for the limited expression following infection of the nus^+ host was not investigated. That the nusA host supports abnormally high levels of β-galactosidase expression is shown in Fig. 3. The amount of enzyme synthesized is about 50-fold higher than that observed during normal induction of the lac operon in E. coli. This expression does not require the presence of an inducer for the lac operon. In fact we have shown that the addition of the inducer IPTG had very little effect on β-galactosidase expression in a nusA host.

The high level of enzyme synthesis is in contrast

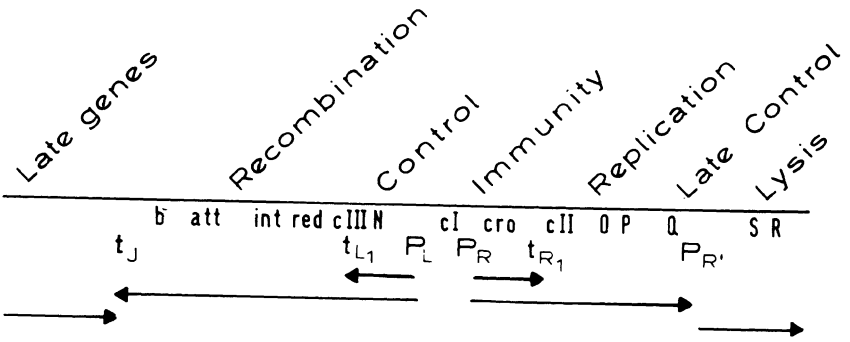


Fig. 1. Genetic map of bacteriophage lambda. In the mature phage the DNA is found as a single linear chromosome 48500 bp long, with single strand cohesive ends. The cohesive ends allow circularization of the phage genome following infection. The genes are clustered according to function. The control region carries regulatory genes. The cI and cro genes encode for repressors of the early promoters P_L and P_R, N gene product permits the extension of the early transcripts beyond t_LI and t_RI. These delayed early elongated transcripts direct the synthesis of gene products involved in replication, recombination and the expression of Q, the late gene activator. Late transcription is directed from the P'_R promoter. It directs the expression of morphogenetic functions involved in head and tail development, DNA packaging and cell lysis. The b region carries dispensible genes. In λ gal and λ lac transducing phages the bacterial operon replaces segments of the b region. The tn9 transposon carrying the chloramphenicol resistance gene present in λ cam is inserted in the b region.

to the absence of any significant expression of λ genes. We reason that the nusA mutations prevents λ transcription antitermination. Transcription can be initiated at P_L and P_R. However, this transcription is not extended by N in the absence of active nusA gene products. The kinetics of expression show a lag of 30 minutes to 1 hour

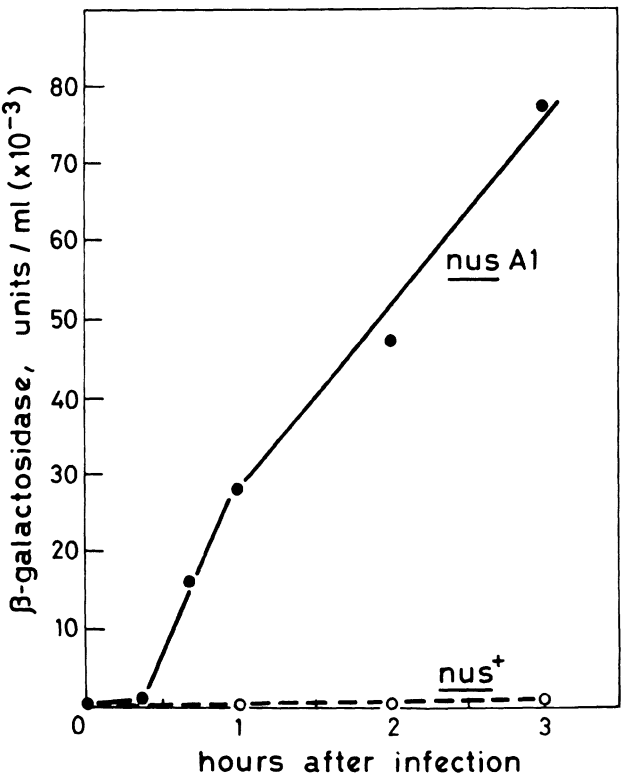


Fig. 2. Kinetics of synthesis of β -galactosidase. Exponentially growing cells were infected by λ Plac5. Samples were assayed at indicated times for β -galactosidase, the product of the lac Z gene as described by Miller (5). Symbols: nus+ (o-o), nusA1 (●-●).

followed by a burst of synthesis that can continue for a few hours. In the case of infection with λ Plac5 the level of synthesis can reach up to 60% of the total cellular proteins.

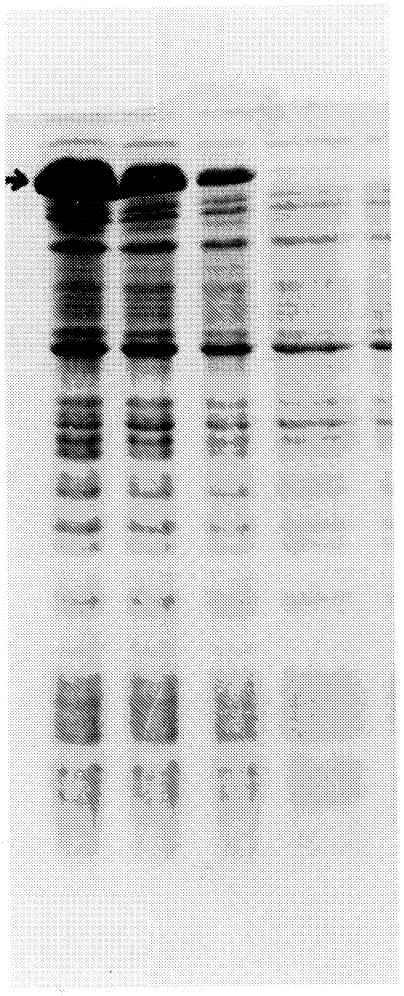


Fig. 3. Synthesis of β-galactosidase in K95 nusA. Samples from the experiment shown in Fig. 2 were centrifuged and analyzed by polyacrylamide gel electrophoresis. The protein bands were visualized by staining with coomassie blue. The β-galactosidase protein band is indicated by an arrow. Samples loaded from left to right were taken at 0, 20, 40, 60 and 120 minutes after infection.

In order to test whether the exclusive enzyme synthesis is unique to the lacZ gene we tested two other genes cloned on phage λ .

The results of infection of K95 with λ phage carrying the chloromphenical acetyltransferase gene (12) are shown in Figs. 4,5 and in Table 2. The rapid synthesis of chloromphenical acetyltransferase is shown in Fig. 4. Analysis of proteins made during the infection period demonstrated that the only protein present in excessive amounts was the chloromphenical acetyltransferase enzyme. The level of CAT synthesis in the infection system was about 30 times higher than that observed in cells carrying the chloromphenical acetyltransferase gene in plasmid pBR325 (Table 2).

Results similar to the above were obtained with galactokinase synthesis following infection with λ gal8 (data not shown).

The reduced expression of λ genes probably allows the transcription and translation machinery of the cell to concentrate on the expression of the cloned gene. In addition the absence of phage endolysin synthesis prevents premature cell lysis and allows an extended period of expression.

The synthesis of β -galactosidase following infection of nusA nusB double mutant (13) was similar to that observed with nusA strains. Infection of nusC mutant also led to a high level of β-galactosidase synthesis.

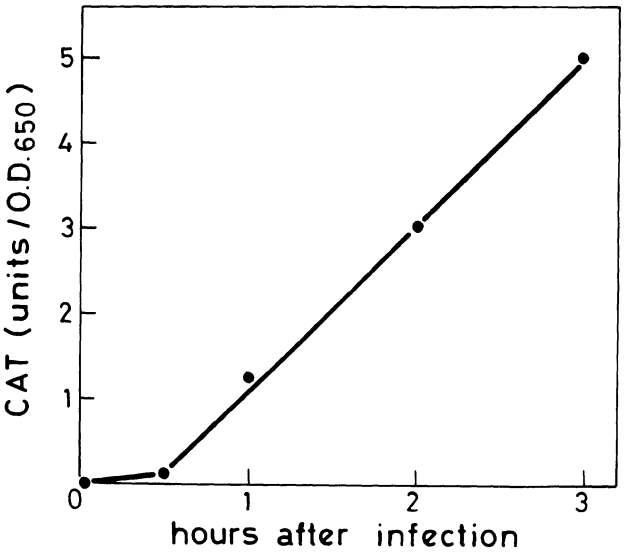


Fig. 4. Kinetics of synthesis of chloramphenicol acetyltransferase. Exponentially growing nusA cells were infected by λ cam. Samples were assayed at indicated times for chloramphenicol acetyltransferase as described in Materials and Methods.

TABLE 2. SYNTHESIS OF CHLORAMPHENICOL ACETYLTRANSFERASE IN nusA1

Host carrying	Units of CAT/OD650	Relative Value
1. Single copy prophage	0.02	1
2. Multicopy plasmid (pBR325)	0.20	10
3. λcl857cam	6.00	300

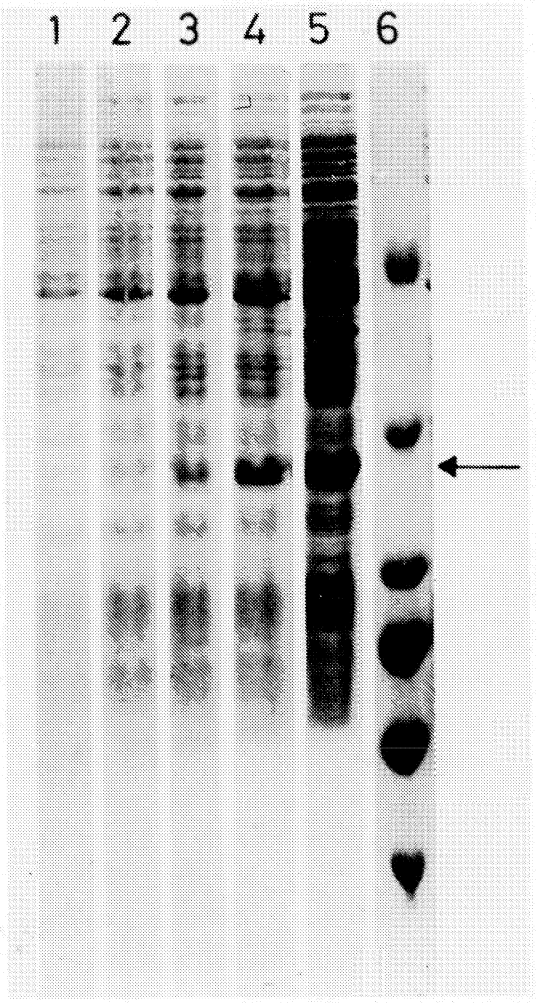


Fig.5. Synthesis of chloramphenicol acetyltransferase analyzed by polyacrylamide gel electrophoresis. Samples from the experiment shown in Fig. 4 were centrifuged and analyzed by polyacrylamide gel electrophoresis. The protein bands were visualised by staining with coomassie blue. The chloramphenicol acetyltransferase band is indicated by an arrow. Samples in slots 1 through 5 were taken at 0, 30, 60, 120, and 180 min after infection. Slot 6 contains protein molecular weight standards: 43,000, Ovalbumin; 25,700, α-chymotrypsinogen; 18,400, β-lactoglobulin; 14,300, Lysozyme; 12,300, Cytochrome C; and 6,200, Bovine trypsin inhibitor.

However, delayed cell lysis was observed.

The nusA negative phenotype is exhibited only at elevated temperatures. Infection at 30°C allows λ phage growth. Table 3 shows a comparison between nus$^+$, wild type parental strain, and the nusA1 strain at the permissive and restrictive temperatures.

In the nus$^+$ host rapid lysis by the phage is induced at 40°C. This is to be expected since λ Plac5 carries a temperature — sensitive repressor mutation. Our results demonstrate that even at the permissive temperature the nusA host allows excessive β -galactosidase expression.

The nusA mutation prevents N function. However, it is possible that the λN gene may affect the level of gene expression. In order to test this possibility, we constructed λPlac5 carrying the Nsus7 and Nsus53 amber mutations. Table 4 summarizes the results of the effect of the nusA and N mutations. It appears that extensive synthesis of β -galactosidase takes place upon infection with Plac5 carrying a defective N gene. The nusA condition is about 2-fold more effective for the synthesis of β -galactosidase. This is a significant improvement as the concentration of β -galactosidase rises from 30 to 60% of the total cellular proteins.

As might be expected, the influence of the N$^-$ mutations and the nusA$^-$ condition on the expression of β -galactosidase are not additive.

TABLE 3. SYNTHESIS OF β-GALACTOSIDASE IN *nus*$^+$ AND *nus*$^-$ HOSTS

Host	Units of β-galactosidase Temperature 30°	40°
nus$^+$	11,050	1,070
nus$^-$	70,110	134,280

Samples were taken 2 hr after infection.

TABLE 4. THE EFFECT OF nusA1 AND N MUTATIONS ON β-GALACTOSIDASE SYNTHESIS

Infecting Phage	β-Galactosidase Units *nus*$^+$	*nus*A1
λc1857P*lac*5	1,070	134,280
λc1857P*lac*5*Nsus*7*Nsus*53	72,600	125,770

Samples were taken 2 hr after infection.

The phage λ *trp lac*, in which the expression of β-galactosidase is under P_L control (14, 15), allowed us to follow the extent of transcription from P_L into the *lacZ* gene in a *nus*$^-$ host. We have previously demonstrated that transcription termination signals prevent β-galactosidase expression from P_L by λ *trp lac* in the absence of the *N* gene product. We found that the *trp lac* phage is incapable of directing extensive synthesis of β-galactosidase from P_L on

infection of the nusA host (data not shown). These findings demonstrate that no N antiterminal activity is present on infection of the nusA1 mutant.

The following experiments were done to test the importance of λ DNA replication on exorbitant gene expression. Two types of experiments were performed. It is known that the λ O and P gene products are essential for the initiation of λ DNA replication (1). A λ Plac5 phage carrying the Psus902 mutation was constructed by genetic cross with i434Psus902. The resulting phage cannot replicate on infections in which the mutation is not suppressed. No increase in β -galactosidase synthesis was observed on infection with this mutant phage (Table 5). These results show that λ DNA replication is required for the elevated expression of β -galactosidase.

Another type of experiment involved direct analysis of λ Plac5 replication following infection of the K95 nusA1 strain (Fig. 6). Total DNA (host and phage DNA) was purified at various times after infection. The purified DNA was cleaved with EcoR1 restriction enzyme,separated on agarose gel, and the bands visualized by ethidium bromide staining. DNA bands specific to the λ plac5 constitute a major part of the DNA. Further proof that the major bands are of λ origin was shown by hybridization to radioactive λ DNA probe (Fig. 6). These results allow us to conclude that excessive λ DNA

TABLE 5. THE EFFECT OF DNA REPLICATION ON β-GALACTOSIDASE SYNTHESIS

Infecting phage	β-Galactosidase units 1 hr	2 hrs
λcI857 plac5	13,092	24,787
λcI857 plac5 P sus 902	426	195

Samples were taken 1 and 2 hr after infection.

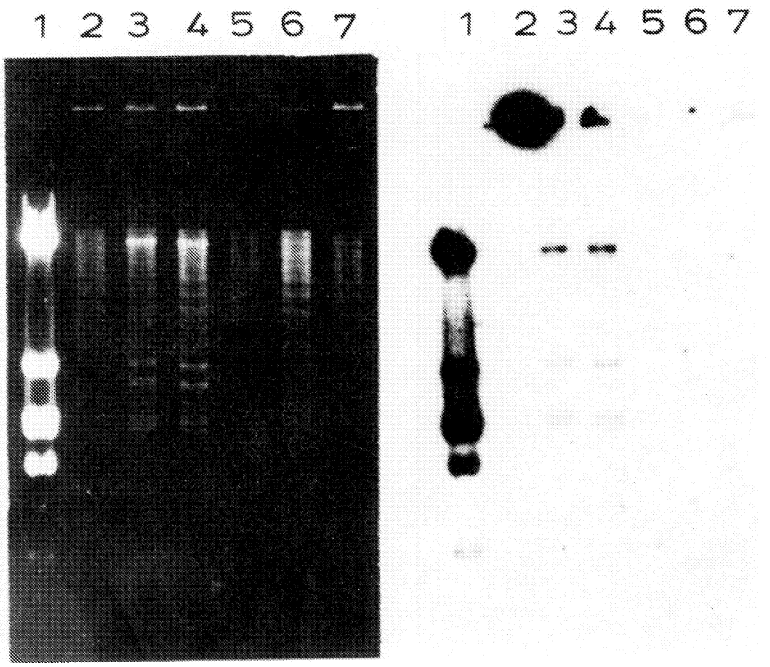


Fig. 6. DNA replication in K95 nusA. NusA cells were infected and at different times after infection, samples were removed for total DNA extraction. The DNA was digested with EcoR1 and the fragments were separated on an agarose gel. The gel was photographed (left panel), dried, and then hybridized to p^{32}-labled λ DNA (right panel). Slot no. 1: EcoR1 digest of λ DNA; slots 2-4: λ cI857Plac5 0, 60, 180 min after infection; slots 5-7: λ cI857Psus902Plac5 0, 60, 180 min after infection.

synthesis takes place following infection of the nusA host and that λ replication under these conditions requires the expression of the λ P gene.

For certain purposes, such as large-scale production, it may be desirable to use an induction protocol rather than infection by the phage. In order to test this procedure we constructed the K95nusA1 strain lysogenic for λ Plac5. As this phage carries a temperature-sensitive repressor, the culture can conveniently be induced by placing the growing cells at 42°C. Only low levels of β-galactosidase were made by this procedure. We suspect that the nusA condition prevents the synthesis of int and xis gene products necessary for the excision of the prophage. The defect in excision probably interferes with phage DNA replication.

We were able to circumvent this problem by integrating the phage in a plasmid. The plasmid used is pWR1, a pBR322 derivative carrying a phage attachment site (16). Cells lysogenic for λ Plac5 integrated into the plasmid can efficiently direct the synthesis of β-galactosidase following heat induction.

In the results presented above we analyzed a general system for achieving exorbitant expression of specific gene products. We demonstrated that this system allows DNA replication, transcription and translation of the cloned gene while inhibiting λ phage gene expression and thus preventing cell lysis. A novel cloning system

that allows efficient isolation of genes by use of antibody probes has recently been described (17). For this system an expression vector λ gtll has been constructed. The cloning site used for insertion of foreign DNA is located within the lacZ gene.

High level expression of β-galactosidase by the nusA system described above can be combined with the λ gtll system for the isolation of unknown native proteins of eukaryotic origin. We are currently developing a simplified way of combining these two systems.

ACKNOWLEDGMENTS

We thank Doctors S. Adhya and David Friedman for the bacterial and phage strains. We thank Mrs. Alisa Shilor for her help in bringing this manuscript to its final form. This work was supported in part by the Israel Academy of Sciences.

REFERENCES

1. Herskowitz,I. and Hagen,D. Annu. Rev. Genet. 14 : 399-445, 1980.
2. Ward, D. and Gottesman, M. Science 216 : 946-951, 1982.
3. Friedman, D., Jolly, C., and Mural,R. Virology 51 : 216-22, 1973.
4. Greenblatt, J. and Li, J. Cell 24 : 421-428, 1981.
5. Miller, J., Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1972.
6. Shaw, W.V., and Brodsky, R.F., J. Bacteriol. 95 : 28-36, 1968.
7. Moir, A. and Brammar, W. Molec. Gen. Genet. 149 : 87-99, 1976.
8. Hershey, A., (Ed.) Bacteriophage Lambda, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1981.

9. Miller, J. and Reznikoff, W. (Eds.) The Operon, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1978.
10. Chow, L. T. and Bukhari, A. I. In : DNA, Insertion Elements, Plasmids and Episomes (Eds. A.I. Bukhari, J. A. Shapiro and S. L. Adhya) Cold Spring Harbor Laboratory, Cold Spring Harbor, 1977, pp. 295-306.
11. MacHattie, L.A. and Jackowski, J.B. In : DNA, Insertion Elements, Plasmids and Episomes (Eds. A. I. Bukhari, J. A. Shapiro and S. L. Adhya) Cold Spring Harbor Laboratory, Cold Spring Harbor, 1977, pp. 219-228.
12. Gottesman, M.M. and Rosner, J.L. Proc.Nat. Acad. Sci. USA 72 : 5041-5045, 1975.
13. Sternberg, N. Virology 73 : 139-154, 1976.
14. Barnes, W. M., Siegal, R.B., Reznikoff, W.S. Molec. Gen. Genet. 129 : 201-215, 1974.
15. Belfort, M. and Oppenheim, A.B., Molec. Gen. Genet. 148 : 171-173, 1976.
16. Hsu, P., Ross,W. and Landy, A. Nature 285 : 85-91, 1980.
17. Young, R.A. and Davis, R.W. Proc. Natl. Acad. Sci. USA 80 :1194-1198, 1983.

17

EXPRESSION OF HERPES SIMPLEX VIRUS TYPE 1 AND TYPE 2 GLYCOPROTEIN D GENES USING THE ESCHERICHIA COLI lac PROMOTER

ROGER J. WATSON, JOHN H. WEIS[1], JOHN S. SALSTROM AND LYNN W. ENQUIST

Molecular Genetics, Inc., 10320 Bren Road East, Minnetonka, MN 55343, U. S. A.

SUMMARY

DNA fragments encoding herpes simplex virus type 1 and type 2 glycoprotein D (gD-1 and gD-2, respectively) have been inserted into plasmid vectors and expressed under the transcriptional control of the Escherichia coli lac promoter-operator. The proteins expressed in this system comprised gD sequences fused to a small bacteriophage λ Cro leader (i.e. Cro-gD). Such Cro-gD fusion proteins were found to be intrinsically unstable in E. coli and accumulated to low levels only. We found that fusion of this Cro-gD coding sequence to the 5' end of a sequence encoding β-galactosidase (β-gal), resulted in high levels of synthesis of Cro-gD-β-gal fusion proteins provided that certain carboxy-terminal gD coding sequences were deleted. These Cro-gD-β-gal proteins accumulated to levels comprising approximately 10% of the total cell protein and were found to form intracellular insoluble aggregates. Insertion of an in-phase amber nonsense codon between the gD and β-gal coding sequences, resulted in synthesis of both Cro-gD and Cro-gD-β-gal proteins in cells containing suppressors. Under these conditions, the Cro-gD amber fragment was stabilized by the Cro-gD-β-gal protein. Chimaeric gD-1 and gD-2 containing proteins were found to be immunologically active and induced antibodies in rabbits which immunoprecipitated authentic gD polypeptides and neutralized the infectivity of both virus types.

INTRODUCTION

Herpes simplex virus type 1 (HSV-1) and type 2 (HSV-2) are related double-stranded DNA containing viruses which cause a

[1] Present address: Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, U.S.A.

spectrum of persistent and latent infections in humans (1). Of these infections, that which has claimed most public attention is herpes genitalis, an often painful and recurrent disease of the urogenital tract. A number of laboratories have reported on the preparation of subunit (2,3,4,5,6), inactivated (7) and attenuated (8) vaccines derived from HSV-1-or HSV-2-infected tissue culture cells. These various vaccines have proven to induce immunity in animals, usually mice, against subsequent HSV-1 and HSV-2 challenge. Recent evidence indicates that such vaccines may be effective, used prophylactically in humans, in limiting the horizontal transfer of herpes genitalis (3,9). Such vaccines may also be effective therapeutically, in lessening the incidence and severity of recurrences of the disease (6,9,10). Despite these positive indications of the benefit of an HSV vaccination program, no such vaccine is licensed for use in the United States, nor is a vaccine in general use in other parts of the world. The reasons for this lack of a vaccination program are manifold, and include problems in obtaining sufficient quantities of a suitable antigen from tissue culture, and in ensuring safety of the product, both from the standpoints of lack of infectivity and of oncogenicity. As an alternative to tissue-culture derived vaccines, we are using recombinant DNA technology to prepare HSV immunogens in microorganisms. It is hoped that, eventually, a recombinant DNA HSV vaccine will satisfy the criteria of safety, cost-effectiveness and clinical efficacy.

It has been established that antisera directed specifically against each of the four membrane envelope glycoproteins of the virus (gB, gC, gD and gE) may neutralize, *in vitro*, infectivity of the homologous HSV type (11,12). Antisera directed against gB and gD neutralize, also, the infectivity of the heterologous HSV type (11). A number of authors have reported that gD is perhaps the major immunogen of the virus, and induces high levels of type-common neutralizing antibodies (13,14). In our initial experiments to produce a subunit HSV vaccine, we have chosen to express the gD genes of HSV-1 and HSV-2 (gD-1 and gD-2 genes, respectively) in *Escherichia coli*. Bacterially produced gD-related proteins have been tested for their ability to induce in

animals an immune response against the virus.

MATERIALS

Expression vectors

The components of the basic plasmid vector, pJS413, used to express gD sequences in E. coli, are represented in Figure 1. Sub-segment a is the 2293 base pair (bp) EcoRl-PvuII fragment from plasmid pBR322 (15) and contains the β-lactamase gene (amp) and the colEl origin of replication (ori). The amp gene confers ampicillin resistance upon transformation of bacteria by the

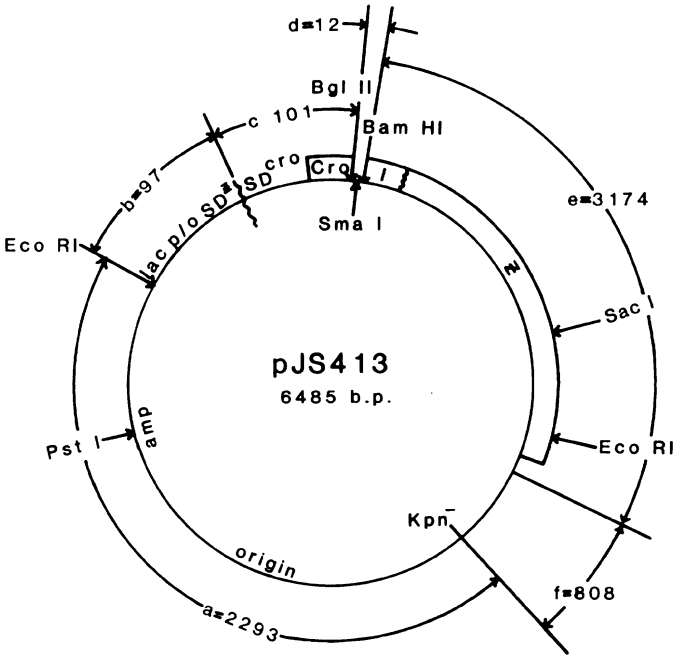


FIGURE 1. Representation of the plasmid expression vector pJS413 (not drawn to scale). The various components of this plasmid are detailed in the text (Materials).

plasmid. Sub-segment b (97 bp) is the AluI fragment from the E. coli lac operon (16) and contains the lac promoter-operator (lac p/o) region and the lacZ ribosome binding site (SD^Z), the so-called Shine-Dalgarno sequence (17). Sub-segment c (101 bp) is from bacteriophage λ, and contains the cro gene Shine-Dalgarno sequence (SD^{cro}) and the first 68 bp of the cro coding sequence (Cro; 16). Subsegment d (12 bp) is a synthetic DNA linker containing BglII, SmaI and BamHI restriction sites. Subsegment e (3174 bp) is a fusion of a small portion of the lacI coding sequence with all but the first 69 bp of the lacZ coding sequence (18). This sequence encodes a functional β-galactosidase (β-gal) activity. Subsegment f (808bp) is from the bacteriophage λ lom gene (18).

The features of the pJS413 structure pertinent to its use as an expression vector are as follows. The cro coding sequences are fused to the lacIZ coding sequence such that they are not in the same translational phase. Insertion of a coding region (for example gD) at one of the three cloning sites (BglII, SmaI and BamHI) contained by subsegment d allows one to correctly phase the cro and lacIZ sequences. It is necessary to ensure, of course, that the inserted coding sequence is in the correct translational phase with those of cro and lacIZ. Translation of this tetra-partite coding sequence should result in synthesis of a chimaeric protein (i.e. Cro-gD-β-gal) having functional β-galactosidase activity. The synthesis of this chimaeric protein can easily be determined by assaying for β-galactoside activity on indicator plates (19). Translation of the chimaeric protein would utilize the SD^{cro} sequence, which is presumed to be a strong ribosome binding site on the basis of its complementarity to the 3' terminus of 16S rRNA (17).

The coding sequences contained by subsegments c and e of pJS413, are under the transcriptional control of lac p/o. The pJS413 lac promoter bears the UV5 mutation that allows high-level transcription initiation in the absence of the E. coli catabolite activator protein binding site. Transcription from the lac promoter is subject to repressor regulation by binding of the lac repressor (the product of the lacI gene) to the operator site.

In bacteria which overproduce the *lac* repressor (those bearing the *lacI*^{Q} mutation), the plasmid *lac* promoters are transcriptionally silent unless induced with allolactose. This provides a convenient way in which synthesis of the β-galactosidase fusion protein may be controlled.

Other expression vectors which have been used for expression of gD (i.e. pHK414) are derivatives of pJS413. These derivatives contain altered subsegment *d* sequences, such that alternative cloning sites were introduced and the relative translational phasing of *cro* and *lacIZ* coding sequences were altered.

The glycoprotein D Genes

The gD-1 and gD-2 genes were first mapped on their respective virus genomes by analysis of HSV-1 x HSV-2 intertypic recombinants (20,21). The two genes were found to be essentially colinear and to map in the short unique (U_S) region of the HSV genome between 0.90-0.945 map units. We have subsequently characterized the gD-1 mRNA and have completely sequenced the gD-1 and gD-2 coding regions (22,23, R.J. Watson, manuscript submitted). We found that gD-1 was encoded by a 3.0 kilobase unspliced mRNA. That the gD-1 gene contained no introns was an important finding, as it allowed us to approach gD expression by cloning genomic DNA directly in an expression vector, rather than having to isolate cDNA first.

Restriction maps of the gD-1 and gD-2 coding sequences are shown in Figure 2. From the DNA sequences, it was deduced that unmodified gD-1 and gD-2 were of 394 and 393 amino acids (aa) in size, respectively. A comparison of deduced amino acid sequences of gD-1 and gD-2 is shown in Figure 3. It has recently been found that both mature proteins initiate with lysine at position 26 (R.J. Eisenberg, D. Long, R. Angeletti and G.H. Cohen, submitted for publication). We presume, therefore, that the amino-terminal 25 amino acids function as a signal peptide for membrane insertion (24) and are removed during translation. A further feature noted in the gD-1 and gD-2 amino acid sequence is a 25 aa hydrophobic region located 30 aa from the carboxy termini (Figs. 2 and 3). This region has the characteristics of a

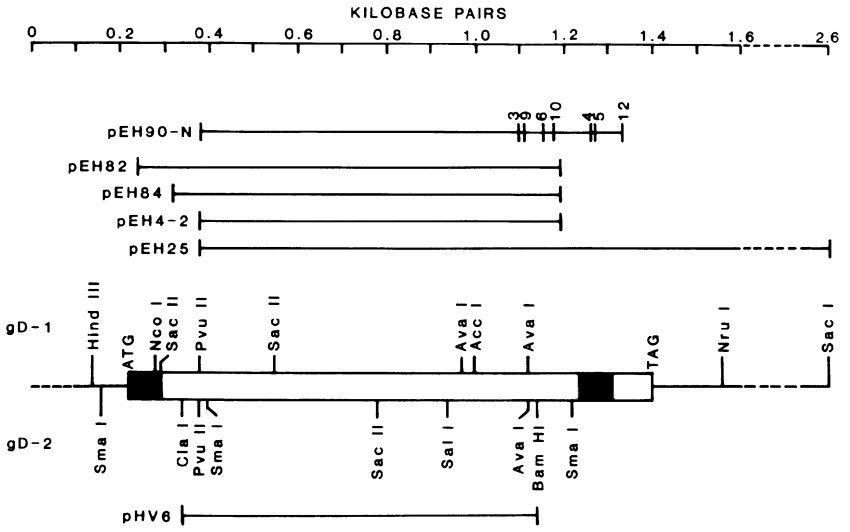


FIGURE 2. Restriction maps of the gD-1 and gD-2 genes. The box represents the coding sequence of these genes: the solid portion of this box indicates the signal peptide sequence and the cross-hatched area the transmembrane sequence. The horizontal bars above and below the restriction maps indicate the gD-1 and gD-2 sequences contained by the various expression plasmids.

transmembrane sequence (25). The carboxy-terminal 30 aa are highly charged and presumably function to anchor the glycoprotein in the membrane. If these assumptions are correct, 314 aa of gD-1 and gD-2 (aa 26-339) would be positioned external to the virion envelope. We assume that the external portions of gD-1 and gD-2 alone contain the antigenic determinants recognized by the immune system. In development of an HSV vaccine, therefore, greater emphasis was placed on synthesis of the external gD region in E. coli, rather than on expression of the entire coding region.

```
          signal peptide
          ____________________
gD-1  MGGTAARLGAVILFVVIVGLHGVRGKYALADASLKMADPNRFRGKDLPVLDQLTDPPGVR  60
gD-2  --RLTSGV-TAA-L--A---RV-CA------P-------------N-------------K

gD-1  RVYHIQAGLPDPFQPPSLPITVYYAVLERACRSVLLNAPSEAPQIVRGASEDVRKQPYNL 120
gD-2  ------PS-E-------I------------------H-------------DEA--HT---
                                                                 ___

gD-1  TIAWFRMGGNCAIPITVMEYTECSYNKSLGACPIRTQPRWNYYDSFSAVSEDNLGFLMHA 180
gD-2  ----Y---D--------------P------V---------S--------------------
      _                 ___

gD-1  PAFETAGTYLRLVKINDWTEITQFILEHRAKGSCKYALPLRIPPSACLSPQAYQQGVTVD 240
gD-2  -------------------------------RA------------A---TSK---------

gD-1  SIGMLPRFIPENQRTVAVYSLKIAGWHGPKAPYTSTLLPPELSETPNATQPELAPEDPED 300
gD-2  -----------------L------------P------------D-T-------V------
                                                    ___

gD-1  SALLEDPVGTVAPQIPPNWHIPSIQDAATPYHPPATPNNMGLIAGAVGGSLLAALVICGI 360
gD-2  -------A---SS-------------V-^-H-A--A-S-P---I--LA--T------G-- 359
                                                  ___________________

gD-1  VYWMHRRTRKAPKRIRLPHIREDDQPSSHQPLFY  394
gD-2  AF-VR--AQM----L------D--A-P-------  393
      ____
```

FIGURE 3. Comparison of the amino acid sequences of gD-1 and gD-2. The single letter amino acid code is used. Where the amino acids of gD-1 and gD-2 are identical, a dash represents the gD-2 amino acid. The symbol (^) indicates the position where there is no equivalent gD-2 amino acid. The three potential glycosylation sites (NXS and NXT) are underlined, as is the putative transmembrane sequence (aa 341-364).

RESULTS

Expression of a Cro-gD protein

Initially, we used the expression vector pJS413 to test whether we could obtain synthesis of a Cro-gD fusion protein in E. coli. In our construction of the Cro-gD expression plasmid pEH25, we took advantage of the PvuII site located 52 codons 3' to the start of the gD-1 coding sequence (Fig. 3). Ligation of the 2.2 kbp PvuII-SacI HSV-1 DNA fragment (Fig. 3) into the SmaI and SacI sites of pJS413 (Fig. 2) resulted in deletion of

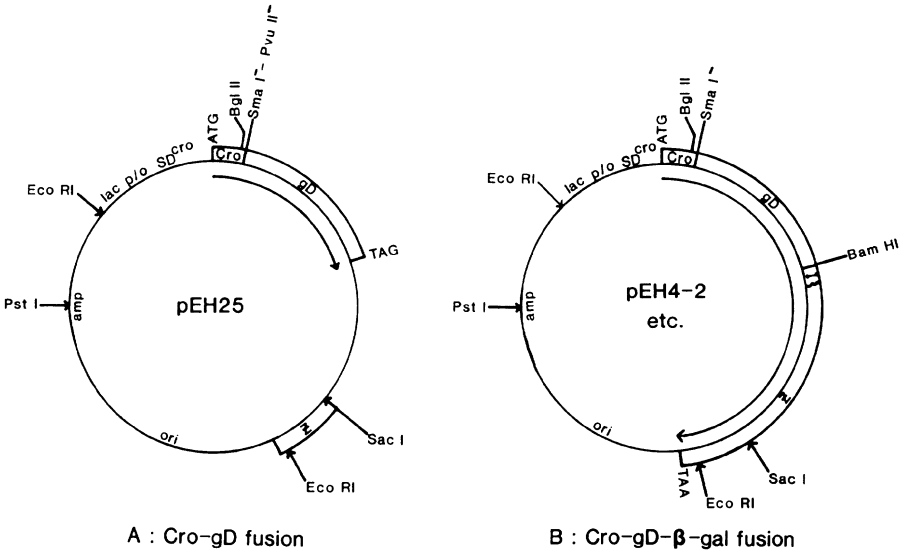


FIGURE 4. Representations of (a) an expression plasmid which specifies a Cro-gD protein (i.e. pEH25) and (B) one which specifies a Cro-gD-β-gal protein (i.e. pEH4-2). The coding sequences which specify the Cro, gD and β-gal portions of these proteins are drawn as open circular boxes. The direction of transcription from the lac promoter (lac p/o) is indicated by the circular arrows, as is the portion of these transcripts which is translated into the respective protein.

sequences encoding the amino-terminal 52 aa of gD-1. As shown in Fig. 4A, this construction brought the remaining 342 gD-1 codons in-frame with 24 codons specifying the amino terminus of Cro. Ligation at the SacI termini resulted in deletion of roughly two-thirds of the I-Z coding sequence. It was predicted that the Cro-gD-1 protein specified by pEH25 would terminate at the natural gD-1 nonsense codon (TAG) and would contain 366 aa; i.e. be of approximately 40300 molecular weight (mw).

To test for expression of a Cro-gD-1 protein, pEH25 was used to transform E. coli NF1829 (a strain which overproduces the lac repressor). Transformants were labelled using ^{35}S-methionine

either in the presence or absence of isopropyl-β-D-thiogalactopyranoside (IPTG), an inducer of the lac promoter. Labelled polypeptides were extracted from induced and uninduced cells, and immunoprecipitation experiments were carried out using

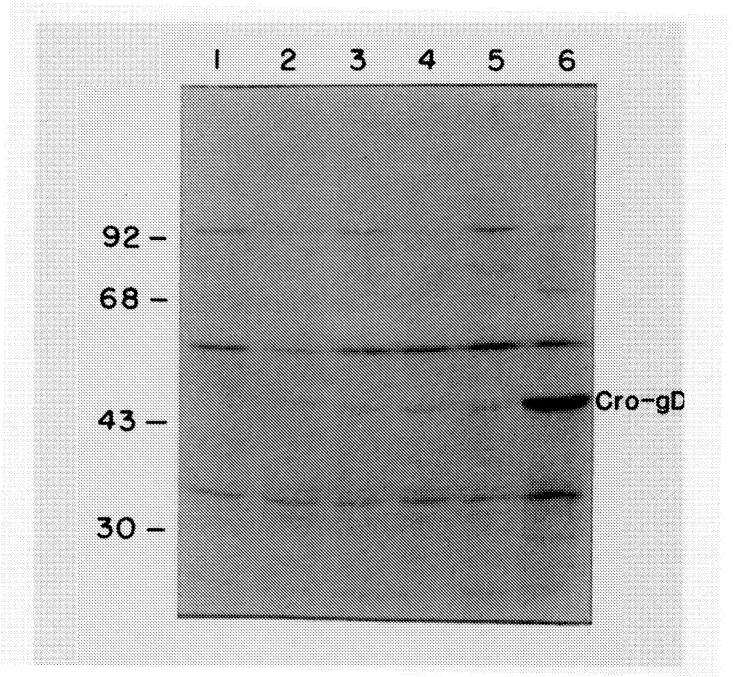


FIGURE 5. Immunoprecipitation of the Cro-gD-1 protein specified by pEH25. Duplicate 1ml cultures of pEH25 transformants in M-9 broth (32) were grown to mid-log phase. IPTG was then added to a final concentration of 1 mM to one of the cultures, and [^{35}S]-methionine was added (50 μCi/ml) to both. After labelling for 60 min. at 37°C, cells were collected by centrifugation, lysed, and the lysates subjected to immunoprecipitation as described previously (23). The immunoprecipitated proteins were analyzed by SDS-PAGE and fluorography. Uninduced (lanes 1, 3 and 5) or IPTG-induced cultures (lanes 2, 4 and 6) were incubated with no added antibody (lanes 1 and 2), serum from unimmunized rabbits (lanes 3 and 4) and rabbit antiserum to HSV-1 (lanes 5 and 6). To the left of the fluorogram are indicated the electrophoretic mobilities of mw standards (92, 68, 43 and 30k). To the right is indicated the position of the Cro-gD-1 protein in lane 6. Reproduced with permission from reference 23: Copyright 1982 by the American Association for the Advancement of Science.

sera from rabbits immunized with HSV-1 virions. As controls, identical immunoprecipitation reactions were performed using unimmunized rabbit serum and no added serum. Immunoprecipitates were analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). It was found that a polypeptide of apparent mw 46000 was specifically precipitated from induced cell lysates by HSV-1-immunized rabbit sera (Fig. 5, lane 6). This protein was not precipitated using control serum (Fig. 5, lane 4). Other experiments have indicated that the 46000 mw protein reacts specifically with a number of monoclonal antibodies (1S, 4S, 55S and 57S) directed against gD-1 (23). These results indicate that a Cro-gD-1 protein was made in pEH25 transformants under the transcriptional control of the *lac* promoter-operator. However, we estimated that this protein represented less than 0.1% of the total cell protein. To obtain a greater level of expression of gD-1 sequences, fusion of the gD-1 coding region to that specifying β-gal was made.

Expression of Cro-gD-β-Gal proteins

A number of plasmids have been constructed which contain gD DNA fused in the region of the carboxy-terminal coding sequence to a β-gal coding sequence. A prototype of these plasmids is pEH4-2 (described in detail in reference 26). The generalized structure of pEH4-2, and of other plasmids encoding a Cro-gD-β-gal protein, is represented in Fig. 4B. In constructing pEH4-2, pEH25 was opened at the *Nru*I site 161 base pairs (bp) 3' to the gD-1 termination codon (Fig. 2), and was processively digested with exonuclease *Bal*31. The *Bal*31 digestion was used to produce a population of DNA fragments which were deleted for the termination codon and approximately a further 40- 240 bp of carboxy-terminal gD-1 coding sequence. The *Bal*31 treated gD-1 coding sequence was ligated to a DNA fragment containing the β-gal coding sequence and used to transform *E. coli* NF1829. Ampicillin-resistant colonies were isolated and tested for -galactosidase activity by replica plating on lactose MacConkey agar indicator plates (19). Isolates producing red colonies on the indicator plates were further tested for synthesis of a

chimaeric Cro- gD-1-β-gal protein by SDS-PAGE analysis of IPTG-induced polypeptides. Of 15 isolates tested, only that containing pEH4-2 was found to produce, on induction, significant quantities of a polypeptide of appropriate size (160,000 mw). By DNA sequence analysis it was found that 205 bp of carboxy-terminal gD-1 coding sequence was deleted in pEH4-2. Hence, the pEH4-2 Cro- gD-1-β-gal protein lacked 69 carboxy-terminal amino

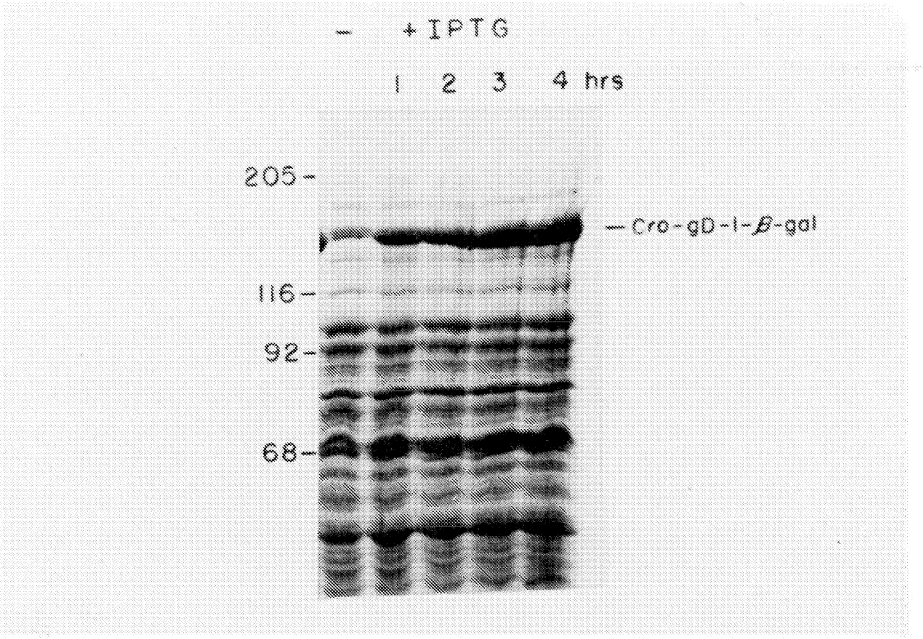


FIGURE 6. Accumulation of the Cro-gD-1-β-gal protein specified by pEH4-2. An overnight culture of pEH4-2 transformants was used to inoculate fresh L broth (in duplicate) at a ratio of 1/20. After incubation at 37°C for 90 min one of the cultures was induced by addition of IPTG to 1 mM. Both cultures were then incubated further at 37°C. 1 ml samples were taken from the induced culture at 1, 2, 3 and 4 hr. after induction, and from the uninduced culture at the latest time point. Cells were collected by centrifugation and lysed by boiling in SDS-PAGE sample buffer (33). Approximately 1/10 of each sample was analyzed by SDS-PAGE (7% gel). The proteins were visualized by Coomassie blue staining. To the left of the gel are indicated the positions of mw standards. The 160000 mw Cro-gD-1-β-gal protein is indicated.

acids of gD-1. This region deleted in pEH4-2 includes the entire transmembrane and anchor sequences of gD-1 (Fig. 2).

The chimaeric pEH4-2 gene contains 273 codons specific for gD-1, thus representing 74% of the mature gD-1 polypeptide sequence. By immunoprecipitation, it was determined that the pEH4-2 chimaeric protein was recognized by the 4S monoclonal antibody (26), demonstrating that the type-common neutralizing epitope of gD defined by this monoclonal (27) was retained in this construction. As shown in Fig. 6, the pEH4-2 chimaeric protein accumulated in cells following induction, and was found to represent as much as 20% of the total cell protein 4 hr after addition of IPTG. We found that disruption of IPTG-induced pEH4-2 transformants by lysozyme and non-ionic detergent treatment resulted in fractionation of the Cro-gD-1-β-gal protein with the insoluble cellular material. This suggested that the pEH4-2 chimaeric protein formed insoluble aggregates within E. coli which may contribute greatly to its stability. As described for insulin-β-gal fusions (28), electron microscopy of IPTG-induced pEH4-2 transformants revealed the presence of large cytoplasmic inclusion bodies (D. Steinberg, personal communication). The insoluble nature of this material provides a relatively simple way to greatly purify the chimaeric protein, as most E. coli proteins can be removed by disruption of cells and repeated washing. These three attributes of the pEH4-2 chimaeric protein, namely, the presence of important antigenic determinants, high yield and ease of purification, make this protein a potentially useful HSV immunogen. The antigenic properties of this protein are described later.

As described above, of 15 isolates tested only the pEH4-2 transformant was found to produce large quantities of a chimaeric protein. The other isolates were found to produce at least 10-fold lower amounts of this protein and to contain plasmids in which only 30-70 bp of the carboxy-terminal gD-1 coding sequence was deleted by Bal31 digestion. These data indicated that deletion of carboxy-terminal gD-1 coding sequences greater than 70 bp from the termination codon was required for efficient expression of a chimaeric protein. To investigate this

phenomenon further, additional plasmids containing fused gD-1 and β-gal coding sequences were constructed (the pEH90-N series). The construction of the pEH90-N plasmids was similar to that described for pEH4-2, except that Bal31-treated gD-1-containing DNA fragments were ligated with a β-gal encoding gene derived from pHK414. In pHK414, subsegment d of pJS413 was replaced by a linker sequence containing BglII-HindIII-SmaI and BamHI restriction sites. The reason for using this vector will become apparent in a later section. The carboxy-terminal gD-1 sequences deleted in 12 pEH90-N plasmids isolated were determined by DNA

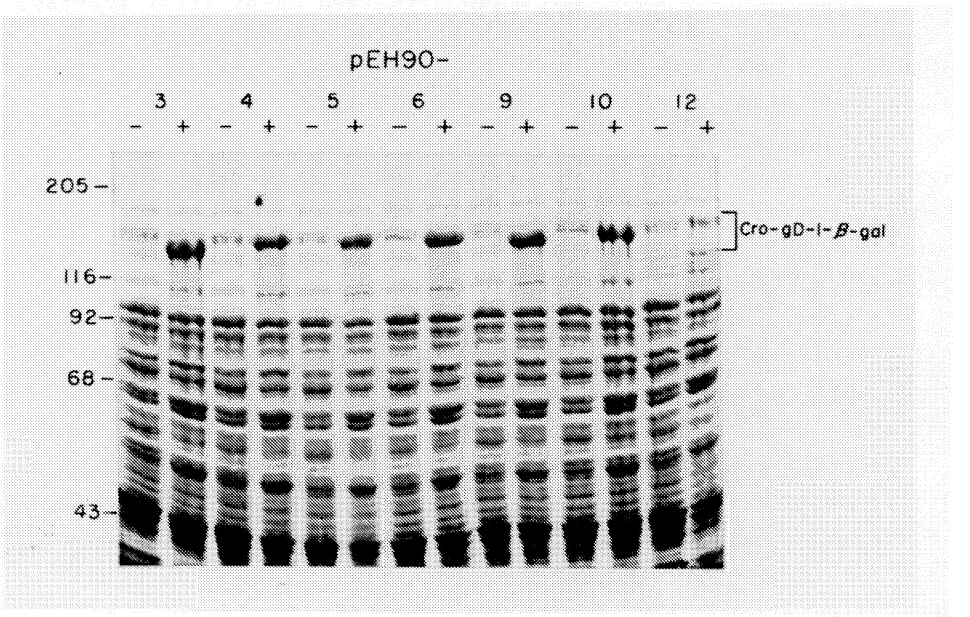


FIGURE 7. Synthesis of the Cro-gD-1-β-gal proteins specified by the pEH90-N series. Duplicate 1 ml L broth cultures, inoculated with 1/20 volume of overnight cultures of pEH90-3, -4, -5, -6, -9, -10 or -12 transformants, were incubated for 90 min at 37°C. IPTG to 1 mM was then added to one of each duplicate and incubation was continued for a further 4 hr. at 37°C. Proteins in uninduced (-) and induced (+) lysates were analyzed as described in the legend to Fig. 6. The variable size Cro-gD-1-β-gal protein is indicated.

sequencing. The gD-1 coding sequences contained by a number of pEH90-N plasmids are represented in Fig. 2.

Polypeptides extracted from IPTG-induced and uninduced cultures of a number of different pEH90-N transformants were analyzed by SDS-PAGE. Coomassie blue staining of the gel (Fig. 7) indicated that the size of the Cro-gD-1-β-gal protein produced upon induction of these isolates varied as predicted from the size of the gD-1 inserts. Hence, pEH90-3 produced the smallest and pEH90-12 the largest chimaeric protein (pEH90-12 carries a deletion of 68 carboxy-terminal gD-1 nucleotides). It was apparent also that whereas pEH90-12 made very small quantities of this protein, these other plasmids specified significant amounts. Of these latter plasmids, pEH90-5 contained the largest gD-1 insert (Fig. 2). This plasmid carries a deletion of 131bp (44 codons) of gD-1 carboxy-terminal coding sequence. It is apparent, then, that at least 11 aa of the hydrophobic transmembrane sequence is tolerated for efficient expression of a Cro-gD-1-β-gal protein. Sequences deleterious for efficient expression must lie in the remainder of the transmembrane sequence or in the anchor sequence.

The gD inserts of some other plasmids specifying Cro-gD-1-β-gal fusion proteins are shown in Fig. 2. Plasmids pEH82 and pEH84 were derived from pEH4-2 and contain additional amino-terminal gD-1 coding sequences. Plasmid pEH82 specifies a protein which includes all but the amino-terminal 6 aa of the signal peptide. Plasmid pEH84 specifies a chimaeric protein lacking 32 aa at the amino-terminus, i.e. it lacks 7 aa of the mature gD-1 amino-terminus. These plasmids express comparable amounts of Cro-gD-1-β-gal proteins to that observed with pEH4-2. Hence, sequences at the amino-terminus do not appear to have such a dramatic influence upon the level of expression of the chimaeric protein as sequences at the carboxy-terminus.

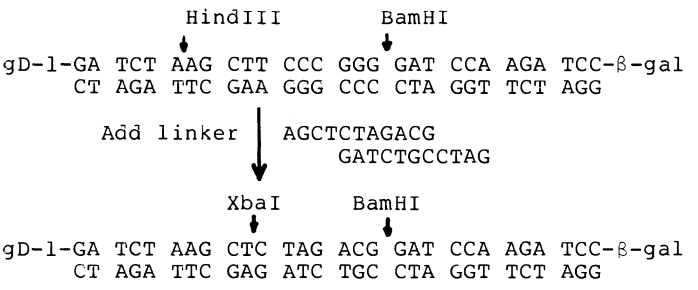
Expression of a chimaeric protein containing gD-2 coding sequences was obtained by inserting a ClaI-BamHI fragment of HSV-2 DNA in the SmaI and BamHI sites of pHK414. The resultant plasmid, pHV6, specifies at high levels (up to 10% of the total cell protein) a Cro-gD-2-β-gal protein which contains 267 aa of

gD-2. This fusion protein lacks 15 aa of mature gD-2 amino-terminal sequence and 86 aa of carboxy-terminal sequence (Figs. 2 and 3). The gD-2 sequences expressed in pHV6 include most of the coding sequences which are strongly conserved between gD-1 and gD-2 (Fig. 3).

High-level expression of a Cro-gD-1 protein

It appeared to us that it would be of greater utility to have a system in which proteins unfused to β-gal could be synthesized and accumulated in E. coli at high levels. To achieve this with gD-1, two factors had to be considered. First, it seemed to be necessary to remove carboxy-terminal coding sequences for efficient expression. Second, a means had to be found to stabilize the Cro-gD-1 protein once synthesized. If the Cro-gD-1 protein were formed, in the presence of Cro-gD-β-gal, it may be reasonable to assume that it would be protected from proteolysis. To satisfy these two aims, we attempted to express in E. coli a truncated gD-1 sequence (lacking the deleterious carboxy-terminal sequences) in concert with a Cro-gD-1-β-gal protein which was known to form intracellular aggregates. To do this, the following construction was made.

Plasmid pEH90-10 (which lacks 218 carboxy-terminal gD-1 nucleotides) was digested with HindIII and BamHI. Both enzymes cut uniquely in the vector linker sequence (subsegment d) which fuses the gD-1 and β-gal coding sequences. A linker sequence containing an XbaI site and, more importantly, an amber (TAG) nonsense codon was then inserted (see below):



The resulting plasmid, designated pEH90-10am, has the amber codon inserted between, and in-phase with, the gD-1 and β-gal coding sequences (see Fig. 4B).

It was predicted that E. coli NF1829 transformed with pEH90-10am should synthesize predominantly a truncated Cro-gD-1 protein. Some Cro-gD-1-β-gal protein may also be made as the amber codon is not a strong terminator of translation in bacteria. Indeed, it was found that these transformants gave a pink phenotype on lactose MacConkey agar plates (pEH90-10 transformants give red colonies), demonstrating a lowel level of β-galactosidase activity. It was further predicted that introduction of pEH90-10am into an amber suppressor cell background would result in elevated synthesis of Cro-gD-1-β-gal and continued synthesis of Cro-gD-1, the relative proportions of these two proteins depending upon the level of suppression (amber suppression is always significantly less than 100%). To test this prediction, two approaches were followed. First, pEH90-10am transformants of E. coli NF1829 were infected with a lysogenic transducing phage, φ80 pSuIII. This phage carries the supF amber suppressor gene. Lysogenised colonies were isolated and found to give red colonies on lactose MacConkey agar plates. Secondly, pEH90-10am was used to transform E. coli LE392, a strain that contains two amber suppressor genes, sup E and sup F. These cells were not tested on indicator plates as LE392 contains a functional lacZ gene.

The proteins produced by the above transformant were analyzed as follows. Duplicate 1 ml L broth cultures were inoculated (1/50) with NF1829 transformed with pEH90-10 (pEH90-10), pEH90-10am NF1829 transformants (pEH90-10am), pEH90-10am NF1829 transformants lysogenized with φ80 pSuIII (pEH90-10am SuIII) and pEH90-10am LE392 transformants (pEH90-10am LE392). To one of each duplicate was added IPTG to 1 mM. After incubation at 37°C for 7 hr, bacteria were collected by centrifugation, lysed and samples of the proteins (approximately 7% of the total) were analyzed by SDS-PAGE. It was observed that pEH90-10 cultures induced with IPTG, and both induced and uninduced pEH90-10am LE392 cultures, expressed high levels (approximately 5% of the

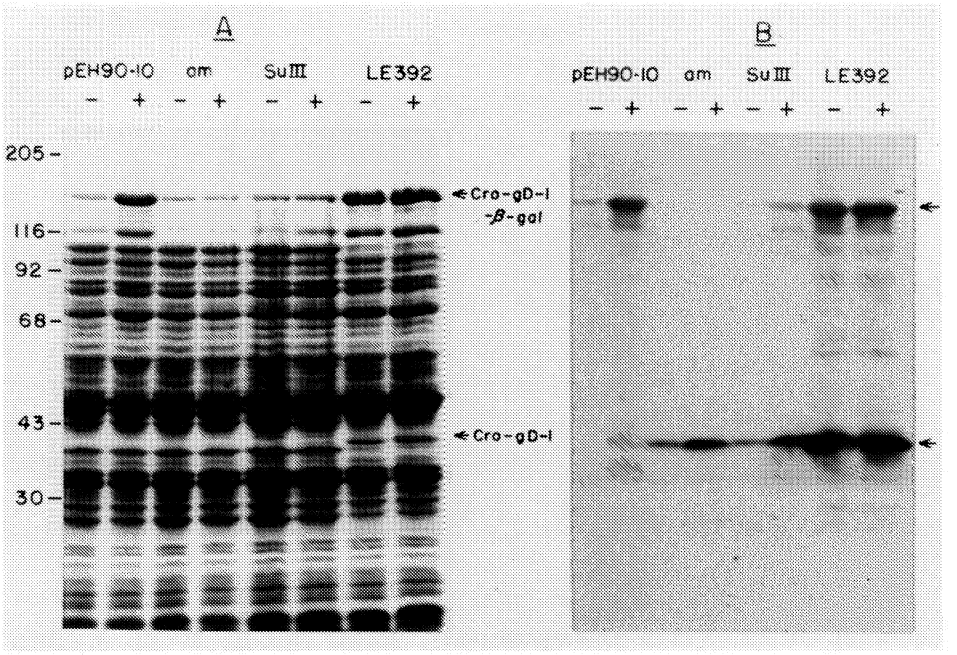


FIGURE 8. Synthesis of the Cro-gD-1 amber fragment. E. coli NF1829 cultures transformed with pEH90-10 (pEH90-10) or pEH90-10am (am), E. coli NF1829 pEH90-10am transformants lysogenized with bacteriophage ϕ80 pSuIII (SuIII) and E. coli LE392 transformed with pEH90-10am (LE392), were incubated for 7 hr at 37°C following inoculation (1/50) with overnight cultures either in the absence (-) or presence (+) of 1 mM IPTG. After incubation, the cells were lysed and run on duplicate 7% - 15% gradient SDS-PAGE gels as described in the legend to Fig. 6. One of the gels was stained with Coomassie blue (A). The positions of mw markers and of the Cro-gD-1-β-gal and Cro-gD-1 proteins are indicated. Proteins resolved on the duplicate gel (B) were transferred to nitrocellulose (29) and reacted successively with anti-HSV-1 rabbit serum and ^{125}I-labelled goat anti-rabbit IgG antibodies essentially as described by Towbin et al. (34). The blot was subjected to autoradiography overnight. Indicated by arrows are the positions of the Cro-gD-1-β-gal and Cro-gD-1 proteins which reacted with the anti-HSV-1 serum.

total cell protein) of the Cro-gD-1-β-gal protein (Fig. 8A). The induced pEH90-10am SuIII cultures produced lesser amounts of this protein. It was apparent also, that pEH90-10am LE392 cultures

synthesized a protein of molecular weight 38000 which was not observed in these other cultures (Fig. 8A). Cultures of LE392 which were not transformed with pEH90-10am were found not to make this 38000 mw protein. We concluded that the 38000 mw protein was likely to be Cro-gD-1. Neither synthesis of this protein nor the chimaeric protein was repressed in the absence of an inducer, as E. coli LE392 does not contain the lacI^{Q} mutation and thus makes insufficient lac repressor to control transcription from multicopy plasmid genes. The putative Cro-gD-1 product represented approximately 1.5% of the total cell protein.

To test the authenticity of the putative Cro-gD-1 protein expressed by pEH90-10am LE392, proteins from the various cultures were separated by SDS-PAGE as before. These proteins were then transferred electrophoretically to nitrocellulose, using the so-called Western procedure (29). The immobilized proteins were reacted successively with rabbit anti-HSV-1 serum and ^{125}I-labelled goat anti-rabbit IgG antibodies. The blot was then washed and subjected to autoradiography. It was found that the 38000 mw protein observed with pEH90-10am LE392 reacted specifically with anti-HSV-1 antibodies (Fig. 8B). In addition, the Cro-gD-1 protein was present in lesser amounts with pEH90-10am and pEH90-10am SuIII. In these cultures, the protein was more abundant following IPTG-induction (in the absence of an inducer, transcription from the lac promoter is greatly reduced, but not entirely shut off). The 38000 mw protein was not observed with pEH90-10.

The results above indicate that the Cro-gD-1 protein is produced at high level in LE392 transformed with pEH90-10am. To test whether this protein was present as an aggregate, 100 ml cultures of pEH90-10, pEH90-10am and pEH90-10am LE392 were IPTG-induced and grown to stationary phase. The cells were collected by centrifugation and disrupted by lysozyme treatment followed by the addition of Triton X100 to 1%. The cells were further disrupted by sonication such that the turbidity decreased and viscosity due to released DNA was reduced. The suspensions were then layered in centrifuge tubes onto 25% sucrose pads, and the aggregated proteins were separated

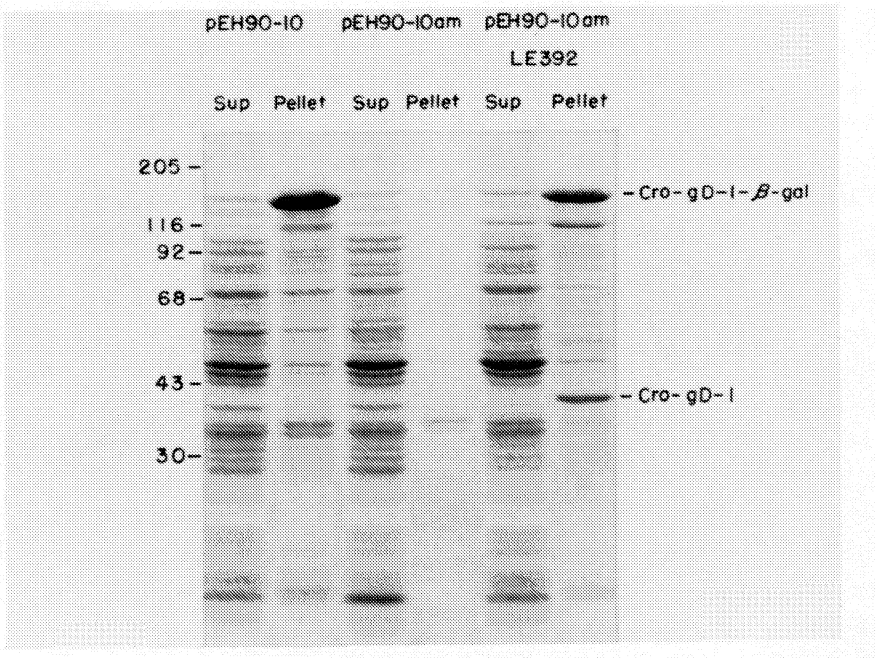


FIGURE 9. Overnight cultures (2 ml) of E. coli NF1829 transformed with pEH90-10 or pEH90-10am and E. coli LE392 transformed with pEH90-10am were used to inoculate 100 ml of fresh L broth containing 100 µg/ml ampicillin. The cultures were incubated at 37°C for 7 hr following addition of IPTG to 1 mM. The cells were collected by centrifugation at 10,000 g for 10 min and resuspended in 5 ml lysis buffer (200 mM NaCl, 2 mM EDTA, 2 mM 2-mercaptoethanol, 50 mM Tris, HCL pH 7.9). 2 mg of lysozyme were then added and the samples were incubated on ice for 20 min. To lyse the cells, Triton X100 was added to a final concentration of 1%. The cells were further disrupted by sonication, such that the viscosity of the samples was greatly reduced. The samples were layered onto 6 ml 25% sucrose in lysis buffer +0.5% Triton X100 contained in Beckman SW28 centrifuge tubes. The samples were centrifuged at 27k for 30 min, the supernatant was taken off and the pellet was resuspend in 2 ml STE (100 mM NaCl, 1 mM EDTA, 10 mM Tris, HCl pH 7.9) by light sonication. 10 µl of each supernatant fraction and 2 µl of each pellet fraction were boiled in sample buffer and subjected to SDS-PAGE (7% - 15% gels). Following electrophoresis, the proteins were visualized by Coomassie blue staining. Indicated are the positions of mw standards and the location of the Cro-gD-1-β-gal and Cro-gD-1 proteins.

from soluble proteins by centrifugation. The pelleted material (containing the aggregates) was then compared to the soluble material present in the supernatant by SDS-PAGE analysis. Equivalent amounts of each fraction were loaded on the gel. Coomassie blue staining of the gel (Fig. 9) revealed that, as expected, the pEH90-10 and pEH90-10am LE392 Cro-gD-1-β-gal proteins were present predominantly in the pelleted material. Additionally, it was found that the Cro-gD-1 protein produced by pEH9-10am LE392 was present almost exclusively, in the pelleted material (Fig. 9). This Cro-gD-1 protein could not be seen in either the soluble or pelleted material from the pEH90-10am culture. We concluded, that the Cro-gD-1 protein was stabilized by the chimaeric Cro-gD-1-β-gal protein. It was estimated on the basis of this experiment, that 1 litre of pEH9-10am LE392 cells would yield approximately 70 mg of the Cro-gD-1 protein and 210 mg of the Cro-gD-1-β-gal protein.

Immune response to the recombinant gene proteins

We have previously reported that the pEH4-2 chimaeric protein elicits an antibody response in rabbits (26). This protein was prepared as an immunogen by SDS-denaturation and separation by SDS-PAGE. The gel slice containing the Cro-gD-1-β-gal protein was macerated and injected sub-cutaneously into rabbits in complete Freund's adjuvant. Antibodies produced in rabbits following inoculation and subsequent boosts were found to immunoprecipitate gD-1 and gD-2 prepared from HSV-infected cells (26). Rabbit sera were also found to neutralize HSV-1 and to a lesser extent HSV-2 infectivity *in vitro*.

We have subsequently repeated the above experiments using SDS-PAGE gel slices containing the PHV6 chimaeric protein. Sera from three rabbits (R159, R160 and R161) were examined after inoculation with three doses (100-200 μg) of the pHV6 protein. It was found that each of these sera immunoprecipitated gD-2 extracted from HSV-2 infected HeLa cells, albeit to different extents (Fig. 10). The two bands observed represent precursor and mature forms of the glycoprotein. Rabbit 018 serum (inoculated with pEH4-2 protein) also immunoprecipitated gD-2,

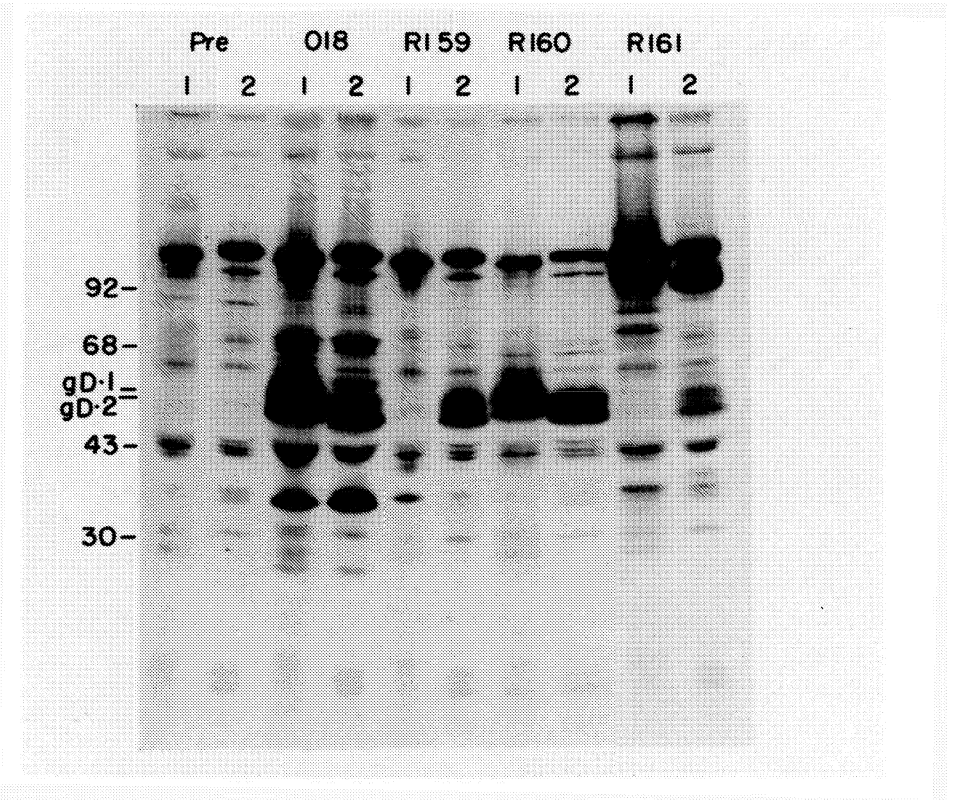


FIGURE 10. Immunoprecipitation of gD-1 and gD-2 by rabbit antiserum directed against the pHV6 chimaeric protein. The pHV6 Cro-gD-2-β-gal protein was prepared for use as an antigen as described previously (26). SDS-PAGE gel slices containing the protein were ground to a paste, and mixed with equal volumes of STE and complete Freund's adjuvant. 100-200 µg amounts of the protein were injected subcutaneously into three rabbits (R159, R160 and R161). 28 days after the initial boost and 10 days thereafter, further boosts (100-200 µg) in complete adjuvant were given. At day 48, serum was obtained by ear bleeding. Immunoprecipitation experiments using these sera were performed as described previously (26) using ^{35}S-methionine-labelled lysates of Vero cells infected with HSV-1 and HeLa cells infected with HSV-2. As controls, pre-immune serum from rabbit R159, R160 and R160 and serum from rabbit 018 (inoculated with the pEH4-2 chimaeric protein: ref. 26) were also tested. Immunoprecipitated proteins were separated by SDS-PaGE and visualized by fluorography. Indicated above are the gD-1 and gD-2 proteins precipitated by these various sera from type 1 and type 2 infected cells.

whereas serum obtained from rabbits R159, R160 and R161 pre-inoculation failed to bring down gD-2. Likewise, it was found that 018 and R160 sera reacted with gD-1 (Fig. 10). Rather unexpectedly, R159 and R161 sera did not immunoprecipitate detectable amounts of gD-1. Hence, there was a rather variable response in rabbits inoculated with pHV6 protein prepared by this means.

These various rabbit sera were also tested for the presence of *in vitro* neutralizing activity against HSV-1 and HSV-2. Approximately 100 plaque-forming units of HSV-1 and HSV-2 were incubated with dilutions of each serum in the presence of active complement, then plated on vero and HeLa cells, respectively. The neutralization titre of each serum was defined as the reciprocal of the highest dilution giving 50% plaque reduction compared to pre-inoculation serum. As shown in Table 1, R159, R160 and R161 sera neutralized HSV-2 infectivity at a lower dilution than 018 serum. The converse applied with neutralization of HSV-1 infectivity. Whereas R159 and R161 sera did not appear to immunoprecipitate gD-1, they did neutralize HSV-1 infectivity, albeit at low dilutions. R160 serum had the highest neutralization activity against HSV-2.

Table 1. Neutralization of HSV-1 and HSV-2 by antisera directed against Cro-gD-β-gal proteins

Antisera	Neutralization[1] HSV-1	HSV-2
018	256	16
R159	32	96
R160	48	128
R161	16	64

[1] Numbers represent the antibody titer of the reciprocal of the serum dilution which reduced plaque numbers by 50%. Assays were performed by incubating approximately 100 plaque forming units of HSV-1 or HSV-2 with serum dilutions in the presence of 5% active guinea pig complement at room temperature for 1 hr. The viruses were then plated on vero cells and the plaques counted 3 days post-infection.

DISCUSSION

We have described the expression of HSV gD-1 and gD-2 gene sequences in E. coli under the transcriptional control of the lac promoter-operator. A number of phenomena relating to the expression of foreign genes in E. coli became apparent from this and other work. Hence, we found that the Cro-gD-1 protein specified by pEH25 did not accumulate in large amounts in cells induced with IPTG. Fusion of the carboxy-terminal gD-1 sequences contained by pEH25 to a sequence encoding β-galactosidase, resulted in a high level of expression of a gD-related protein in plasmid transformants, providing that as much as 131 bp encoding the gD-1 carboxy-terminus were deleted. Deletion of carboxy-terminal sequences without concurrent fusion to β-gal (as with pEH90-10am) did not result in accumulation of the resultant Cro-gD-1 amber fragment in large amounts. These data point to at least two factors determining the level of accumulation of the gD-related product. These factors are discussed below.

Sequences near the gD-1 carboxy-terminus are deleterious for high level expression of the recombinant genes. This may reflect a detrimental influence of this region upon transcription or translation of the recombinant gene. Alternatively, the expression of the gD carboxy-terminus may influence the stability of the plasmid in transformants. For example, expression of this sequence may be poisonous to transformed cells, and IPTG-induction may selectively favour those cells containing few or no plasmids (the selective drug, ampicillin, becomes cleared from the growth medium relatively quickly upon cell growth). Sequences at the gD-1 carboxy-terminus include the transmembrane and anchor sequences. The presence of 11 aa of the transmembrane sequence (as in pEH90-5) did not have a deleterious effect upon accumulation of the Cro-gD-1-β-gal fusion protein. It is possible, however, that the presence of greater than 11 aa of this hydrophobic region may have such an effect. It is possible, also, that the highly charged anchor sequence of gD-1 contributes towards this effect. However, deletion of all but 7 aa of the anchor sequence (as in pEH90-12) did not result in high level expression of the chimaeric protein. It should be noted that of

the 7 aa of the anchor sequence retained in pEH90-12, 5 are basically charged. Parenthetically, it is interesting to note that whereas expression of the 14 aa hydrophobic amino- terminal signal peptide of the vesicular stomatitis virus glycoprotein gene was lethal in E. coli (30), expression of the 13 aa hydrophobic core of the gD-1 signal peptide (as in pEH82) was permissive.

The second factor influencing the level of accumulation of the gD-related product is stability of the protein. We have found by pulse-chase experiments that the Cro-gD-1 proteins specified by both pEH25 and by pEH90-10am (in E. coli NF1829) were unstable and turned-over with half-lives of approximately 15 mins and 5 mins, respectively. However, the pEH90-10am Cro-gD-1 protein was stabilized in E. coli LE392, by the presence of Cro-gD-1-β-gal. It has been reported previously (31) that amber fragments of β-gal are stabilized when overproduced under the transcriptional control of the lac promoter-operator carried on a multicopy plasmid. Aggregation of these Cro-gD-β-gal proteins thus appears to be an inherent property of the β-gal moiety. Presumably, the formation of insoluble aggregates protects the proteins against the activity of proteolytic enzymes.

The Cro-gD-β-gal proteins produced here have been tested as immunogens in rabbits following solubilization with SDS. We found that the chimaeric proteins treated in this way elicited antibodies which reacted with authentic gD-1 or gD-2 and which neutralized in vitro the virus infectivity. We assume that the antibodies directed against the SDS-denatured protein recognize linear rather than conformational antigenic determinants. The antibodies produced were both type-specific and type-common. It is likely, therefore, that the best immunogen against either HSV-1 or HSV-2 infections would be one which contains the homologous gD sequence. It is probable that SDS denaturation does not represent the best mode in which to prepare the antigen. We are currently developing other ways in which to solubilize the fusion protein aggregate so that, hopefully, the integrity of additional antigenic determinants will be retained.

REFERENCES

1. Nahmias, A.J., Dannenbarger, J., Wickliffe, C. and Muther, J. In: The Human Herpesviruses: an Interdisciplinary Perspective (Eds. A.J. Nahmias, W.R. Dowdle and R.F. Schinazi), Elsevier, New York, 1981,pp. 3-9.
2. Cappel, R., De Cuyper, F. and Rikaert, F. Arch. Virol. 65: 12-23, 1980.
3. Skinner, G.R.B., Woodman, C.B.J., Hartley, C.E., Buchan, A., Fuller, A., Durham, J., Synnott, M., Clay, J.C., Melling, J., Wiblin, C. and Wilkins, J. Br. J. Vener. Dis. 58: 381-386, 1982.
4. Hilleman, M.R., Larson, V.M., Lehman, E.D., Salerno, R.A., Conard, P.G. and McLean, A.A. In: The Human Herpesviruses: an Interdisciplinary Perspective (Eds. A.J. Nahmias, W.R. Dowdle and R.F. Schinazi), Elsevier, New York, 1981, pp. 503-506.
5. Klein, R.J., Buimovici-Klein, E., Moser, H., Moucha, R. and Hilfenhaus, J. Arch. Virol. 68: 73-80, 1981.
6. Slichtova, V., Kutinova, L. and Vonka, V. Arch. Virol. 66: 207-214, 1980.
7. Dundarov, S., Andonov, P., Bakalov, B., Nechev, K. and Tomov, C. In: Developments in Biological Standardization, Vol. 52, S. Karger, Basel, 1982, pp. 351-358.
8. Roizman, B., Warren, J., Thuning, C.A., Fanshaw, M.S., Norrild, B. and Meignier, B. Ibid. pp. 287-304.
9. Skinner, G.R.B., Woodman, C., Hartley, C., Buchan, A., Fuller, A., Wiblin, C., Wilkins, G. and Melling, J. Ibid. pp. 333-344.
10. Cappel, R., Sprecher, S. and de Cuyper, F. Ibid. pp. 345-350.
11. Norrild, B. Curr. Topics Microbiol. Immun. 90:67-106, 1980.
12. Para, M.F., Baucke, R.B. and Spear, P.G. J. Virol. 41:129-136, 1982.
13. Sim, C. and Watson, D.H. J. Gen. Virol. 19: 217-133, 1973.
14. Cohen, G.H., Katze, M., Hydrean-Stern, C. and Eisenberg, R.J. J. Virol. 27:172-181, 1978.
15. Bolivar, F., Rodriguez, R.L., Betlach, M.C. and Boyer, H.W. Gene 2:75-93, 1977.
16. Roberts, T. and Lauer, G. In: Methods in Enzymology, Vol. 68 (Ed. R.Wu), Academic Press, New York, 1979, pp. 473-482.
17. Shine, J. and Dalgarno, L. Proc. Natl. Acad. Sci. U.S.A. 71: 1342-1346, 1974.
18. Guarente, L., Lauer, G., Roberts, T.M. and Ptashne, M. Cell 20: 543-553, 1980.
19. Miller, J.H. In: The Operon (Eds. J.H. Miller and W.S. Reznikoff), Cold Spring Harbor Laboratory, New York, 1978, pp. 31-88.
20. Marsden, H.S., Stow, N.D., Preston, V.G., Timbury, M.C. and Wilkie, N.M. J. Virol. 28:628-642, 1978.
21. Ruyechan, W.T., Morse, L.S., Knipe, D.M. and Roizman, B. J. Virol. 29:677-697, 1979.

22. Watson, R.J., Colberg-Poley, A.M., Marcus-Sekura, C.J., Carter, B.J. and Enquist, L.W. Nucleic Acids Res. 11: 1507-1522, 1983.
23. Watson, R.J., Weis, J.H., Salstrom, J.S. and Enquist, L.W. Science 218:381-384, 1982.
24. Kreil, G. Annu. Rev. Biochem. 50:317-348, 1981.
25. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. Nature (London) 288:236-241, 1980.
26. Weis, J.H., Enquist, L.W., Salstrom, J.S. and Watson, R.J. Nature (London) 302:72-74, 1983.
27. Showalter, S.D., Zweig, M. and Hampar, B. Infect. Immun. 34:684-692, 1981.
28. Williams, D.C., Van Frank, R.M., Muth, W.L. and Burnett, J.P. Science 215:687-689, 1982.
29. Burnett, W.N. Anal. Biochem. 112:195-203, 1981.
30. Rose, J.K. and Shafferman, A. Proc. Natl. Acad. Sci. U.S.A. 78:6670-6674, 1981.
31. Cheng, Y.-S., E., Kwoh, D.Y., Kwoh, T.J., Soltvedt, B.C. and Zipser, D. Gene 14:121-130, 1981.
32. Miller, J.H. In: Experiments in Molecular Genetics, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1972.
33. Laemmli, U.K. Nature (London) 227:680-682, 1970.
34. Towbin, H., Staehelin, T. and Gordon, J. Proc. Natl. Acad. Sci. U.S.A. 76:4350-4354, 1979.

18

CLONING AND EXPRESSION OF FOOT AND MOUTH DISEASE VIRUS GENES

D J ROWLANDS and F BROWN
Wellcome Biotechnology Ltd, FMD Division, Ash Road, Pirbright, Woking, Surrey GU24 ONQ, U.K.

SUMMARY

Foot and mouth disease virus is roughly spherical with a diameter of c. 25nm containing 30% RNA and 70% protein. It consists of one molecule of positive sense ssRNA, molecular weight c. 2.6×10^6 and 60 copies of each of four proteins VP1-4. Proteins VP1-3 have molecular weights of c. 24×10^3 and VP4 has a molecular weight of c. 8×10^3. The RNA has a covalently linked protein VPg, molecular weight c. 3×10^3 at the 5' end and a poly-adenylic acid tract at the 3' end. In addition there is a poly-cytidylic acid tract of variable length (100-250 residues) starting from a position 390 residues from the 5' end.

cDNA clones corresponding to the translatable region of the genome have been attained and sequenced. Expression has been confined largely to the capsid protein region of the genome, concentrating on polypeptide VP1 because this carries the immunising antigen of the virus particle. Good levels of expression have been attained in E. coli cells. Furthermore immunogenic fragments of VP1 have been identified and expressed in E. coli cells. Both the expressed protein and fragments containing the amino acid tract 130-157 protect cattle against challenge infection with the homologous virus.

THE DISEASE

Foot and mouth disease is the most important virus disease of farm animals. Cattle, sheep, pigs and goats are susceptible and the loss of productivity following infection approaches 25%. The most serious effects are seen in dairy cattle where the loss of milk yield, abortion, chronic mastitis and lameness are commonplace results of infection. Moreover, in addition to the direct losses in productivity, there are indirect losses caused by the interruption of trading in meat and dairy products.

The disease has a worldwide distribution, being endemic in every continent except Australia and North America. Those countries that are free from the disease apply strict importation and quarantine measures to minimise the

possibility of importing the virus. Control of the disease is by slaughter in those countries where it does not normally occur. Vaccination is used where the disease is endemic. This is a costly operation because it is required at least annually and usually the practice is to vaccinate at more frequent intervals.

ANTIGENIC VARIATION

Vaccination is complicated by the occurrence of the virus as seven serotypes O, A, C, SAT 1, SAT 2, SAT 3 and Asia 1. The difference between the serotypes is such that an animal which has recovered from infection with virus of one serotype is still fully susceptible to infection with the others, although it is protected against re-infection with the homologous virus. A further complicating issue is the occurrence within the serotypes of subtypes which are sufficiently different from each other that they do not confer the degree of protection which is obtained with the homologous subtype. In practical terms this means that where the level of immunity is low, as for example when it is waning several months after vaccination, infection can occur with virus of another subtype, although protection would still be afforded against the homologous virus.

PREPARATION OF VACCINES

In addition to the problems associated with antigenic variation, there are other difficulties in the preparation of foot and mouth disease vaccines.

1. Handling the virus requires the exercise of strict disease security measures because of the potential danger to the environment. With the very large quantities of virus involved, this is a very serious problem and in most countries the virus is handled only under strict supervision.
2. Complete inactivation of the virus is necessary. However, it is impossible to be certain that this has been achieved without testing the entire production batch. In practice less than 0.1% of a production batch is tested for innocuity. It is known that on occasions the presence of residual live virus in vaccines has resulted in infection of animals (1).
3. The virus particle carrying the immunising protein is rather fragile, thus making handling and storage difficult. At the present time a cold chain is required to ensure that the vaccines retain their potency. Such a cold chain is difficult to guarantee, particularly in the field in tropical countries.

Consequently, studies have been made in several laboratories to identify the structural features necessary for eliciting a good immune response. There has been

a steady progression during the last 25 years, starting with the recognition that the immunising activity was associated with the intact virus particle (2). The observation that the activity was apparently confined to a single virus protein (3) and the isolation of this protein in an active form (4-7) stimulated several workers to apply recombinant DNA technology to its biosynthesis. This line of approach has been highly successful and good yields have been obtained (8, 9) but the immunising activity of the product is very low compared with that of the intact virus particle.

In an extension of this work, fragments of the immunising protein have been shown to possess immunogenic activity (10). This has led to the identification of the immunogenic site of the protein and the synthesis of peptides containing this site. These peptides in turn have been shown to possess immunising activity (11, 12). These observations are allowing the structural basis for immunising activity to be examined at the fundamental level of amino acid sequence.

STRUCTURE OF THE VIRUS

The virus belongs to the family Picornaviridae. It is roughly spherical with a diameter of c. 25nm and contains 30% RNA and 70% protein. The virus consists of one molecule of positive sense ssRNA with a molecular weight of c. 2.6×10^6 and 60 copies of each of four proteins VP1-VP4 (13). The precise molecular weights of VP1-VP3 of virus serotype A, subtype 10 (strain 61) have been predicted from the nucleotide sequence of the RNA to be (VP1), (VP2) and (VP3) (14). The molecular weight of VP4 is not known with the same degree of accuracy because the N-terminus is blocked, so that its precise position on the genome has not been identified. However, it is considered to be c. 8×10^3.

The RNA possesses some unusual features (13). Unlike most messenger RNAs, the 5' terminus is uncapped but instead has a small protein (VPg - virus protein, genome linked) of molecular weight c. 3×10^3 covalently linked via a tyrosine residue to the 5'-terminal uridine. At a position 390 residues from the 5' terminus there is a tract of cytidylic acid residues which varies from 100 to 250 in length between different isolates. This unusual structural feature has also been found in the members of the cardiovirus genus of the picornavirus family but it has not been reported for any other virus. As with most other messenger RNAs, the genome of FMDV has a tract of adenylic acid residues at its 3' terminus. This tract is variable in length, even within one virus preparation. This variation in poly A length is intriguing since it is derived by copying of a poly U tract in the template strand RNA and not added post-transcriptively in response to a poly-adenylation signal sequence as is the case for eucaryotic mRNAs.

Two pieces of evidence indicate that only VP1 of the four structural proteins is immunogenic. The first is that treatment of the virus particle with trypsin and other proteolytic enzymes reduces its infectivity and, with several strains, also reduces the activity in eliciting neutralising antibody. In these particles VP1 is cleaved but the other proteins remain unaffected (3). The second, more direct evidence, is that the isolated protein will also elicit neutralising antibody, whereas the remaining three proteins are devoid of activity (4-7).

With the information obtained by biochemical mapping of the genome (Figure 1) (13), it became possible to locate the position of VP1 on the virus genome, thus allowing the DNA complementary to the RNA coding for VP1 to be cloned and inserted into plasmids.

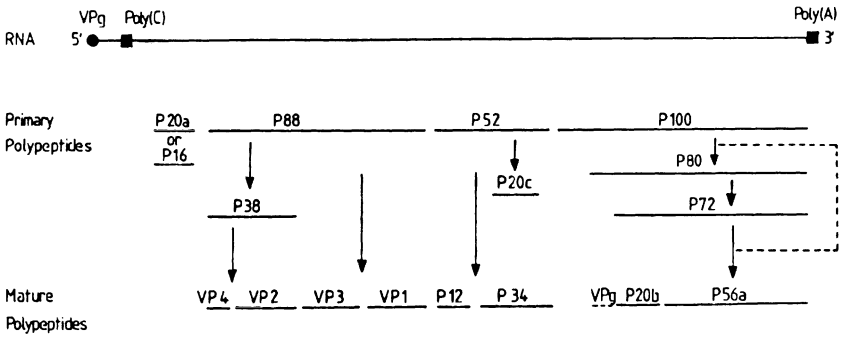


Fig. 1 Biochemical map of the FMDV genome (ref. 13). The positions of the primary and secondary cleavage protein products are shown in relation to the genomic RNA.

CLONING OF THE GENOME

The first reports of the molecular cloning of the FMDV genome appeared in the literature in 1981 (8, 15, 16). The technique used by all the groups working on this problem was the now standard procedure of making cDNA copies of RNA isolated from purified FMDV particles using avian myelobastosis virus reverse transcriptase (17). Synthesis of the cDNA was initiated at the 3' poly A tract of the RNA with an oligo dT primer. Second strand synthesis was carried out using

DNA polymerase or reverse transcriptase after removal of the original template RNA. No primer was added and second strand synthesis was initiated by the loop back mechanism to produce a hairpin structure covalently linked at one end. The single-stranded loops were digested with S1 nuclease and the resulting flush-ended, double-stranded DNA molecules extended with homopolymeric C tails using terminal transferase. The DNAs were inserted in the Amp gene of plasmids pBR322 or pAT153 by hybridisation to Pst1 cut, poly G-tailed molecules. Using these techniques various groups have obtained clone banks which together represent the majority of the coding sequences of the virus RNA (8, 15, 16). In all reports the cloned inserts represented segments from different regions of the genome, presumably due to interruption of transcription at both the first and second strand synthesis stages. The longest continuous cloned insert reported was 5,500 bp in length and extended from the 3' poly A tract through the VP1 coding region (16). Overlapping clones could be aligned by restriction mapping and their orientation with respect to the virus genome determined by restriction mapping of first strand cDNA copies or by hybridisation to RNase T_1 oligonucleotides whose position on the virus RNA had been determined.

Comparison of the restriction enzyme maps determined for the cloned DNA with the biochemical map referred to earlier made it possible to predict the regions of cDNA containing the sequences corresponding to specific virus proteins. The structural proteins, and VP1 in particular, were of primary interest because they allowed the study and exploitation of the antigenic determinants present on these proteins. Earlier studies by Strohmaier et al (18) and Bachrach (19) on limited N terminal and C terminal amino acid sequences of the structural proteins facilitated the identification of the sequences coding for these proteins.

EXPRESSION OF VP1

Encouraged by the demonstration that VP1 isolated from virus particles could elicit neutralising antibodies in a variety of species and would protect animals from the disease (5), several groups attempted to express VP1 in bacteria. Plasmids were constructed in which VP1 sequences were inserted down stream from DNA coding for a prokaryotic protein and expressed as a chimaeric protein under the control of an inducible promoter. Kupper et al (15) inserted VP1 sequences behind a cloned DNA representing part of the coat protein coding region of the phage MS2. The insert contained sequences representing the translation control signals of the MS2 protein and transcription of the chimaeric construction

was placed under the control of the P_L promoter. Expression was initiated by thermal inactivation of a ts repressor constitutively produced by the host cells. Another chimaeric protein composed of part of the trp E protein linked to VP1 has been expressed under the control of the trp promoter (8) which is switched on at low tryptophan concentrations. Both of these constructions were successful in producing large amounts of VP1, up to 17% of bacterial protein in some experiments. The chimaeric proteins are insoluble within the bacteria and precipitate as aggregates. This insolubility may be one of the reasons why bacteria are able to produce such large amounts of the proteins without toxic effects and without extensive proteolytic degradation of the product.

Kleid et al (8) showed that after purification by SDS polyacrylamide gel electrophoresis, two injections at 24 day intervals of 250µg of chimaeric protein protected cattle placed in contact with infected animals from contracting the disease. This observation constitutes the first demonstration that a vaccine produced by genetic engineering can protect animals from virus disease.

Despite the satisfaction of producing a biologically active antigen by genetic manipulation in E. coli., there remain several serious obstacles to the commercial exploitation of this work. These drawbacks primarily revolve around the immunological inefficiency of protein subunit antigens of FMDV. The dramatic loss of c. 90-99% of the neutralising antibody stimulating potential of FMDV particles when these are disrupted into 12S pentameric subunits consisting of VP1-3 demonstrates the exacting requirement for presentation of the important antigenic determinants in a native form. Isolated VP1 protein is orders of magnitude less efficient than the 12S subunit in eliciting a primary neutralising antibody response and the genetically engineered product seems to be no better.

An exception to the general finding that VP1 has very low activity as an immunogen has been reported recently by Moore et al (20). These workers found that their preparations of VP1 were almost as effective as an equivalent weight of intact virus particles in eliciting neutralising antibodies. This observation is contrary to previous reports from the same laboratory and remains unexplained.

A second problem associated with the use of VP1 or a VP1-containing chimaeric product in bacterial cells is the necessity to purify the product. Current FMDV vaccines are essentially inactivated filtered tissue culture harvests which are used without further treatment but the requirement for purification of the genetically engineered product could present severe commercial constraints on the marketing of such a product.

The cloning of FMDV cDNA provided the opportunity for enormous advances in the understanding of the structure of the virus genome at the nucleotide sequence level. Prior to the application of gene cloning and DNA sequencing techniques nucleotide sequence data was restricted to a hundred or so nucleotides from each end of the RNA (21, 22). These sequences were within the non coding regions at the extremities of the genome and so were of no value in obtaining amino acid sequence data for the virus proteins. The only amino acid sequences available were the N terminal 20 or so of structural proteins VP1, 2 and 3 and one or two residues at their C termini (18, 19). This situation has now changed dramatically and the complete coding sequences of viruses from serotypes A and O have been determined (23, 24). Initial sequencing work was concentrated on the structural proteins because of interest in the antigenic structure of the virus and there have been reports on the sequence of VP1 from several viruses (8, 15, 16, 25, 26) and the entire structural protein precursor for a type A virus (14). A number of interesting features have emerged from these data but for the purposes of this review we will limit ourselves to observations which are relevant to the immunogenicity of the virus.

LOCATION OF IMMUNOGENIC SITES

Identification of the amino acids comprising the epitope(s) responsible for eliciting a neutralising antibody response has been a prime objective in FMD research in the past few years. The value of this work is the furthering of our understanding of antigenic variation at the structural level and the provision of information essential for the design of novel synthetic or biosynthetic peptide vaccines.

The first and most direct approach to this problem was made by Strohmaier et al (10) who investigated the immune response of mice to fragments of VP1 generated by a variety of proteolytic or chemical cleavages. The exact location of these fragments on the VP1 protein was obtained by comparing limited N and C terminal sequences derived directly from the peptides with the total VP1 sequence predicted from nucleotide sequences. By aligning overlapping peptides on the VP1 sequence and scoring them for ability to elicit a neutralising antibody response these workers predicted that residues 146-154 and 200-213 would be included in neutralising antibody eliciting antigenic epitopes (Figure 2). This of course provides a minimum estimate of antigenic sites since it is possible that conformational determinants are present on VP1 within the constraints of the virus

particle which are lost when the protein or fragments of the protein are tested in isolation. Nevertheless this work was an invaluable step forward and subsequent studies have confirmed and rationalised the initial observations.

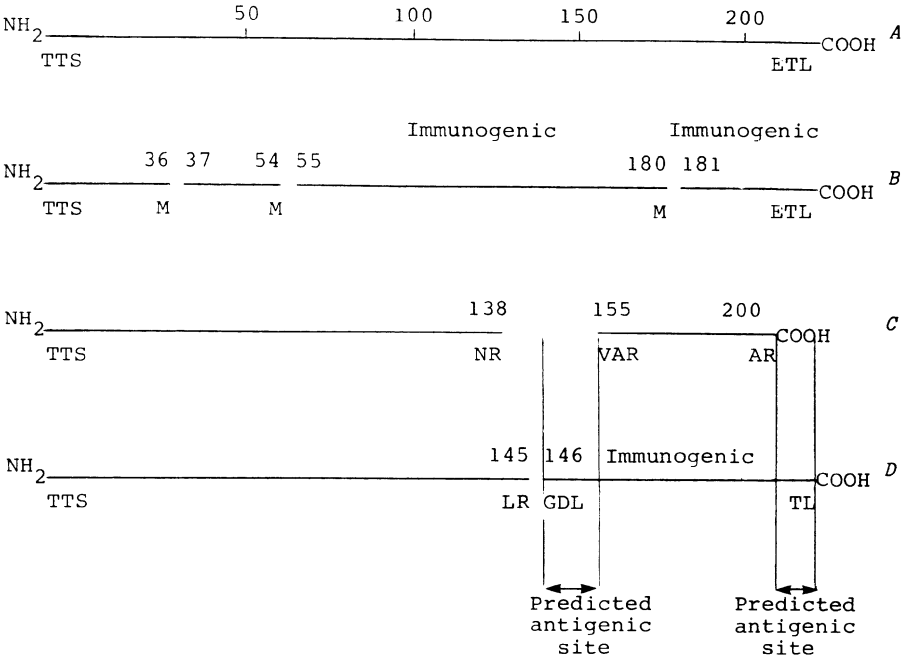


Fig. 2 Location of immunogenic sites on VP1 of foot and mouth disease virus (adapted from ref. 10).

A Intact VP1.

B Cleaved at methionine residues with cyanogen bromide.

C Cleaved in situ at exposed arginine and lysine residues with trypsin.

D Cleaved in situ at exposed arginine residues with mouse submaxillary gland protease.

Amino acids are given in the single letter notation.

As nucleotide sequences of the part of the genome coding for VP1 from different FMD viruses became available it was possible to compare the predicted amino acid sequences of this protein. Studies of sequence variation between the VP1 proteins of viruses of known serological relationships would be expected to help in the identification of antigenically important regions of the protein since antigenic differences must correlate with sequence differences and such sequence differences are most likely to occur within antigenic epitopes themselves. Some examples of sequence variation between viruses of different degrees of antigenic relatedness are shown in Figure 3. Several interesting features emerged from such comparative studies. The first is that sequence variation is not uniformly distributed over the whole molecule. Three major regions of variation are apparent when VP1 proteins from three serotypes (O, A and C) are compared although there are several smaller variable regions. These major variable regions lie between residues 42-61, 129-160 and 193-204 towards the C terminus of the protein. Surface-orientated regions of proteins are likely to be hydrophilic in character and when predicted hydrophilicity plots were superimposed on amino acid sequence variation plots, it was found that the only regions of high variability which coincided with a conserved hydrophilic character were those between 129-160 and 193-204, which coincides well with the location of immunogenic sites predicted by Strohmaier et al (10).

Another technique for the prediction of potentially antigenic regions of VP1 was used by Pfaff et al (12). They reasoned that a good candidate structure would be a strong α helical region which displayed hydrophilic and hydrophobic zones on opposite sides of the helix. Such a structure was found, using the secondary structure prediction rules of Chou and Fasman (28), between residues 144-159 in the sequence of type O_1 (Kaufbeuren) virus. However, a large helical content is not a conserved feature of this region from other viruses (unpublished observations).

ACTIVITY OF SYNTHETIC PEPTIDES

In recent years much progress has been made in the use of chemically synthesised peptides to define antigenic regions present on intact proteins and for a review of this subject the reader is referred to an article by Lerner (29). It has been demonstrated with a number of systems that synthetic peptides of about 20 amino acids or less in length can, when coupled to a carrier protein and injected

A_{12}	Thr	Thr	Ala	Thr	Gly	Glu	Ser	Ala	Asp	Pro	Val	Thr	Thr	Thr	Val	Glu	Asn	Tyr	Gly	Gly	20
A_{10}			Thr																		
A_5				Val																	
A_{24}																					
O_1			Ser	Ala																	
C_3			Thr																		
A_{12}	Glu	Thr	Gln	Val	Gln	Arg	Arg	His	His	Thr	Asn	Val	Ser	Phe	Ile	Met	Asp	Arg	Phe	Val	40
A_{10}													Glu								
A_5	Asp			Thr				Tyr	Tyr	Met			Gly								
A_{24}				Ile								Ile	Gly								
O_1			Glu	Ile				Gln													
C_3				Ile									Ala		Val	Leu					
A_{12}	Lys	Ile	Lys	Ser	Leu	Asn	Pro	Thr	His	Val	Ile	Asp	Leu	Met	Gln	Thr	His	Gln	His	Gly	60
A_{10}			Asn			Ser												Lys			
A_5			Asn			Ser															
A_{24}			Gln			Ser															
O_1		Val	Thr	Pro	Gln		Gln	Ile	Asn	Ile	Leu					Val	Pro	Ser		Thr	
C_3		Val	His	Val	Ser	Gly	Asn	Gln		Thr	Leu		Val			Val		Lys	Asp	Ser	
A_{12}	Leu	Val	Gly	Ala	Leu	Leu	Arg	Ala	Ala	Thr	Tyr	Tyr	Phe	Ser	Asp	Leu	Glu	Ile	Val	Val	80
A_{10}	Ile																				
A_5																					
A_{24}									Thr												
O_1									Ser										Ala		
C_3	Ile								Ala										Ala		
A_{12}	Arg	His	Asp	Gly	Asn	Leu	Thr	Trp	Val	Pro	Asn	Gly	Ala	Pro	Glu	Ala	Ala	Leu	Ser	Asn	100
A_{10}																					
A_5																					
A_{24}			Glu													Ser		Leu			
O_1	Lys		Glu	Arg	Asp											Lys		Asp			
C_3	Thr		Thr		Lys										Val	Ser		Asp			
A_{12}	Thr	Gly	Asn	Pro	Thr	Ala	Tyr	Asn	Lys	Ala	Pro	Phe	Thr	Arg	Leu	Ala	Leu	Pro	Tyr	Thr	120
A_{10}		Ser																			
A_5		Ser								Val											
A_{24}		Ser																			
O_1		Thr						His				Leu									
C_3		Ala						His		Gly		Leu									
A_{12}	Ala	Pro	His	Arg	Val	Leu	Ala	Thr	Val	Tyr	Asn	Gly	Thr	Asn	Lys	Tyr	Ser	Ala	Ser	Gly	140
A_{10}																				Asp	
A_5																		Thr	Gly		
A_{24}						Pro								Ser			Ala	Val	Gly		
O_1													Glu	Cys	Arg			Arg	Asn	Ala	
C_3									Thr	Thr				Thr	Ala		Thr			Ala	

A_{12}	Ser	Gly	-	Val	Arg	Gly	Asp	Phe	Gly	Ser	Leu	Ala	Pro	Arg	Val	Ala	Arg	Gln	Leu	Pro	160
A_{10}		-	Arg	Ser	-			Leu			Ile		Ala				Thr				
A_5	Pro	-	Arg	-				Met			Ala		Ala		Ala		Lys				
A_{24}			Arg					Met					Ala			Val	Lys				
O_1	Val	Pro	Asn	Leu				Leu	Gln	Val			Gln	Lys				Thr			
C_3	-	-	Arg	-				Leu	Ala	His			Ala	Ala	His			His			

A_{12}	Ala	Ser	Phe	Asn	Tyr	Gly	Ala	Ile	Lys	Ala	Glu	Thr	Ile	His	Glu	Leu	Leu	Val	Arg	Met	180
A_{10}									Gln		Gln	Ala									
A_5									Arg		Ile										
A_{24}											Asp	Ala									
O_1	Thr										Thr	Arg	Val	Thr				Tyr			
C_3	Thr			Phe			Val				Glu			Thr							

A_{12}	Lys	Arg	Ala	Glu	Leu	Tyr	Cys	Pro	Arg	Pro	Leu	Leu	Ala	Ile	Glu	Val	Ser	Ser	Gln	Asp	200
A_{10}									Lys						Lys		Thr				
A_5																					
A_{24}																					
O_1					Thr										His	Pro	Thr	Glu	-	Ala	
C_3											Val		Pro	Val	Gln	Pro	Thr	Gly	-		

A_{12}	Arg	His	Lys	Gln	Lys	Ile	Ile	Ala	Pro	Gly	Lys	Gln	Leu	Leu	214
A_{10}		Tyr								Ala					
A_5										Ala	Arg				
A_{24}										Ala					
O_1							Val			Val		Thr			
C_3					Pro	Leu				Ala					

Fig. 3 Comparison of amino acid sequences of VP1 of three subtypes of serotype A virus and one example each of viruses of serotypes O and C. Only those amino acids which are different from the A_{12} sequence are shown. Dashes indicate the deletion of an amino acid. The source of the sequences shown are A_{12} (ref. 8), A_{24}, O_1 and C_3 (ref. 25), A_{10} (ref. 16) and A_5 (ref. 32).

into animals, induce the formation of a spectrum of antibodies, some of which recognise epitopes on the native intact protein from which the sequence originated. With this approach, Bittle et al (11) and Pfaff et al (12) demonstrated that synthetic peptides representing each of the two potential antigenic regions of VP1 discussed above induced, in experimented animals, high levels of antibodies which recognised intact virus particles. However, the virus neutralising activity of antisera to peptides representing the antigenic site located in the region of amino

acids 141-160 was several orders of magnitude higher than that of antisera to the C terminal site. In fact, Bittle et al (11) showed that a single injection of peptide 141-160 coupled to keyhole limpet haemocyanin was sufficient to protect guinea pigs from challenge with 10,000 infectious units of virulent virus.

More recent work has shown that with closely related FMD viruses major antigenic variation can be correlated with amino acid substitutions at only two positions in the 141-160 region at amino acids 148 and 153 (30). Since peptides of the 141-160 sequence present in these viruses induced antibodies with discriminating properties in neutralisation tests similar to those found in response to vaccines prepared from the variant viruses themselves, it is apparent that the synthetic 141-160 peptides mimic a major neutralising epitope of the virus.

The preparation of antigenic fragments to mimic the antigenic features of a protein is not restricted to the chemical synthesis of such peptides. Yansura et al (31) have expressed in E. coli a construction consisting of part of the trpLE protein linked to amino acids 130-157 of FMDV type A, subtype 24. Fusion proteins containing the trp LE protein are insoluble in the bacterial cell and are resistant to proteolysis. As a consequence of this property and by the placing of this construction under the control of a strong inducible promoter (the trp E promoter), up to 20% of the cell protein consisted of the desired product. Two inoculations of 250µg of fusion protein (equivalent to 25µg of VP1 peptide) protected cattle from disease when the animals were placed in contact with control animals infected with the homologous virus.

CONCLUSION

The application of genetic engineering techniques to the study of FMDV during the last 3-4 years has led to enormous advances in our understanding of the location and structure of the important antigenic sites of the virus. The practical application of this understanding to produce a new generation of vaccines which do not suffer the inherent disadvantages of the current chemically inactivated virus vaccines is possibly more advanced than with any other virus. However, there is still a lot to learn. In particular the fine structural details of the major immunogenic site of the virus need to be understood in order to mimic the structure more accurately in novel vaccine preparations. Obviously high resolution X-ray crystallography data from virus particle will be invaluable for this purpose.

At present it would appear that the entire VP1 molecule will be of little value as a practical vaccine and short peptides representing the major antigenic

site seem to be more promising. Should such materials become practical candidates as alternative vaccines it is difficult to predict at the moment whether chemical synthesis or biosynthesis by genetic engineering will be the best means of production. At the experimental research stage chemical synthesis of potential antigens has the practical advantage that large numbers of different sequences for antigenic evaluation can be produced more easily than by the biosynthetic route.

In addition to mimicing of antigenic sites there is still a lot to be learnt about the influence of the form of presentation of peptide antigens to the immune system. It may be possible to influence such factors as the magnitude, duration and immunoglobulin type of the immune response to peptide antigens by modification of methods by which the antigen is presented. However, these studies are in their early stages and so far no clear picture is emerging.

REFERENCES

1. King AMQ, Underwood BO, McCahon D, Newman JWI and Brown F. (1981) Nature 293:479-480.
2. Brown F and Crick J. (1959) J. Immunol. 82:444-447.
3. Wild TF, Burroughs JN and Brown F. (1969) J. gen. Virol. 4:313-320.
4. Laporte J, Groschande J, Wantyghem J, Bernard S and Rouze PC. (1979) Compt. rendu. hebd. Seanc. Acad. Sci., Paris 276:3399-3401.
5. Bachrach HL, Moore DM, McKercher PD and Polatnick J. (1975) J. Immunol. 115:1636-1641.
6. Meloen RH, Rowlands DJ and Brown F. (1979) J. gen. Virol. 45:761-763.
7. Kaaden OR, Adam K-H and Strohmaier K. (1977) J. gen. Virol. 34:397-400.
8. Kleid DG, Yansura D, Small B, Dowbenko D, Moore DM, Grubman MJ, McKercher PD, Morgan DO, Robertson BH and Bachrach HL. (1981) Science 214:1125-1129.
9. Hofschnider PH, Burgelt E, Kauzmann M, Mussgay M, Franze R, Ahl R, Bohm H, Strohmaier K, Kupper K and Otto B. (1981) In: Biological Products for Viral Diseases (Munich Symposia on Microbiology) Ed. P A Bachmann, Taylor & Francis, London p105-113.
10. Strohmaier K, Franze R and Adam K-H. (1982) J. gen. Virol. 59:295-306.
11. Bittle JL, Houghten RA, Alexander H, Shinnick TM, Rowlands DJ and Brown F. (1982) Nature 298:30-33.
12. Pfaff E, Mussgay M, Bohm HO, Schalz GE and Schaller H. (1982) EMBO J. 1:869-874.
13. Sangar DV. (1979) J. gen. Virol. 45:1-13.
14. Boothroyd JC, Highfield PE, Cross GAM, Rowlands DJ, Lowe PA, Brown F and Harris TJR. (1981) Nature 290:800-802.
15. Kupper H, Keller W, Kurz C, Forss S, Schaller H, Franze R, Strohmaier K, Marquardt O, Zaslavsky VG and Hofschneider PH. (1981) Nature 289:555-559.
16. Boothroyd JC, Harris TJR, Rowlands DJ and Lowe PA. (1982) Gene 17:153-161.
17. Williams JG. (1981) In: Genetic Engineering I (Ed. R Williamson) Academic Press, p1-59.
18. Strohmaier K, Wittmann G, Liebold B and Geissler AW. (1978) Biochem. Biophys. Res. Comm. 85:1640-1645.

19. Bachrach HL, Swaney JB and van de Woude GF. (1973) Virology 52:520-528.
20. Moore DM. (1983) In: Genetic Engineering: Applications to Agriculture (Beltsville Symposium, Ed. L D Owens) Rowman & Allanheld, Totowa, USA, p89-106.
21. Ponter AG, Fellner P, Black DN, Rowlands DJ, Harris TJR and Brown F. (1978) Nature 276:298-301.
22. Harris TJR. (1979) Nucl. Acid Res. 7:1765-1785.
23. Carroll AR, Rowlands DJ and Clarke BE. (1984) Nucl. Acid Res. (Submitted).
24. Forss S. (Personal communication).
25. Makoff AJ, Paynter CA, Rowlands DJ and Lowe PA. (1982) Nucl. Acids Res. 10:8285-8295.
26. Villanueva N, Davila M, Ortin J and Domingo E. (1983) Gene 23:185-194.
27. Cheung A, de Lamarter J, Weiss S and Kupper H. (1983) J. Virol. 48:451-459.
28. Chou PY and Fasman GD. (1978) Adv. Enzymol. 47:45-148.
29. Lerner R. (1982) Nature 299:592-596.
30. Rowlands DJ, Clarke BE, Carroll AR, Brown F, Nicholson BH, Bittle JL, Houghten RA and Lerner RA. (1983) Nature 306:694-697.
31. Yansura DG, Dowbenko D, Weddell GN, Hoathin ME, Shire SJ, Bock LA, Patzer EJ, Kleid DG, Moore DM, Robertson BH, Grubman MJ, Morgan DO and McKercher PD. (1983) In: Advances in Gene Technology: Molecular Genetics of Plants and Animals. (Miami Winter Symposium 20) Academic Press, New York.
32. Beck E, Feil G and Strohmaier K. (1983) EMBO J. 2:555-559.

INDEX