1985

The Oncogene of Avian Sarcoma Virus UR2 and its Cellular Homologue: Structure, Sequence and Expression

Wendi S. Neckameyer

The Oncogene of Avian Sarcoma Virus UR2 and its Cellular Homologue:

Structure, Sequence and Expression

A thesis submitted to the faculty of The Rockefeller University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Wendi S. Neckameyer

1 April 1985
The Rockefeller University
New York, NY

Table of Contents

V. Discussion

Acknowledgements

There are several people who have contributed to this dissertation, most of whose names will never appear on any publications resulting from this work, but to whom I am indebted. I would like to thank my parents and other members of my family for their understanding, support and encouragement through some very trying times.

I am grateful to Dr. William Sawyers for his encouragement during my first trials with research.

I want to thank Wilma Friedman and Celeste Simon for their very necessary support during our various trials by fire.

There are several members of the Hanafusa laboratory who have been very generous in sharing their ideas, techniques, and friendship. Working with them has been a genuine pleasure. They are Ellen Garber, Shinji Iijima, Sally Kornbluth, Boris Lin, Bruce Mayer and Masabumi Shibuya. I am also indebted to Drs. Teruko and Hidesaburo Hanafusa for their suggestions and guidance during my time in the laboratory.

Lastly, I want to express my gratitude to my advisor, Dr. Lu-Hai Wang. Webster's New Collegiate Dictionary defines the word support as follows:

1. to endure bravely or quietly
2. to promote the interests or cause of
3. to uphold or defend: assist, help
4. to provide with substantiation
5. to hold up or serve as a foundation or prop for
6. to keep from fainting, yielding or losing courage
7. to keep (something) going

This certainly describes Lu-Hai's role over the past four years, and I am very

grateful for all he has done.

" Whatever you do, don't embarrass us"

-from Buckaroo Bonsai

Summary

Avian sarcoma virus UR2 is a replication-defective virus that can induce sarcomas *in vivo* and transforms chicken embryo fibroblasts in culture to a characteristic, extremely elongated morphology. The genome of UR2 contains a 1.2 kb transformation-specific sequence, v-*ros*, which has a homologous counterpart, c-*ros*, in normal chicken cellular DNA. The 5' end of v-*ros* is fused to the helper virus UR2AV-related sequencing coding for one of the viral *gag* proteins, p19. The fused p19 and *ros* sequences in UR2 code for a polyprotein of 68 kd, $P68^{gag\text{-}ros}$, which has an associated tyrosine kinase activity. To elucidate the basis of the functional conservation as well as differences between *ros* and other oncogenes, I sequenced the entire genome of UR2 and compared the predicted amino acid sequence of P68 with other members of the tyrosine kinase family.

The results show that *ros* is 1273 nucleotides in length, including a 65 bp 3' noncoding stretch. *ros* is joined at its 5' and 3' ends to the 3' region of p19, and the 3' region of gp37, respectively, and replaces the sequences of UR2AV in between. The deduced amino acid sequence for P68 gives a molecular weight of 61,113 daltons and shows that it is closely related to the oncogene family coding for tyrosine protein kinases. However, P68 contains two distinctive hydrophobic regions that are absent in most of the other tyrosine kinases and it has unique amino acid changes and insertions within the conserved domain of the kinases.

I also determined the sequence of cellular *ros* and compared it to viral *ros* to determine the changes between them that may be responsible for their differential oncogenicity; in addition I have analysed the expression of the cellular gene in both embryonic and adult chickens. The 1.2 kb v-*ros* sequence is remarkably well conserved when compared with the corresponding region of c-*ros*. The v-*ros* sequences are distributed in nine exons of c-*ros* over a range of 12 kb of DNA.

The two differ only by a 9 bp duplication, a single base change not resulting in an amino acid change, and the divergence of their 3' ends. c-*ros* and v-*ros* abruptly diverge 36 bp upstream of the v-*ros* termination codon. The open reading frame of c-*ros* continues after this divergence and may terminate 34 amino acids downstream, or, more likely, the reading frame is spliced to further 3' coding sequences using a splice donor site 24 nucleotides downstream of the divergence. The v-*ros* sequence 3' to the divergence was not found in the 3' c-*ros* sequences in the lambda clone or in helper virus-related sequences. Comparison of the nucleotide sequences of viral and cellular *ros* suggests the viral *ros* and Δ*gag* junction in UR2 was formed by splicing.

A 3.1 kb c-*ros* transcript has been detected in adult muscle tissue and in kidney from chickens only after long exposure of the Northern filters. It appears that the c-*ros* transcript in kidney is preferentially degraded relative to *src*, and this may account for the seeming lack of expression in this tissue. In all other tissues examined, and in several cell lines, c-*ros* transcripts are not detectable.

Although the exact function of the cellular *ros* protein is not known, evidence based on the sequence of the viral and structural genes has enabled us to detect its structural similarity with the EGF and insulin receptors and postulate that *ros* is a member of a family of growth factor receptors.

List of Figures

Abbreviations

AEV avian erythroblastosisvirus

AMV avian myeloblastosis virus

ALV avian leukosis virus

BaEV baboon endogenous virus

bp base pair

CEF chicken embryo fibroblasts

EDTA ethylenediamine tetraacetic aid

FSV Fujinami sarcoma virus

kb kilobase

kd kilodaltons

LLV lymphoid leukosis virus

LTR long terminal repeat

MMTV mouse mammary tumor virus

PB primer binding site

PR-C Prague C strain of Rous sarcoma virus

RAV Rous associated virus

SDS sodium dodecyl sulfate

SR-A Schmidt-Ruppin A strain of Rous sarcoma virus

TBR tumor-bearing rabbit

UR2 University of Rochester virus isolate number 2

UR2AV UR2 associated virus

Nomenclature for oncogenes

onc	virus strain
abl	Abelson murine leukemia virus
*erb*B	AEV-H
fes	Snyder-Theilen and Gardner-Arnstein FeSV
fgr	Garner-Rasheed FeSV
fms	McDonough FeSV
fps	FSV and others
fos	Finkel-Biskis-Jenkins murine sarcoma virus
mos	Molony murine sarcoma virus
myb	AMV
myc	MC29 and others
sis	Simian sarcoma virus
src	RSV
ros	UR2
yes	Y73, Esh sarcoma virus

Section I. Introduction

Oncogenic viruses provide simplified systems to study tumorigenesis and cell transformation: introduction of the virus into a normal cell is capable of transforming the cell to a neoplastic state. Tumor viruses may contain RNA or DNA as their genomes; however, the DNA tumor viruses are diverse and encompass several classes which differ in genomic structure and methods of replication. The RNA tumor viruses are more unified in their genomic organization and mode of replication. These viruses replicate through a double-stranded DNA intermediate synthesized by an RNA-dependent DNA polymerase which is encoded by one of the viral genes. The more uniform structure and method of replication of the RNA tumor viruses reduces the number of variables to be considered when studying the mechanisms involved in cellular transformation by this group of viruses. RNA tumor viruses induce rapid and reproducible transformation of cells in culture and tumor formation *in vivo*, and the product(s) encoded by the viral genome are responsible for the induction and maintenance of the transformed phenotype. RNA tumor viruses have been isolated from a wide spectrum of vertebrate species, mostly murine, feline and avian. These isolates share many structural similarities, yet are capable of causing a wide variety of neoplasms in their respective hosts.

Taxonomy.

Retroviridae (so named for the RNA-dependent DNA polymerase, or reverse transcriptase) consists of three subfamilies: the oncogenic RNA-containing viruses

(or oncoviruses), and two non-oncogenic subfamilies. The non-oncogenic subfamilies include the Lentiviruses, which produce disease only after prolonged latency, and the spumaviruses, or foamy viruses, which chronically infect many mammalian species without obvious manifestation of disease.

The viral genome of all three subfamilies is composed of a dimer of linear positive-sense single stranded RNA molecules linked at their 5' ends. The genome is replicated through a DNA intermediate via the reverse transcriptase encoded by one of the viral genes. These viruses are enveloped, with glycoprotein surface projections.

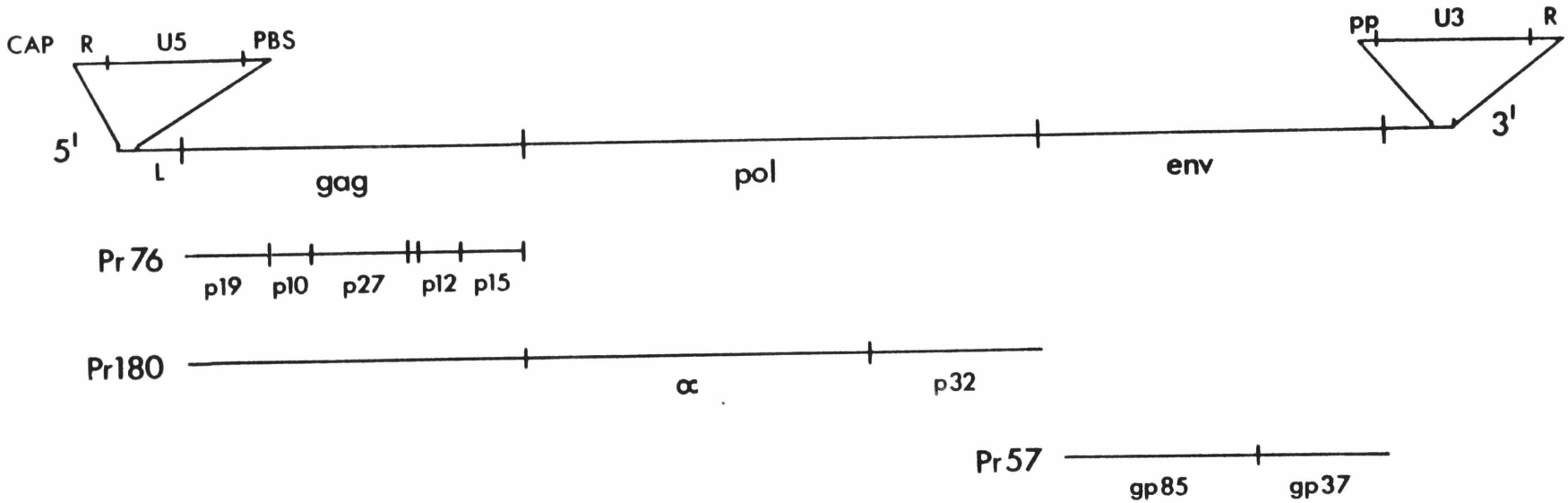
Four categories of oncovirus have been described by morphological classification of physical structure in electron micrographs: A-, B-, C-, and D-type particles. Most oncoviruses have been classified as type C, and can be divided into two categories: leukosis viruses with low oncogenic potential, and leukemia and sarcoma viruses with high oncogenic potential. Both kinds of virus have been isolated from murine, feline, avian and primate species, but this discussion will be restricted to viruses of avian origin.

Lymphoid leukosis viruses (LLVs).

Avian leukosis virus (of which the Rous-associated viruses, RAV-1 and RAV-2, and the UR2-associated virus, UR2AV, are examples) does not transform chicken cells in culture, and induces tumors, mostly lymphomas, *in vivo* only after a long latency period of three to four months after infection of chickens. The disease originates primarily in the bursa. Infected birds may show proliferation of immature red blood cells, and/or infiltration of neoplastic cells into the liver or spleen.

The genome is composed of three genes important for viral replication: *gag*, *pol*, and *env*, as well as the long terminal repeat regions, or LTRs (Fig. 1). The

Fig. 1. Genome of ALV. The upper lines show the terminal regions expanded as indicated. The precursor protein is shown as a solid bar underneath the genome, and the molecular weights of the processed proteins are indicated. L, *gag*, *pol*, and *env* are defined in the text.


CAP
R
U5
PBS
5'
L
gag
pol
env
pp
U3
R
3'
Pr 76
p19
p10
p27
p12
p15
Pr180
α
p32
Pr 57
gp85
gp37

ALV genome is bounded on its 5' end by a cap, and a short sequence, R, which is repeated at both the 5' and 3' ends of the RNA. The binding site for the initiation of the tRNA primer (PB) for negative-strand DNA synthesis follows the region denoted U5. The *gag* gene is 2.1 kb in length and codes for the internal structural proteins of the virion, which are first synthesized as a 76,000 dalton precursor ($Pr76^{gag}$) from the 35S genomic RNA. Pr76 is subsequently cleaved by proteolysis to several lower molecular weight *gag* proteins. The genomic order of these proteins is H_2N-p19-p10-p27-p12-p15-COOH. The *gag* gene is preceded by an untranslated sequence (L) of approximately 250 nucleotides which is part of the 5' 380 nucleotide leader sequence that is spliced to the subgenomic mRNAs. The L region may include sequences important for RNA packaging. The *pol* gene has a length of 2.7 kb and codes for the reverse transcriptase, which is synthesized as a *gag-pol* precursor, Pr180, from the 35S genomic RNA (or an as yet unidentified spliced mRNA). Pr180 is cleaved to produce a 92,000 dalton polypeptide known as β which is subsequently cleaved into the 58,000 dalton α, and 32,000 dalton p32, proteins. The reverse transcriptase is composed of α and β subunits. A DNA endonuclease activity was found to be associated with p32 and is proposed to be involved in the process of viral DNA integration. The *env* gene codes for the glycoproteins coating the surface of the virion. Unlike *gag*, which is translated from 35S genomic RNA, *env* is translated from a 22S subgenomic mRNA. The primary *env* product, a 57,000 dalton protein, is glycosylated to form the precursor, $gPr92^{env}$ which is later cleaved into two polypeptides, gp85 and gp37. The region immediately preceding U3 contains a polypurine tract (pp) believed to be important for initiation of positive-strand DNA synthesis. U3 precedes the 3' terminal repeat region, R, and contains promoter sequences necessary for transcription of the genome. During replication of the viral RNA to form proviral DNA, U3 and U5 are duplicated to form the LTR

(consisting of U3-R-U5) which appears at each end of the provirus.

After nonspecific adsorption of the virus to the host cell surface, a penetration step follows that is dependent on the envelope glycoproteins of the virus and on specific cell receptors. A cellular tRNA acts as primer by binding near the 5' end of the genome (PB) and initiating synthesis of (-) strand DNA. Concommitant synthesis of (+) strand DNA yields a double stranded linear provirus. The reverse transcriptase coded by the *pol* gene is responsible for the synthesis of the DNA intermediate. Duplication of the U3 and U5 regions during DNA synthesis leads to formation of the LTRs. The linear DNA is synthesized in the cytoplasm; after transport to the nucleus, a population of the linear molecules is converted to a closed circular form. Recent work by Panganiban and Temin (48) has shown that the 5 bp terminal inverted repeats at the ends of the LTRs from unintegrated linear DNA, as well as 3-7 adjacent nucleotides, are required for integration. After integration into the host genome, the provirus will replicate with the cellular DNA (Fig. 2). Transcription of the proviral template yields RNA that can be packaged and released by budding as infectious virions after assembly on the cell membrane. However, integration of the retrovirus DNA is not absolutely required for retroviral gene expression of the spleen necrosis virus(48).

The leukosis viruses do not contain oncogenic sequences. However, these viruses may become acutely transforming by transduction of cellular sequences, now known as proto-oncogenes or cellular oncogenes. These cellular genes acquired by the retrovirus are called viral oncogenes. Acquisition of the viral oncogene usually involves the sacrifice of some of the viral replicative genes, so that the newly-derived acutely transforming virus needs the parental leukosis virus as "helper" to replicate.

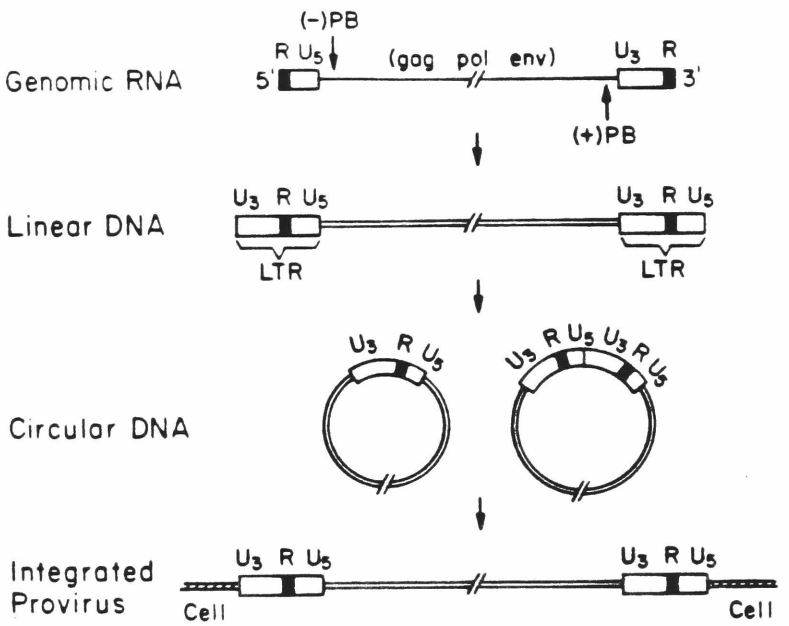


Fig. 2. Structure of retroviral RNA and proviral DNA. R, terminal redundancy on genomic RNA; U5 and U3, sequences unique to the 5' and 3' ends of the genomic RNA; LTR, long terminal repeat.

Acutely transforming viruses.

Rous sarcoma virus, isolated in 1910 by Peyton Rous, was the first retrovirus later found to contain a genetic sequence essential for neoplastic transformation that was independent of the genes necessary for viral replication. The transforming sequence of RSV, known as *src*, is inserted in nondefective strains of the virus between *env* and the LTR, and the resulting virus retains all the replicative functions (Fig. 3). (In Fig. 3, all viruses shown are presented as the proviral DNA intermediate.) In some isolates of RSV, e.g. the Bryan high-titer strain, *src* has replaced *env* and the virus then requires an associated leukosis virus to provide replicative functions in trans. Although the nondefective strains of RSV were at one time considered to be prototypical of the sarcoma viruses, they have turned out to be exceptional among all known acutely transforming retroviruses in that acquisition of the *src* sequences has not been at the expense of some of the viral replicative genes. The *src* gene product is translated from a subgenomic mRNA and encodes a 60,000 dalton phosphoprotein, $pp60^{src}$. With the help of *src* deletion and temperature-sensitive mutants, coupled with transfection of *src* DNA, it was demonstrated that both *in vitro* cell transformation and *in vivo* tumorigenicity of RSV require a functional *src* product. $pp60^{src}$ is phosphorylated on serine and tyrosine, and has been shown to have tyrosine kinase activity. Phosphotyrosine is extremely rare in normal cells, but cells transformed by RSV have a 10-fold elevated level of phosphotyrosine (58). It is believed that transformation by RSV is dependent on the tyrosine kinase activity of $pp60^{src}$.

In 1976, Dominic Stehelin experimentally demonstrated the presence of cellular sequences in chicken related to viral *src* (66). Later studies showed that *src*-related sequences are present in fish, birds, mammals, and even *Drosophila*, indicating these sequences have been highly conserved during evolution. The gene product of the *src* cellular homologue was identified in normal chicken cells as a

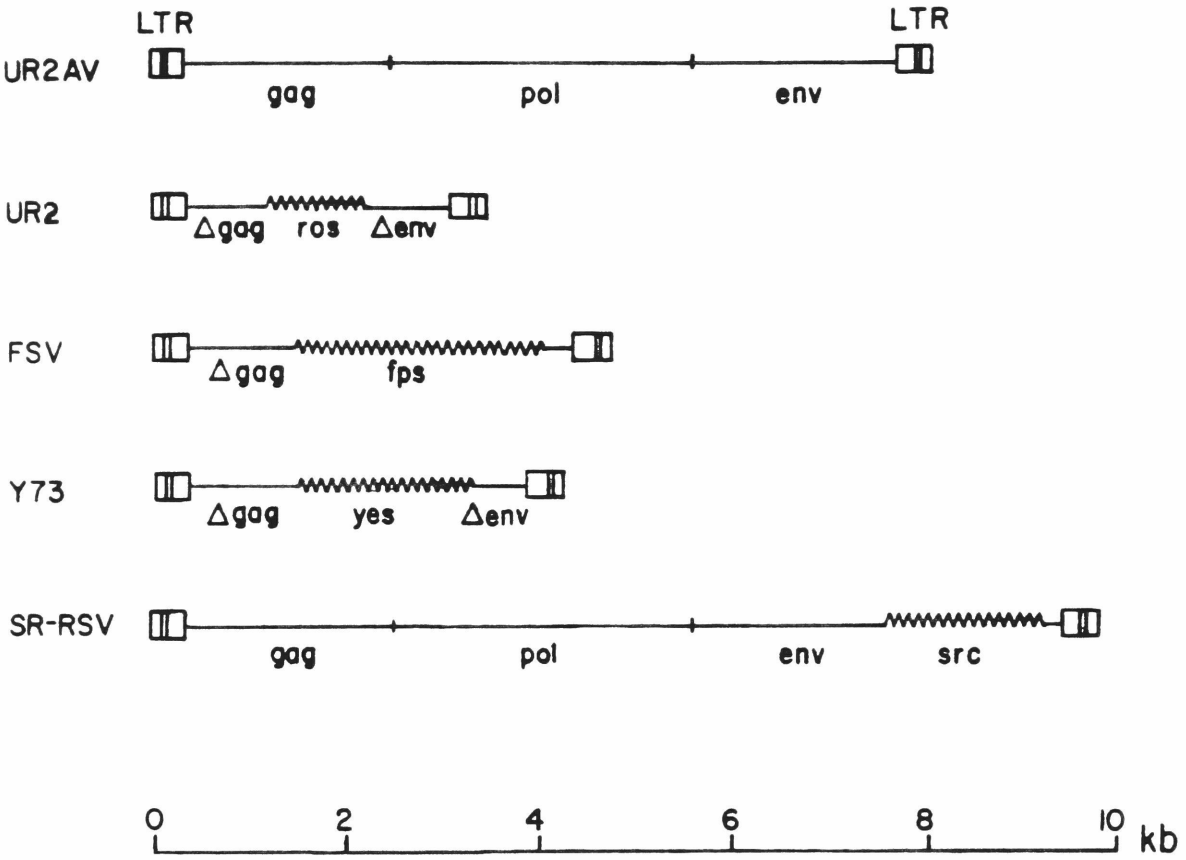


Fig. 3. Genomic structure a representative member of each class of avian sarcoma virus. UR2AV, the UR2-associated virus, is shown to compare the structure of an avian leukosis virus with the ASVs. All viruses are shown as the proviral DNA form. UR2, University of Rochester viral isolate number 2; FSV, Fujinami sarcoma virus; Y73, Yamaguchi-73; and SR-RSV, Schmidt-Ruppin strain of Rous sarcoma virus. *ros*, *fps*, *yes*, and *src* (denoted by jagged line) are the transforming genes of UR2, FSV, Y73, and RSV, respectively.

cells as a 60,000 dalton phosphoprotein which also has tyrosine kinase activity; the viral and cellular proteins are highly homologous.

RSV and its derivatives were thought to be the only avian sarcoma viruses until 1980, when the Fujinami sarcoma virus was shown to contain a transforming gene distinct from *src* (24, 33). This was originally shown by partial sequence analysis of FSV RNA by oligonucleotide fingerprinting. The transforming sequence of FSV is now known as *fps*. Several viruses which have independently transduced portions of the cellular *fps* gene have been isolated and include FSV, PRCII, PRCIV, 16L, and UR1. It appears that *fes*, the transforming sequence of the Snyder-Theilen and Gardner-Arnstein strains of feline sarcoma virus, is the feline homologue of avian *fps* (59).

Acquisition of *fps* by FSV has led to the loss of the 3' region of *gag* and the entire *pol* and *env* genes (Fig. 3). FSV encodes a 140,000 dalton *gag*-fusion product that also has an associated tyrosine kinase activity.

A third class of avian sarcoma virus is comprised of viral isolates that have acquired the transforming sequence known as *yes*, and includes the Yamaguchi 73 (Y73) and Esh sarcoma virus (ESV) strains. Y73 encodes a 90,000 dalton phosphoprotein, P90, which, like FSV P140, has 5' *gag* sequences fused to the amino-terminus of the specific sequence.

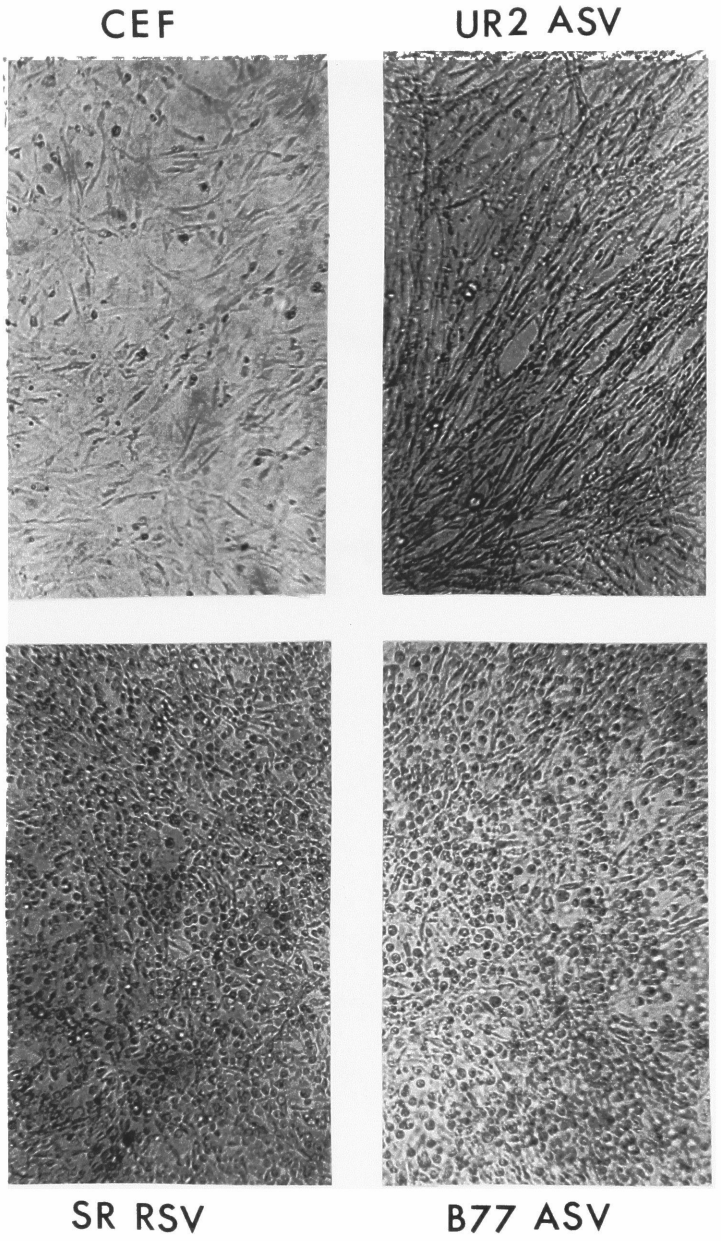
The fourth and most recently characterized class of avian sarcoma virus is the UR2 virus, so called because it is the University of Rochester avian sarcoma virus isolate number two. UR2 was originally isolated with its associated helper virus, UR2AV, in 1963 from a fibrosarcoma of a 9 month old White Rock chicken (4). UR2 has lost most of both the *gag* and *env* genes and all of the *pol* gene during acquisition of *ros* (Fig. 3). UR2 is able to induce sarcomas in chickens and efficiently transform chicken embryo fibroblasts (CEF) in culture (4). CEF transformed by UR2 are characterized by an extremely elongated morphology (4).

Fig. 4 shows the characteristic UR2 transformed morphology, relative to the refractile, rounded morphology of cells transformed by two strains of RSV, Schmidt-Ruppin and B77. Previous analysis of the UR2 RNA genome showed that it shares with its associated helper virus, UR2AV, about 2.1 kb of 5' and 3' sequences and contains at the middle of the genome about 1.2 kb of *ros*-specific sequence (82). Studies of the *ros* sequence by hybridization and oligonucleotide fingerprinting have shown that it is distinct from the transforming genes of other known ASVs and acute leukemia viruses (61, 82). The normal cellular DNA homologue of viral-*ros* has been detected in chickens, quail and ducks (61). However, the expression of c-*ros* in various tissues and organs of 10- to 14-day old chickens is very low (less than 1 copy per cell), except in the kidney (2.5 copies per cell) (61).

The UR2-infected cells produced only 24S genomic RNA (82) which was shown to encode for a 68,000 dalton *gag-ros* fusion protein, P68, that was associated with a tyrosine-specific protein kinase activity (17) (Fig. 5). The enzymatic properties of P68 are distinct from other ASV protein kinases, i.e., Rous sarcoma virus (RSV) p60, Fujinami sarcoma virus (FSV) P140, and Y73 ASV P90, in cation preference, pH optimum, and phosphate donors (17). Similar to other ASV-transformed cells, organization of microfilament bundles in UR2-transformed CEF is significantly decreased compared to uninfected cells (2). A common feature of most of the acutely transforming viruses exemplified by three classes of avian sarcoma virus shown in Fig. 3 is that the acquired cellular sequences are fused to viral sequences, in these cases, *gag*, to form a fusion polypeptide. All virus isolates described thus far in the literature which have acquired either *fps*, *yes*, or *ros* are replication-defective.

In addition to the sarcoma viruses, it was shown that certain avian leukosis viruses, for example, AMV (avian myeloblastosis virus), AEV (avian erythroblas-

Fig. 4. Transformation of CEF in culture by ASV. CEF, uninfected chicken embryo fibroblasts; B-77 and SR-RSV, CEF transformed by the Bratislava or Schmidt-Ruppin strains of RSV; UR2, CEF transformed by UR2.


CEF
UR2 ASV
SR RSV
B77 ASV

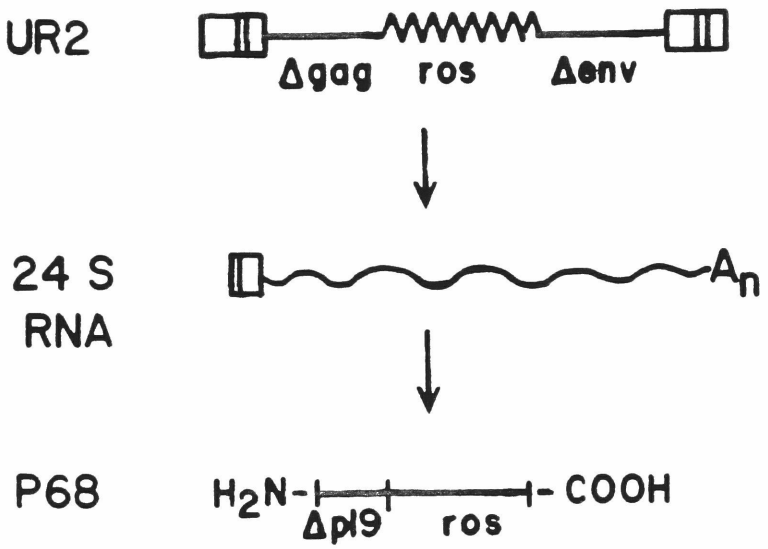


Fig. 5. Production of P68$^{gag\text{-}ros}$ from UR2 proviral DNA. The integrated provirus is transcribed into 24S RNA, which is translated to yield P68. Δ*gag*, *ros*, and Δ*env* are defined in the text.

tosis virus), and MC29, have also acquired cellular proto-oncogene sequences that permit the rapid induction of leukemias *in vivo* and the transformation of specific hematopoietic cells *in vitro*. These acute leukemia viruses are defective for replication as well, due to the loss of viral replicative sequences during transduction of the proto-oncogene.

Scope of this thesis.

Clearly, *ros* is a member of the retroviral oncogene family coding for tyrosine-specific protein kinases, despite differences in nucleotide sequences among these genes. In order to understand the basis for similarities and differences in the transforming functions of *ros* and other oncogenes, a detailed analysis of the genetic structure of UR2 and *ros* sequence is necessary.

I have molecularly cloned the full-length genomes of UR2 and UR2AV. Both were shown to be biologically active. Cross-hybridization among *ros* and three other ASV transforming genes (*src* of RSV, *fps* of FSV, and *yes* of Y73 ASV), showed that *ros* shared some sequence homology with *fps*, and little or no homology with *src* and *yes*. I have also sequenced the entire genome of UR2 and compared the predicted amino acid sequence of P68 with other members of the tyrosine-specific protein kinase family. This work will be detailed in Section III.

The strong evolutionary conservation of most viral oncogenes and the fact that several of the cellular oncogenes are expressed in both embryonic and adult tissues suggests they play an important role in growth and/or differentiation of normal cells. Intensive characterization of known oncogenes has in a few cases enabled us to infer their function in the normal cell; for example, *sis*, an oncogene transduced by a simian retrovirus, appears to code for part of the platelet-derived growth factor. *erb*B, an oncogene of avian erythroblastosis virus, appears to represent part of the EGF receptor. In the majority of cases, however, little is

known about the normal function of the cellular proto-oncogene, and what changes during or after acquisition of the oncogene by the virus confer transforming ability to these genes. It is possible that transformation by retroviruses is dependent on the sustained and abundant expression of the oncogene, or that mutation of the viral oncogene during transduction by the virus in some way effects transformation.

To address these fundamental questions, I have analysed the structure and expression of the chicken cellular homologue of viral *ros*. The cellular and viral *ros* genes appear to have been highly conserved during transduction of the specific sequence by the virus and subsequent passages of the virus in culture. In addition, the cellular *ros* locus is under very tight developmental control. This work is described in Section IV.

Section II. Materials and Methods

Cells and viruses.

The preparation of CEF, UR2, and the subgroup A UR2-associated virus (UR2AV) followed the published procedure (4, 82). A methylcholanthrene-transformed quail cell line, QT6 (42), was cultured similarly to CEF.

Isolation of closed circular proviral DNA.

A pilot experiment was performed to determine conditions for maximal yield of unintegrated circular proviral DNA for cloning. QT6 or CEF cells were seeded at 5 x 10^6 cells per dish and infected with 5 ml of UR2(UR2AV) containing medium from transformed CEF cultures in the presence of 17 ug per ml DEAE-dextran. Medium was replaced 5 h later. 6 10 cm dishes each (approximately 5 x 10^7 cells) of UR2(UR2AV)-infected QT6 or CEF were harvested at 24, 48 and 72 hour timepoints. At each timepoint, total cellular DNA was isolated by the Hirt procedure (25), and 20 ug of the Hirt supernatant DNA containing unintegrated provirus was electrophoresed through a 0.8% agarose gel, blotted to nitrocellulose paper, and probed with ^{32}P cDNA made from 24S UR2 RNA. The highest yield of proviral DNA was obtained from UR2(UR2AV)-infected QT6 cells harvested 24 h after infection. Therefore, QT6 cells were seeded as described above and infected with UR2(UR2AV) at a multiplicity of one in the presence of DEAE-dextran and were harvested 24 h postinfection. After Hirt precipitation, the DNA in the supernatant was extracted with phenol-chloroform and concentrated by ethanol precipitation. The DNA was further purified by treatment with RNase A, additional phenol-chloroform extraction, and ethanol precipitation. To enrich for closed circular DNA, the DNA was then extracted with acid-phenol as

described by Zasloff (86). A yield of 960 ug of DNA was obtained from a total of 95 8.5-cm dishes after acid-phenol extraction and were applied to a Bio Rad A5m (100-200 mesh) gel column to remove small (less than 1 kb), contaminating linear DNA fragments. The column was washed and the DNA was eluted with a buffer containing 20 mM Tris-HCl (pH 7.2), 0.1 N NaCl, and 5 mM EDTA. A total of 20 ug of closed, circular DNA were recovered in the peak fractions.

Molecular cloning and subcloning of UR2 and UR2AV DNAs.

One ug each of *Eco*RI- or *Sst*I-cut UR2 and UR2AV closed circular proviral DNAs were ligated to two ug of λgtWESλB (32) *Eco*RI or *Sst*I arms, respectively. One ug of the ligation mix was packaged *in vitro* into lambda phage particles (32, 62). Lambda arms were purified from the internal fragment by sucrose gradient centrifugation (36). The recombinant phages were titered on *E. coli* ED8654 (44). Packaging efficiencies of 2.6 x 10^5 and 8 x 10^3 plaque forming units (pfu) per ug of DNA were obtained for the *Sst*I and *Eco*RI cut and ligated DNAs, respectively. Screening of the recombinant phages with ^{32}P-labelled cDNA made from 24S UR2 RNA was according to the procedure of Benton and Davis (6).

Subcloning of the *Eco*RI insert from the UR2-lambda recombinant phage clone into pBR322 was according to Bolivar et. al. (8). The UR2 plasmid clone was called pUR2. Similarly, a 850 bp *Eco*RI - *Pvu*II v-*ros* specific DNA fragment was cloned using the 2293 bp *Pvu*II - *Eco*RI fragment of pBR322 DNA. This clone was called p*ros*1.

A recombinant clone containing cellular sequences homologous to v-*ros* was previously isolated by Masabumi Shibuya in our laboratory from a library composed of *Alu*I- and *Hae*III-partially digested chicken genomic DNA cloned in Charon 4A with *Eco*RI linkers (15). The three *Eco*RI fragments comprising λc-*ros* were freed from the lamda vector DNA and subcloned into the *Eco*RI site of

pBR322. *E. coli* C600 cells (3) were used for all transformations.

Transfection of cloned UR2 and UR2AV DNAs .

The UR2 and UR2AV DNAs were freed from the vector DNAs by restriction enzyme digestion and were purified by agarose gel electrophoresis. They were transfected without prior ligation onto CEF using the calcium phosphate method (22) with the following modifications. One ug each of UR2 and UR2AV DNA were mixed with with salmon sperm DNA to a total of 20 ug and added to 500 ul of 20 mM HEPES (pH 7.0), 137 mM NaCl, 5 mM KCl, 0.59 mM Na_2HPO_4. One-tenth volume of 1.25 M $CaCl_2$ was added and the mixture was allowed to sit at room temperature for 15 minutes. The solution was then added to cells in 5 ml of growth medium and the culture was incubated at 37°C. The cells were seeded one day before at 7 x 10^5 cells per 6 cm plate. Medium was changed five hours after the addition of viral DNAs. The cells were transferred to 8.5 cm dishes as they became confluent and overlayed with soft agar medium the next day to enhance the growth of transformed cells. Virus production was assayed by determining the presence of reverse transcriptase activity in the supernatants of transfected cells (80).

Protein analysis.

UR2/UR2AV transfected cells grown in 6 cm dishes were labeled with 500 uCi per plate of [^{35}S]-methionine for 5 h. Extraction of cold or ^{35}S-labeled cellular proteins, immunoprecipitation, protein kinase assay and sodium-dodecyl sulfate (SDS)-polyacrylamide gel (5 to 15%) electrophoresis followed the described methods (16). Anti-*gag* serum and sera from RSV-infected tumor bearing rabbit (TBR) were kindly provided by Ricardo Feldman.

RNA isolation, blotting and hybridization.

Total cellular poly(A)-containing RNAs from UR2AV and UR2/UR2AV DNA transfected cells were isolated according to the described procedures (81, 83). The RNAs were denatured with glyoxal, separated on 1% agarose gels, transferred to nitrocellulose paper (81, 83), and hybridized to a 5' probe containing the "leader" sequence (see below), and to a probe derived from the 850 bp *ros*-specific DNA described above. Conditions for hybridization and washing of the filters have been described previously (81).

Polyadenylic acid-containing RNAs were isolated from brain, eye, stomach and intestine, heart, and liver of 12, 14, 16, 18, and 20 day chicken embryos and 2 day and 7 day old chicks; from lung tissue of 18 and 20 day embryos and 2 day and 7 day old chicks; from extraembryonal membranes at all embryonic timepoints described above; from kidney of 20 day embryos and 2, 4, 7, 10, 14, 21, 28, and 56 day old chickens; and from muscle tissue of all embryonic and adult timepoints described above. Tissues were removed from freshly-killed birds and frozen by immersion in liquid nitrogen, and were stored at -70 ° C. Alternatively, the tissue was immediately suspended in a solution containing 5 M guanidine thiocyanate, 25 mM Tris-hydrochloride (pH 7.0), 10 mM dithiothreitol, 1% sarkosyl, and 5 mM EDTA and homogenized at high speed in a Waring blender for two 30-second pulses. Unbroken tissue debris was removed by low speed centrifugation. The clear homogenate was layered on top of a solution containing 5.7 M CsCl and 0.1 M EDTA (pH 7.0) and centrifuged at 26,000 rpm for 18 hr. The RNA pellets were resuspended in 10 mM Tris-hydrochloride (pH 7.2), 10 mM NaCl, 1 mM EDTA, and 0.05% SDS, extracted with phenol-chloroform and ethanol precipitated. Poly(A)-positive RNAs were twice selected and electrophoresed as described above. The Northern blots were hybridized to the 850 bp *ros*-specific DNA and to a 3' *src*-specific probe (described below).

Preparation of ^{32}P-labeled DNA probes.

24S UR2 and 35S UR2AV RNA was isolated from purified virus as described previously (79, 80), twice poly(A)-selected and used separately as templates in the *in vitro* reverse transcriptase reaction. The 200 ul reaction mixture contained 50 mM Tris-HCl (pH 8.0), 40 mM KCl, 4 mM $MgCl_2$, 2 mM DTT, 200 uM each dATP, dGTP and TTP, 60 ug of calf thymus DNA primers (72), 20 ug Actinomycin D, 300 uCi (3000 Ci/mmole) [^{32}P]-dCTP, 150 units avian myeloblastosis virus reverse transcriptase (provided by J. Beard, Life Sciences, St. Petersburg, Florida, through the courtesy of J. Gruber, the Resource Program, National Cancer Institute), and 0.3 ug of template. The reaction mixture was incubated at 37 ° C for 2 h, and the reaction was quenched by the addition of a solution containing 0.5 M LiCl, 10 mM Tris-HCl (pH 7.2), 10 mM EDTA, and 0.2% SDS. After ethanol precipitation, the sample was base-hydrolyzed with 0.2 N NaOH containing 1 mM EDTA at 100 ° C for one hour, neutralized, and passed through a Sephadex G-50 column. The DNA from the void volume fractions were precipitated in ethanol.

Probe containing 80-90% *ros*-specific sequences was made by annealing $cDNA_{UR2}$ to total chicken embryo fibroblast RNA and to viral RNA isolated from cells infected with UR2AV to remove cDNAs unrelated to *ros* using the procedure described by Shibuya et. al. (62).

Preparation of the following probes specific to various regions of the genome of the Schmidt-Ruppin strain of RSV was as described (83). The 5' probe is a 500 bp long fragment spanning the *Eco*RI site within the U3 region of the left-hand long terminal repeat (LTR) to the *Bam*HI site in the 5' *gag* region; the 5' *gag* probe is 1.3 kb and spans the two *Bam* HI sites within the *gag* region; and the 3' *gag* probe spans the second *Bam*HI site in *gag* to the downstream *Eco*RI site and is 400 bp in length. *pol* and *pol-env* probes were prepared from pSR2

(13). The *pol* probe is 1.45 kb and spans the *Hind*III to *Bgl*II site in *pol*; the *pol-env* probe is 1.8 kb, extends from the *Bgl*II site in *pol* to the *Eco*RI site in *env*, and covers the 3' portion of *pol* plus more than 1 kb of *env* sequence. The *c* probe covers the 450 bp extending from the *Pvu*II to the *Eco*RI site in the U3 region of Schmidt-Ruppin B.

A probe specific to 3' v-*fps* was derived from pBRFO4, a plasmid containing a 400 bp *Bam*HI DNA fragment from FSV *fps* (62). The 3' v-*yes* probe used was the 1.1 kb *Pst*I fragment from Y73 (29), and the 3' v-*src* probe used was a 900 bp *Pvu*II fragment from pTT107 (68). Two *ros*-specific probes were used: one was the 850 bp *Eco*RI - *Pvu*II fragment which covers the 3' two-thirds of *ros*, and the other was an internal *ros* sequence, the 300 bp *Ava*I D fragment (Fig. 14).

Restriction mapping.

Two approaches were employed. In the first approach, end-labeled DNAs were used for mapping. UR2 DNA isolated from *Eco*RI cleaved pUR2 was labeled at its 5' end with ^{32}P using T4 polynucleotide kinase (BRL) after digestion with bacterial alkaline phosphatase (BRL). UR2AV DNA from one of the recombinant clones was isolated from its lambda vector by digesting with *Sst*I. The 3' protruding *Sst*I ends were labeled using 3'-dATP ([^{32}P]-cordycepin 5' triphosphate, NEN) and terminal transferase (NEN) according to the conditions provided by the manufacturer. The end-labeled DNAs were digested with a single-cut enzyme to generate two fragments of unequal size, each containing a single ^{32}P-labeled end. These fragments were subjected to various restriction enzyme digestions, aliquots containing 10^4 cpm DNA were removed 3 min, 6 min, 10 min, and 60 min after initiation of the reaction and stopped by addition of EDTA to 10 mM followed by heating at 65 ° C for 3 minutes. About 5000 cpm of each restricted, end-labeled fragment were electrophoresed through 5%

acrylamide and 1.4% agarose gels using ^{32}P-labeled *Hind*III-digested lambda DNA and *Hinf*I-digested pBR322 DNA as markers. The gels were dried on DE81 paper (Whatman) and exposed for 8 to 12 hours with an intensifying screen (Cronex). In this manner, the order of restrictions sites from the labeled end could be precisely mapped.

In the second approach, cold UR2 and UR2AV DNAs were digested with the same set of enzymes used for partial mapping, electrophoresed through 0.8% agarose gels and blotted onto nitrocellulose (65). The approximate gene boundaries were determined by hybridizing the blots with 2 - 5 x 10^5 cpm of the gene-specific probes mentioned above. Hybridization and washing conditions have been described elsewhere (61).

Enzymes were purchased from NEB or BRL.

Hybridization of v-ros to 3'-specific v-onc probes.

UR2 DNA was digested with enzymes that cut within *ros*, generating 5'- and 3'-*ros* specific fragments, which were blotted and hybridized to 2 x 10^6 cpm of probe under low (35% formamide, 5X SSC) or moderate (50% formamide, 3X SSC) stringency at 37 ° C for 2 days (1X SSC equals 0.15 M NaCl and 0.015 M Na citrate). Washing conditions for moderate stringency hybridizations were similar to those described before (61); for low stringency conditions, blots were washed 3 times for 20 min each in 300 ml of 26 mM Tris-HCl (pH 7.4), 2X SSC, 1 mM EDTA and 0.1% SDS at 55 ° C followed by a similar wash at 60 ° C. For sequential hybridization, the previous probe was eluted at 68 ° C in 40 ml of 50% formamide, 1X SSC, 50 mM Tris-HCl (pH 7.4), 1 mM EDTA and 0.1% SDS. The blot was exposed to an X-ray film overnight to determine the extent of elution.

Heteroduplex mapping.

In collaboration with Ming-Ta Hsu at this University, heteroduplex mapping of the c-*ros* clone with pUR2 was performed and visualized by electron microscopy.

DNA sequencing.

pUR2 was used for nucleotide sequencing of the UR2 genome. The DNA sequence was determined by the methods of Maxam and Gilbert (39) and Sanger (55). For dideoxy sequencing, DNA fragments from pUR2 (Fig. 14) were subcloned into the *Eco*RI site of M13mp8 (41) with *Eco*RI linkers (Fig. 16). M13mp8 recombinant clones of both polarities were sequenced at least twice by the dideoxy method.

The DNA sequence of the intron/exon boundaries and exons in λc-*ros* as well as the 1.4 kb of 3' noncoding sequence in the recombinant clone was determined by the method of Sanger (55). DNA fragments (Fig. 23) from the c-*ros* clone were gap-filled and blunt-end ligated into the *Sma*I site of M13mp8 (41) (Fig. 23).

Hydrophilicity analysis.

The hydrophilicity profile of the 400-amino acid *ros* region of $P68^{gag\text{-}ros}$ was determined by using the computer program of Hopp and Woods (26) with the hydrophilicity values for each amino acid determined by Levitt (35).

Section III. Results

(v-ros)

Restriction enzyme analysis of circular UR2 and UR2AV DNAs.

In general, although the total yield of DNA from UR2(UR2AV)-infected CEF was greater at all three timepoints than the DNA yield from infected QT6 cells, Southern hybridization analysis revealed a higher proportion of UR2 and UR2AV proviral DNAs from infected QT6 cells (Fig. 6). The DNA does not migrate as discrete bands. This may be due to the presence of degraded viral DNAs, which would be detected by the hybridization probe. Contaminating cellular DNA and protein might not have been completely excluded by the Hirt procedure and might also have contributed to the anamolous pattern of migration. Maximal yield of material hybridizing to the viral-specific probe was obtained 24 hours postinfection from UR2(UR2AV)-infected QT6 cells. Linear UR2AV proviral DNA migrates at approximately 8 kb, and the circular form of the UR2AV provirus migrates more rapidly at 4.3 kb. The ratio of linear to circular DNA is 1:1. However, UR2 linear and circular molecules are not clearly seen, indicating a higher ratio of UR2AV to UR2 molecules.

DNA from all timepoints was pooled and further enriched for circular DNA by acid-phenol extraction. Approximately 25% of DNA recovered from the aqueous phase was analysed by Southern hybridization with a probe representative of the UR2 genome, $cDNA_{UR2}$, which includes UR2AV-derived sequences (Fig. 7A), or to a *ros*-specific probe (Fig. 7B). Probe containing sequences common to UR2 and UR2AV detected the UR2AV circular DNA doublets containing one or two LTRs migrating at 4.3kb as well as the corresponding two species of UR2 circular DNA migrating at 2 kb. There is no evidence of the linear forms, which would migrate at 8 kb for UR2AV and 3.5 kb for UR2, indicating a

Fig. 6. Isolation of proviral UR2 and UR2AV DNA. UR2(UR2AV)-infected chicken embryo fibroblasts (CEF) or QT6 cells were harvested 24, 48 or 72 hours post infection. Proviral DNA was isolated from the Hirt supernatant, and 20ug from each timepoint was electrophoresed through a 0.8% agarose gel, blotted onto nitrocellulose paper and probed with [^{32}P]cDNA made from 24S UR2 RNA.

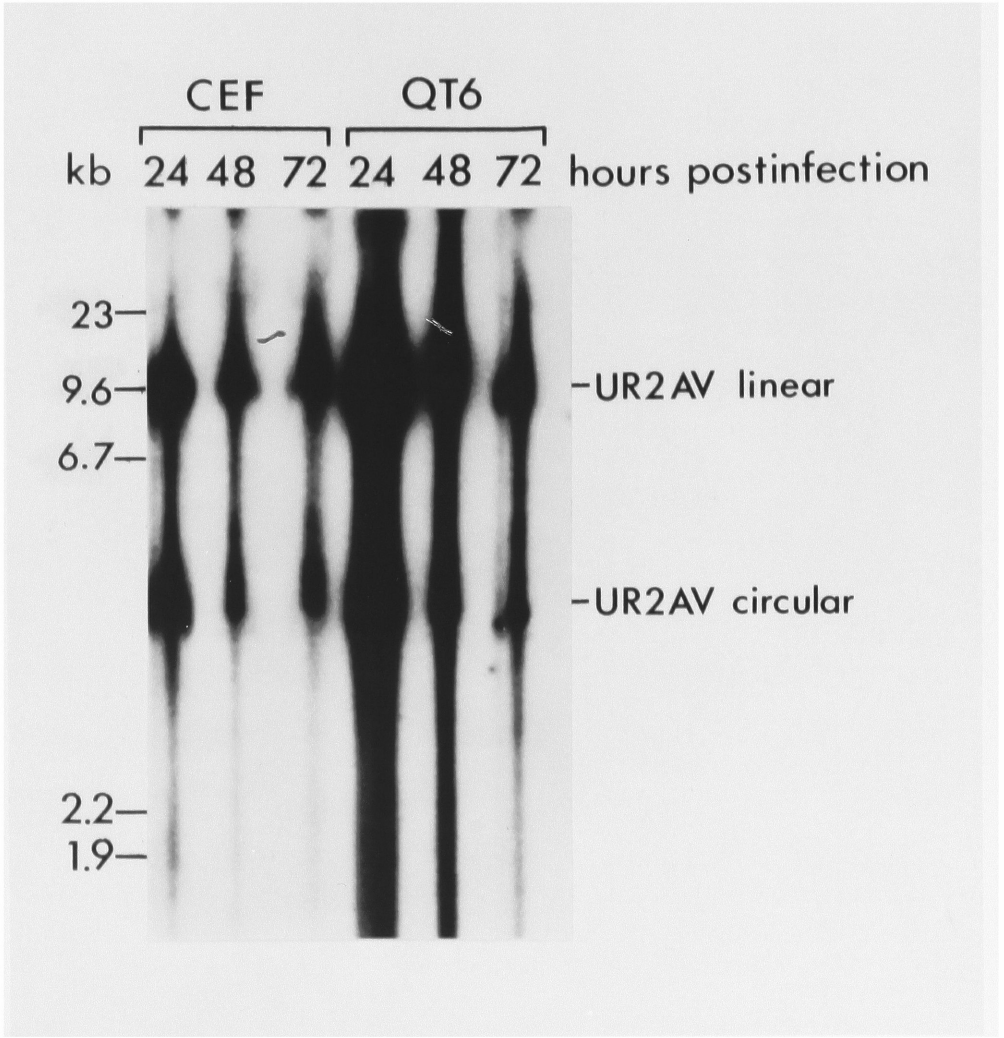
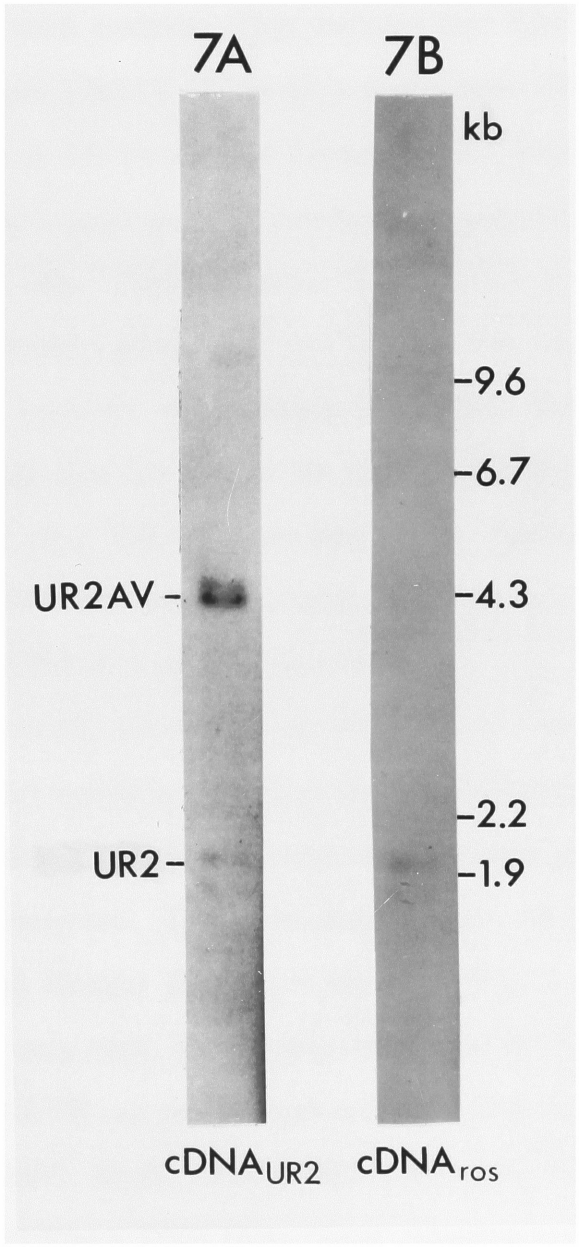

CEF
QT6
kb
24 48 72 24 48 72
hours postinfection
23—
9.6—
6.7—
2.2—
1.9—
–UR2AV linear
–UR2AV circular

Fig. 7. Enrichment for circular proviral DNA. Proviral DNA from the timepoints in Fig. 6 were pooled and extracted with acid phenol. 25% of the recovered aqueous phase was loaded onto a 0.8% agarose gel, blotted onto nitrocellulose, and probed with $cDNA_{UR2}$ (A) or $cDNA_{ros}$ (B).


7A
7B
kb
9.6
6.7
4.3
2.2
1.9
UR2AV
UR2
cDNA UR2
cDNA ros

high degree of enrichment for the circular form after acid-phenol extraction.

Unintegrated viral DNA isolated from UR2(UR2AV)-infected QT6 cells was subjected to digestion by various restriction enzymes to determine which endonucleases cleaved at a single site and thus would be suitable for cloning. Fig. 8 shows the results of such analysis. They indicate that *Kpn*I, *Sst*I, and *Xho*I each cleaves both UR2 and UR2AV DNAs at a single site. The linearized UR2AV DNA doublet bands at 8.0 kb and the linearized UR2 doublet at 3.5 kb. These sizes compare well with previous estimates by electrophoretic analysis of UR2 and UR2AV viral RNAs (82). *Bam*HI cleaves UR2AV DNA several times, and UR2 DNA once. The *Hind*III pattern indicates at least one recognition site in both UR2 and UR2AV; however, the presence of partially-digested DNA obscures further analysis. *Sal*I does not cut either UR2 or UR2AV DNA. The doublet migrating at 9.5 kb after *Sal*I digestion most likely represents randomly nicked circles, and could also be seen after incubation of the circular proviral DNA in restriction enzyme buffer with no enzyme added.

To remove the small, contaminating linear DNAs (most likely representing degraded viral DNAs) visible in each lane, the large preparation of proviral DNA for cloning was further purified by chromatography through an A5m gel column after acid-phenol extraction. The restriction patterns of *Eco*RI and *Sst*I digestions of the purified circular DNAs are shown in Fig. 9. *Sst*I cuts UR2 and UR2AV DNAs each only once, and the linearized viral DNA containing either one or two copies of the LTR run at positions corresponding to about 3.5 kb for UR2 and 8.0 kb for UR2AV. *Eco*RI cleaved UR2 DNA once, liberating the linearized UR2 DNA, but cut UR2AV DNA three times, releasing several subgenomic fragments. The majority of the acid phenol- and column-purified viral DNAs appeared to be of the circular forms. However, the majority of the DNA preparation was of cellular origin, because the viral DNA could be detected only by

Fig. 8. Restriction enzyme analysis of circular UR2 and UR2AV DNAs. 20 ug of partially-purified proviral circular DNA was digested with the indicated enzyme, electrophoresed through a 0.8% agarose gel, blotted onto nitrocellulose and probed with $cDNA_{UR2}$.

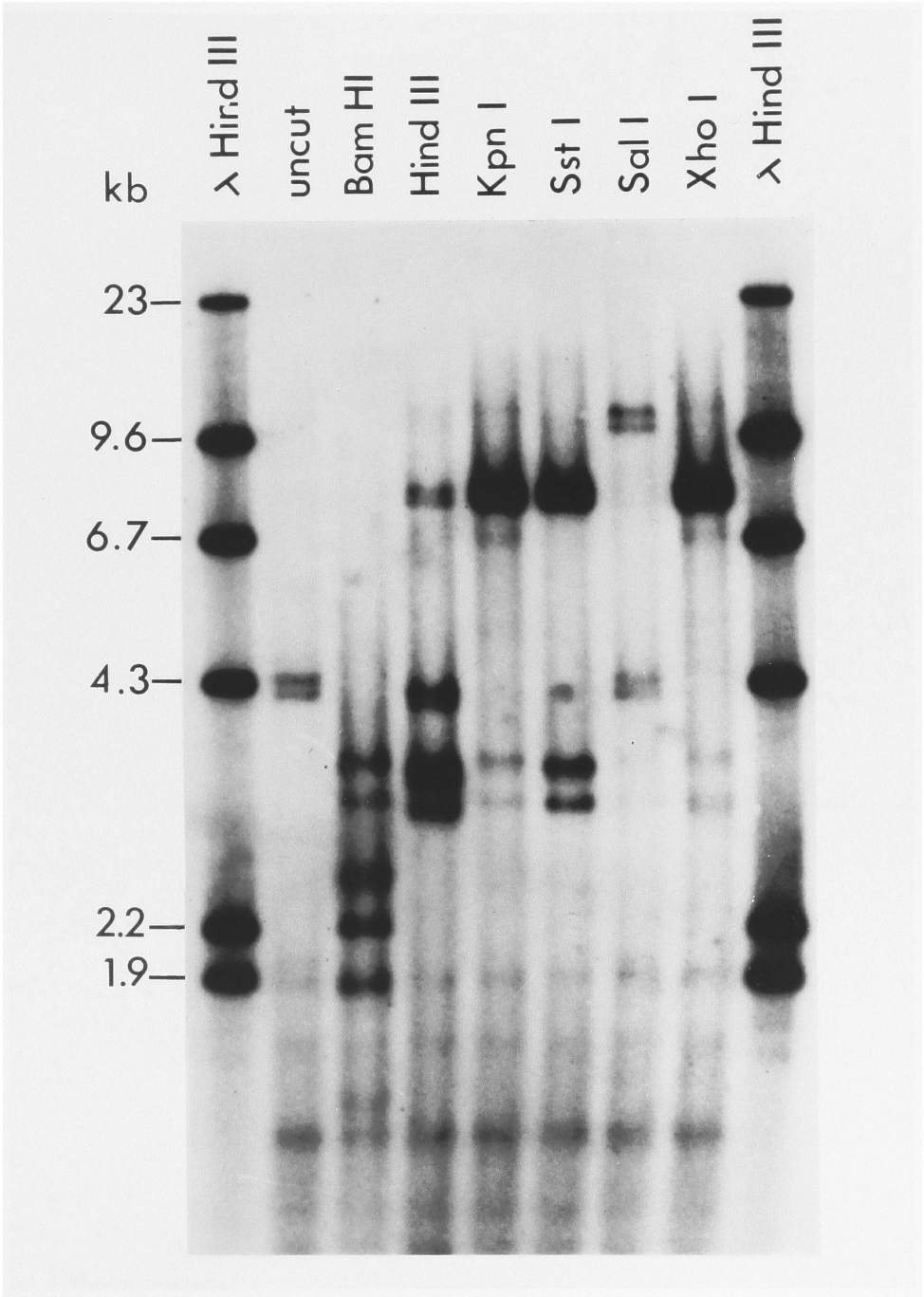
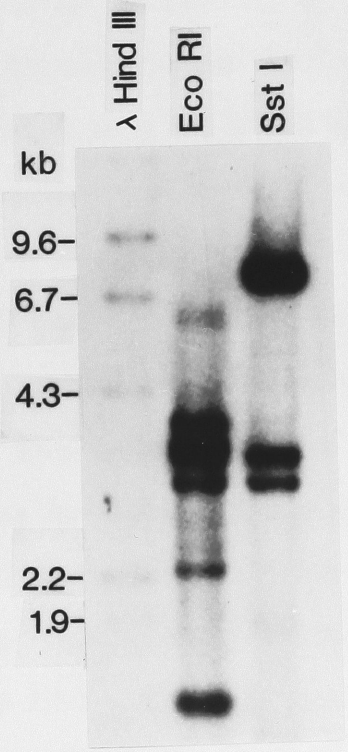

kb
λ Hind III
uncut
Bam HI
Hind III
Kpn I
Sst I
Sal I
Xho I
λ Hind III
23
9.6
6.7
4.3
2.2
1.9

Fig. 9. Restriction enzyme analysis of isolated circular UR2 and UR2AV DNAs. 5 ug of material containing unintegrated proviral DNA were restricted with either *Sst*I or *Eco*RI, and 2 ug of each digest were loaded per well, electrophoresed through a 0.8% agarose gel, blotted onto nitrocellulose papers, and probed with [^{32}P]cDNA made from 24S UR2 RNA.

Restriction Enzyme Analysis of
Isolated Circular Proviral UR2 and UR2AV DNAs



hybridization to the cDNA probe, and not by ethidium bromide staining of the 2 ug of DNA loaded per well in the agarose gel.

Molecular cloning of UR2 and UR2AV DNAs.

Remaining DNA from the *Eco*RI and *Sst*I digests described above were used for cloning, using purified λgtWESλB *Eco*RI and *Sst*I arms. cDNA probe made from 24S UR2 RNA was used for the screening. Eight full length helper viral DNA clones were obtained using the *Sst*I site for the cloning. Eleven λUR2AV recombinant phages, each containing more than one UR2AV *Eco*RI fragment, and a λUR2 clone were isolated using the *Eco*RI site for cloning. About 0.15% (12 clones out of 8000 plaques screened) of the recombinant phages contained viral sequences using *Eco*RI arms of the lambda DNA for the cloning.

The number of UR2AV clones isolated was roughly tenfold the number of UR2 clones. This can be expected from the ratio of UR2AV to UR2 DNA as shown in Figs. 7, 8 and 9. This probably reflected the ratio of the helper to the UR2 virus in the stock used for the infection of the QT6 cells. This has been previously observed in several UR2(UR2AV) virus stocks (4). Additionally, the packaging efficiency of the lambda vector for fragments the size of the UR2 genome is quite low. Each UR2 and UR2AV lambda recombinant phage clone was purified by 3 to 4 cycles of single plaque isolation.

To confirm the identity of the UR2 clone, the viral DNA insert was hybridized with probes specific to various regions of the RSV genome according to the procedure of Southern (65). As expected, the UR2 DNA hybridized only to the 5', 5' *gag*, and *c* probes in addition to the cDNA made from UR2 genomic RNA, but not to *pol* and *pol-env* probes. It has been previously determined (82) that the *pol* gene is deleted and the *gag* and *env* genes are truncated in the UR2 genome. Because the size of the insert from this recombinant clone, 3.4 kb, is

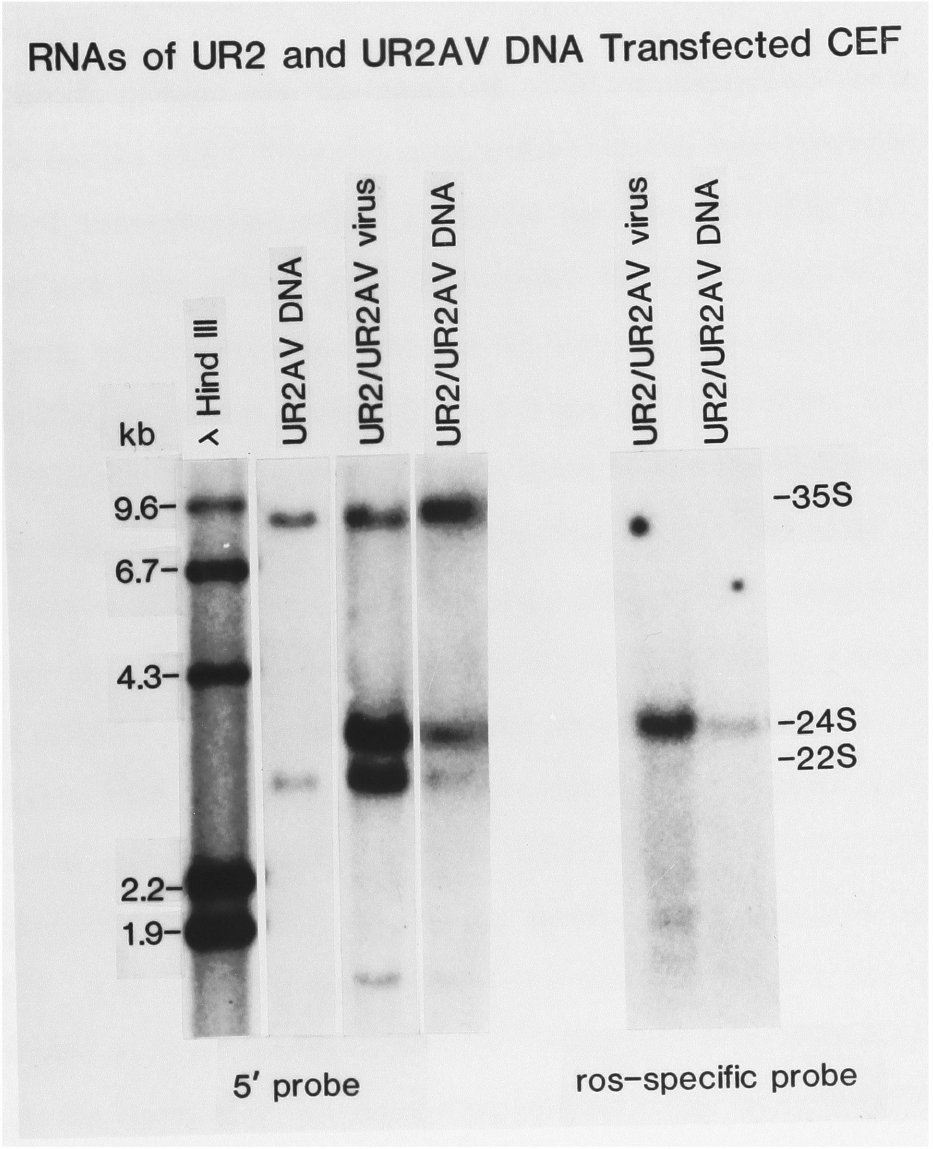
equivalent to that determined for the UR2 genomic RNA (82), it is most likely that this clone contains a full-length copy of the UR2 genome. However, this clone contained, in addition to the 3.4 kb UR2 insert, a tandemly linked 3.7 kb DNA fragment of non-viral sequence. Therefore, the 3.4 kb UR2 DNA was purified and subcloned into the *Eco*RI site of pBR322. All further studies of the UR2 genome were done with this recombinant plasmid clone, called pUR2.

Transfection assays of UR2 and UR2AV DNAs.

Insert DNAs isolated from pUR2 and from one of the UR2AV lambda clones, 4B2, were introduced together onto CEF. The transfected cells displayed typical UR2 transformed elongated morphology and produced UR2(UR2AV) pseudotype virus about two weeks after transfection. This demonstrated that both UR2 and UR2AV clones tested were biologically active. In addition, 5 independently isolated UR2AV-lambda clones (A21, D1, 1C-1, 4-1, 14-1) were cut with *Sst*I to free the insert from the lambda DNAs, and were introduced similarly onto CEF. Assays for reverse transcriptase activity in the culture fluid of transfected cells 10 days post transfection were positive for all but clone 1C-1. The map of the UR2AV DNA insert of clone 1C-1 was identical to that of 14-1, which was biologically active, except that the 1C-1 insert was missing a portion of the *Sst*I - *Hin*dIII right-hand region (see Fig. 13 for the restriction sites). This region contains the LTR and deletion of this segment might account for the loss of biological activity.

To confirm that the transformation was induced by UR2, total poly(A)-containing cellular RNA was isolated from the transfected cells, and analysed by RNA blotting and hybridization. Fig. 10 shows that the transfected cells display a subgenomic and genomic RNA pattern identical to that of UR2(UR2AV)-infected CEF. Hybridization of the RNAs to the 5' probe detected 35S genomic

Fig. 10. RNAs of UR2 and UR2AV DNA transfected CEF. 10 ug of poly(A)-containing RNAs isolated by SDS-proteinase K extraction of UR2 and UR2AV transfected cells were denatured with 1 M glyoxal and fractionated on 1% agarose gels. The RNAs were transferred to nitrocellulose paper and hybridized with a 5' leader or *ros*-specific probe. RNAs isolated from UR2(UR2AV)-infected cells were used as controls. ^{32}P-labeled *Hind*III cut lambda DNA was denatured and run is parallel for molecular weight markers.


RNAs of UR2 and UR2AV DNA Transfected CEF
kb
λ Hind III
UR2AV DNA
UR2/UR2AV virus
UR2/UR2AV DNA
9.6–
6.7–
4.3–
2.2–
1.9–
5′ probe
UR2/UR2AV virus
UR2/UR2AV DNA
–35S
–24S
–22S
ros–specific probe

RNA and 22S subgenomic envelope mRNA of UR2AV, as well as the 24S genomic UR2 RNA. Cells transfected with UR2AV alone produced only 35S and 22S RNAs. A *ros*-specific probe hybridized only with the 24S UR2 RNA. The RNA patterns of all UR2AV-transfected cells were identical; however, as expected from the reverse transcriptase assay, no viral RNAs were detected in clone 1C-1-transfected CEF.

Viral specific proteins were also analyzed. CEF transformed by the molecular clones were labeled with [^{35}S]-methionine, and cell lysates were precipitated with RSV-infected tumor-bearing rabbit (TBR) or anti-*gag* sera (Fig. 11). DNA-transfected and virus-infected cells gave essentially similar patterns with the same antisera, which precipitated the *gag-ros* fusion product, P68. Recognition of P68 by TBR serum was apparently via the *gag* peptide in P68. The appearance of P68 as a doublet in virus-infected cells has been observed previously (17), and apparently reflects the existance of a variant in the UR2 stock. The V8 digestion pattern of the two species are identical (Ellen Garber, unpublished), and only a single species is seen after a kinase assay. In addition, only a single species migrating at 68 kd is detected after metabolic labeling or kinase assay of proteins extracted from CEF transformed by the molecular clone. For assaying the kinase activity, cold cell lysates prepared from UR2/UR2AV-transfected CEF were immunoprecipitated with TBR or anti-*gag* serum and incubated *in vitro* with [γ^{32}P]-ATP. Fig. 12 shows the results of this assay, which indicates that P68 was associated with protein kinase activity. The much lower level of P68 kinase activity in the transfected cell lanes compared with that in virus-infected cells was due to the fact that only about 20% of the transfected cells were transformed at the time of the experiment, as opposed to complete transformation in the virus-infected culture. The radioactivity in the control TBR lane is due to phosphorylation of IgG by a contaminating kinase in the antiserum (Ellen

Fig. 11. ^{35}S labeling of proteins from UR2 and UR2AV DNA transfected cells. Cell lysates prepared from UR2 and UR2AV transfected CEF that had been labeled for 5 h with 500 uCi per 6 cm plate of [^{35}]S-methionine were immunoprecipitated with TBR or anti-*gag* serum. The immune complexes were subjected to electrophoresis on an SDS 5-15% polyacrylamide gradient gel. ^{35}S-labeled proteins from cell lysates of uninfected and UR2(UR2AV)-infected CEF were analysed in parallel.

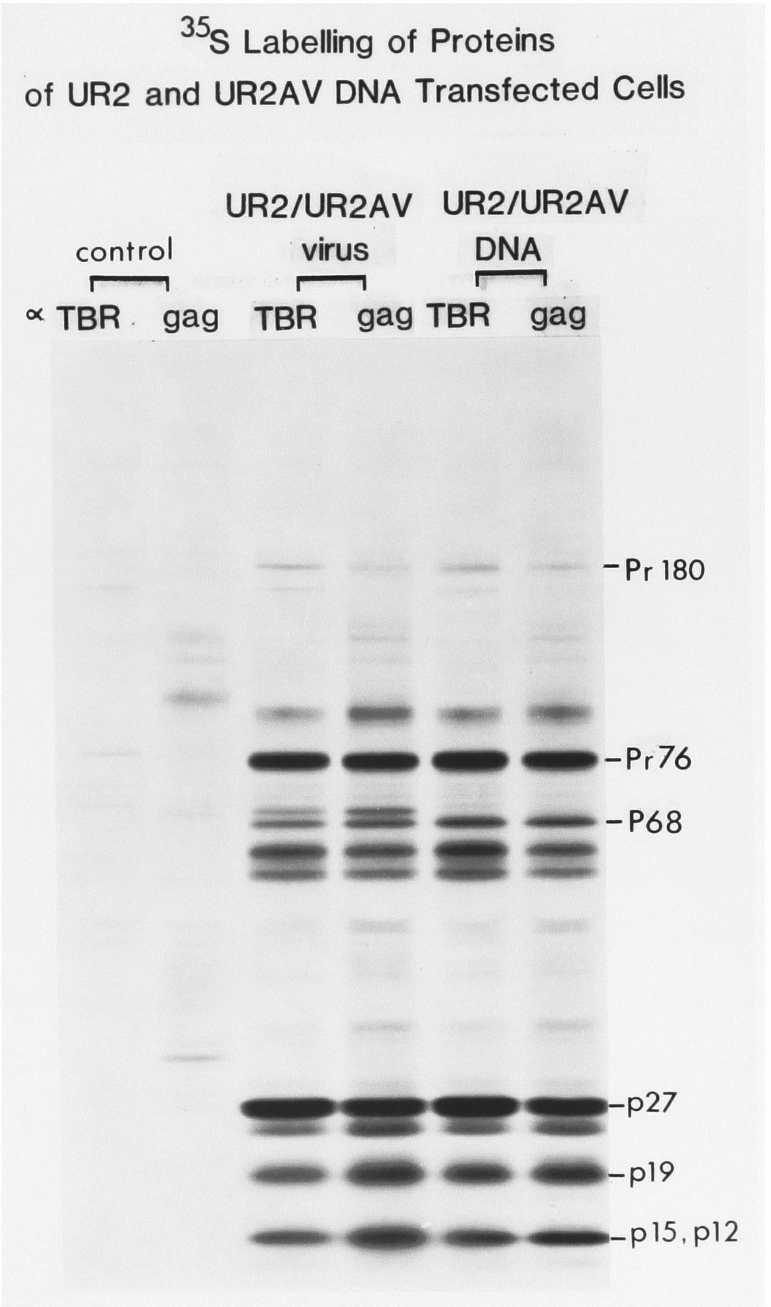
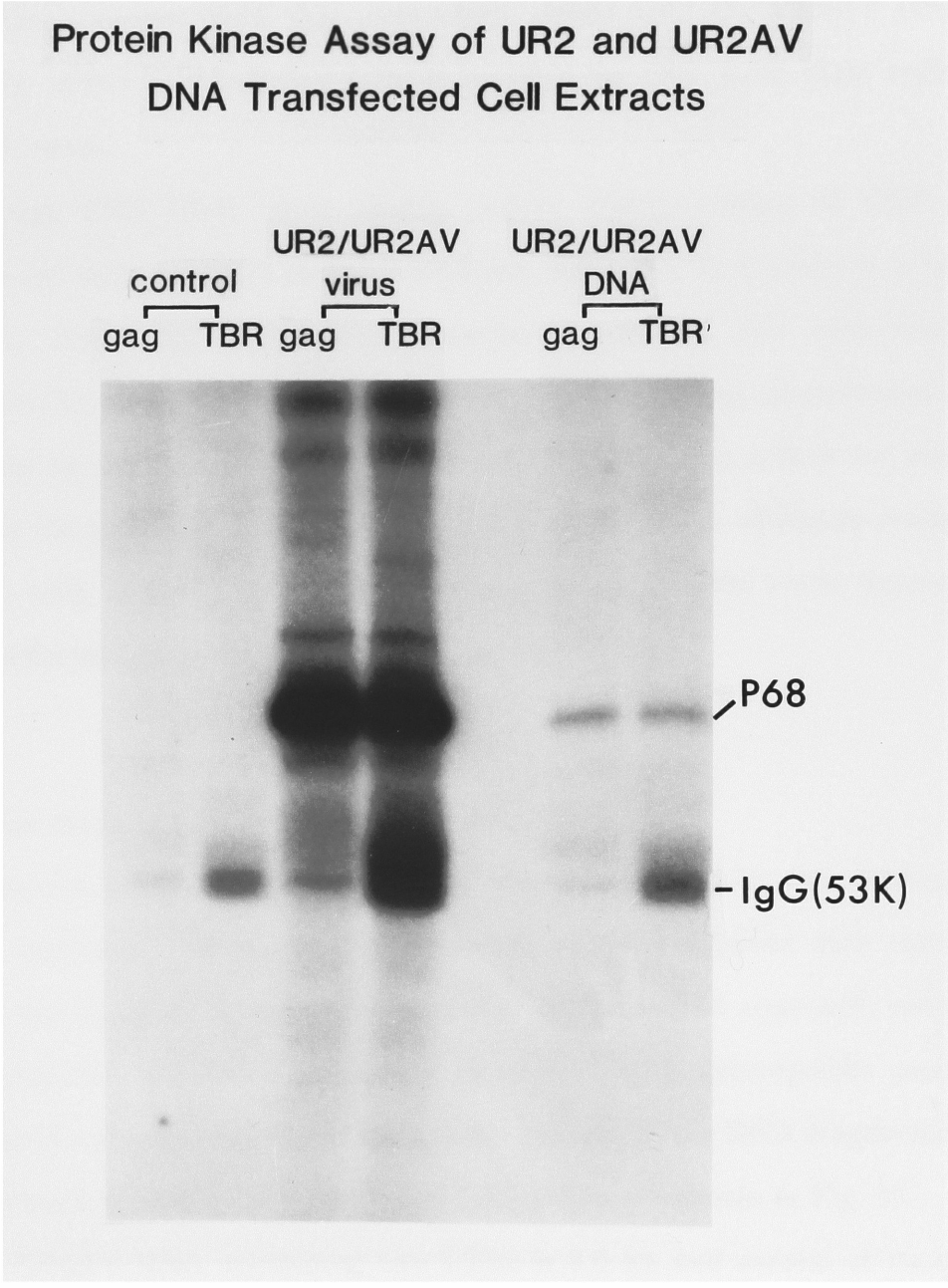

35S Labelling of Proteins
of UR2 and UR2AV DNA Transfected Cells
UR2/UR2AV
virus
UR2/UR2AV
DNA
control
∝ TBR
gag
TBR
gag
TBR
gag
Pr 180
Pr 76
P68
p27
p19
p15, p12

Fig. 12. Protein kinase assay of UR2 and UR2AV DNA transfected cell extracts. Cell lysates from CEF transformed by the molecular clones were precipitated with TBR or anti-*gag* serum and incubated *in vitro* with [γ^{32}P]ATP. The phosphorylated proteins were analysed on an SDS 8.5% polyacrylamide gel. Cell lysates prepared from uninfected and UR2(UR2AV)-infected CEF were treated similarly and used as controls.


Protein Kinase Assay of UR2 and UR2AV
DNA Transfected Cell Extracts
control
gag
TBR
UR2/UR2AV
virus
gag
TBR
UR2/UR2AV
DNA
gag
TBR
P68
IgG(53K)

Garber, unpublished). The non-P68 bands seen in the virus-infected culture in Fig. 12 were most likely due to incomplete washing of the immunocomplex. It was observed before that IgG could be phosphorylated by P68, in particular when TBR serum was used in the immunoprecipitation (17).

The above studies show that molecularly cloned UR2 and UR2AV DNA are biologically active and indistinguishable in effect on CEF from their respective parental viruses.

Although UR2 DNA could readily induce transformation of CEF when cotransfected with UR2AV, so far I have not been able to transform a rat cell line, 3Y1, with UR2 DNA. Under similar conditions, 3Y1 cells could be transformed by either SR-A or FSV DNA (38, 62, L.-H. Wang, unpublished data). The reason for the failure of UR2 DNA to transform 3Y1 rat cells is not clear. It is possible the promoter sequence in UR2 DNA might not be efficiently recognized in the rat cells, or that a second oncogene may be required for *ros* to induce complete transformation of rat cells (31).

Restriction maps of UR2 and UR2AV DNAs.

Restriction maps of UR2 and UR2AV were constructed by two methods. In the first approach, ^{32}P-end-labeled DNA was partially digested with restriction enzymes and analysed by gel electrophoresis. In the second approach, restriction enzyme-digested, unlabeled DNA was hybridized with gene-specific probes to determine the physical order and the genetic content of the DNA fragments. The results of such mapping for UR2AV and UR2 DNAs are shown in Fig. 13.

The UR2AV DNA containing two LTRs is 8.0 kb and several of its restrictions sites are similar to those of Rous-associated virus-2 (RAV-2) (51). The *Kpn*I, *Sst*I, and *Bam*HI sites in RAV-2 are conserved in UR2AV. A *Hind*III site in RAV-2 is also conserved in UR2AV, although UR2AV contains two extra *Hind*III

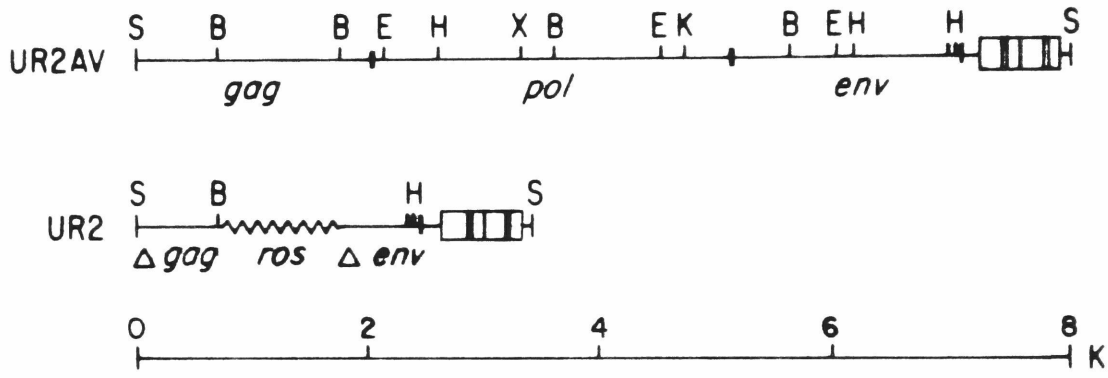


Fig. 13. Restriction map of cloned UR2AV DNA. UR2AV proviral DNA was purified from clone 14-1, and restriction enzyme sites were determined. S, *Sst*I; B, *Bam*HI; E, *Eco*RI; X, *Xba*I; K, *Kpn*I. The genetic structure was determined by hybridization of UR2AV restriction fragments to probes specific for various regions of the RSV genome. The UR2 DNA was included here for comparison to show the helper virus-related and *ros* sequences.

sites in the *env* region. Similarly, UR2AV shares with RAV-2 two *Eco*RI sites, yet has an additional *Eco*RI site in the middle of the UR2AV genome. It does not contain an *Eco*RI site in the LTR, as does RAV-2.

By comparing the restriction maps of UR2AV and UR2 DNAs, the *gag-ros* and *ros-env* borders may be determined. Construction of a detailed map of UR2 (Fig. 14) allowed a more precise definition of the gene borders than those determined before (82). UR2 DNA has a single *Eco*RI site. There is no corresponding *Eco*RI site in UR2AV DNA, suggesting that this site is located within the *ros*-specific sequence. The *gag-ros* boundary appears to be located very close to the *Hae*II site immediately upstream from the right-hand *Eco*RI site, because the 460 bp *Hae*II B fragment hybridized strongly to a UR2 representative probe and weakly to a UR2AV representative probe and the 5' *gag* probe (data not shown). The *Pvu*II B fragment appeared to contain only *ros* sequence since it hybridized only to the UR2, but not the UR2AV, cDNA probe. The *ros-env* boundary was mapped between the *Pvu*II site and the second *Ava*I site from the left, because the *Ava*I B fragment hybridized to both the UR2 and UR2AV cDNA probes (data not shown). Since the *Ava*I B fragment could not hybridize with *gag*, *pol*, and *c* probes, I inferred that the sequence present in this DNA that was hybridizable with UR2AV cDNA must be the *env* sequence. The mapping is consistent with the previous finding that two highly conserved *env*-specific oligonucleotides, spots 11 and 12a, located at 96 and 594 nucleotides, respectively, upstream from the termination codon of gp37 of the SR-A genome, were present in the UR2 genomic RNA (82). Given the estimates of these borders, *ros* is about 1.2 kb in length.

From the results of the restriction enzyme analysis, I concluded that the UR2 genome is 3.4 kb in length, containing in the middle about 1.2 kb of transforming sequence, *ros*. UR2 shares with UR2AV 0.8 kb of 5' leader and Δgag and 1.4 kb of 3' Δenv and *c* sequences.

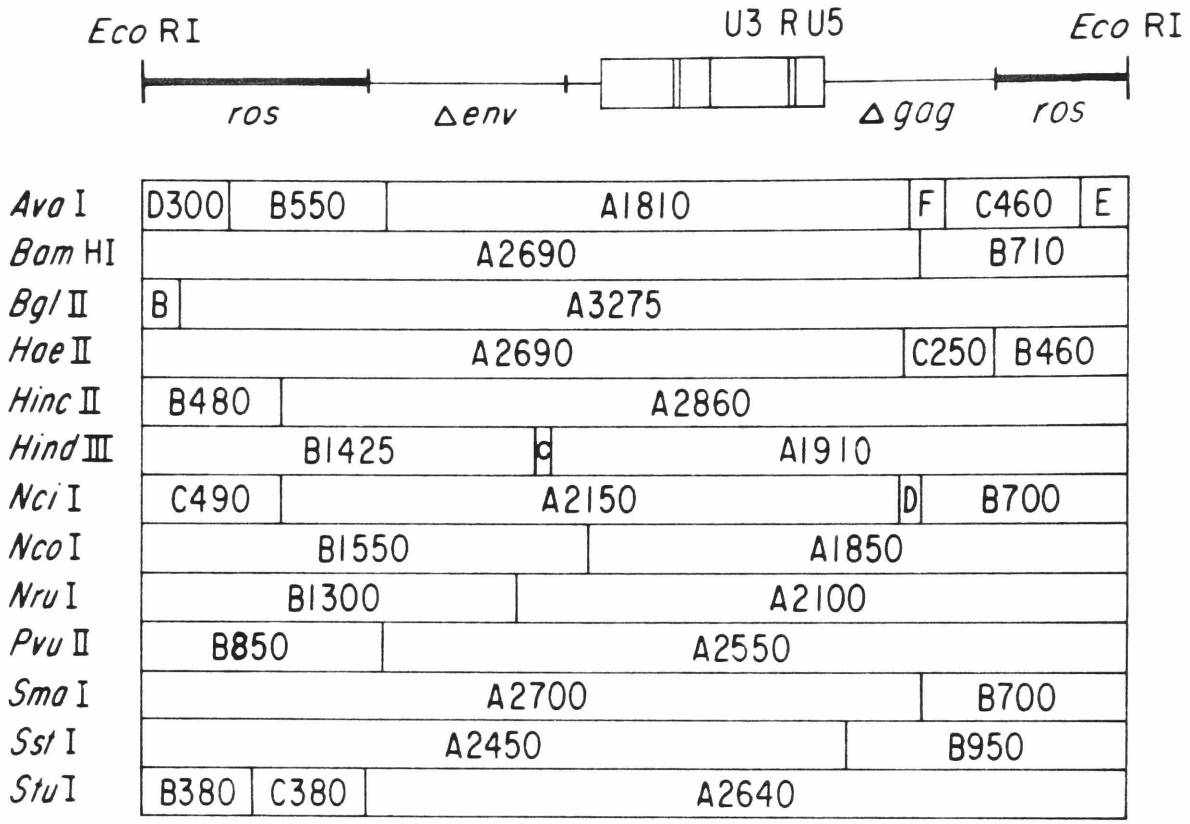
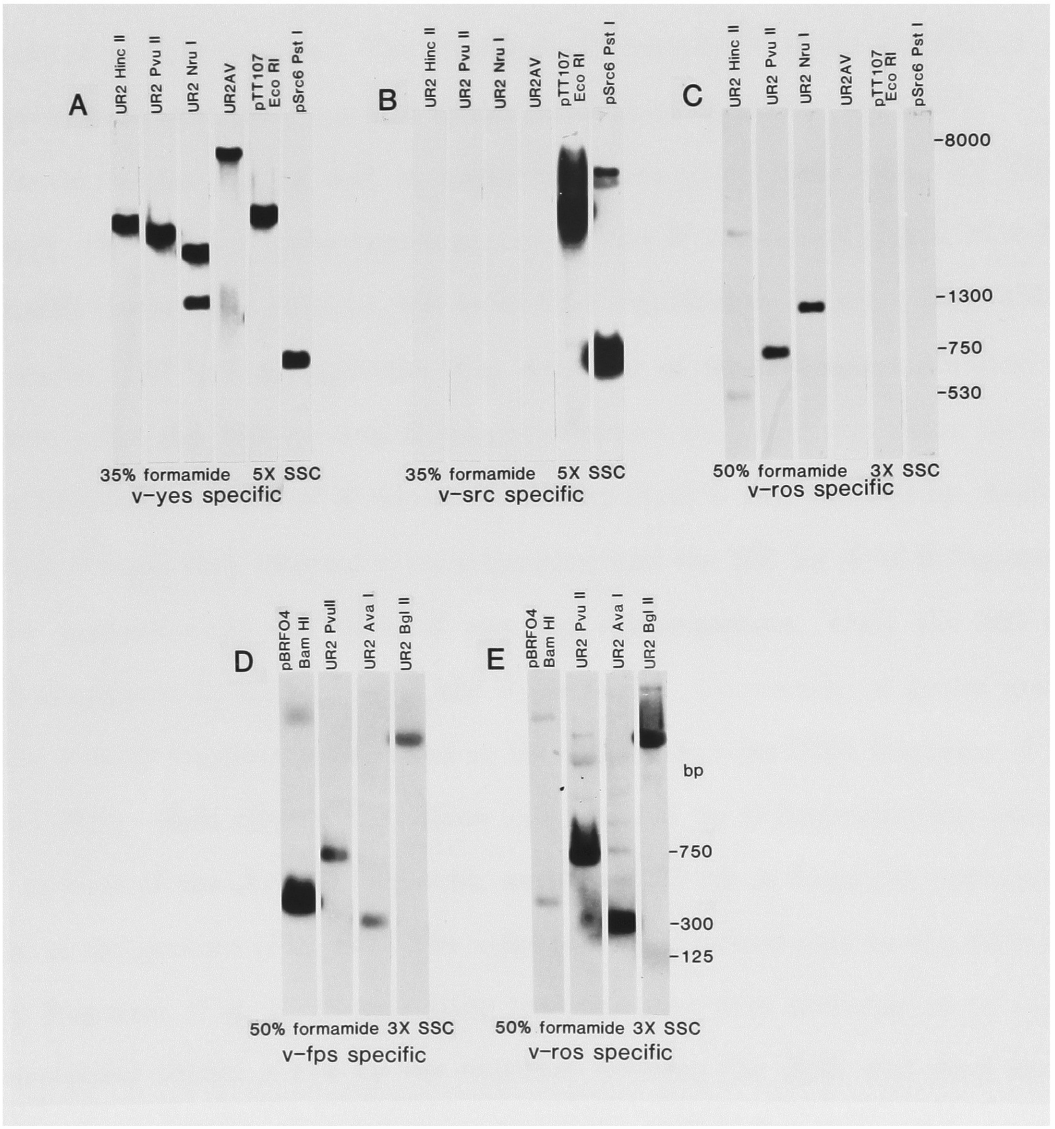


Fig. 14. Restriction map of cloned UR2 DNA. The 3.4 kb UR2 DNA insert was isolated from pUR2 and restriction sites were determined as described in Materials and Methods and the Results sections. The numbers in the boxes indicate the length in bp of the individual restriction fragments. The sizes of restriction fragments *Ava*I E, *Ava*I F, *Bgl*II B, *Nci*D and *Hind*III C are 160 bp, 120 bp, 125 bp, 125 bp, 60 bp, and 67 bp, respectively.

Homology between v-ros and other transforming genes.

Previous comparison of the *ros* sequence with those of *src*, *fps*, and *yes* by liquid hybridization between viral RNAs and cDNAs specific to individual v-*onc* sequences detected no significant homology between *ros* and the rest of the ASV transforming genes (61). However, it is possible that the homology may be restricted to a small region of *ros* and thus can not be easily detected by hybridization between total viral RNA and the cDNA probe representing the entire domain of a v-*onc* gene. To test this possibility, we undertook the approach of subdividing the *ros* sequence by digestion with enzymes known to cut within *ros* and hybridizing the individual fragments with probes prepared from other v-*onc* DNAs. Probes derived from the 3' regions of *src*, *fps* and *yes* were used since these regions were shown to contain the sequences conserved among those transforming genes (29, 60, 62). The results are shown in Fig. 15. No significant hybridization was detected between v-*ros* and v-*yes* even under conditions of low stringency (35% formamide, 5X SSC), although the v-*yes* probe cross-hybridized with the 3.1 kb *Eco*RI fragment of SR-A (from pTT107) containing the entire v-*src* sequence as well as with the 0.75 kb *Pst*I 3'-*src* specific fragment (from p*src*6). The high degree of amino acid sequence homology within the C-terminal half of $pp60^{src}$ and $pp90^{yes}$ has been shown previously (9, 29). Hybridization of this v-*yes* probe to UR2 DNA fragments containing *gag* and *env* sequences and to UR2AV DNA is apparent in Fig. 15A. This was due to contamination of the *yes*-specific probe with helper virus-related sequences of Y73, since the 1.1 kb 3' v-*yes* fragment was slightly contaminated with the 1.4 kb *gag*-5' *yes* fragment and the 1.5 kb fragment containing the *env* and LTR region of Y73 (29). The 1300 bp *Nru*I B fragment (Fig. 14) hybridized to this 3' *yes* probe, apparently due to the *env* sequences present in this DNA. Neither of the *ros*-specific fragments (*Hinc*II B and *Pvu*II B) hybridized with the 3'-*yes* probe.

Fig. 15. Homology between v-*ros* and 3' v-*onc* regions. UR2 DNA insert was digested with restriction enzymes cleaving within *ros* to yield 5'- and 3' *ros*-specific fragments. pTT107 and p*src*6 *src*-containing plasmids (see text) and the pBRFO4 *fps*-containing plasmid DNA were cut with the appropriate enzyme to free the *onc* insert. The amount of each DNA added per well was adjusted to yield approximately 50 ng of each *onc* DNA fragment containing the putative region of homology. The DNAs were fractionated on 1% agarose gels and transferred to nitrocellulose filters. The probes used in each hybridization were: (A) 1.1 kb *Pst*I fragment from λ-Y73; (B) 0.9 kb *Pvu*II fragment from pTT107, (C) 0.75 kb *Pvu*II - *Eco*RI from pUR2, (D) 0.4 kb *Bam*HI insert from pBRFO4, and (E) 0.3 kb *Ava*I - *Eco*RI fragment of pUR2.


A
UR2 Hinc II
UR2 Pvu II
UR2 Nru I
UR2AV
pTT107 Eco RI
pSrc6 Pst I
35% formamide 5X SSC
v-yes specific
B
UR2 Hinc II
UR2 Pvu II
UR2 Nru I
UR2AV
pTT107 Eco RI
pSrc6 Pst I
35% formamide 5X SSC
v-src specific
C
UR2 Hinc II
UR2 Pvu II
UR2 Nru I
UR2AV
pTT107 Eco RI
pSrc6 Pst I
-8000
-1300
-750
-530
50% formamide 3X SSC
v-ros specific
D
pBRFO4 Bam HI
UR2 PvuII
UR2 Ava I
UR2 Bgl II
50% formamide 3X SSC
v-fps specific
E
pBRFO4 Bam HI
UR2 Pvu II
UR2 Ava I
UR2 Bgl II
bp
-750
-300
-125
50% formamide 3X SSC
v-ros specific

As shown in Fig. 15B, no hybridization between v-*src* and v-*ros* sequences could be detected, although intense hybridization of the v-*src* probe to pTT107 *Eco*RI and p*src*6 *Pst*I was seen. The upper band of the doublet in the p*src*6 lane represented the partially digested, linearized p*src*6; the lower band, which hybridized with much less intensity, was the pBR322 vector. The insert *src* DNA used as probe might not have been completely purified from the pBR322 vector after one cycle of gel purification. The *Eco*RI-*Pvu*II *ros*-specific probe hybridized to the expected *ros*-containing but not to the *src*-containing DNAs (Fig. 15C).

However, under low as well as under moderate (50% formamide, 3X SSC) stringency, the 400 bp probe representative of the 3' conserved region of v-*fps* (the *Bam*HI insert from pBRF04, 62) hybridized significantly to v-*ros* (Fig. 15D). *Ava*I cleaves UR2 into 6 fragments (Fig. 14), four of which contain portions of the entire v-*ros*: the 460 bp *Ava*I C fragment covers the *gag-ros* junction including approximately 300 bp of 5' v-*ros*; the 160 bp *Ava*I E and the 300 bp *Ava*I D fragments contain only internal v-*ros* sequences; and the 550 bp *Ava*I B fragment contains approximately 450 bp of 3' *ros* plus *env* sequences. Only the 300 bp *Ava*I D fragement hybridized with the v-*fps* probe. Conversely, a probe made from the *Ava*I D fragment hybridized to the 400 bp 3' v-*fps* DNA fragment (Fig. 15D and 15E). *Bgl*II cleaves UR2 DNA into the 125 bp B fragment that covers the 5' portion of the *Ava*I D fragment, and the 3275 bp A fragment containing the rest of the genome (Fig. 14). The v-*fps* probe hybridized only to the 3275 bp *Bgl*II A fragment (Fig. 15D), indicating that the sequences homologous to v-*fps* were contained within a 175 bp *ros* sequence between the *Bgl*II and *Ava*I sites. However, the possibility that failure to detect the *Bgl*II B fragment might be due to the inefficiency of DNA transfer cannot be excluded.

The location of *fps*-related sequence within v-*ros* is reminiscent of that of v-*abl* (52) among the transforming genes coding for tyrosine protein kinases. We

checked the homology between v-*ros* and v-*abl*. Using a probe derived from the 5' 1.2 kb of v-*abl* DNA (54), no significant hybridization could be detected (data not shown).

Nucleotide sequence of the UR2 genome.

The sequencing strategy of pUR2 is shown in Fig. 16. Since pUR2 was derived from a UR2 circular DNA molecule cut with *Eco*RI, the UR2 genome in this clone is permuted with respect to the *Eco*RI site in *ros* (Fig. 16). However, the complete UR2 genomic sequence (shown in lower case letters) is presented in Fig. 17 as colinear to the viral RNA genome which begins at the predicted cap site (nucleotide 1) and ends at the polyadenylation site (nucleotide 3165). The *gag-ros* fusion protein P68 begins with the initiator amino acid methionine of p19 at nucleotides 380-382. The *ros*-specific sequence is fused to *gag* at nucleotide position 831. The coding sequence of the transforming protein P68 ends with an ochre stop codon at nucleotides 2036-2038. The v-*ros* non-coding region continues down to nucleotide 2103, and the Δgp37 envelope sequence resumes afterwards. The UR2 LTR is 326 bp long.

Previous restriction enzyme analysis of UR2 DNA showed that UR2 shares with its helper virus, UR2AV, 0.8 kb of 5' and 1.4 kb of 3' sequences (46). Since I have not sequenced UR2AV, I compared the UR2 sequence with those of PR-C RSV (57) in the regions of *gag* and *env* in order to determine possible sites of recombination between UR2AV and cellular *ros* in the generation of UR2. As can be seen in Fig. 18, the UR2 sequence diverges from that of PR-C RSV downstream from nucleotide 830 in the 5' region of *ros* and converges after nucleotide 2103 in the 3' *ros* region. I concluded that the 5' and 3' recombination sites between UR2AV and c-*ros* must be at or very close to nucleotides 831 and 2103 in the p19 and gp37 regions, respectively. The total length of *ros* is 1273

Fig. 16. Restriction enzyme cleavage map and nucleotide sequencing strategy of the UR2 genome. The transforming sequence, *ros*, of the permuted UR2 molecular clone, pUR2, is defined by the heavy line. The arrows indicate the direction and approximate extent of sequence determined by either the Maxam-Gilbert (39) or Sanger (55) method. Abbreviations for the restriction enzymes are as follows: A, *Ava*I; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; Ha, *Hae*II; N, *Nci*I; P, *Pvu*II; S, *Sst*I.

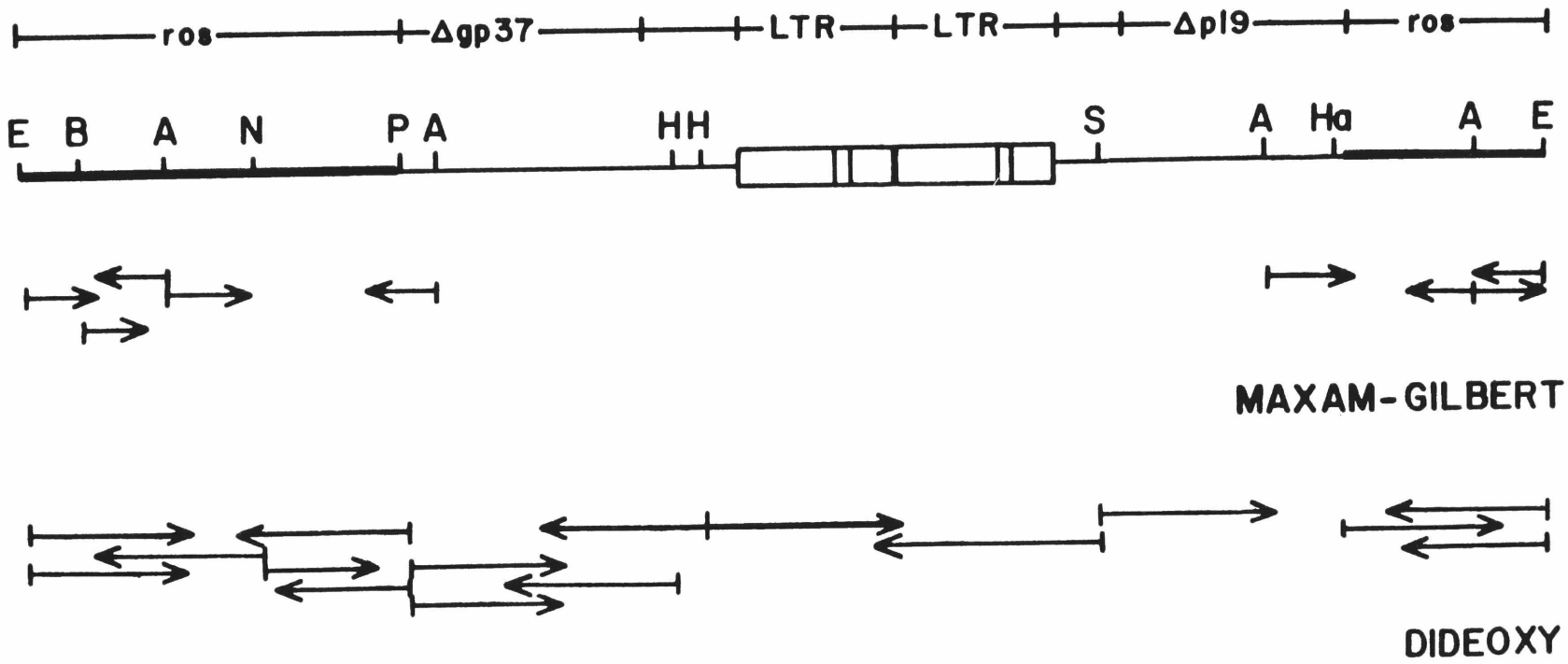
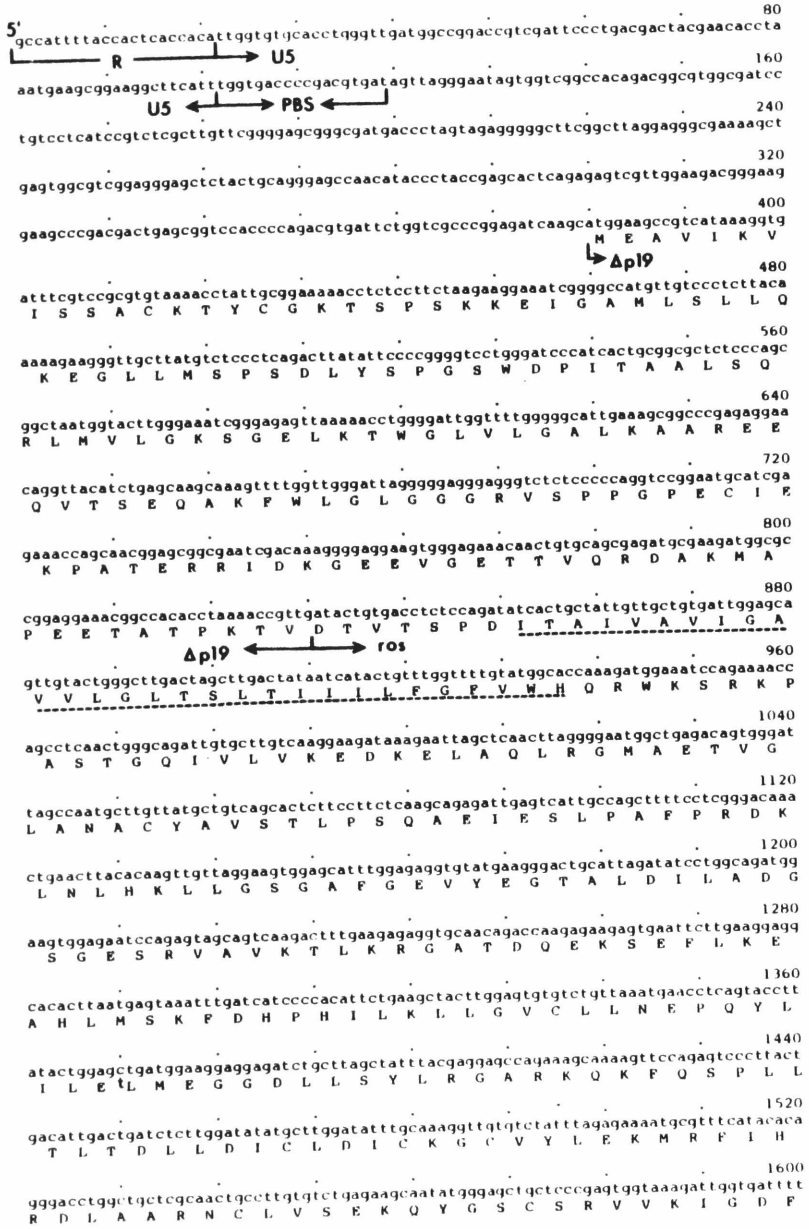
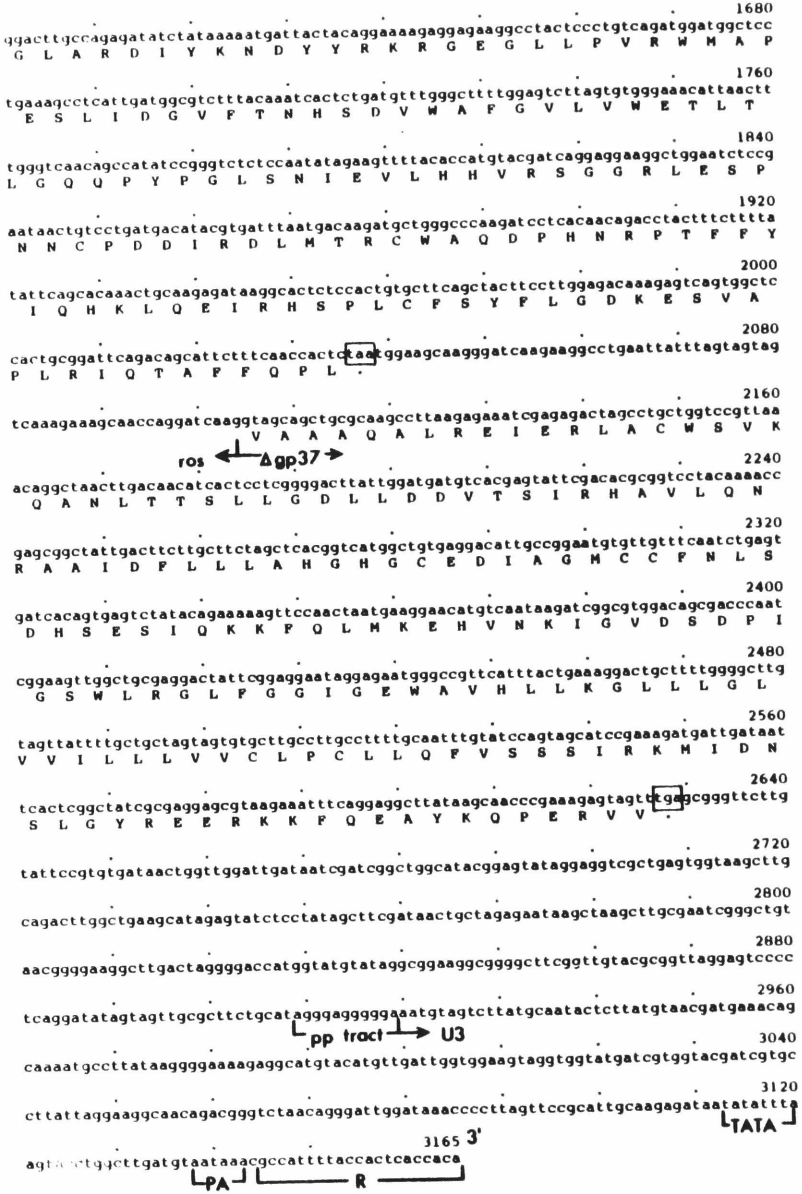
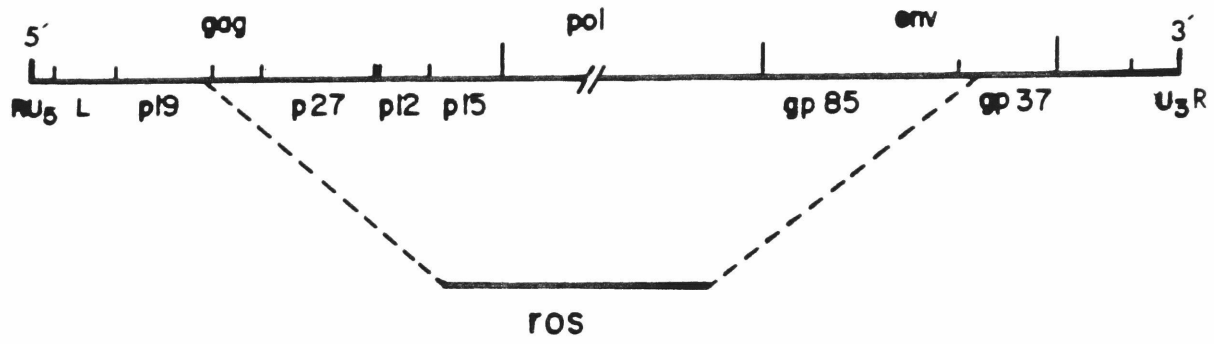

ros
Δgp37
LTR
LTR
Δp19
ros
E B A N P A HH S A Ha A E
MAXAM-GILBERT
DIDEOXY

Fig. 17. Complete nucleotide sequence of the UR2 genome and deduced amino acid sequence of P68 and Δgp37. The nucleotide sequence from the predicted cap site (nucleotide 1) to the polyadenylation site (nucleotide 3165) in UR2 DNA is shown by the lower case letters. Each amino acid is represented by a capital letter shown under the second nucleotide of the triplet codon. The R, U5 and U3 regions of the LTR, the primer binding site (PBS), the polypurine tract (pp tract), the poly(A) signal (PA), and the TATA box are indicated by heavy lines underneath these sequences. The dashed line under amino acid nos. 8 through 36 of the *ros* region indicates the hydrophobic region of the *ros* polypeptide presumed to be important in membrane association. Nucleotide number is given at the end of each row of sequence, and every ten nucleotides are marked by a period over the tenth nucleotide. The one letter symbols for the amino acids are used: A, ala; C, cys; D, asp; E, glu; F, phe; G, gly; H, his; I, ile; K, lys; L, leu; M, met; N, asn; P, pro; Q, gln; R, arg; S, ser; T, thr; V, val; W, trp; Y, tyr.


5'
R
U5
U5
PBS
Δp19
Δp19
ros


ros
Δgp37
pp tract
U3
TATA
3165 3'
PA
R



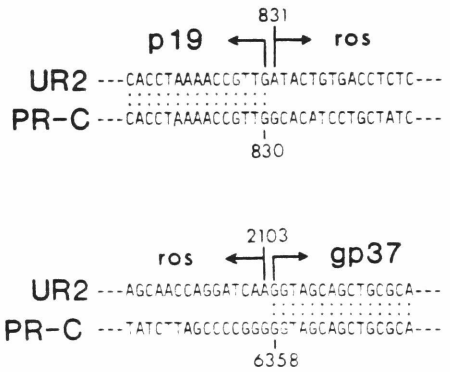


Fig. 18. Junctions of *ros* and UR2AV-derived sequences. The numbers refer to the nucleotide positions of UR2 sequence shown in Fig. 17 and to the published PR-C RSV sequence (57).

nucleotides.

Extensive homology was found between the UR2 sequences outside *ros* and the corresponding sequences of other avian viruses. Within the 5' 380 bp noncoding sequence of UR2, there are 22, 30 and 19 scattered single base changes relative to RAV-2 (7), Y73 (29), and PR-C (57), respectively. These changes include one deletion at nucleotide 371. This change (A in UR2 replaces TG in PR-C) occurs immediately upstream from the initiation site for translation. It has been suggested (7) that this upstream region is important for efficient mRNA translation; we do not know whether this change in any way affects the translation of UR2 genomic RNA. Sequences within the binding site for the tRNA primer (PBS) are completely identical to other avian viruses.

The p19 amino acid sequence of UR2 differs from that of PR-C and Y73 by three and two amino acids, respectively. It is not clear whether these amino acid changes affect any of the p19 antigenic determinants; however, P68 could still be efficiently precipitated by the anti-*gag* serum apparently by the remaining p19 peptide (see Fig. 21). The gp37 sequence of UR2 is well conserved relative to PR-C, Y73 and RAV-2 except for the carboxy-terminal region. The termination codons of gp37 in UR2 (at nucleotides 2627-2629), Y73 and RAV-2 occur 10 bp further downstream than that in PR-C, and there is considerable divergence in this region, especially in the stretch spanning the 13 carboxy-terminal amino acids of gp37 in UR2.

In the U3 region, 18 to 31 base changes were found when compared with other avian viruses including RAV-2, Y73, PR-C RSV and SR-A RSV (67). 12 of these changes are in positions unique to UR2, and 8 of these occur in the 3' third of U3. In addition, there are stretches of deletions and insertions. In particular, a stretch of 17 bp in the middle of U3 (nucleotides 3023 to 3040 in UR2) is marked by variable deletions in Y73, PR-RSV and SR-RSV, but not in RAV-2

and UR2. These changes in UR2 apparently have no pronounced effect on the efficiency of transcription since UR2 RNA is produced in comparable amounts in UR2-infected cells relative to the amounts of viral RNAs produced in other ASV transformed cells (46, 82, unpublished data). The universal signal sequences for transcription and processing of mRNAs are conserved in all viruses. Sequences in the inverted repeat regions found at both ends of the LTR, and in the polypurine tract, are also conserved in UR2. In addition, a region corresponding to the sequence suggested to be involved in RNA packaging is also present in UR2 (64).

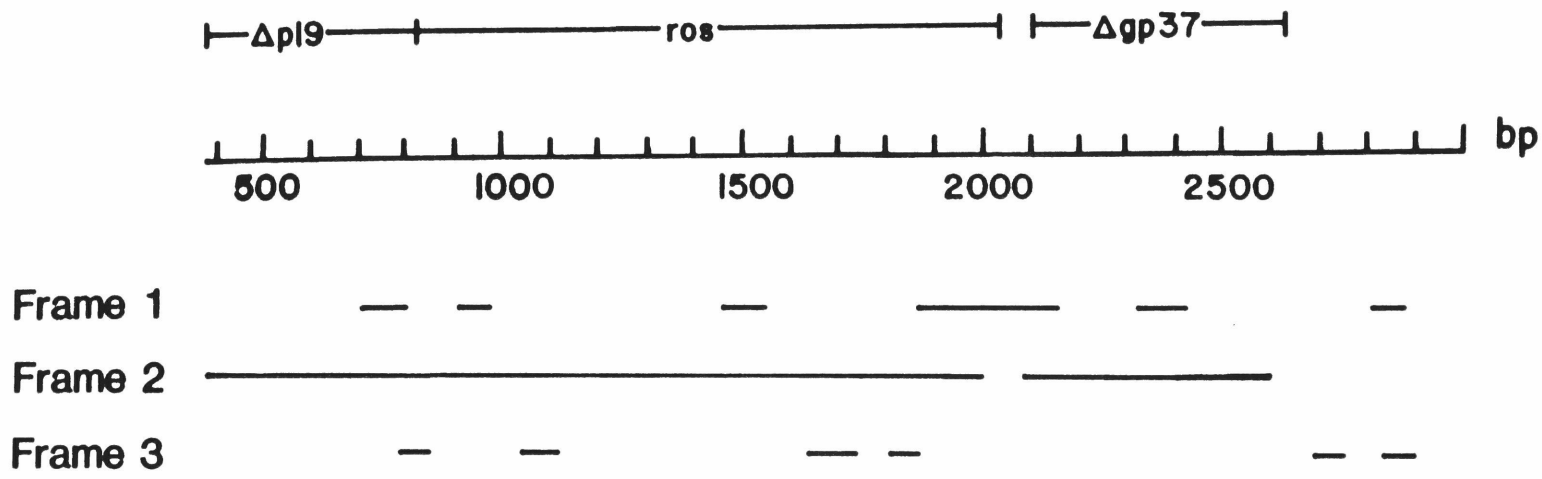
An additional *Hind*III site 67 bp upstream from the *Hind*III site 3' to *ros* detected previously by restriction enzyme mapping was discovered after sequencing the entire 3' region of the UR2 genome (Fig. 17).

Structural domains and reading frames in the UR2 genome.

The location of open reading frames in the region spanning nucleotides 380-3000 of UR2 is shown in Fig. 19. A long open reading frame in frame 2, beginning with an AUG codon at nucleotide 380 and ending in an ochre stop codon at nucleotide 2036, encodes the UR2 transforming protein P68. The other open frame in frame 2 represents the Δgp37 sequence. The next largest sequence with an open reading frame is the 111 amino acid stretch in frame 1; all other open frames are less than 40 amino acids in length. All the deduced amino acid sequences of these reading frames except Δgp37 begin with an initiator methionine. It is unlikely that those open reading frames other than the one coding for P68 are translated because subgenomic mRNAs have not been detected in UR2-transformed cells (82).

The size of the *ros* coding region is 1205 bases. A non-coding sequence of 65 bp follows the ochre stop codon of *ros* and contains four termination codons, three of which are immediately adjacent to each other.

Fig. 19. Location of sequences with open reading frames within the UR2 genome. Considering the predicted cap site as nucleotide 1 of the sequence presented in Fig. 17, the genome was translated in all three frames. Only the region spanning nucleotides 380-3000 is shown here.


Δp19
ros
Δgp37
bp
500
1000
1500
2000
2500
Frame 1
Frame 2
Frame 3

As deduced from the DNA sequence, P68 contains 552 amino acids, of which 150 are encoded by *gag* and 402 are encoded by *ros*. The transforming protein has an apparent molecular weight of 68 kd as estimated from SDS-polyacrylamide gel electrophoresis (17), but the predicted molecular weight of the deduced amino acid sequence is 61,113 daltons. It is unlikely that this discrepancy is due to glycosylation of the protein, because *in vitro* translation of the genomic 24S UR2 RNA yielded a protein product of 68 kd (17). Unusual amino acid composition such as a high percentage of cysteine or proline could cause extensive folding and change the relative mobility of the protein in SDS-polyacrylamide gels (20, 78). The amino acid composition of $P68^{gag\text{-}ros}$ shows that proline comprises only 5.2% and cysteine 2.35% of the protein, so this probably does not account for the 6.8 kd higher apparent molecular weight. It is possible that phosphorylation of specific amino acid residues *in vivo* may alter the migration of the protein.

The hydrophilicity profile of the *ros* domain of the fused protein $P68^{gag\text{-}ros}$ (Fig. 20) revealed one amino- and one carboxy-terminal hydrophobic regions (shaded area in Fig. 20). The amino-terminal hydrophobic region is 29 amino acids long (underlined in Fig. 17) and is characterized by a stretch of 13 consecutive hydrophobic amino acids, and is flanked by an acidic residue (aspartic acid) immediately upstream and a basic residue (arginine) two amino acis downstream of this sequence. The carboxy-terminal hydrophobic region (from amino acids 448 to 456, Fig. 21) contains nine hydrophobic amino acids flanked by an aspartic acid and a glutamic acid at N- and C-termini, respectively. With the exceptions of the *ros* and *abl* (54) proteins, this hydrophobic region is interrupted by serine in all other tyrosine kinase oncogene products (Fig. 21). P68 in UR2-transformed CEF remained associated with the plasma membrane during cellular fractionation even under conditions of high salt (300 mM) (E. Garber et. al., personal

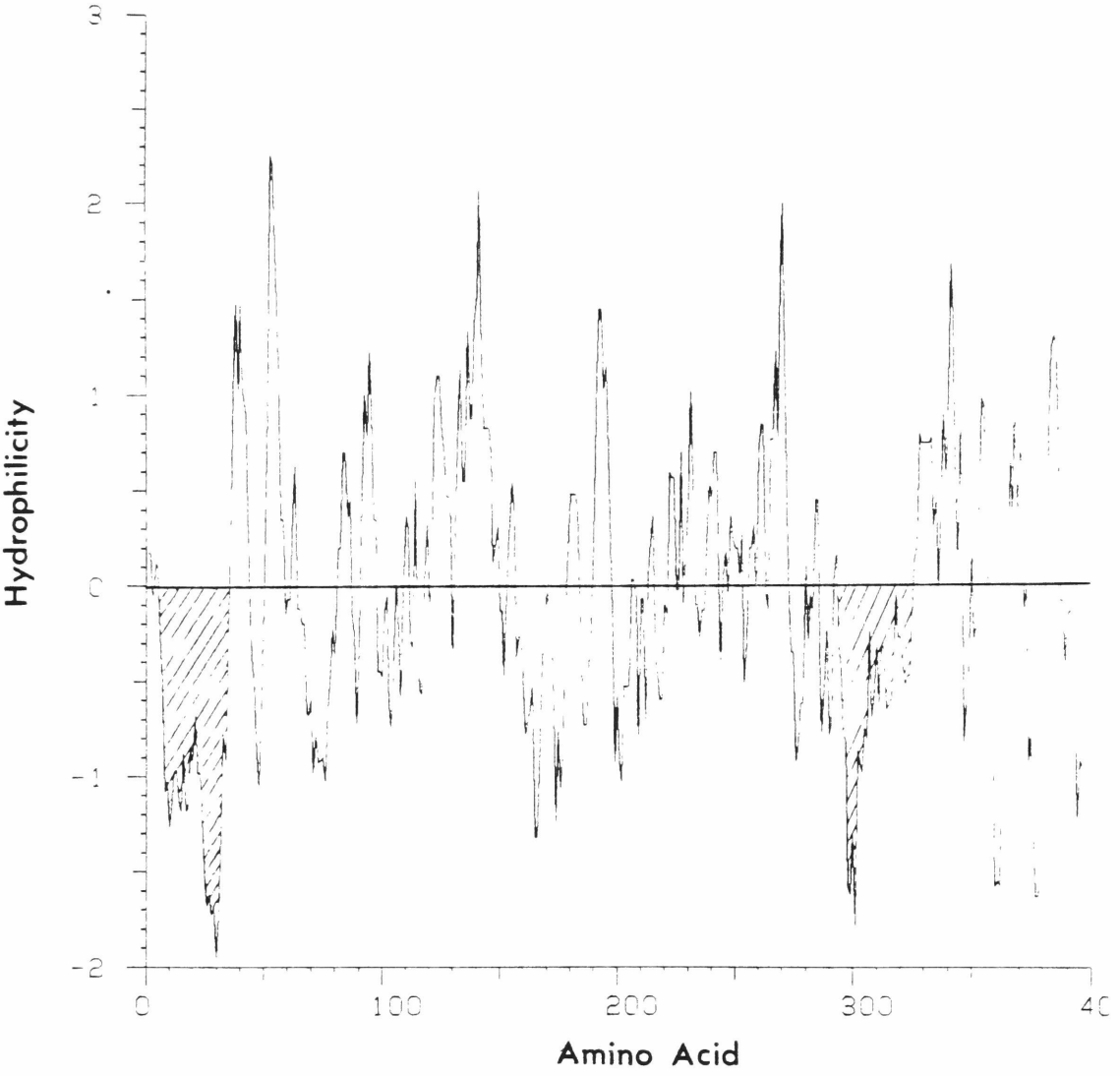


Fig. 20. Hydrophilicity analysis of the deduced amino acid sequence of *ros*. Positive values on the Y axis denote hydrophilic stretches of amino acids, and negative values denote hydrophobicity. The two cross-hatched areas indicate the extensively hydrophobic regions of *ros*.

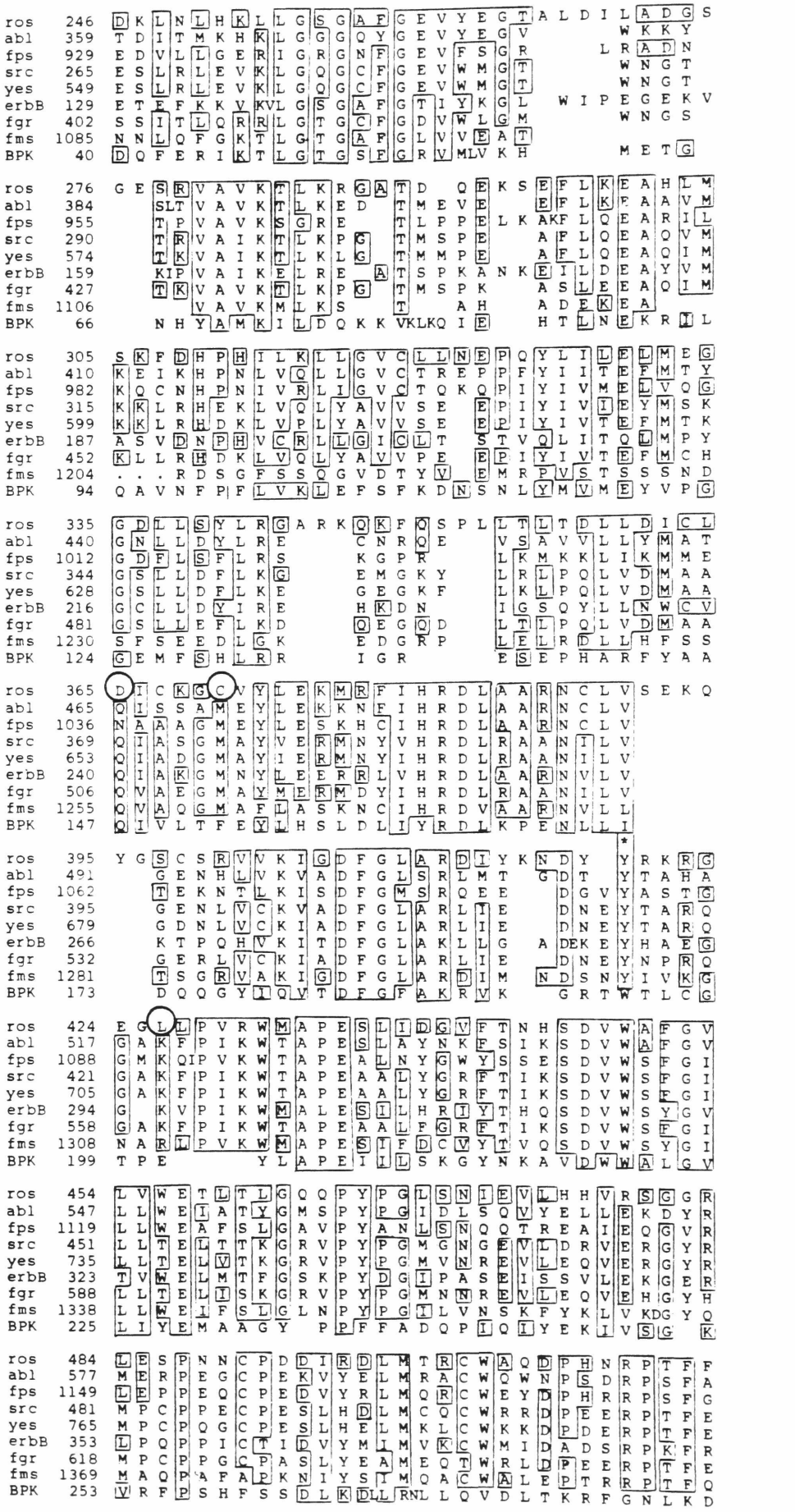
communication). The amino-terminal hydrophobic region is long enough to span the membrane, and we speculate that this region may be responsible for the association of P68 with the plasma membrane.

Comparison of P68$^{gag\text{-}ros}$ with other protein kinases.

The region of the deduced amino acid sequence of P68 was compared to the analogous domains of the proteins encoded by the viral oncogenes *src* (11, 12, 69), *yes* (29), *fps* (60), *fms* (23), *erb* B (87), *fgr* (45) and *abl* (54) as well as the catalytic subunit of the cyclic AMP-dependent bovine protein kinase (5, 63) (Fig. 21). Certain highly conserved regions are immediately apparent. The lysine residue which is the proposed ATP-binding site of BPK is conserved in *ros* (amino acid no. 283), and so is the characteristic sequence Leu-Gly-X-Gly-X-Phe-Gly-X-Val 15 to 20 residues upstream from the binding site (5). This region is highly conserved in all the kinases under comparison and is likely to be an important structural domain. Two other highly conserved regions include a 26 residue stretch 28 amino acids upstream from the putative phosphotyrosine acceptor site (amino acid no. 419, marked with an asterisk), and the 28 amino acid stretch 6 amino acids downstream from this site (found in row. 5-7 in Fig. 21). Among the kinase related oncogenes, *fgr* and *raf* (not shown in Fig. 21) lack demonstrable kinase activity (37, 45) although they share regions of conserved amino acid sequences. It is interesting to note that the circled amino acids of *ros*, which denote amino acids different from the rest of the tyrosine protein kinases, occur at the N-terminal regions of both of the two highly conserved sequences. Furthermore, *ros* contains a unique 6 amino acid insertion (amino acid no. 391-396) within the highly conserved region, in addition to two other insertions at positions 344 to 346 and 352 to 353.

I have aligned the tyrosine residue at amino acid no. 419 of *ros* with the

Fig. 21. Comparison of the amino acid sequences within the conserved region of protein kinases encoded by *ros* and other oncogenes and the catalytic subunit of the bovine protein kinase gene. The 3' no. 246 through 513 amino acids of $P68^{gag\text{-}ros}$ have been aligned to show homology with the other protein kinases. They are: *abl* (Abelson murine leukemia virus, 54); *fps* (Fujinami sarcoma virus, 60); *src* (SR-A Rous sarcoma virus, 11, 12, 69); *yes* (Y73 sarcoma virus, 29); *erb*B (avian erythroblastosis virus AEV-H, 87); *fgr* (Gardner-Rasheed feline sarcoma virus, 45); *fms* (McDonough strain of feline sarcoma virus, 23); and BPK (catalytic subunit of the cyclic AMP-dependent protein kinase from bovine cardiac muscle, 63). Amino acid number is given preceding each row of sequence. The putative phosphotyrosine acceptor site of v-*ros* (amino acid no. 419) is marked with an asterisk. Regions of conserved amino acids among the proteins are boxed. Amino acids of *ros* at positions 365, 370 and 426 are circled to show the unique differences from the rest of the tyrosine protein kinases.


ros 246
abl 359
fps 929
src 265
yes 549
erbB 129
fgr 402
fms 1085
BPK 40
ros 276
abl 384
fps 955
src 290
yes 574
erbB 159
fgr 427
fms 1106
BPK 66
ros 305
abl 410
fps 982
src 315
yes 599
erbB 187
fgr 452
fms 1204
BPK 94
ros 335
abl 440
fps 1012
src 344
yes 628
erbB 216
fgr 481
fms 1230
BPK 124
ros 365
abl 465
fps 1036
src 369
yes 653
erbB 240
fgr 506
fms 1255
BPK 147
ros 395
abl 491
fps 1062
src 395
yes 679
erbB 266
fgr 532
fms 1281
BPK 173
ros 424
abl 517
fps 1088
src 421
yes 705
erbB 294
fgr 558
fms 1308
BPK 199
ros 454
abl 547
fps 1119
src 451
yes 735
erbB 323
fgr 588
fms 1338
BPK 225
ros 484
abl 577
fps 1149
src 481
yes 765
erbB 353
fgr 618
fms 1369
BPK 253

known tyrosine phosphoacceptors of the other protein kinases. The region surrounding the tyrosine phosphoacceptor site appears to be somewhat conserved, although its essentiality is not clear, in several tyrosine protein kinases (Y73 P90, PR-CII P105, $p60^{src}$ and ST-FeSV P85) in that the phosphotyrosine is located seven residues to the carboxy-terminal side of a basic amino acid (arginine or lysine) and either four residues to the carboxy-terminal side of or adjacent to a glutamic acid residue (50). However, the likely phosphoacceptor tyrosine residue at position 419 of *ros* is four residues and two residues to the carboxy-terminal side of a lysine and an aspartic acid, respectively. No tyrosine-containing peptides in the deduced amino acid sequence of P68 concur with the above proposed consensus sequence. This implies that there is some degree of flexibility with respect to primary amino acid sequence at the phosphotyrosine acceptor site. Although $P68^{gag\text{-}ros}$ is phosphorylated on serine residues *in vivo* (17), I cannot speculate which residues in particular are possible phosphorylation sites.

The N-terminal 95 amino acids of *ros* upstream from the region compared in Fig. 21 vary greatly from the corresponding domains of other protein kinases, and are characterized by a long stretch of hydrophobic sequence mentioned above.

Previous hybridization of *ros* DNA to *fps*, *src*, *yes* and *abl* DNAs detected significant homology only to *fps* (46). However, there is greater amino acid homology among *ros*, *abl* and *src* than between *ros* and *fps*. This is due to degeneracy of the triplet code. The longest stretch of nucleotide homology between *ros* and *fps* has been localized to the region upstream of the tyrosine phosphoacceptor site, where the sequences are identical for 40 nucleotides, with the exception of a single base change at the 18th position within this sequence.

It is evident from the deduced amino acid sequence that P68 is a member of the tyrosine protein kinase family. Aside from the unique amino acid changes and insertions within the conserved domain of the protein kinases compared here,

P68 contains two distinctive highly hydrophobic regions, in particular in the 5' region of *ros*. Any combination of these changes could be responsible for modulating the P68 activity and for the specific interactions between P68 and its cellular targets that lead to the unique elongated transformed CEF morphology.

Section IV. Results

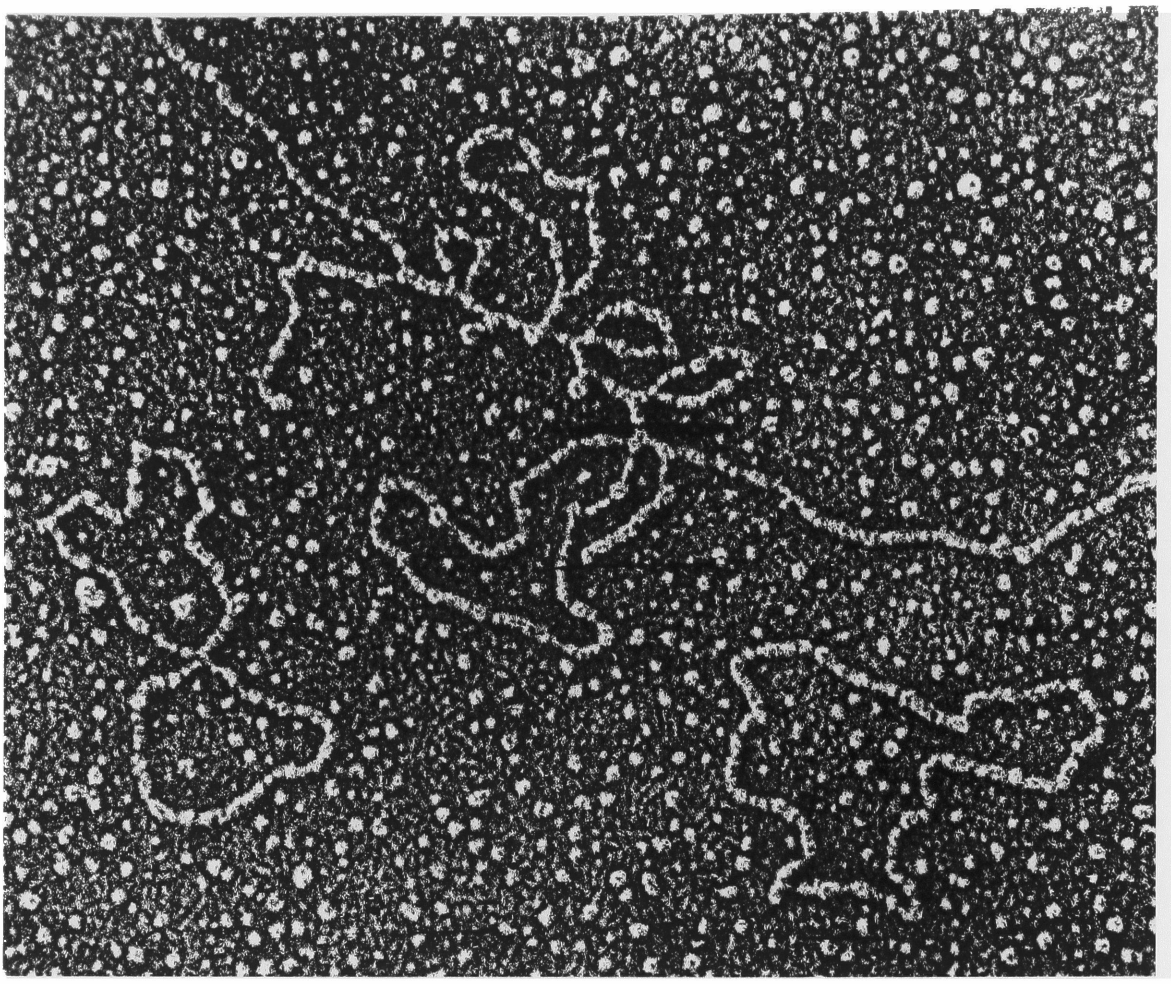
(c-ros)

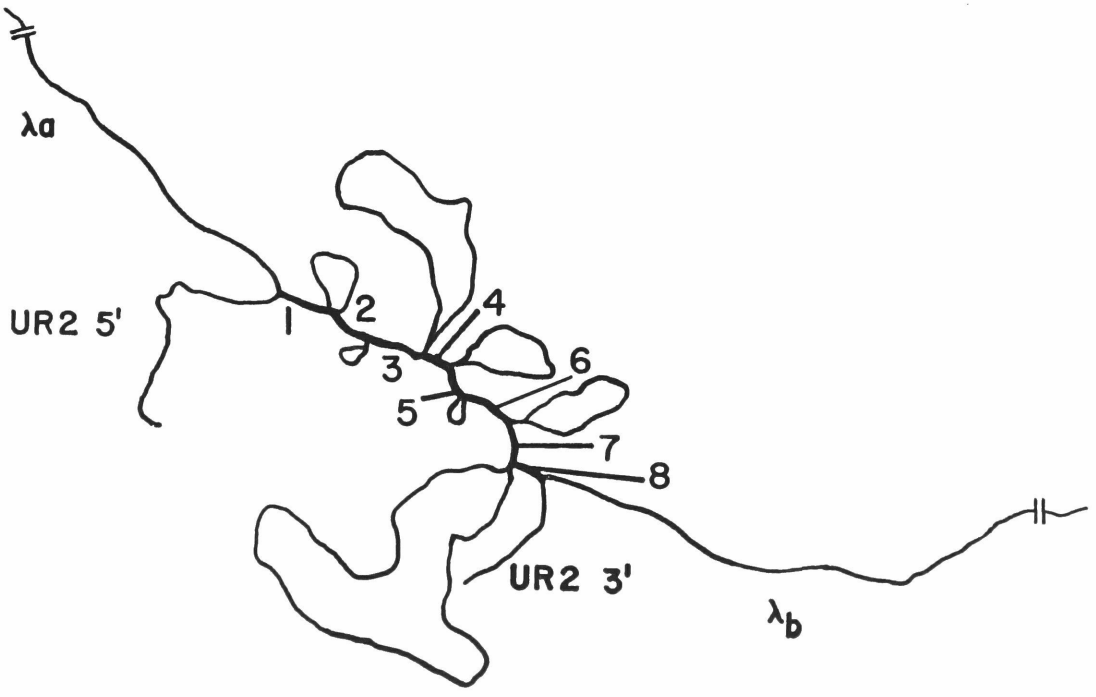
Heteroduplex and restriction mapping of c-ros.

A recombinant DNA clone containing sequences homologous to v-*ros* was isolated from a library composed of partially-digested chicken genomic DNA cloned into Charon 4A with *Eco*RI linkers by Masabumi Shibuya in our laboratory. v-*ros* is permuted at the *Eco*RI site in pUR2, and an uninterrupted *ros* sequence was desired for heteroduplex mapping with the lambda c-*ros* clone. Therefore, the *Eco*RI insert from pUR2 was freed from the vector, ligated, cut with *Hind*III and subcloned into the *Hind*III site of pBR322. Heteroduplex mapping of the lambda c-*ros* DNA with the *Hind*III insert DNA, done in collaboration with Ming-Ta Hsu at this University, showed that the v-*ros* homologous sequences in c-*ros* are distributed in eight exons, ranging in size from .11 to .19 kb, and are interrupted by seven introns, ranging in size from .3 to 3.8 kb (Fig. 22). The polarity of c-*ros* was oriented by the unequal length of the the 5' 1.3 kb and 3' 0.9 kb UR2 sequences not in the complex. The size of the exons and introns were calculated as the average values from the measurement of 30 heteroduplexed molecules (data not shown).

When CEF DNA was analysed by digestion with various restriction enzymes and subsequent hybridization of the restriction fragments with probes 150-250 bp in length derived from various regions of v-*ros*, the cellular *ros* gene appeared to be distributed over 11.7 kb, which was in good agreement with the results from electron microscopy. Analysis of the lambda recombinant clone, using the same restriction enzymes and hybridization probes as for CEF DNA, agreed fairly well with the results from heteroduplex mapping and CEF analysis. The presence of

Fig. 22. Heteroduplex mapping of pUR2 with λc-*ros* DNA. The *Eco*RI insert of pUR2 was freed from the vector, ligated, digested with *Hind*III, recloned into the *Hind*III site of pBR322 and cut with *Sst*I to provide a non-permuted v-*ros* gene. Heteroduplex mapping was done in collaboration with Ming-Ta Hsu at this Institute.




λa
UR2 5'
1
2
3
4
5
6
7
8
UR2 3'
λb

partially-digested DNA after restriction enzyme digestions of CEF DNAs obscured a complete analysis of the c-*ros* DNA, but analysis of the lambda c-*ros* clone by ^{32}P end-labeling and partial restriction enzyme digestion, and by Southern blotting analysis using the same v-*ros* DNA probes mentioned above, enabled me to construct a restriction map to orient the c-*ros* clone relative to v-*ros* and to determine the approximate intron/exon boundaries (Fig. 23).

Sequencing of cellular ros.

The c-*ros* insert from the lambda clone contains two *Eco*RI sites (Fig. 23), and can be digested with *Eco*RI into 5' 5.2 kb, internal 9.7 kb, and 3' 1.7 kb fragments. These *Eco*RI DNA fragments were subcloned into pBR322. Each fragment prepared from the pBR322 clones was further divided into several restriction fragments and subcloned into an M13 phage vector for Sanger dideoxy sequencing (Fig. 23). I am currently sequencing the region I believe to contain the 65 bp exon 4.

Fig. 24 details the differences found between viral and cellular *ros*. In exon one, there is a 9 bp insertion in v-*ros* which adds 3 hydrophobic amino acid residues (specifically serine, leucine and threonine) to the carboxy end of the amino-terminal hydrophobic region of viral *ros*. Therefore, v-*ros* contains a stretch of 29 and c-*ros* a stretch of 26 hydrophobic residues. The sequence inserted in v-*ros*, GCTTGACTA, is a direct repeat of the immediately preceding 9 bases. The 26 amino acid hydrophobic stretch in c-*ros* is still of sufficient length to span the plasma membrane.

Additionally, a single base pair change has been noted in exon nine: C in c-*ros* has been changed to an A in v-*ros*; however, the amino acid, arginine, is conserved.

The 3' ends of viral and cellular *ros* diverge before the termination codons

Fig. 23. Restriction enzyme cleavage map and nucleotide sequencing strategy of λc-*ros*. Exons are denoted by the black boxes. The arrows indicate the direction and extent of sequencing determined by the Sanger method (55). Abbreviations for the restriction enzymes are as follows: A, *Acc*I; B, *Bam*HI; C, *Cla*I; H, *Hind*III; Ha, *Hae*III; Hi, *Hinf*I; N, *Nco*I; P, *Pst*I; Pv, *Pvu*II; R, *Eco*RI; RV, *Eco*RV; S, *Sst*I; St, *Stu*I.

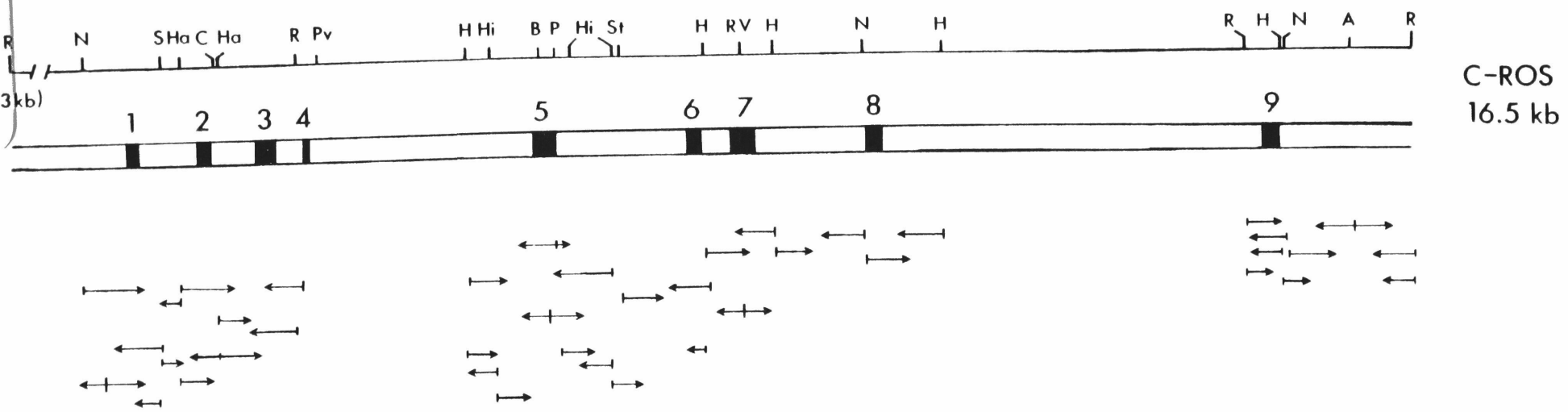
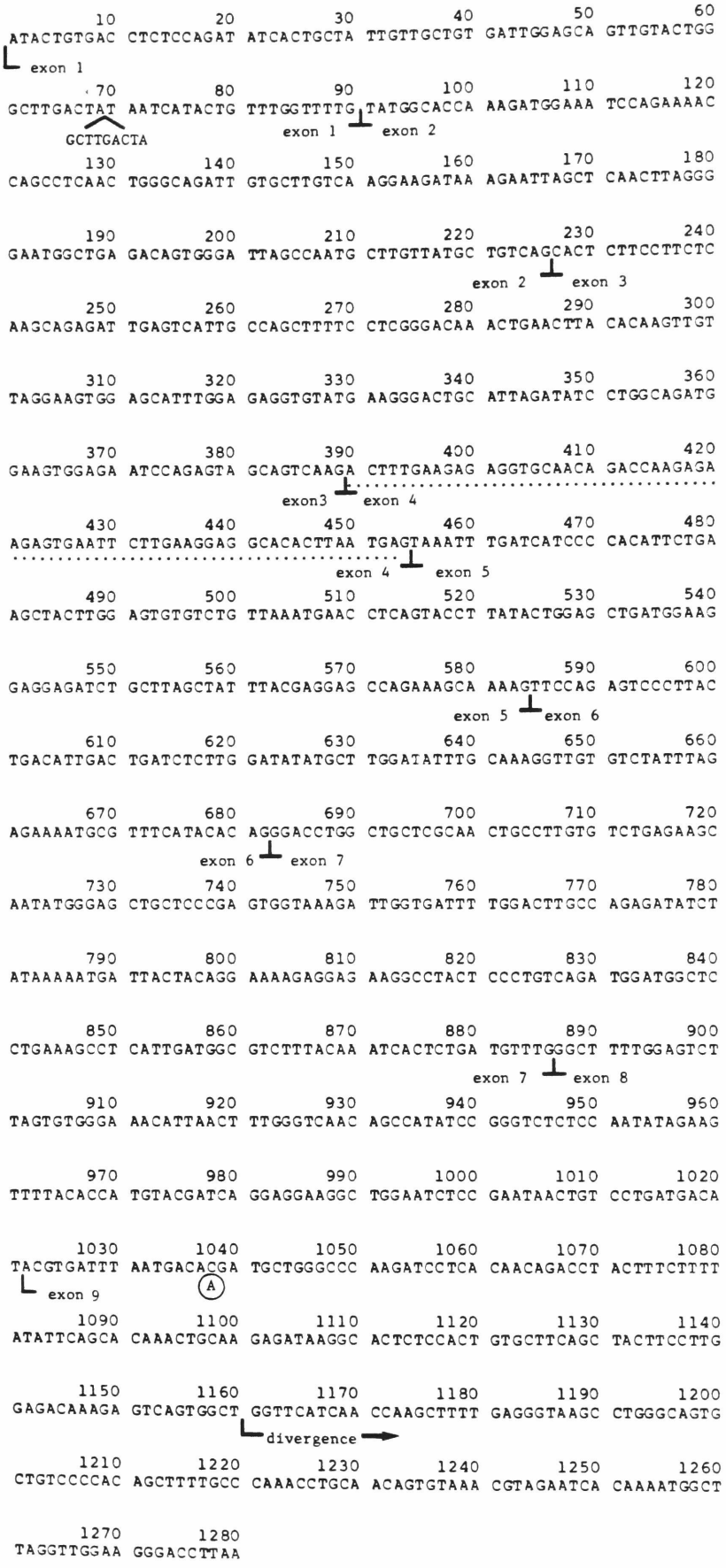

R
N
S Ha C Ha
R Pv
H Hi
B P Hi St
H RV H
N
H
R H N A R
3kb)
1
2
3
4
5
6
7
8
9
C-ROS
16.5 kb

Fig. 24. Sequence of v-*ros* homologous sequences in c-*ros*. Exon boundaries are denoted below the sequence in addition to any changes occurring in v-*ros*. Ellipses under the exon 4 sequence indicate it has not yet been sequenced.


10 20 30 40 50 60
ATACTGTGAC CTCTCCAGAT ATCACTGCTA TTGTTGCTGT GATTGGAGCA GTTGTACTGG
exon 1
70 80 90 100 110 120
GCTTGACTAT AATCATACTG TTTGGTTTTG TATGGCACCA AAGATGGAAA TCCAGAAAAC
GCTTGACTA
exon 1 exon 2
130 140 150 160 170 180
CAGCCTCAAC TGGGCAGATT GTGCTTGTCA AGGAAGATAA AGAATTAGCT CAACTTAGGG
190 200 210 220 230 240
GAATGGCTGA GACAGTGGGA TTAGCCAATG CTTGTTATGC TGTCAGCACT CTTCCTTCTC
exon 2 exon 3
250 260 270 280 290 300
AAGCAGAGAT TGAGTCATTG CCAGCTTTTC CTCGGGACAA ACTGAACTTA CACAAGTTGT
310 320 330 340 350 360
TAGGAAGTGG AGCATTTGGA GAGGTGTATG AAGGGACTGC ATTAGATATC CTGGCAGATG
370 380 390 400 410 420
GAAGTGGAGA ATCCAGAGTA GCAGTCAAGA CTTTGAAGAG AGGTGCAACA GACCAAGAGA
exon3 exon 4
430 440 450 460 470 480
AGAGTGAATT CTTGAAGGAG GCACACTTAA TGAGTAAATT TGATCATCCC CACATTCTGA
exon 4 exon 5
490 500 510 520 530 540
AGCTACTTGG AGTGTGTCTG TTAAATGAAC CTCAGTACCT TATACTGGAG CTGATGGAAG
550 560 570 580 590 600
GAGGAGATCT GCTTAGCTAT TTACGAGGAG CCAGAAAGCA AAAGTTCCAG AGTCCCTTAC
exon 5 exon 6
610 620 630 640 650 660
TGACATTGAC TGATCTCTTG GATATATGCT TGGATATTTG CAAAGGTTGT GTCTATTTAG
670 680 690 700 710 720
AGAAAATGCG TTTCATACAC AGGGACCTGG CTGCTCGCAA CTGCCTTGTG TCTGAGAAGC
exon 6 exon 7
730 740 750 760 770 780
AATATGGGAG CTGCTCCCGA GTGGTAAAGA TTGGTGATTT TGGACTTGCC AGAGATATCT
790 800 810 820 830 840
ATAAAAATGA TTACTACAGG AAAAGAGGAG AAGGCCTACT CCCTGTCAGA TGGATGGCTC
850 860 870 880 890 900
CTGAAAGCCT CATTGATGGC GTCTTTACAA ATCACTCTGA TGTTTGGGCT TTTGGAGTCT
exon 7 exon 8
910 920 930 940 950 960
TAGTGTGGGA AACATTAACT TTGGGTCAAC AGCCATATCC GGGTCTCTCC AATATAGAAG
970 980 990 1000 1010 1020
TTTTACACCA TGTACGATCA GGAGGAAGGC TGGAATCTCC GAATAACTGT CCTGATGACA
1030 1040 1050 1060 1070 1080
TACGTGATTT AATGACACGA TGCTGGGCCC AAGATCCTCA CAACAGACCT ACTTTCTTTT
exon 9
A
1090 1100 1110 1120 1130 1140
ATATTCAGCA CAAACTGCAA GAGATAAGGC ACTCTCCACT GTGCTTCAGC TACTTCCTTG
1150 1160 1170 1180 1190 1200
GAGACAAAGA GTCAGTGGCT GGTTCATCAA CCAAGCTTTT GAGGGTAAGC CTGGGCAGTG
divergence
1210 1220 1230 1240 1250 1260
CTGTCCCCAC AGCTTTTGCC CAAACCTGCA ACAGTGTAAA CGTAGAATCA CAAAATGGCT
1270 1280
TAGGTTGGAA GGGACCTTAA

of both genes (Fig. 25). In UR2, as described above, the stop codon of viral *ros* is followed by a noncoding region of 65 bp that includes four more in frame termination codons. However, c-*ros* and v-*ros* abruptly diverge 36 bp or 12 codons upstream of the v-*ros* stop codon. c-*ros* continues for 102 bp or 34 amino acids before reaching a termination codon (Fig. 25A). However, there is a splice donor sequence, AGG-GTAAGC (dashed box in Fig. 25A) 21 bp (or 7 amino acids) after the divergence, and a splice acceptor, CTTCAATTTCCCCATTAG-C, approximately 525 nucleotides from the GT dinucleotide of the donor site. If these sites are used for splicing, then c-*ros* would continue for an additional 9 amino acids before termination (Fig. 25B). There are two additional possible splice donor sites just upstream of the possible splice acceptor; however, no other possible splice acceptor consensus sequences were found. The 3' 1.4 kb downstream of the divergence in c-*ros* does not contain an AAUAAA sequence; although the variant poly(A) signal AGTAAA, found in MTV and BaEV (71, 74), is found in a single copy, the consensus sequence YGTGTTYY (40) was not found downstream. No other sequences corresponding to the two other known variant poly(A) signals, AUUAAA and AAUAUA (1, 35, 73), were found. The 3' 500 bp of the c-*ros* clone is marked by stretches of As and Ts.

There is no homology between the 3' divergent region of v-*ros* and the 1.4 kb 3' noncoding sequences in the c-*ros* clone, nor is there any apparent homology between the v-*ros* divergent region and any viral or chicken genes sequenced to date. Hybridization of a probe containing only the v-*ros* divergent sequence to CEF DNA is under progress to determine whether this region might be part of c-*ros* sequences found further downstream, or whether it might have arisen from another locus in the chicken genome.

These results suggest the c-*ros* protein differs from v-*ros* for the 12 carboxy-terminal amino acids of v-*ros*, although it is unclear how much larger the

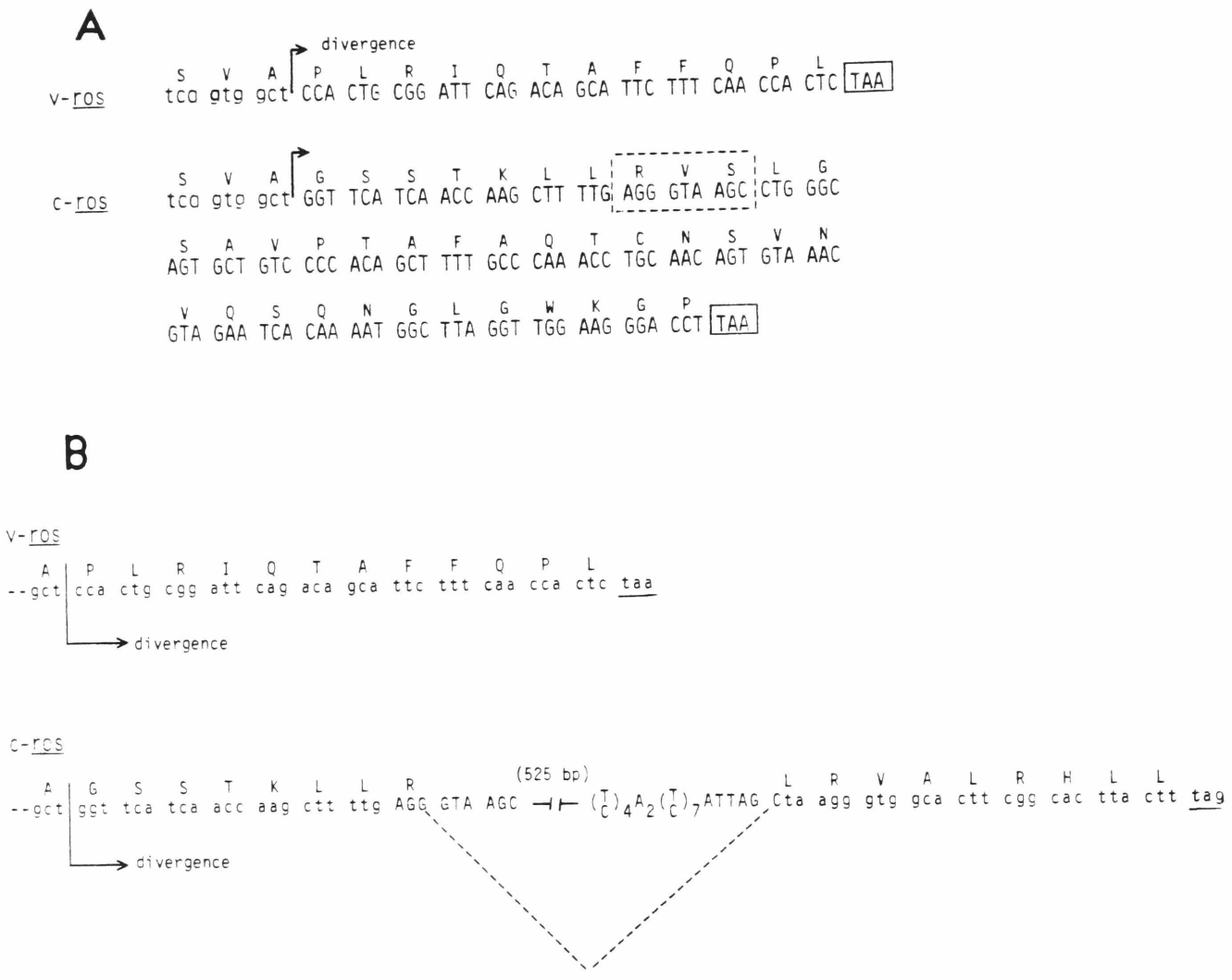


Fig. 25. Comparison of the 3' ends of viral and cellular *ros*. A. Deduced amino acid sequence of 3' c-*ros* if ochre stop codon (boxed) is used. Potential splice donor is shown in dashed box. B. Deduced amino acid sequence of 3' c-*ros* if splice donor (upper case letters) is used, showing potential splicing to the acceptor signal (upper case letters) approximately 525 bp downstream.

carboxy-terminus of c-*ros* may be. Several oncogenes, including *src*, *myc*, *myb*, and *fos*, differ from their viral counterparts in the C-terminal region.

In UR2, the 5' end of v-*ros* is fused to the 3' coding region for p19, and P68 uses the UR2 promoter and the initiator methionine of *gag*. The beginning of v-*ros* corresponds to a consensus splice acceptor sequence in c-*ros* (Fig. 26). 340 bp upstream of this splice acceptor has been sequenced, and there are stop codons both in and out of frame relative to v-*ros*. In addition, there do not appear to be any promoter sequences, indicating there exists an additional upstream exon (or exons) in c-*ros*. Preliminary sequencing of the corresponding region of the UR2AV molecular clone 14-1 reveals a splice donor signal (Fig. 26). This suggests that the viral *ros* and Δ*gag* junction in UR2 was formed by splicing.

Fig. 27 shows the splice acceptor, the size of the exon, and the splice donor for each known exon of c-*ros*. The exons range in size from 65 to 204 bp. The sizes of the exons and introns even within the conserved kinase domain do not appear to be conserved when compared with those of c-*fps* and c-*src* (27, 70).

Expression of cellular ros.

To study the expression of c-*ros*, I extracted poly(A)-containing RNA from chicken and quail embryo fibroblasts as well as from HeLa, 3Y1, 3T3, and QT6 cell lines (human, rat, mouse and quail origin, respectively). The RNA was denatured by treatment with glyoxal, and 20 ug of each RNA species was analysed by Northern blotting using using a *ros*-specific probe (*Pvu*II B fragment, Fig. 14) under conditions of low stringency (35% formamide, 5X SSC). No *ros*-specific transcripts were detected; however, hybridization with a *src*-specific probe showed strongly hybridizing bands corresponding to expected c-*src* transcripts from CEF, QEF, and QT6, and fainter bands from 3Y1 and 3T3 cells not expressed, or expressed at an

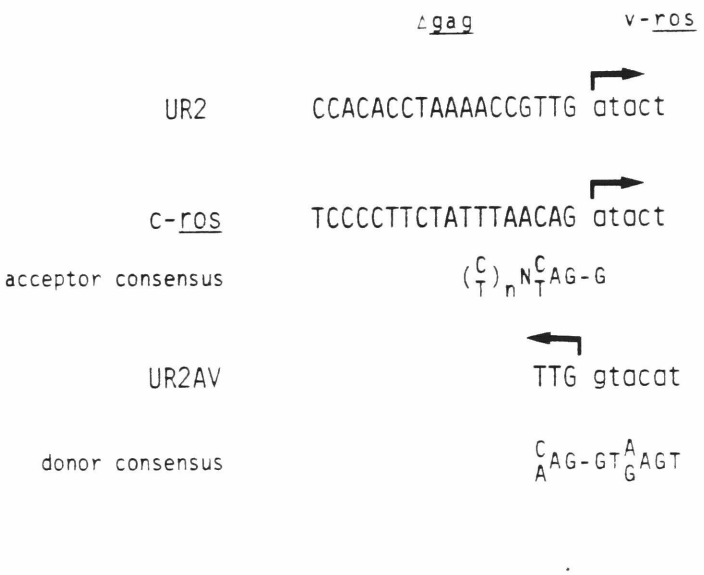


Fig. 26. Comparison of the 5' regions of the junctions of viral and cellular *ros*. The consensus splice signals are shown, as well as the corresponding region from UR2AV.

Fig. 27. Exon size, splice acceptor and splice donor signals in c- *ros*.

EXON	ACCEPTOR	BP ROS (SIZE IN BP)	DONOR
CONSENSUS	$(^{T}_{C})_{n}N^{T}_{C}AG\text{-}G$		$^{C}_{A}AG\text{-}GT^{A}_{G}AGT$
1	$(^{T}_{C})_{12}A(T)_{3}AACAG\text{-}A$	1 - 90 (90)	TTG-GTATGT
2	$(^{T}_{C})_{11}ATTCCAG\text{-}T$	91 - 226 (136)	CAG-GTATGT
3	$(^{T}_{C})_{7}GCAG\text{-}C$	227 - 389 (163)	AAG-GTAATC
4		390 - 454 (65)	
5	$(T)_{3}A(^{T}_{C})_{7}TTAG\text{-}T$	455 - 584 (130)	AAG-GTAGGA
6	$(^{T}_{C})_{5}A(^{T}_{C})_{7}GTAG\text{-}T$	585 - 682 (98)	CAG-GTACAG
7	$(^{T}_{C})_{5}ACCATAAAATTCAG\text{-}G$	683 - 886 (204)	TTG-GTGAGC
8	$(^{T}_{C})_{4}TTAG\text{-}G$	887 - 1021 (135)	CAT-GTAAGT
9	$(^{T}_{C})_{6}AATAA(^{T}_{C})_{10}CCAG\text{-}A$	1022 - (139+)	

extremely low level (less than one copy per cell) in those cells. However, the possibility that the c-*ros* transcript is very unstable cannot be excluded since I have been looking at steady-state RNA. Nevertheless, previous analysis using liquid hybridization and unselected total cellular RNA did not detect c-*ros* transcripts in most tissues examined from two week old chickens (61).

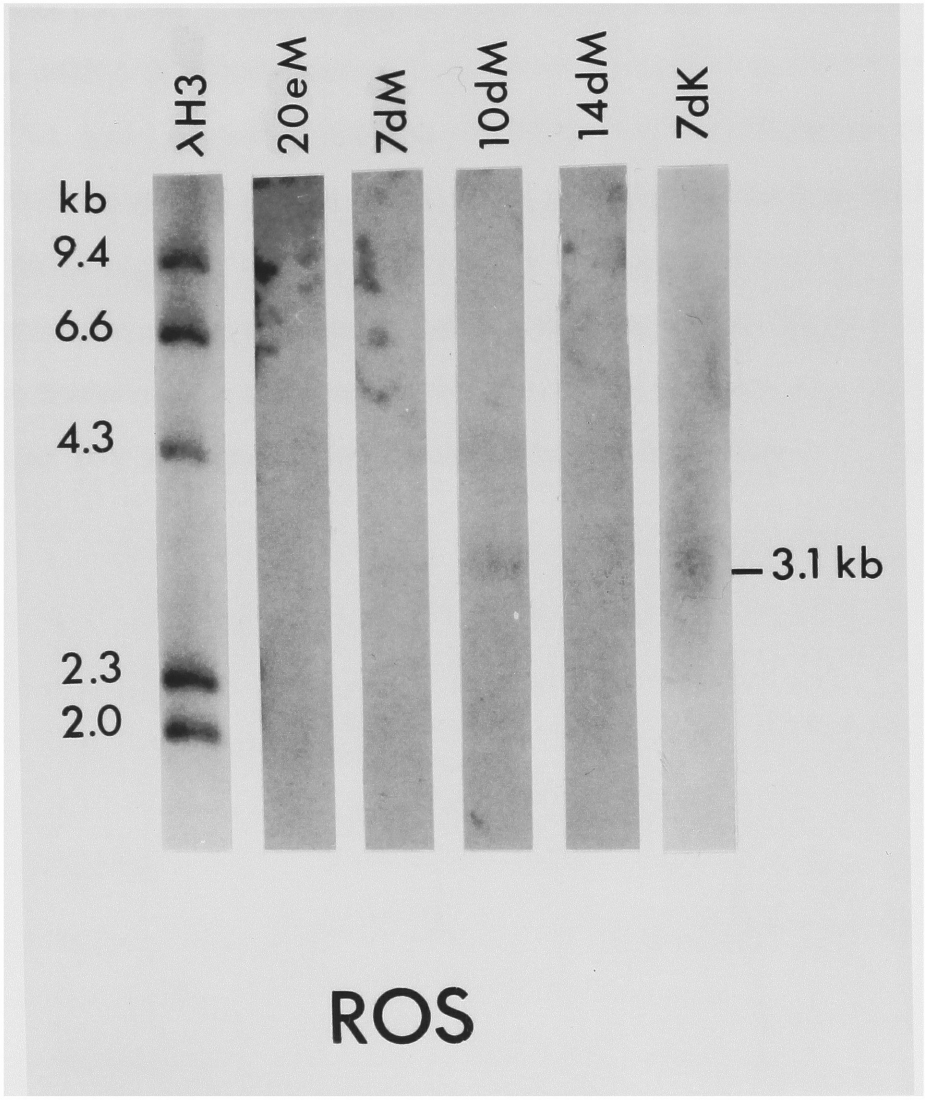
I also examined several tissues from chicken at different embryonic and adult developmental stages to determine the pattern of cellular *ros* expression in chicken. Twice poly(A)-selected RNA was isolated from brain, eye, muscle, gut, heart, liver, and lung, and from extraembryonal membranes, from 12, 14, 16, 18, and 20 day old chick embryos (lung tissue was not isolated from 12, 14, or 16 day embryos), and from 2 day and 7 day old chicks. In addition, kidney mRNA was isolated from 4 and 7 day old chicks. The RNAs were hybridized with a *src*-specific probe (Fig. 28) or a *ros*-specific probe (Fig. 28). Fig. 28 depicts only the 12 and 14 day embryo and 2 day and 7 day chick timepoints. Hybridization with the *src*-specific probe as a positive control showed transcripts of the expected size in the appropriate tissues: a 4 kb transcript was seen in all tissues where *src* is expressed except for muscle, where the multiple forms of smaller RNAs were seen as previously observed by Iijima and Wang (unpublished data). These *src* transcripts can be detected after a 17-20 hour exposure of the blot with an intensifying screen. However, even after a 96 hour exposure, no obvious evidence of *ros* expression is seen after hybridization of the Northern filters to the *ros*-specific probe.

A faint signal was detected in kidney tissue at the 7 day timepoint; even fainter was a possible band in adult muscle tissue. To verify these results, poly(A)-selected mRNA was isolated from kidney at 20 day embryo and 3, 7, 10, and 14 day old chicks (Fig. 29). A 3.1 kb transcript from muscle was detectable in 7 day old chick and gave its strongest signal at 10 days. This transcript was

Fig. 28. Expression of c-*ros*. Twice poly(A)-selected mRNA from the tissue indicated was denatured with glyoxal, electrophoresed through a 1% agarose gel, blotted onto nitrocellulose and hybridized with a *src*-specific or a *ros*-specific probe.

brain
eye
muscle
gut
heart
liver
kidney
λH3
brain
eye
muscle
gut
heart
liver
kidney
6.6
4.3
2.3
2.0
12e
14e
2d
7d
SRC
ROS

Fig. 29. Expression of c-*ros*. Twice poly(A)-selected mRNA from kidney and muscle tissues from chicken was denatured with glyoxal and subjected to electrophoresis througha 1% agarose gel and transferred to nitrocellulose. Hybridization was with a *ros*-specific probe.


λH3
20eM
7dM
10dM
14dM
7dK
kb
9.4
6.6
4.3
2.3
2.0
3.1 kb
ROS

not detectable in 20 day embryo or 14 day chick, suggesting the c-*ros* gene is activated only briefly in this tissue. Muscle mRNA from chickens one month or older appears negative for *ros* expression (data not shown). A similar sized transcript can be found in kidney from 20 day embryo to 14 day old chicks, although this signal was not seen in mRNA isolated from chickens one month and older.

The *ros* mRNA in kidney appears to be more degraged than the corresponding *src* mRNA, and I cannot exclude the possibility that the cellular *ros* message is preferentially degraded in this tissue. It is possible that lack of detectable transcripts in kidney with the *ros*-specific probe may not reflect a very low level of transcription but rather a greater instability of the message. These results are preliminary; however, it is apparent that in the tissues screened, the cellular *ros* locus is under very stringent control, most likely due to repression of transcription.

Section V. Discussion

Avian sarcoma virus UR2 is a replication-defective virus that can induce sarcomas *in vivo* and transforms chicken embryo fibroblasts in culture to a characteristic, extremely elongated morphology (4). The genome of UR2 contains a 1.2 kb transformation-specific sequence, v-*ros*, which has a homologous counterpart, c-*ros*, in normal chicken cellular DNA (61). UR2 was presumably generated by recombination between UR2AV and c-*ros* at the expense of certain replicative sequences in UR2AV. As a result, *ros* was fused to the 5' region of the UR2AV sequence which codes for part of the viral structural protein p19 (82). The fused p19 and *ros* sequences in UR2 code for a polyprotein of 68 kd, called P68, which was found to be associated with a tyrosine-specific protein kinase activity (17). P68 has similar biochemical properties to the protein kinases encoded by several other oncogenic viruses, and this functional conservation implies that *ros* must be closely related to the tyrosine protein kinase-encoding oncogene family. However, P68 differs from the kinases encoded by FSV, ASV Y73 and RSV in many enzymatic properties. To elucidate the basis of the functional conservation as well as the differences between *ros* and other oncogenes, I sequenced the entire genome of UR2 and compared the predicted amino acid sequence of P68 with other members of the tyrosine protein kinase family.

The results show that *ros* is 1273 nucleotides in length, including a 65 bp 3' noncoding stretch. The deduced amino acid sequence for the UR2 transforming protein P68 gives a molecular weight of 61,113 daltons and shows that it is closely related to the oncogene family coding for tyrosine protein kinases. However, P68 contains two distinctive hydrophobic regions that are absent in most of the other tyrosine kinases and it has unique amino acid changes and insertions within the conserved domain of the kinases. Fig. 30 summarizes the structural

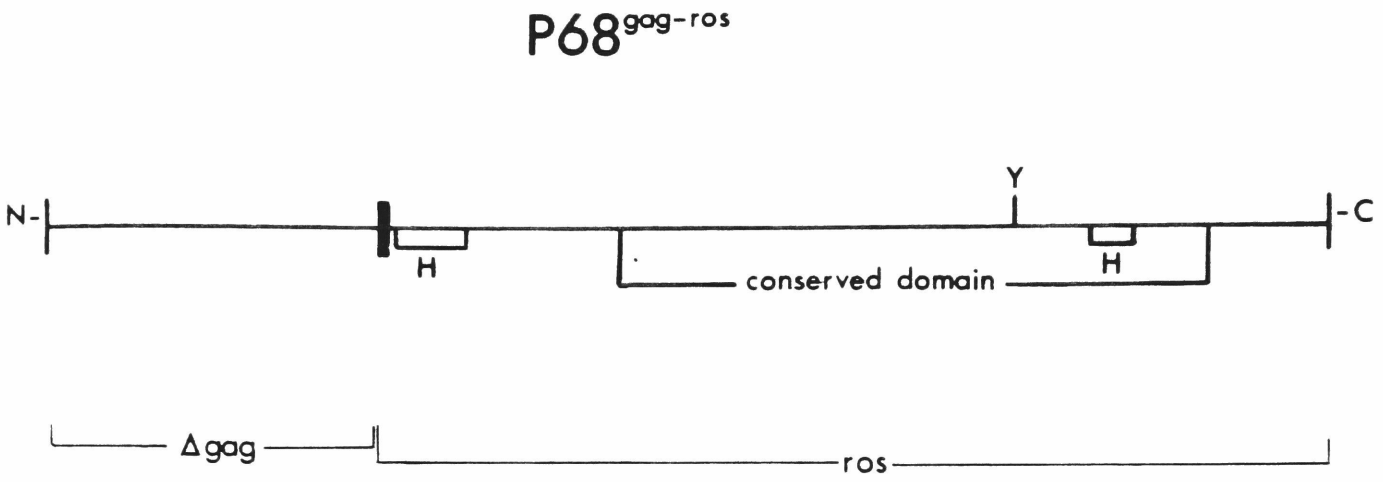


Fig. 30. Structural domains of P68$^{gag-ros}$. A portion of the p19 region of *gag* is fused to *ros* at the amino terminus. The amino- and carboxy-terminal hydrophobic regions (H) are marked with small boxes. The possible phosphotyrosine acceptor site (Y) is indicated within the conserved kinase domain of *ros*.

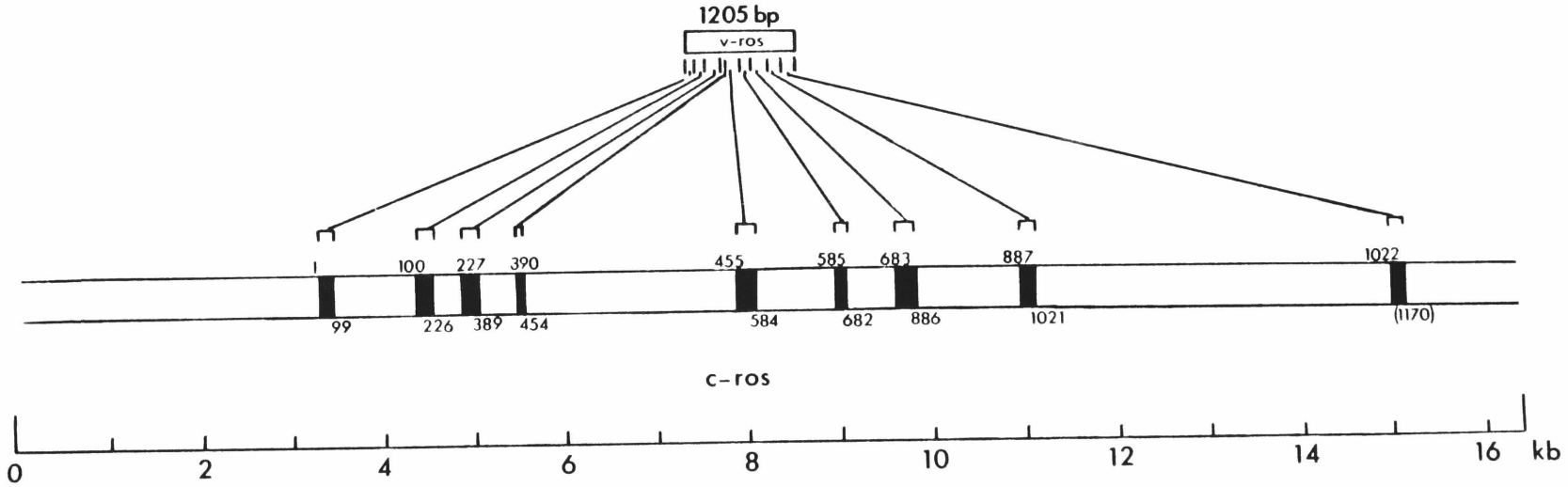
domains of $P68^{gag\text{-}ros}$.

The cellular homologues of the retroviral transforming genes, called cellular oncogenes or proto-oncogenes, have been highly conserved during the course of evolution. Several of the cellular oncogenes are expressed in a tissue-specific manner, and in many cases the sizes of these transcripts have been maintained across widely divergent species. It is currently believed these genes play an important role in cellular growth and/or differentiation, and appear to have an oncogenic potential that can be manifested after transduction by a retrovirus. The process of conversion from a normal proto-oncogene to a transforming oncogene can involve either mutation of the gene or deregulation of the gene, or both. I have determined the sequence of cellular *ros* and compared it to that of viral *ros* to determine the changes between them that may be responsible for their differential oncogenicity. In addition, I have analysed the expression of the cellular gene in both embryonic and adult chickens in an attempt to understand the normal function of cellular *ros*.

Conservation of the viral and cellular ros genes.

Viral oncogenes and their corresponding cellular homologues are closely related. In fact, excluding the 3' divergence, the 1.2 kb v-*ros* sequence is remarkably well conserved when compared with the corresponding region of c-*ros*. The v-*ros* sequences are distributed in nine exons of c-*ros* over a range of 12 kb of DNA (Fig. 31). c-*ros* and v-*ros* differ only by a 9 bp duplication, a single base change not resulting in an amino acid change, and the divergence of their 3' ends (see Fig. 24). The viral reverse transcriptase has been shown to have a high frequency of mismatch (21) and it has been speculated that the scattered nucleotide changes between viral and cellular oncogenes are a consequence of faulty reverse transcription during replication (30). UR2 was molecularly cloned from virus

Fig. 31. Derivation of c-*ros* exon sequences from v-*ros*. Numbers above and below the exons (solid black boxes) denote the corresponding nucleotide in v-*ros*.


1205 bp
v-ros
1
100
227
390
455
585
683
887
1022
99
226
389
454
584
682
886
1021
(1170)
c-ros
0
2
4
6
8
10
12
14
16
kb

that had been through a limited number of passages since its isolation from the original tumor. The striking degree of conservation between viral and cellular *ros* may result from limited opportunity for the reverse transcriptase to induce mutations in the viral gene. A similar degree of conservation between viral and cellular oncogenes for which sequences are known has been observed in ASV PR-CII where v-*fps* differs from c-*fps* by only 4 out of 1700 nucleotides; there are no amino acid changes (27).

Possible mechanism of transduction of c-ros.

Comparison of the nucleotide sequences of viral and cellular *ros* suggests the viral *ros* and Δ*gag* junction in UR2 was formed by splicing (Fig. 32). This is consistent with insertion of the provirus upstream from the region of c-*ros* corresponding to the beginning of v-*ros*, followed by transcription of the hybrid DNA and splicing of the transcribed RNA to join Δ*gag* to the *ros* sequence. Since the beginning of v-*ros* corresponds to an excellent splice acceptor site in c-*ros*, it is more logical to assume the Δ*gag-ros* junction has been formed by splicing. However, the possibility exists that the junction was formed by fortuitous recombination at that position between the virus and c-*ros* either at the DNA level or through an RNA intermediate. Although the UR2AV sequence at the junction corresponds to a potential splice donor site, this site could be generated by joining the viral sequence to c-*ros* through recombination at the DNA level.

c-*ros* and v-*ros* abruptly diverge 36 bp upstream of the v-*ros* termination codon. The open reading frame of c-*ros* continues after this divergence and may terminate 34 amino acids downstream, or, more likely,the reading frame is spliced to further 3' coding sequences using an excellent splice-donor consensus site 27 amino acids downstream of the divergence. The v-*ros* sequence 5' to the divergence was not found in the 3' c-*ros* sequences in the lambda clone or in helper

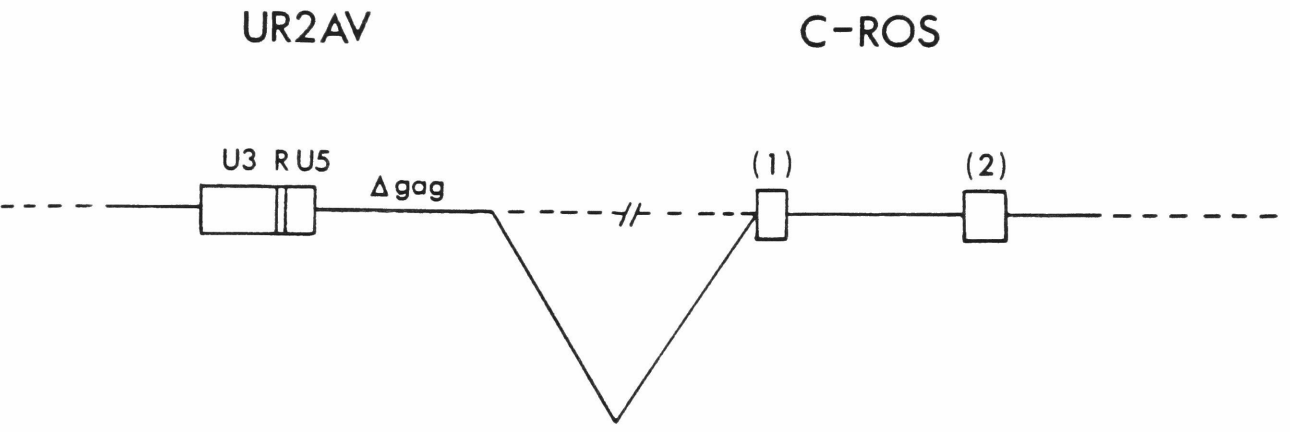


Fig. 32. Mechanism for the generation of the 5' v-*ros* junction with UR2AV. c-*ros* exons are denoted as open boxes. The splice is indicated by the "V" joining UR2AV and c-*ros*.

virus-related sequences. It may represent further downstream sequences in c-*ros* that were joined to the upstream *ros* sequence via an abnormal splicing of the virus-c-*ros* hybrid RNA, or via a recombinational event at the DNA level that deleted the c-*ros* sequence in between. This sequence may also represent another locus fused to v-*ros* by a recombinational event subsequent to the transduction of c-*ros*. Several viral oncogenes and their cellular counterparts are discontinuous at their 3' ends. The carboxy-termini of viral and cellular *fos* differ due to a 104 bp deletion in the 3' v-*fos* sequence (77). The 3' end of v-*myb* is contained within c-*myb*, which continues in frame a short distance before termination, or which, like c-*ros*, may use a splice-donor signal that lies just upstream of the stop codon to extend the coding sequence (30). The carboxy-terminal 12 amino acids of viral *src* are derived from chicken genomic sequences found 1 kb downstream of the c-*src* termination codon (49). The interposing sequence in c-*src* contains a poor splice donor consensus, and this sequence may be fused to the 3' end of viral *src* by an infrequent splicing event, a DNA:DNA recombination (70), or an RNA:RNA recombination during the process of reverse transcription (49). This situation is analagous to that of c-*ros*, although the origin of the 3' v-*ros* divergent sequence is not yet clear.

The exact mechanism of transduction of an oncogene by a retrovirus remains to be further explored, but the accumulating evidence, including the transduction of c-*ros*, suggests that it involves nonhomologous recombination between the virus and the cellular gene, occurring in one or multiple steps at the DNA level followed by splicing of the viral-c-*onc* RNA transcript. The final junctions may not reflect the original site of recombination. Formation of the 3' viral-c-*ros* junction would involve a second recombination between UR2AV and the viral-c-*ros* hybrid RNA at the step of reverse transcription.

Although capture of c-*ros* by UR2AV has resulted in few changes within the

transduced region of the cellular gene, it is possible that truncation of the 5' sequences and/or alteration of the 3' end may have activated the transforming potential of c-*ros*. Additionally, fusion of Δp19 to c-*ros* could affect the conformation of the resulting viral protein. Prywes et. al. (53) have shown that the p15 region of *gag* in Abelson murine leukemia virus is required to stabilize the transforming protein in lymphoid cells. However, fusion to *gag* sequences is dispensable for transformation; although the *myc* transforming protein of MC29 is a *gag*-fusion product, *myc* is independently expressed in MH2. In addition, Δ*gag* is not necessary for transformation by v-*fps* (18). It is entirely possible, however, that transformation by v-*ros* is a consequence of the 5' and/or 3' alterations to the cellular gene.

Expression of c-ros.

Many of the cellular oncogenes are expressed during prenatal and early postnatal development and show patterns of stage and tissue specificity unique for each gene. However, c-*mos* (the cellular counterpart of the viral transforming gene of Molony murine sarcoma virus) transcripts have yet to be identified in normal tissues or cell lines although induction of c-*mos* has been reported in neoplastic cells resulting from a genomic rearrangement that substituted an LTR element for 5' *mos* sequences (for review see Muller and Verma, 43). c-*mos* sequences are capable of transforming NIH-3T3 cells when linked to a viral LTR. These observations lend credence to the hypothesis that transformation by the viral gene is a result of its heightened expression. A 3.1 kb c-*ros* transcript has been detected in adult muscle tissue and in kidney from chickens only after long exposure of the Northern filters. It appears that the c-*ros* transcript in kidney is preferentially degraded relative to *src*, and this may account for the seeming lack of expression in this tissue. However, in all other tissues, and in several cell lines,

c-*ros* transcripts are not detectable. Therefore, it is entirely possible that the high level of expression of viral *ros* in infected cells may be a causitive factor in effecting transformation. Alternatively, the increased stability of the viral message may be responsible for its neoplastic effects.

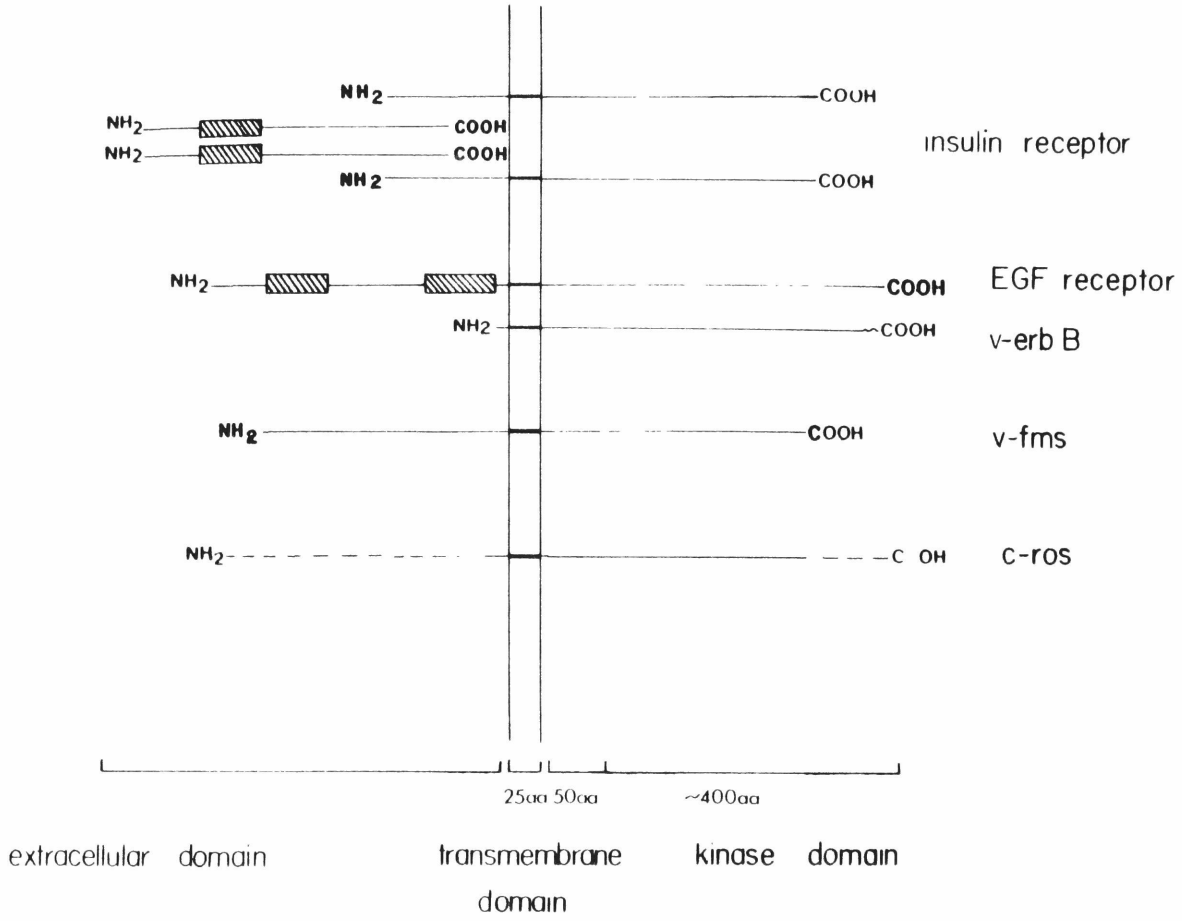
Function of cellular ros in normal cells.

The transcript detected for c-*ros* is 3.1 kb, and the portion of c-*ros* transduced by UR2AV is 1.2 kb of coding sequence. There is insufficient data to speculate how the remaining 1.9 kb of the transcribed sequences would distribute among the 5' and 3' ends of the gene.

Comparison of the deduced amino acid sequence of viral and cellular *ros* indicates structural and possibly functional similarities with both the EGF and insulin receptors. The transmembrane and cytoplasmic domains of the EGF receptor appear to have been transduced by a retrovirus to form v-*erb*B (14, 75). The v-*fms* protein product (of McDonough feline sarcoma virus), like v-*erb*B, is glycosylated, membrane associated (presumably through the hydrophobic domains of the proteins), and shares striking homology with the family of tyrosine protein kinases (23, 56). Recent reports (19, 56) have shown tyrosine kinase activity to be associated with the v-*erb*B and v-*fms* protein products.

P68 has been shown to be an integral membrane protein (Ellen Garber, unpublished), presumably due to the 29 amino acid amino-terminal hydrophobic stretch; the protein is not glycosylated (Ellen Garber, unpublished). v-*ros* shares greater homology in the conserved kinase domain with the insulin receptor than any of the other tyrosine protein kinase-encoding oncogenes (76). Comparison of the structural domains of the EGF and insulin receptors with v-*erb*B, v-*fms*, and the region of cellular *ros* sequenced thus far leads to the tentative conclusion that the protein product of c-*ros* is a member of a family of growth factor receptors

Fig. 33. Family of growth factor receptors. Cysteine-rich residues found in the extracellular domains of the proteins are shown by hatched boxes. A heavier solid line shows the region of the protein spanning the membrane. The transmembrane domain is 25 residues in length.


NH2
COOH
insulin receptor
EGF receptor
v-erb B
v-fms
c-ros
25aa 50aa
~400aa
extracellular domain
transmembrane domain
kinase domain

(Fig. 33).

The transmembrane domain in c-*ros* and v-*fms* is 26 residues in length compared to 23 for the EGF and insulin receptors. On the cytoplasmic side of the membrane, the insulin and EGF receptors have 3 basic amino acids (Arg-Lys-Arg and Arg-Arg-Arg, respectively). v-*fms* is marked by Lys-Tyr-Lys-Gln-Lys and c-*ros* by Gln-Arg-Trp-Lys, both of which are highly basic sequences of the kind expected to occur at the cytoplasmic junction of the membrane domain. The insulin and EGF receptors contain a Pro-Ser flanking the extracellular side of the membrane; in c-*ros*, a Ser-Pro sequence is present. The tyrosine kinase domains of v-*fms* and the EGF and insulin receptors are located 50 amino acids downstream of the transmembrane domain; in *ros*, the kinase domain begins 68 residues downstream.

The 5' domain of c-*ros* has not been characterized. In the EGF and isulin receptors, the transmembrane domain is located within the middle of the protein, and the 5' extracellular domain is marked by cysteine-rich regions. It would be interesting to see whether this is true for cellular *ros*.

The EGF receptor is not oncogenic, yet v-*erb*B has transforming ability. Their sequences are highly conserved except for the missing 5' extracellular domain in v-*erb*B, and replacement of some of the 3' cellular sequences with a few amino acids of viral *env*. Perhaps truncation of the gene activates its oncogenic potential, and this might be analagous to v-*ros*, which has only 7 amino acid residues upstream of the transmembrane domain before fusion with viral *gag* sequences.

P68 is a *gag*-fusion product. It is not known whether the *gag* moiety is essential for transformation by *ros*. It is conceivable that replacement of the extracellular domain of the c-*ros* protein by Δp19 (and perhaps alteration of the carboxy-terminus) may dramatically alter the receptor's specificity of interaction

with the growth factor.

P68 has been proposed to be capable of phosphorylating phosphatidylinositol to phosphatidylinositol-4-phosphate (4a), resulting in an increase in diacylglycerol. Diacylglycerol activates C kinase, and activation of this protein may be linked to mitogenesis. C kinase has been strongly implicated in a receptor-mediated cell-signalling cascade. If v-*ros* is an altered receptor, it is possible that its oncogenic potential results from an acquired ability to phosphorylate, and thus activate, components of this cascade. However, recent results from our laboratory (Sumio Sugano and Hidesaburo Hanafusa, Mol. Cell Biol., in press) suggests that lipid kinase activity is not intrinsic to P68, and that the previously described phosphatidylinositol kinase activity is due to contaminating cellular kinases.

Although the exact function of the cellular *ros* protein is not known, evidence based on the sequence of the viral and cellular genes has enabled us to detect its structural similarity with the EGF and insulin receptors and postulate that *ros* is a member of a family of growth factor receptors.

Section VI. References

1. Ahmed, C., R. Chanda, N. Stow, B. Zain. 1982. The nucleotide sequence of mRNA for the M_r 19000 glycoprotein from early gene block III of Adenovirus 2. Gene **20**:339-346.

2. Antler, A., M. E. Greenberg, G. M. Edelman and H. Hanafusa. 1984. Increased phosphorylation of tyrosine in vinculin does not occur upon transformation by some avian sarcoma viruses. Mol. Cell. Biol. **5**:263-267.

3. Appleyard, R. K. 1954. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. Genetics **39**:440-552.

4. Balduzzi, P. C., M. F. D. Notter, H. R. Morgan, and M. Shibuya. 1981. Some biological properties of two new avian sarcoma viruses. J. Virol. **40**:268-275.

4a. Macara, I. G., G. V. Marinetti, and P. C. Balduzzi. 1984. Transforming protein of avian sarcoma virus UR2 is associated with phosphatidylinositol kinase activity: Possible role in tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. **81**:2728-2732.

5. Barker, W. C. and M. O. Dayhoff. 1982. Viral *src* gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. U. S. A. **79**:2836-2839.

6. Benton, W. D. and R. W. David. 1977. Screening gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180-182.

7. Bizub, D., R. A. Katz and A. M. Skalka. 1984. Nucleotide sequence of non-coding regions in Rous-associated virus-2: comparisons delineate conserved

regions important in replication and oncogenesis. J. Virol. **49**:557-565.

8. Bolivar, F. R., L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95-113.

9. Brugge, J. and D. Darrow. 1984. Analysis of the catalytic domain of phosphotransferase activity of two avian sarcoma virus transforming proteins. J. B. C. **259**:4550-4557.

10. Cooper, J. A. and T. Hunter in **Current Topics in Microbiology and Immunology**. 1983. **107**:125-161.

11. Czernilofsky, A. P., A. Levinson, H. Varmus, J. Bishop, E. Tischer and H. Goodman. 1983. Corrections to the nucleotide sequence of the *src* gene of Rous sarcoma virus. Nature **301**:736.

12. Czernilofsky, A. P., A. Levinson, H. Varmus, J. Bishop, E. Tischer and H. Goodman. 1980. Nucleotide sequence of an avian sarcoma virus oncogene (*src*) and proposed amino acid sequence for gene product. Nature **287**:198-203.

13. Delorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus and J. M. Bishop. 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**: 50-61.

14. Downward, J. et. al. 1984. Close similarity of EGF receptor and v-*erb*B oncogene protein sequences. Nature **307**: 521-527.

15. Dodgson, J. B., J. Strommer and J. D. Engel. 1979. Isolation of the chicken β-globin gene from a chicken DNA recombinant library. Cell **17**:879-887.

16. Feldman, R. A., T. Hanafusa, and H. Hanafusa. 1980. Characterization of

protein kinase activity associated with the transforming gene product of Fujinami sarcoma virus. Cell **22**: 757-765.

17. Feldman, R. A., L.-H. Wang, H. Hanafusa, and P. C. Balduzzi. 1982. Avian sarcoma virus UR2 encodes a transforming protein which is associated with a unique protein kinase activity. J. Virol. **42**:228-236.

18. Foster, D. and H. Hanafusa. 1983. A *fps* gene without *gag* sequences transforms cells in culture and induces tumors in chickens. J. Virol. **48**:744-751.

19. Gilmore, T., J. DeClue and G. Martin. 1985. Protein phosphorylation at tyrosine is induced by the v-*erb*B gene product *in vivo* and *in vitro*. Cell **40**:609-618.

20. Gingeras, T., D. Sciaky, R. Gelinas, J. Bing-Dong, C. Yen, M. Kelly, P. Bullock, B. Parsons, K. O'Neill and R. Roberts. 1982. Nucleotides sequences from the Adenovirus-2 genome. J. Biol. Chem. **257**:13475-13491.

21. Gopinathan et. al. 1979. Mutagenesis *in vitro* by DNA polymerase from an RNA tumor virus. Nature **278**:857-858.

22. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456-467.

23. Hampe, A., M. Gobet, C. J. Sherr, and F. Galibert. 1984. Nucleotide sequence of the feline retroviral oncogene v-*fms* shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. Proc. Natl. Acad. Sci. U. S. A. **81**:85-89.

24. Hanafusa, T., L.-H. Wang, S. M. Anderson, R. E. Karess, W. S. Hayward and H. Hanafusa. 1980. Characterization of the transforming gene of Fujinami sarcoma virus. Proc. Natl. Acad. Sci. U. S. A. **77**:3009-3013.

25. Hirt, B. 1967. Selective extraction of Polyoma DNA from infected mouse cell cultures. J. M. B. **26**:365-369.

26. Hopp, T. P. and K. R. Woods. 1982. Prediction of antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U. S. A. **78**:3824-3828.

27. Huang et. al. 1985. Nucleotide sequence and topography of chicken c-*fps*: genesis of a retroviral oncogene encoding a tyrosine specific protein kinase. J. M. B. **181**:175-186.

28. Jung, A., A. Sippel, M. Grez and G. Schultz. 1980. Exons encode functional and structural units of chicken lysozyme. Proc. Natl. Acad. Sci. **77**:5759-5763.

29. Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida. 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. Nature **297**:205-208.

30. Klempnauer, K.-H., T. Gonda and J. M. Bishop. 1982. Nucleotide sequence of the retroviral leukemia gene v-*myb* and its cellular progenitor c-*myb*: the architecture of a transduced oncogene. Cell **31**:453-463.

31. Land, H., L. F. Parada, and R. A. Weinberg. 1983. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature **304**:596-602.

32. Leder, P., D. Tiemeier, and L. Enquist. 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. Science **196**:175-177.

33. Lee, W.-H. et. al. 1980. FSV: an avian RNA tumor virus with a unique transforming gene. Proc. Natl. Acad. SCi. U. S. A. **77**:2018-2022.

34. Levinson, A. D., S. A. Courtneidge, and J. M. Bishop. 1981. Structural and functional domains of the Rous sarcoma virus transforming protein ($p60^{src}$). Proc. Nat. Acad. Sci. USA **78**:1624-1628.

35. Levitt, M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. **104**:59-107.

36. Maniatis, T., R. C. Hardison, E. Lacy, J. Cauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis. 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**:687-701.

37. Mark, G. E. and U. R. Rapp. 1984. Primary structure of v-*raf*: relatedness to the *src* family of oncogenes. Science **224**:224-226.

38. Mathey-Prevot, B., H. Hanafusa, and S. Kawai. 1982. A cellular protein is immunologically cross-reactive with and functionally homologous to the Fujinami sarcoma virus transforming protein. Cell **28**:897-906.

39. Maxam, A. M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-560.

40. McLauchlan, J., D. Gaffney, J. L. Whitton and J. B. Clements. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. N. A. R. **13**:1347-1368.

41. Messing, J. and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digested restriction fragments. Gene **19**:269-276.

42. Moscovici, C., M. G. Moscovici, H. Jimenez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt. 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell **11**:95-103.

43. Muller, R. and I. M. Verma in **Current Topics in Microbiology and Immunology**. 1984. **112**:73-115.

44. Murray, N. E., W. J. Brammar, and K. Murray. 1977. Lamboid phages that simplify the recovery of *in vitro* recomninants. Molec. gen. Genet. **150**:53-61.

45. Naharro, G., K. C. Robbins and E. P. Reddy. 1984. Gene product of v-*fgr onc*: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. Science **223**:63-66.

46. Neckameyer, W. S. and L.-H. Wang. 1984. Molecular cloning and characterization of avian sarcoma virus UR2 and comparison of its transforming sequence with those of other avian sarcoma viruses. J. Virol. **50**:914-921.

47. Notter, M. F. D. and P. C. Balduzzi. 1984. Cytoskeletal changes induced by two avian sarcoma viruses: UR2 and Rous sarcoma virus. Virology **136**:56-68.

48. Panganiban, A. T. and H. M. Temin. 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. Nature **306**:155-160.

49. Parker, R. C., R. Swanstrom, H. E. Varmus and J. M. Bishop in **Cancer Cells**. 1984. **2**:19-25.

50. Patschinsky, T., T. Hunter, F. Esch, J. Cooper and B. Sefton. 1982. Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation. Proc. Natl. Acad. Sci. U. S. A. **79**:973-977.

51. Payne, G. S., S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop and H. E. Varmus. 1981. Analysis of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for maintenance of the tumor

state. Cell **23**: 311-322.

52. Prywes, R., J. G. Foulkes, N. Rosenberg, and D. Baltimore. 1983. Sequences of the A-MuLV protein needed for fibroblast and lymphoid cell transformation. Cell **34**:569-579.

53. Prywes, R., J. Hoag, N. Rosenberg and D. Baltimore. 1985. Protein stabilization explains the *gag* requirement for transformation of lymphoid cells by Abelson MuLV. J. Vir. **53**:123-132.

54. Reddy, E. P., M. J. Smith, and A. Srinivasan. 1983. Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene to other *onc* gene products with tyrosine-specific kinase activity. Proc. Natl. Acad. Sci. USA **80**:3623-3627.

55. Sanger, F., S. Nicklen and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. **74**:5463-5467.

56. Scherr, C. J., S. J. Anderson, C. W. Rettinmier and M. F. Roussel in **Cancer Cells**. 1984. **2**:329-338.

57. Schwartz, D. E., R. Tizard and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853-869.

58. Sefton, B. M., T. Hunter, K. Beemon and W. Eckhart. 1980. Evidence that the phosphorylation of tyrosine is essential for cellular transformation by RSV. Cell **20**:807-816.

59. Shibuya, M., T. Hanafusa, H. Hanafusa and J. R. STephenson. 1980. Homology exists among the transforming sequences of avian and feline sarcoma viruses. Proc. Natl. Acad. Sci. U. S. A. **77**:6536-6540.

60. Shibuya, M. and H. Hanafusa. 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **30**: 787-795.

61. Shibuya, M., H. Hanafusa, and P. C. Balduzzi. 1982. Cellular sequences related to three new *onc* genes of avian sarcoma virus (*fps*, *yes*, and *ros*) and their expression in normal and transformed cells. J. Virol. **42**:143-152.

62. Shibuya, M., L.-H. Wang, and H. Hanafusa. 1982. Molecular cloning of the Fujinami sarcoma virus genome and its comparison with sequences of other related transforming viruses. J. Virol. **42**:1007-1016.

63. Shoji, S., D. Parmelee, R. Wade, S. Kumar, L. Ericsson, K. Walsh, H. Neurath, G. Long, J. Demaille, E. Fischer and K. Titani. 1981. Complete amino acid sequence of the catalytic subunit of bovine cardiac muscle cyclic AMP-dependent protein kinase. Proc. Natl. Acad. Sci. U. S. A. **78**:848-851.

64. Sorge, J., W. Ricci and S. Hughes. 1983. *cis*-Acting RNA packaging locus in the 115 nucleotide direct repeat of Rous sarcoma virus. J. Virol.**48**:667-675.

65. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503-517.

66. Stehelin, D., H. E. Varmus, J. M. Bishop and P. K. Vogt. 1976. DNA related to the transforming gene(s) of ASV is present in normal avian DNA. Nature **260**:170-173.

67. Swanstrom, R., W. J. DeLorbe, J. M. Bishop and H. E. Varmus. 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U. S. A. **78**:124-128.

68. Takeya, T., H. Hanafusa, R. P. Junghans, G. Ju, and A. M. Skalka. 1981. Comparison between the viral transforming gene (*src*) of recovered avian sarcoma virus and its cellular homolog. Molec. and Cell Biol. **1**:1024-1037.

70. Takeya, T. and H. Hanafusa. 1983. Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus. Cell **32**:881-890.

71. Tamura, T., M. Noda and T. Takano. 1981. Structure of the baboon endogenous virus genome: nucleotide sequence of the long terminal repeat. N. A. R. **9**:6615-6626.

72. Taylor, J. M., R. Illmensee, and J. Summers. 1976. Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. Biochim. Biophys. Acta **442**:324-330.

73. Tosi, M., R. Young, O. Hagenbuchle, U. Schibler. 1981. Multiple polyadenylation sites in a mouse α-amylase gene. N. A. R. **9**:2313-2323.

74. Ucker, D. A., G. Firestone and K. Yamamoto. 1983. Glucocorticoids and chromosomal position modulate mouse mammary tumor virus transcription by affecting efficiency of promoter utilization. Mol. Cell. Biol. **3**:551-561.

75. Ullrich, A. et. al. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature **309**:418-425.

76. Ullrich, A. et. al. 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature **313**:756-761.

77. Verma, I. et. al. in **Cancer Cells**. 1984. **2**:309-321.

78. Verma, I. 1984. From c-*fos* to v-*fos* Nature **303**:317.

79. Wang, L.-H., and P. Duesberg. 1974. Properties and location of poly(A) in Rous sarcoma virus RNA. J. Virol. **14**:1515-1529.

80. Wang, L.-H., P. Duesberg, K. Beemon, and P. K. Vogt. 1975. Mapping RNase T_1-resistant oligonucleotides of avian tumor virus RNAs: sarcoma-specific oligonucleotides are near the poly(A) end and oligonucleotides common to sarcoma and transformation-defective viruses are at the poly(A) end. J. Virol. **416**:1051- 1070.

81. Wang, L.-H, R. Feldman, M. Shibuya, H. Hanafusa, M. F. D. Notter, and P. C. Balduzzi. 1981. Genetic structure, transforming sequence, and gene product of avian sarcoma virus UR1. J. Virol. **40**: 258-267.

82. Wang, L.-H., H. Hanafusa, M. F. D. Notter, and P. C. Balduzzi. 1982. Genetic structure and transforming sequence of avian sarcoma virus URB. J. Virol. *41*:833-841.

83. Wang, L.-H., M. Beckson, S. M. Anderson, and H. Hanafusa. 1984. Identification of the viral sequence required for the generation of recovered avian sarcoma viruses (rASVs) and characterization of a series of replication-defective rASVs. J. Virol. **49**:881-891.

85. Wang, L.-H., B. Edelstein, and B. J. Mayer. 1984. Induction of tumors and generation of recovered sarcoma viruses from two molecularly cloned **src**-deletion mutants and mapping of the deletions in these mutants. J. Virol. **50**:904-913.

86. Weinmaster, G., E. Hinze, and T. Pawson. 1983. Mapping of multiple phosphorylation sites within the structural and catalytic domains of the Fujinami avian sarcoma virus transforming proteins. J. Virol. **46**:29-41.

87. Yamamoto, T., T. Nishida, N. Miyajima, S. Kawai, T. Ooi and K. Toyoshima. P983. The *erb*B gene of avian erythroblastosis virus is a member of the *src* gene family. Cell **35**:71-78.

88. Zasloff, M., G. Ginder and G. Felsenfeld. 1978. A new method for the purification and identification of covalently closed circular DNA molecules. N. A. R. **5**:1139-1152.

End