EXPRESSION OF EUKARYOTIC VIRAL AND CELLULAR GENES

RALF F. PETTERSSON LEEVI KÄÄRIÄINEN
HANS SÖDERLUND NILS OKER-BLOM

Expression of Eukaryotic Viral and Cellular Genes

Eighth Sigrid Jusélius Foundation Symposium: Helsinki, Finland. June 1980

Edited by

RALF F. PETTERSSON
Department of Virology, University of Helsinki, Helsinki, Finland

LEEVI KÄÄRIÄINEN
Department of Virology, University of Helsinki, Helsinki, Finland

HANS SÖDERLUND
Department of Biochemistry, University of Helsinki, Helsinki, Finland

NILS OKER-BLOM
Department of Virology, University of Helsinki, Helsinki, Finland

1981

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers
London · New York · Toronto · Sydney · San Francisco

ACADEMIC PRESS INC. (LONDON) LTD
24/28 Oval Road
London NW1 7DX

United States Edition published by
ACADEMIC PRESS INC.
111 Fifth Avenue
New York, New York 10003

British Library Cataloguing in Publication Data
Expression of eukaryotic viral and cellular genes.
1. Gene expression – Congresses
2. Eukaryotic cells – Congresses
I. Pettersson, R. F.
574.87'322 QH450
ISBN 0-12-553120-6

Filmset in 'Monophoto' Times New Roman by Eta Services (Typesetters) Ltd., Beccles, Suffolk
Printed in Great Britain by Thomson Litho Ltd., East Kilbride, Scotland

ORGANIZING COMMITTEE

Nils Oker-Blom
Chairman

David Baltimore Leevi Kääriäinen Ralf Pettersson
Lennart Philipson Hans Söderlund

Anja Nykänen
Editorial Secretary

LIST OF PARTICIPANTS

Full addresses of contributors appear on the first page of each chapter

AMINOFF, C. G., Helsinki, Finland
BALTIMORE, D., Cambridge, U.S.A.
VON BONSDORFF, B., Helsinki, Finland
VON BONSDORFF, C.-H., Helsinki, Finland
BOURNE, A., London, U.K.
BROWNLEE, G. G., Cambridge, U.K.
CANTELL, K., Helsinki, Finland
CARBON, J., Santa Barbara, U.S.A.
CHAMBON, P., Strasbourg, France
CHANG, A. C. Y., Stanford, U.S.A.
DE LA CHAPELLE, A., Helsinki, Finland
CLARKE, L., Santa Barbara, U.S.A.
COHEN, S. N., Stanford, U.S.A.
DARNELL, J. E., New York, U.S.A.
EK, J., Pori, Finland
ENARI, T. M., Espoo, Finland
FARRAND, R. A., London, U.K.
FIERS, W., Gent, Belgium
GAHMBERG, C. G., Helsinki, Finland
GAROFF, H., Heidelberg, West Germany
GILBERT, W., Cambridge, U.S.A.
GRAF, T., Heidelberg, West Germany
HALKKA, O., Helsinki, Finland
HALONEN, P., Turku, Finland
HOVI, T., Helsinki, Finland
HUANG, A. S., Boston, U.S.A.
JÄNNE, J., Helsinki, Finland
JÄNNE, O., Oulu, Finland
JÄRNEFELT, J., Helsinki, Finland
JÄRVI, O., Turku, Finland
KÄÄRIÄINEN, L., Helsinki, Finland
KAESBERG, P., Madison, U.S.A.
KAFATOS, F. C., Cambridge, U.S.A.
KALKKINEN, N., Helsinki, Finland
KERÄNEN, S., Helsinki, Finland
KNOWLES, J., Espoo, Finland
KOURILSKY, P., Paris, France
KRUG, R. M., New York, U.S.A.
KULONEN, E., Turku, Finland
LEHTOVAARA P., Helsinki, Finland
LEINIKKI, P., Tampere, Finland
MÄENPÄÄ, P., Kuopio, Finland
MÄKELÄ, O., Helsinki, Finland
O'MALLEY, B. W., Houston, U.S.A.
MÄNTYJÄRVI, R., Kuopio, Finland
MIKOLA, J., Jyväskylä, Finland
VAN MONTAGU, M., Gent, Belgium
NATHANS, D., Baltimore, U.S.A.
NEVALAINEN, H., Espoo, Finland
OKER-BLOM, N., Helsinki, Finland
PALVA, I., Helsinki, Finland
PENTTINEN, K., Helsinki, Finland
PERSSON, H., Uppsala, Sweden
PETTERSSON, R. F., Helsinki, Finland
PETTERSSON, U., Uppsala, Sweden
PFEIFER, S., Helsinki, Finland
PHILIPSON, L., Uppsala, Sweden
PYHTILÄ, R., Oulu, Finland
RAINA, A., Kuopio, Finland
RANKI, M., Helsinki, Finland
SALKINOJA-SALONEN, M., Helsinki, Finland
SALMI, A., Turku, Finland
SARVAS, M., Helsinki, Finland
SCHIMKE, R. T., Stanford, U.S.A.
SHATKIN, A. J., Nutley, U.S.A.
SÖDERLUND, H., Helsinki, Finland
SORSA, M., Helsinki, Finland
SORSA, V., Helsinki, Finland
STEHELIN, D., Lille, France
STEITZ, J. A., New Haven, U.S.A.
TUOHIMAA, P., Tampere, Finland
VAHERI, A., Helsinki, Finland
VARMUS, H. E., San Francisco, U.S.A.
WEISSMANN, C., Zürich, Switzerland
WIKGREN, B.-J., Åbo, Finland

PREFACE

Fifty years have now elapsed since Fritz Arthur Jusélius, a prosperous businessman in the city of Pori on the south-west coast of Finland, created the Sigrid Jusélius Foundation in memory of his beloved daughter Sigrid who died of tuberculosis after a measles infection in 1898 at the age of only eleven years. The aim of the Foundation is to support, irrespective of race or language, international medical research in the fight against diseases which are particularly harmful to mankind, with an emphasis on microbiology. It is understandable that microbiology was central to the ideas of F. A. Jusélius, partly because at the turn of the century infectious diseases played an important role in medicine and partly because his daughter had succumbed to one. Unfortunately, owing to the very rigorous stipulations concerning the growth of the Foundation's capital, the distribution of grants could not start until 1948. Since then, though, most of the basic research in medicine in Finland has been supported by the Sigrid Jusélius Foundation.

In his will F. A. Jusélius stressed the internationality of the Foundation. In accordance with this wish, the Foundation also provides grants for foreign scientists to work in hospitals and laboratories in Finland and, from 1965 on, it has supported many international symposia. Most of the symposia so far have been concerned with basic biological problems like "Control of Cellular Growth in Adult Organisms", "Regulatory Functions of Biological Membranes", "Cell Interactions and Receptor Antibodies in Immune Response", "Biology of Fibroblasts" and, most recently, "Cell Interactions in Differentiation". Two symposia have been devoted to more clinical problems; one on "Amyloidosis" and the latest concerning "Population Genetic Studies on Isolates".

The principle of all the symposia has been to choose a specific problem and to try to elucidate this problem from as many angles as possible. The same is also true of this symposium. Under the common heading "Expression of Eukaryotic Viral and Cellular Genes" this meeting combines recent advances in molecular virology, molecular genetics and recombinant DNA research. Viruses are excellent models for studying eukaryotic gene structure and expression, and much of our current knowledge in certain areas comes from experiments carried out with animal viruses, e.g. the presence and synthesis of poly A tracts at the 3′ end of messenger RNAs, the structure and synthesis of the 5′ end present in eukaryotic messengers, synthesis and processing of heterogenous nuclear RNA, ribosome-binding

to messenger RNAs, the splicing phenomena, the complete determination of the nucleotide sequences of certain DNA viruses, etc.

Recombinant DNA research has again expanded tremendously during the last two years. Virology in the future will be intimately connected to recombinant DNA research, and here again it is enough just to mention a few fields like production of vaccines against, for instance, influenza and hepatitis B, diagnostic procedures, production of interferon, etc.

In some fields of science the progress is so fast that publishing monographs does not seem warranted. This also holds true for certain aspects of this symposium. In spite of this, it was felt that since our approach has been to focus on one problem complex from as many angles as possible, the compilation of these papers could give new impulse to the discussion. Thanks to the collaboration of the participants of the symposium allowing fast publication, their papers retain their original freshness and it is therefore hoped that this book will be of value for those working in this interesting and thrilling field.

Helsinki, June 1980

Nils Oker-Blom
The Rector,
Professor,
University of Helsinki

CONTENTS

III Structure and Expression of RNA Virus Genes

IV Structure and Expression of Eukaryotic Genes

I

Structure and Expression of DNA Virus Genes

Molecular Analysis of Adenovirus Transformation

M. PERRICAUDET,[1] G. WESTIN,[2] J. M. LE MOULLEC,[1] L. VISSER,[3] J. ZABIELSKI,[2] P. ALESTRÖM,[2] G. AKUSJÄRVI,[2] A. VIRTANEN[2] and U. PETTERSSON[2]

[1] *Institut Pasteur, Unite de Genie Genetique, 28, rue du Dr Roux, 75724 Paris Cedex 15, France*

[2] *Department of Microbiology, The Biomedical Center, Box 581, S-751 23 Uppsala, Sweden*

[3] *State University of Utrecht, Laboratory for Physiological Chemistry, Vondellaan 24a, 3521 GG Utrecht, The Netherlands*

INTRODUCTION

The adenovirus system is often used as a model to study eukaryotic gene expression. The adenovirus group has in addition, attracted a great deal of attention due to the oncogenic potential of its members. In 1962, Trentin and his colleagues discovered that human adenovirus type 12 (ad12) causes tumors in newborn hamsters. Since then a number of adenovirus serotypes have been shown to be oncogenic for rodents and the human adenoviruses are subdivided into highly, weakly, and non-oncogenic serotypes. Almost all human adenoviruses, including the non-oncogenic serotypes, are, however, able to transform rat or hamster cells in tissue culture. Like many other DNA tumor viruses, the adenoviruses preferentially transform cells which are non-permissive for viral replication. The ultimate consequence of viral transformation is the integration of viral DNA sequences in the genome of the transformed cell. The sequences present in cells transformed by adenovirus type 2 (ad2) have been studied in great detail, using nucleic acid hybridization methods. Sharp *et al.* (1974) demonstrated for the first time that rat cells transformed by ad2 do not contain a complete copy of the viral genome. Subsequently it has been shown that the presence of

subgenomic fragments of viral DNA in cells transformed by adenovirus types 2 and 5 is a rule rather than an exception (Gallimore *et al.*, 1974). Graham and coworkers (1974) have identified the transforming genes of adenovirus type 5; by the use of specific restriction enzyme fragments they were able to show that genes mandatory for transformation are located at the left-hand end of the adenovirus genome. Subsequently van der Eb *et al.* (1977) showed that the smallest fragment required to achieve transformation comprises the leftmost 4.5% of the adenovirus type 5 genome. However, cells transformed by this small fragment have a different phenotype than cells transformed by virus or large fragments. The minimum fragment which has been used to obtain complete transformation includes the leftmost 15% of the viral genome (van der Eb *et al.*, 1979). Promotor mapping studies have identified two major early promotors in the leftmost early region of ad2, also known as region E1 (Fig. 1). In this way region E1 is subdivided into the two transcription units, E1A and E1B (Wilson *et al.*, 1979). In addition, the mRNA for the quasi-late polypeptide IX is transcribed from a separate promotor in region E1 (Alestrom *et al.*, 1980).

Studies on the organization of the adenovirus genome have progressed very rapidly. Many genes have been mapped by hybridization coupled with *in vitro* translation (Halbert *et al.*, 1979; Lewis *et al.*, 1979) and large regions of the genome have been analyzed by DNA sequencing methods (for review see Tooze, 1980). The sequenced regions include the left-hand end of the ad5 genome, carrying all genes required for transformation.

In this chapter we describe a molecular analysis of mRNAs from the transforming region of the adenovirus genome and a preliminary study of integrated viral sequences in two adenovirus-transformed cell lines.

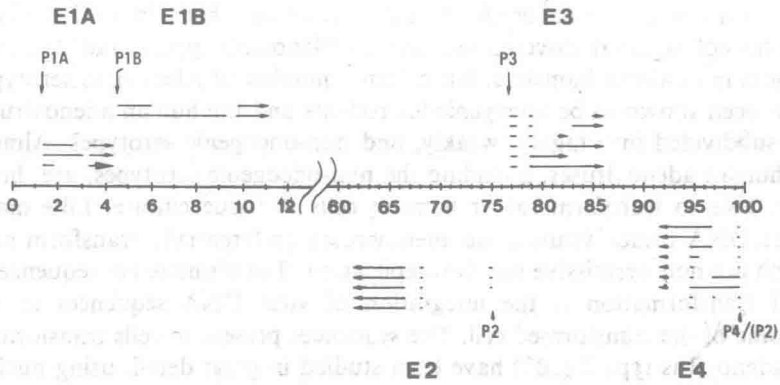

FIG 1. Schematic drawing, indicating regions which are transcribed early after adenovirus infection. Region E1 contains two promoters (P1A, P1B) which subdivide it into transcription units E1A and E1B.

THE mRNAs FROM REGION E1A

Region E1A comprises the leftmost 4.5% of the ad2 genome. Electron microscopy and S1 nuclease mapping have identified three distinct mRNAs in this region, a 13S, a 12S and a 9S mRNA (Berk and Sharp, 1978; Spector *et al.*, 1978; Chow *et al.*, 1979). The 9S mRNA appears to be synthesized exclusively late after infection and is presumably not involved in transformation. All the E1A mRNAs are believed to be transcribed from a single promotor, located about 475 nucleotides from the left-hand end of the ad5 genome (Baker and Ziff, 1980). They are spliced and carry identical 5′ and 3′ ends, but have intervening sequences of different sizes. In order to examine the mRNAs from the transforming region of subgroup C human adenoviruses we have used molecular cloning techniques. Double-stranded cDNA copies were synthesized with early mRNA as a template and inserted into the Pst I cleavage site of the pBR322 plasmid by dG/dC tailing. Recombinant plasmids corresponding to the 12S and the 13S mRNAs from region E1A have been identified and part of their sequences have been determined. By comparing the established sequences with the sequence of region E1A of the closely related ad5 DNA (van Ormondt *et al.*, 1978) it was possible to deduce the structure of two spliced mRNAs (Fig. 2). The splice in the 12S mRNA deletes sequences between nucleotides 975 and 1228 and the small splice in the 13S mRNA deletes sequences between nucleotides 1113 and 1228. The two mRNAs thus use a common acceptor site for splicing but different donor sites. In both mRNAs the splice results in a frame shift which allows the DNA sequence to be utilized in the most efficient way and a very AT-rich region containing numerous termination codons is excluded from both mRNAs. From the structure of the two mRNAs it is possible to predict the amino acid sequence of the corresponding protein products. The polypeptide specified by the 13S mRNA will be 288 amino acids long (32K) whereas the polypeptide corresponding to the 12S mRNA will be 242 amino acids long (26K), assuming that the first AUG following the cap is used for initiation of translation. Both polypeptides have unusual amino acid compositions. They are particularly rich in proline and glutamic acid, and a region with alternating glutamic acid and proline residues is present in both polypeptides. Due to the structure of their mRNAs, the 26K and 32K polypeptides will be completely overlapping, their only difference being a deletion of 46 amino acids from the shorter polypeptide.

The polypeptides encoded by region E1A have also been analyzed by several investigators using mRNA selection coupled with *in vitro* translation (Halbert *et al.*, 1979; van der Eb *et al.*, 1979). In this way it has been possible to show that four polypeptides in the molecular weight range

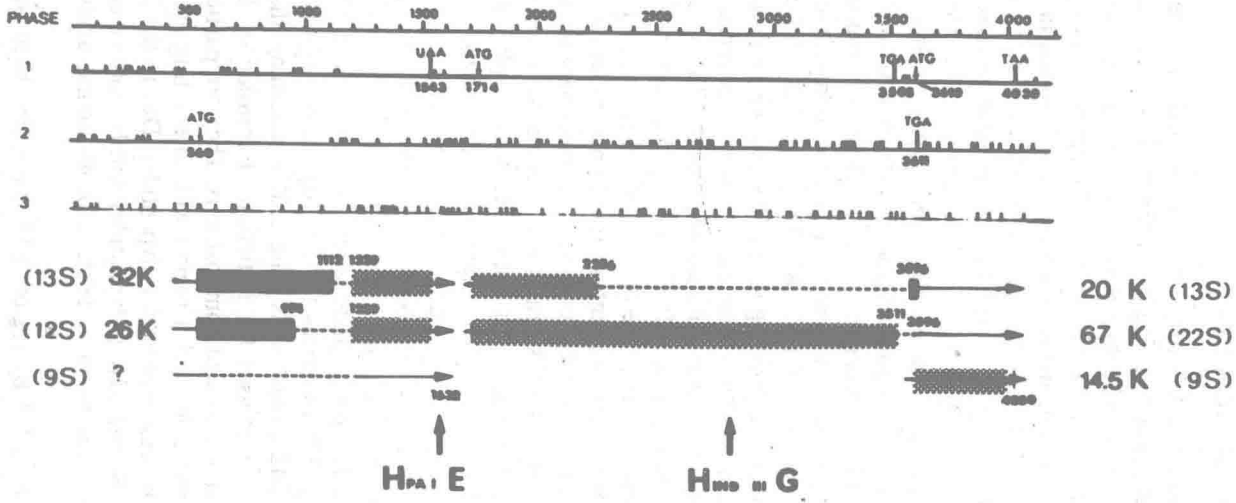

FIG. 2. Schematic drawing indicating the mRNA sequences which are transcribed from the transforming region of the adenovirus type 5 genome. The positions of splices as well as reading frames are indicated.

42–53 000 are encoded by this region. The discrepancy between the molecular weight estimated by SDS polyacrylamide gel electrophoresis and that predicted from the DNA sequence may be due to the unusual amino acid composition of the polypeptides. The difference between the number of mRNAs and the number of polypeptide chains encoded by this region is puzzling. Studies using tryptic fingerprint analysis of the proteins have shown that the four polypeptides in the 42–53K size range all are related and that the 12S and the 13S mRNAs from region E1A each gives rise to two polypeptide species, presumably through post-translation modification. However, we cannot yet exclude the possibility of having additional mRNAs in this region although neither electron microscopy nor S1 nuclease mapping give support for this.

THE mRNAs FROM REGION E1B

The mRNAs of region E1B were analyzed in a similar way. Bacterial clones containing sequences from region E1B were identified by hybridization, using the Hind III-C fragment (map position 8.0–17.0) as a probe. Two different clones were identified, one containing, and one lacking a cleavage site for endonuclease Bgl II (located at map position 9.0 in the ad2 genome). Results from S1 nuclease mapping studies and electron microscopy, although not completely consistent, predict the presence of two differently spliced mRNAs from region E1B (Chow *et al.*, 1979; Berk and Sharp, 1978). These mRNAs have common 5′ and 3′ ends and differ by the size of their intervening sequences. They also have a common promoter which has been mapped around nucleotide 1675 by sequencing the 5′ ends of the E1B mRNAs (Baker and Ziff, 1979). By sequence analysis of selected regions of the clones it was possible to correlate our results with the established DNA sequence for region E1B (Maat and van Ormondt, 1979; Maat *et al.*, 1979) and in this way predict the structure of a 13S and a 22S mRNA from region E1B (Fig. 2). The 22S mRNA is spliced between nucleotides 3512 and 3595. It is interesting to note that the coding potential of the 22S mRNA will not be altered by the splice since the termination codon occurs before the splice and the polypeptide will thus be translated in an uninterrupted reading frame. Thus, this mRNA resembles the mRNA for the small t-antigen of SV40 which also is read in a contiguous sequence from a spliced mRNA (Reddy *et al.*, 1979; Fiers *et al.*, 1978). The 13S mRNA is spliced in a different fashion; the acceptor site is shared with the 22S mRNA but the donor site is located at position 2257 in the DNA sequence. From the sequence we can predict that the 13S mRNA will be translated before as well as after the splice although in different reading frames. Only five codons