Current Topics in Microbiology and Immunology

103

Retroviruses 1

Edited by P.K. Vogt and H. Koprowski

Current Topics in Microbiology and Immunology

103

Prof. Dr. P.K. Vogt
University of Southern California
School of Medicine
Dept. of Microbiology
2011 Zonal Avenue HMR 401
Los Angeles, California 90033
U.S.A.

Prof. Dr. Hilary Koprowski
The Wistar Institute
36th Street at Spruce
Philadelphia, PA 19104
U.S.A.

ISBN 3-540-12167-6 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-12167-6 Springer-Verlag New York Heidelberg Berlin

Library of Congress Catalog Card Number 15-12910
Printed in Germany

Typesetting: Fotosatz Service Weihrauch, Würzburg.
Printing and binding: Universitätsdruckerei H. Stürtz AG, Würzburg.
2121/3321-543210

Retroviruses 1

Edited by P.K. Vogt and H. Koprowski

With 16 Figures

Springer-Verlag
Berlin Heidelberg New York 1983

Table of Contents

Indexed in Current Contents

Avian Endogenous Viral Genes

UGO G. ROVIGATTI* AND SUSAN M. ASTRIN*

1 Introduction

Endogenous viruses are defined as germline genes that code for the components of a retrovirus. These genes are present in all cells of all tissues of an animal and are inherited by progeny in a Mendelian manner. Two general methods have been used to detect the presence of endogenous viral sequences in DNA of a given species. First, sequence homology with the genome of a known retrovirus can be demonstrated for the chromosomal DNA. Second, production of viral particles or viral components can be demonstrated for cells of the species. The first evidence for the existence of endogenous viruses came from studies on spontaneous leukemia in the murine system (for review see *Gross* 1958a). There it was also shown that in lymphoid tumors induced by X rays, a murine leukemia virus was produced which caused similar tumors when injected into unirradiated mice (*Gross* 1958b; *Lieberman* and *Kaplan* 1959). Endogenous viral genes have now been shown to be essentially ubiquitous in vertebrate species including man. The purpose of this article will be to describe and discuss the current state of information with respect to the endogenous viral genes of the domestic chicken, *Gallus gallus*. We

* Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA

will discuss DNA structure, gene products, and some current ideas about the relationship between endogenous viral genes and disease.

An avian retroviral particle contains an RNA genome composed of two indentical 8-kilobase subunits, which are complexed with reverse transcriptase (a 154000-dalton protein composed of α and β subunits of 92000 and 62000 daltons, respectively). Four other proteins, termed group-specific antigens (gs), are internal to the viral particle (for reviews see *Vogt* and *Hu* 1977; *Eisenman* and *Vogt* 1978). The molecular weights of the four gs antigens are 27000, 19000, 15000, and 12000 daltons, and they are processed from a precursor of 76000 daltons. The viral envelope contains glycoproteins of 85000 daltons and 37000 daltons. In the avian system the envelope glycoprotein, which is the product of endogenous viral genes, is often termed chf or chick helper factor because production of envelope protein by cells complements or helps the production of the envelope-defective Bryan strain of Rous sarcoma virus (*Weiss* 1969; *H. Hanafusa* et al. 1970).

Upon infection of a susceptible cell, the viral RNA is transcribed into a double-stranded DNA provirus, which integrates into the host genome and serves as a template for the synthesis of messenger RNA and RNA to be encapsulated into virions. When such proviral sequences are present in DNA of the germ cells of a species, they are termed endogenous viral genes.

Endogenous viral genes were first demonstrated to be present in chicken cells in the 60s. A complement-fixation assay showed gs antigens to be present in uninfected cells from chicken embryos (*Dougherty* and *DiStefano* 1966). Subsequently, gs expression was shown to segregate as an autosomal dominant gene in crosses between a line which produced the gs antigen and a line which lacked the antigen (*Payne* and *Chubb* 1968). Another endogenous viral gene product, viral envelope or chf, was also identified in embryo cells and could be rescued by infection and recombination with sarcoma or leukosis viruses (*H. Hanafusa* et al. 1970; *Weiss* and *Payne* 1971). Both gs and chf antigen appeared to be present in several different flocks of white leghorn chickens and also in wild jungle fowls (*Weiss* and *Biggs* 1972). It was proposed that these antigens were the products of defective viral genomes which resided in the cellular DNA. However, it soon became clear that cells could also carry intact viral genomes, and that, in certain instances, viral particles could be released spontaneously or after induction by chemical and physical agents. A virus with a distinctive envelope and host range was shown to be the product of endogenous viral genes (*T. Hanafusa* et al. 1970; *Vogt* and *Friis* 1971). This virus was called RAV-0 and later shown to be encoded by a specific genetic locus, *ev 2* (*Astrin* et al. 1980a). A similar virus could be induced from normal chicken cells after treatment with ionizing radiation or chemical carcinogens or mutagens (*Weiss* et al. 1971).

Additional evidence for the presence of endogenous viral genes came from biochemical data that indicated several copies of viral DNA were present in the cellular genomes of cells expressing gs antigens (gs^+ cells), as well as of cells lacking these antigens (gs^- cells) (*Rosenthal* et al. 1971; *Baluda* 1972; *Varmus* et al. 1972; *Neiman* 1973). However, RNAs specific for the gs and chf antigens were present only in gs^+ chf^+ cells, suggesting a transcriptional regulation over the expression of these genes (*Hayward* and *Hanafusa* 1973; *Wang* et al. 1977). However, gs and chf were not always coordinately regulated (*T. Hanafusa* et al. 1972). Another viral function, which does not appear to be expressed in a coordinate fashion with other viral genes, is the reverse transcriptase (*Weissbach* et al. 1972; *Eisenman* et al. 1978). The genetics of the induction and spread of RAV-0 virus have been studied extensively (*Crittenden* et al. 1974, 1977). RAV-0 propaga-

tion and horizontal transmission are complicated by the fact that the cells of only a very few lines of chickens are susceptible to its infection. The possibility of control by two different loci, one specifying the receptors for RAV-0 and the other acting as an epistatic inhibitor of its infectivity, has been studied (*Crittenden* et al. 1974), and recently elucidated (*Robinson* et al. 1981). It had been speculated that the genes controlling expression of gs, chf, and RAV-0 production were regulatory, rather than structural genes (for review see *Tooze* 1973). That this is not the case has been shown only recently (*Astrin* 1978; *Astrin* et al. 1980b).

2 The *ev* Loci

With the advent of restriction endonucleases and the Southern blot (*Southern* 1975), it first became possible to look at the structure of a set of unique sequences in a eucaryotic genome, providing one had a probe. Using radiolabeled avian leukosis or sarcoma virus genomic RNA or radiolabeled cDNA (complementary DNA made by reverse transcription), several groups began to look at the structure of the endogenous viral sequences in chickens of different phenotypes. It was immediately obvious that the situation was very complicated. The standard enzymes gave very complex patterns, often containing five or more bands, even with DNA from gs$^-$ chf$^-$ chickens, and there did not appear to be obvious correlations of pattern with phenotype. The situation was eventually elucidated by a combination of both a biochemical and a genetic approach. Using several restriction enzymes to screen a group of more than 150 birds of various phenotypes, *Astrin* (1978) clearly demonstrated that certain gs$^-$chf$^-$ birds gave the most simple pattern of bands (three bands with *Eco*RI, *Bam*HI, or *Hind*III). In addition, these bands were included in the pattern of all the other birds, and it was clear from the molecular weights that these fragments represented internal fragments of the viral genome, as well as junction fragments containing viral and cellular sequences. On the assumption that the DNA yielding the simple three-band pattern contained only a single endogenous provirus, more than 20 enzymes were screened and an enzyme, *Sac*I, was found which gave but a single band with the test DNA. When *Sac*I was used to cleave the other DNAs, simplified patterns emerged although, in most cases, multiple bands were still present. It was not clear whether each and every band represented an individual provirus, or whether some proviruses yield more than one fragment. However, it was possible at this stage to correlate individual bands with certain phenotypes, and it was also clear that virtually all the DNAs tested contained a common proviral element (*Astrin* 1978).

Further clarification of the situation required a genetic approach. Matings were set up in which homozygous gs$^-$chf$^-$ birds were mated with birds homozygous for the gs$^+$chf$^+$, V$^+$, or gs$^-$chf$^+$ phenotypes, and the F_1 progeny were backcrossed to the gs$^-$chf$^-$ parent. The phenotypes and genotypes (DNA restriction patterns) of the parents' F_1 progeny and backcross progeny were determined (*Astrin* et al. 1980b). From these experiments several major conclusions were possible. First, it was apparent that each *Sac*I fragment represented a separate genetic locus for endogenous viral sequences. In addition, there was an unexpectedly large number (> 10) of such loci, and one locus, designated *ev* 1, was common to all the white leghorns tested. Finally, it was possible to get an absolute correlation between segregation of a particular band (endogenous virus locus) and segregation of a particular phenotype. Thus, a good case could be made for the

Table 1. Distribution of endogenous retroviral loci in white leghorn chickens

Chicken flock	No. of birds tested	% of birds containing specific *ev* loci											
		*ev*1	*ev*2	*ev*3	*ev*4	*ev*5	*ev*6	*ev*7	*ev*8	*ev*9	*ev*10	*ev*11	*ev*12
Inbred lines													
6_3	24[a]	100	0	100	0	0	0	0	0	0	0	0	0
7_2	26[a]	100	100	0	0	0	0	0	0	0	0	0	0
15_B	60[a]	100	0	0	0	0	0	100	0	0	0	0	0
15_1	15	100	0	100	0	0	100	0	0	0	0	7	100
C	11	100	0	0	0	0	0	100[b]	0	0	100[b]	0	0
15_{14}	12	100	0	0	0	0	89	0	0	0	100	100	0
15_{15}	10	100	0	0	0	0	100	0	0	0	90	20	0
Noninbred lines													
SPAFAS line 11 gs^-chf^-	200	99	0	0	56	25	0	0	ND[c]	0	0	0	0
SPAFAS line 11 gs^-chf^+	10	100	0	100	30	70	20	0	0	0	0	0	0
SPAFAS line 11 gs^-chf^+	6	100	0	0	83	0	33	0	17	67	0	0	0
Heisdorf and Nelson gs^+chf^+	2	100	0	100	0	0	0	0	0	0	0	0	0
Heisdorf and Nelson gs^-chf^-	3	100	0	0	33	67	0	0	0	0	0	0	0
K(−)	34[a]	100	0	0	12	12	0	0	0	0	0	0	0
K16	32[a]	100	0	100	38	0	25	0	ND	0	0	0	0
K18	24[a]	100	0	0	71	0	100	0	79	58	0	0	0
K28	39[a]	100	0	0	0	0	0	0	0	0	0	0	0

[a] Birds chosen so as to completely represent the gene in the breeding stock of these lines;
[b] Identified by *Sac*I and *Bam*HI digestion, not confirmed by genetic analysis;
[c] ND, not determined, additional restriction endonuclease analysis required for identification

Table 2. Associated phenotypes and identifying restriction fragments for 12 *ev* loci

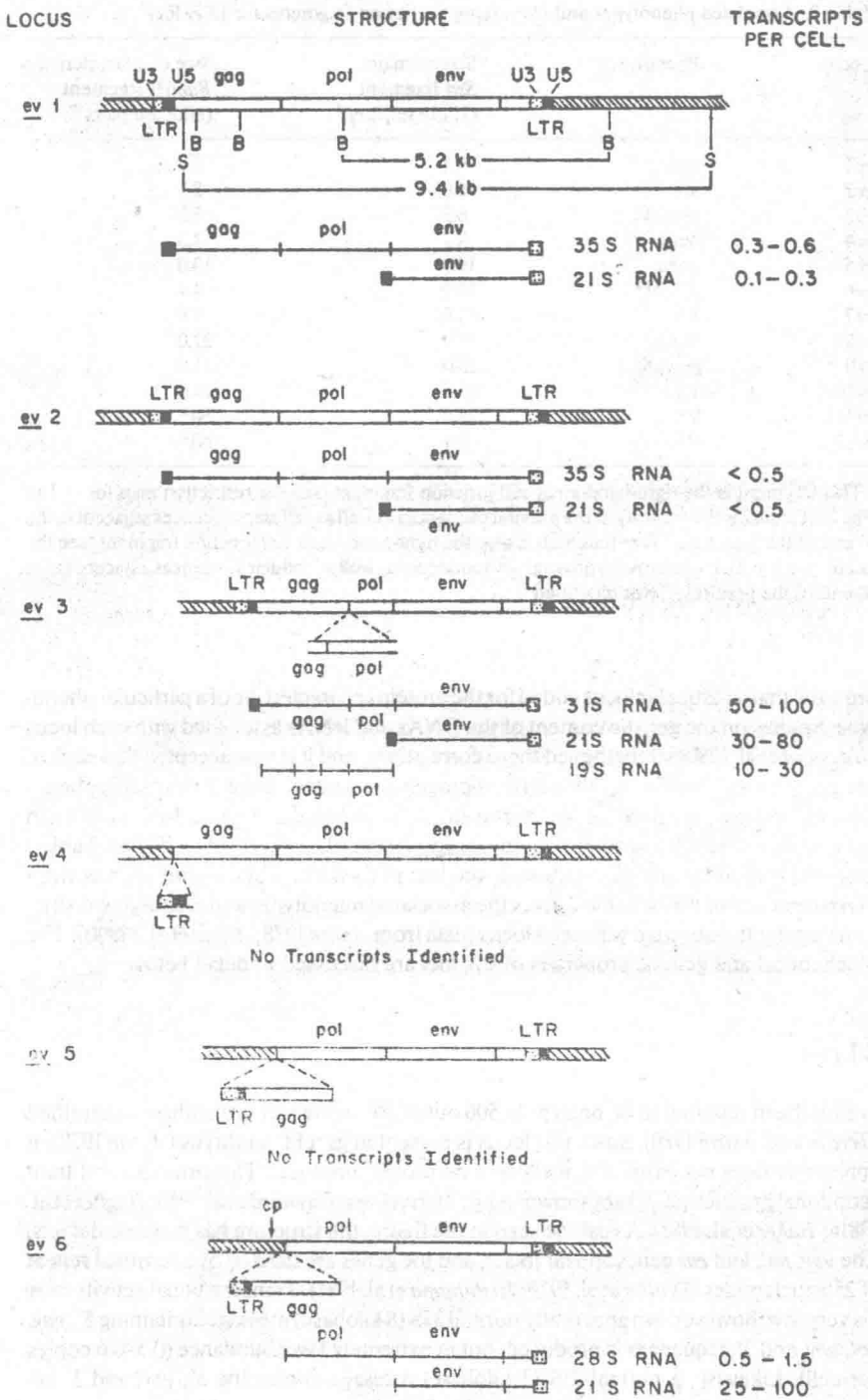
Locus	Phenotype	Size of major *Ssi*I fragment (kilobase pairs)[a]	Size of characteristic *Bam*HI fragment (kilobase pairs)[b]
ev 1	none	9.4	5.2
ev 2	V+	6.0	8.2
ev 3	gs+chf+	6.3	7.3
ev 4	none	8.7	7.3
ev 5	none	19.0	13.0
ev 6	gs−chf+	21.0	4.4
ev 7	V+	13.0	7.6
ev 8	none	18.0	23.0
ev 9	gs−chf+	23.0	11.0
ev 10	V+	21.0	14.0
ev 11	V+	13.0	NI[c]
ev 12	V+	8.1	NI[c]

[a] This fragment is the right-hand virus cell junction fragment (see the restriction map for *ev* 1 in Fig. 1). It contains the majority of the proviral sequences as well as cellular sequences adjacent to the 3′ end of the provirus; [b] This fragment is also the right-hand virus cell junction fragment (see the *ev* 1 map in Fig. 1). It contains the proviral *env* sequences as well as cellular sequences adjacent to the 3′ end of the provirus; [c] Not identified

proposal that a particular locus coded for the proteins characteristic of a particular phenotype. Studies on the genetic content of the DNAs and RNAs associated with each locus (*Hayward* et al. 1980) strengthened these correlations, and it is now accepted that each of the genetic loci identified by the above approach does indeed code for a specific phenotype of endogenous viral gene expression. The individual genetic loci have been designated *ev* loci and numbered sequentially (*Astrin* 1978; *Astrin* et al. 1980b). Table 1 gives the frequency and distribution of the loci in flocks of white leghorns (data from *Tereba* and *Astrin* 1980). Table 2 gives the associated phenotypes and identifying restriction fragments associated with each locus (data from *Astrin* 1978; *Astrin* et al. 1980b). The biochemical and genetic properties of the loci are discussed in detail below.

2.1 *ev* 1

ev 1 has been reported to be present in 506 out of 508 white leghorn embryos examined (*Tereba* and *Astrin* 1980). Since this locus is present in gs−chf− embryos (*Astrin* 1978), it apparently does not express detectable viral protein products. The structure and transcriptional products of *ev* 1 are shown in Fig. 1 (data from *Hayward* et al. 1980; *Hughes* et al. 1981a; *Baker* et al. 1981). As can be seen in the figure, the structure has no gross defects. The *gag*, *pol*, and *env* genes appear intact, and the genes are flanked by a terminal repeat of 250 nucleotides (*Skalka* et al. 1979; *Hishinuma* et al. 1981). Transcriptional activity of *ev* 1 is very low, however. An apparently normal 35S (8 kilobase) message containing 5′, *gag*, *pol*, *env*, and 3′ sequences is produced, but in extremely low abundance (0.3–0.6 copies per cell). Likewise, a normal 21S (3 kilobase) message containing 5′, *env*, and 3′ se-



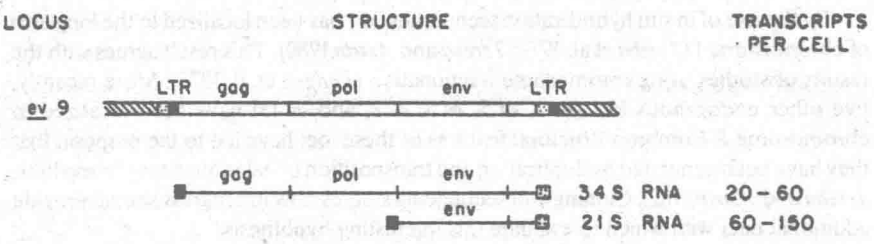


Fig. 1. Structure and transcriptional products for *ev* 1, *ev* 2, *ev* 3, *ev* 4, *ev* 5, *ev* 6, and *ev* 9. ▭, cellular sequences; ▪▪, long terminal repeat (LTR), composed of unique 3′ sequences (▪ U3) and unique 5′ sequences (▪ U5); *cp*, cellular promoter sequence. Restriction sites for *Sst* I (S) and *Bam* HI (B) are indicated for *ev* 1. The 5.2- and 9.4-kb fragments are the identifying fragments listed for *ev* 1 in Table 2

quences is also produced in small quantities (0.1–0.3 copies per cell). Both messages have been found in nuclei and cytoplasm (*Hayward* et al. 1980; *Baker* et al. 1981). Recent work strongly suggests that methylation is in part responsible for the low level of expression of *ev* 1. An embryo has been identified which spontaneously expresses *ev* 1. Cells of this embryo produce a noninfectious particle containing *gag* polypeptides but lacking reverse transcriptase and envelope antigen (*Conklin* et al. to be published). In addition, other *ev*-1-containing embryos can be induced to express such particles by treatment with 5-azacytidine, an inhibitor of DNA methylation. The induced phenotype is stable for many generations in culture. The appearance of DNase-I-hypersensitive sites in the chromatin domain of *ev* 1 is correlated with its expression after 5-azacytidine treatment (*Groudine* et al. 1981). These results imply that the association of a gs⁻chf⁻ phenotype with *ev* 1 is a result of lack of efficient expression at the RNA level, due at least in part to methylation, as well as to probable structural defects in the *pol* and *env* genes.

The *ev* 1 provirus and flanking cellular sequences have been cloned using recombinant DNA techniques (*Hishinuma* et al. 1981). Using these clones, a comparison has been made between the structure of endogenous viral sequences and the structure of an exogenously acquired provirus. As mentioned above, the basic structure of the two types of provirus is very similar. The following common features are found. First, long terminal repeats (LTRs) flank the viral genes. The repeats flanking *ev* 1 are shorter than any other LTRs studied so far – only 273 base pairs, as compared to 325–1300 base pairs for other viruses. However, the *ev* 1 LTRs contain most of the regulatory sequences known to be present in other viral LTRs, such as a $tRNA^{trp}$ primer binding site, an AT-rich region similar to the Hogness box consensus sequence 32 nucleotides before the 5′ cap site, and polyadenylation signals about 20 nucleotides upstream from a CA dinucleotide. A second common feature is that a 6-base-pair sequence, present only once in the cellular DNA lacking *ev* 1, is repeated once at either end of the provirus in DNA containing *ev* 1. Such a repeat flanks exogenously acquired proviruses, as well as procaryotic and eucaryotic transposable elements. It is by virtue of these common structural features (LTRs and a repeat of flanking host sequences) that transposable elements and endogenous and exogenous retroviruses have been proposed to share a common origin and a common integration mechanism (*Hishinuma* et al. 1981; *Ju* and *Skalka* 1980; *Shimotohno* et al. 1980; *Dhar* et al. 1980; *Majors* and *Varmus* 1981; *Roeder* et al. 1980; *Dunsmuir* et al. 1980; *Levis* et al. 1980).

By the use of in situ hybridization techniques, *ev* 1 has been localized to the long arm of chromosome 1 (*Tereba* et al. 1979; *Tereba* and *Astrin* 1980). This result agrees with the results of studies using chromosome fractionation (*Padgett* et al. 1977). More recently, five other endogenous loci (*ev* 4, *ev* 5, *ev* 6, *ev* 8, and *ev* 13) have been localized to chromosome 1. Common structural features of these loci have led to the proposal that they have been generated by duplication and transposition of *ev* 1 sequences (*Tereba* 1981; *Tereba* and *Astrin* 1982). Cloning and sequencing studies now in progress should provide additional data with which to evaluate this interesting hypothesis.

2.2 *ev* 2

Locus *ev* 2 has been found exclusively in RPRL lines 7_2 and 100 (Astrin 1978; *Tereba* and *Astrin* 1980). Genetic studies have demonstrated that this locus codes for RAV-0 produced by these lines (*Astrin* et al. 1980a). The proviral structure and transcriptional products for *ev* 2 are shown in Fig. 1. The locus appears structurally complete (*Hayward* et al. 1980; *Hughes* et al. 1981a), but is associated with an extremely low transcriptional activity. Between 0.1 and 0.5 copies per cell of apparently normal 35S and 21S messages are produced (*Hayward* et al. 1980). This leads to the production of extremely small quantities of RAV-0 virus. In the cultures of line 7_2 embryos, the cells of which lack receptors for subgroup E virus and therefore cannot be infected by the RAV-0 they produce, no additional virus is obtained. However, in cultures of embryos such as line 100, which by virtue of being susceptible to infection acquire additional RAV-0 proviruses derived from *ev* 2, the quantity of virus produced is 10^3- to 10^4-fold greater (*Robinson* 1978; *Crittenden* et al. 1979). Thus, although *ev* 2 is extremely poor in transcriptional activity, the virus it produces forms transcriptionally active proviruses. Several proposals have been made to account for this phenomenon.

Early studies using a transfection assay to analyze infectivity of DNA from high and low producers of RAV-0 indicated that DNA infectivity parallels virus production. It was proposed that *cis*-acting regulatory sequences inhibited transcription of the *ev* 2 locus, but not of other RAV-0 proviruses acquired by virus infection of susceptible cells (*Cooper* and *Temin* 1976). These conclusions were supported by the observation that infectivity of the *ev* 2 locus could be increased by shearing the cellular DNA to approximately the size of the provirus (*Cooper* and *Silverman* 1978). In further experiments, clones of cells containing *ev* 2 as well as exogenously acquired RAV-0 proviruses were investigated for sites of proviral integration, DNA infectivity, and expression of RNA and virus (*Jenkins* and *Cooper* 1980; *Humphries* et al. 1979, 1981). Each exogenously acquired provirus had a different site of integration; RNA expression, virus production, and DNA infectivity varied 30- to 100-fold when different clones were compared. It was concluded that differences in flanking cellular sequences and/or in modifications such as methylation of proviral DNAs were responsible for the observed differences in expression. The methylation hypothesis is supported by recent experiments in which the *ev* 2 locus has been shown to be activated by 5-azacytidine to produce high levels of RAV-0 virus (*Eisenman* et al. to be published). The activation was shown to correlate with decreased methylation of the locus.

In situ hybridization to metaphase chromosomes has indicated that *ev* 2 is located near the middle of the long arm of chromosome 2 (*Tereba* et al. 1981). Thus far, *ev* 2 is the only endogenous viral locus localized to chromosome 2.

The genome of RAV-0, the gene product of the *ev* 2 locus, has been compared with the genomes of Rous sarcoma virus (RSV), the avian leukosis viruses (ALV), and other exogenous retroviruses of the avian system. The major techniques used in these comparisons have been oligonucleotide maps of the RNA genomes and restriction enzyme maps and sequencing studies of cloned proviral DNA. Of course, the acute transforming viruses such as RSV contain a gene which codes for a transforming protein. These transforming genes are not present in the genomes of the leukosis viruses, such as RAV-2, or of RAV-0 and other endogenous viruses. Comparison of the *gag* and *pol* regions reveals a strong homology for all the viruses (*Shank* et al. 1981). However, in the *env* region and the U3 region of the LTR, considerable divergence occurs between the endogenous and exogenous viruses (*Skalka* et al. 1979; *Tsichlis* and *Coffin* 1980; *Hishinuma* et al. 1981). The envelope glycoproteins of the exogenous viruses have been classified as A, B, C, or D by genetic and biochemical assays. RAV-0 and all other endogenous chicken viruses isolated to date have a subgroup E envelope (for review see *Vogt* 1969). Oligonucleotide and restriction enzyme maps show a corresponding divergence in the envelope regions of the genome. A second difference occurs in the region of the viral LTR, termed U3. This region corresponds to several hundred nucleotides at the 3′ end of the viral RNA and has also been called the C, or constant region. Whereas the U3 or C regions of all the exogenous retroviruses show a close homology, the C region of RAV-0 and other endogenous viruses is distinctly different from that of the exogenous viruses (*Tsichlis* and *Coffin* 1980). The sequence difference manifests itself biologically in two significant ways. First, the difference in C has been observed to be the major determinant for a difference in growth rate between RAV-0, which grows comparatively slowly, and the exogenous viruses which show a more rapid growth (*Tsichlis* and *Coffin* 1980). Second, the difference in C region has been implicated in the failure of the endogenous viruses to cause disease, whereas the exogenous viruses cause a wide variety of acute as well as long latent period neoplasms (*Crittenden* et al. 1980; *Robinson* et al. 1980). Acute disease is caused by the presence of transforming genes such as *myc*, *src*, *erb*, and *myb*, genes which are not present in the leukosis viruses or the endogenous viruses. However, the long latent period neoplasms, such as bursal lymphoma, are readily induced by the leukosis viruses and also by transformation-defective sarcoma viruses, but are never seen in connection with infection by RAV-0 or any other endogenous virus. This striking difference has been shown to be correlated with the difference in the C regions of the viruses (*Crittenden* et al. 1980; *Robinson* et al. 1980). Since the U3 region encodes a promoter for transcription by RNA polymerase II, it is possible to explain both the inefficient growth of RAV-0 and the lack of disease-producing capability by postulating that the RAV-0 promoter is markedly less efficient than that of the exogenous viruses. This hypothesis is made very plausible by the finding that diseases such as bursal lymphoma, which are associated with infection with leukosis viruses, are a result of activation of a cellular oncogene by insertion of the viral promoter sequences (*Neel* et al. 1981; *Payne* et al. 1981; *Hayward* et al. 1981). If the RAV-0 promoter were inefficient, it might lack the capacity to activate cellular oncogene expression to a level suitable for tumor formation.

2.3 *ev* 3

ev 3 codes for the proteins characteristic of the gs$^+$chf$^+$ phenotype (expressing *gag* and *env*). *ev* 3 was originally identified as being present in more than 60 gs$^+$chf$^+$ birds from

four different flocks (*Astrin* 1978). Genetic experiments using RPRL line 6_3 and Kimber line K16, both of which are homozygous for *ev* 3, were used to analyze segregation of *ev* 3 and of the gs$^+$chf$^+$ phenotype (*Astrin* et al. 1979b; *Astrin* and *Robinson* 1979). F_1 progeny of a cross between a line 6_3 or K16 parent and a gs$^-$chf$^-$ parent were backcrossed to the gs$^-$chf$^-$ parent. Seventy-nine progeny of the backcross matings were analyzed, and an exact correlation between the segregation of *ev* 3 and of the gs$^+$chf$^+$ phenotype was observed.

The proviral structure and transcriptional products of *ev* 3 are shown in Fig. 1. The *ev* 3 provirus is defective, lacking part of the *gag* region, as well as a portion of the *pol* region (*Hayward* et al. 1980; *Hughes* et al. 1981a). A 31S (6.5 kilobase) transcript is produced at levels of 50–150 copies per cell and is found in both nuclei and cytoplasm (*Hayward* et al. 1980; *Baker* et al. 1981). This transcript contains an internal deletion in the *gag-pol* region. This defect correlates with the fact that *ev* 3 cells do not contain the normal 180 000-dalton *gag-pol* precursor protein, but instead contain a 120 000-dalton polyprotein which lacks the *gag* determinants specific for P15, as well as some polymerase determinants. The 120 000-dalton polyprotein is not cleaved to yield functional P27, P19, and P12 gag components or functional reverse transcriptase (*Eisenman* et al. 1978). A second transcript of 3 kilobases (21S) is also found in *ev*-3-containing cells. This transcript appears identical to a normal *env* message, and is present at a level of 30–80 copies per cell in both nuclei and cytoplasm. A 19S (2.3 kilobase) transcript containing *gag* sequences is present at 10–30 copies per cell, but is restricted to the nucleus and may represent a residual product of processing of the *env* message (*Hayward* et al. 1980; *Baker* et al. 1981).

Work on the chromatin structure of the *ev* 3 locus (*Groudine* et al. 1981) has demonstrated that the *ev* 3 sequences are under-methylated as compared to *ev* 1 sequences. In addition, the locus is preferentially sensitive to DNAse 1 digestion, and contains nuclease-hypersensitive sites in each of its two LTRs. These features correlate nicely with the transcriptional activity of *ev* 3.

In situ hybridization experiments using metaphase chromosomes from line 6_3 have indicated that *ev* 3 is located on a microchromosome (*Tereba* 1981).

2.4 *ev* 4, *ev* 5, and *ev* 8

Each of these three loci has been found in gs$^-$chf$^-$ cells (*Astrin* 1978), an indication that they do not express detectable viral protein products. Structures for the proviruses are shown in Fig. 1. No transcriptional products have been detected for *ev* 4 or *ev* 5 (*Hayward* et al. 1980). Lack of transcriptional activity of *ev* 4 and *ev* 5 is not surprising, since each locus has a deletion of 5′ sequences and lacks the 5′ LTR, the putative promoter for viral transcription (*Hayward* et al. 1980; *Hughes* et al. 1981a; *Baker* et al. 1981). No information on transcriptional activity or proviral defectiveness has been reported for *ev* 8.

ev 4 and *ev* 5 have been shown to be genetically linked in mating experiments where recombination between the two loci could be analyzed (*Astrin* et al. 1979b). Results of in situ hybridization experiments confirm this linkage and indicate that the two loci lie in proximity on the long arm of chromosome 1 (*Tereba* 1981; *Tereba* and *Astrin* 1982). As has been alluded to above, *ev* 4, *ev* 5, *ev* 6, *ev* 8, and *ev* 13 have all been localized to chromosome 1 by in situ hybridization. This finding has served as a basis for the hypothesis that *ev* 4, *ev* 5, *ev* 6 and *ev* 8 have been generated by multiple duplications of *ev* 1. Structural

similarities between retroviral proviruses, *ev* 1 included, and transposable elements lend this model a certain credence.

2.5 *ev* 6

ev 6 was originally identified in 14 birds of the gs⁻chf⁺ phenotype (expressing *env* but not *gag*) (*Astrin* 1978). It is one of two loci which code for this phenotype, the other being *ev* 9. Genetic experiments using Kimber line K18, which expresses the phenotype, have been used to analyze segregation of chf expression and *ev* 6. An examination of more than 30 progeny of backcross matings which were segregating for the expression of chf revealed an exact correlation between the segregation of *ev* 6 and segregation of the gs⁻chf⁺ phenotype (*Astrin* et al. 1980b).

The proviral structure and transcriptional products of *ev* 6 are shown in Fig. 1. The structure of the *ev* 6 provirus, like that of *ev* 4 and *ev* 5, is defective. *ev* 6 lacks both the 5′ LTR and *gag* sequences (*Hayward* et al. 1980; *Hughes* et al. 1981a). Surprisingly, although this locus lacks the viral promoter sequences present in the 5′ LTR, transcriptional products are observed. A low level (0.5–1.5 copies per cell) of a 28S (5.3 kilobase) RNA-containing *pol*, *env*, and U3 sequence is observed in the nucleus, and a moderately high level (25–100 copies per cell) of a 21S (RNA-containing *env* and U3 sequence is observed in both nucleus and cytoplasm (*Hayward* et al. 1980; *Baker* et al. 1981). The 21S RNA most likely serves as message for the production of *env* protein characteristic of *ev*-6-containing cells. Neither of these RNAs contains viral U5 sequences (*Hayward* et al. 1980; *Baker* et al. 1981). This finding is further evidence that the RNAs are not transcribed from a viral promoter. It has been proposed that the *ev*-6-encoded 21S *env* message is initiated within a cellular promoter located adjacent to the left end of the *ev* 6 provirus. Initiation within the cellular promoter and transcription of adjacent cellular sequences followed by transcription of viral sequences might provide a message which contains a cellular leader sequence covalently linked to viral information (*Hayward* et al. 1980; *Baker* et al. 1981). Further experimentation will be required to confirm this attractive hypothesis.

As mentioned above, *ev* 6 has been localized to chromosome 1 of the chicken by in situ hybridization (*Tereba* 1981; *Tereba* and *Astrin* 1982).

2.6 *ev* 7

ev 7 was originally identified in 15 embryos of RPRL line 15_b (*Astrin* 1978). This line of chickens has an interesting phenotype. Cells grown in the presence of bromodeoxyuridine are induced to express a noninfectious avian leukosis virus. The virus is detected as particles containing reverse transcriptase activity (*Robinson* et al. 1976; *Robinson* 1978). *ev* 7 segregates with this phenotype in backcross matings and apparently codes for the particles produced by line 15_b cells (*Robinson* et al. 1979b).

The *ev* 7 provirus does not appear to contain any gross deletions (*Baker* et al. 1981). No analysis of transcriptional products has been reported.

The genetic information of *ev* 7 has been reported to undergo recombination with *ev* 1 genetic information to produce infectious subgroup E virus. These viruses were shown to have P27 and P19 components which were characteristic of RAV-0 (*Robinson* et al.

1979a). This finding supports the conclusion that the viruses are formed by recombination of endogenous viral genetic information. The parents of the recombinant viruses were postulated to be the *ev*-7-encoded virus particles and *ev*-1-encoded RNA.

Genetic experiments utilizing line 15_b in crosses with lines 7_2 and 6_3 have shown that *ev* 7 segregates with the male (z) chromosome (*Smith* and *Crittenden* 1981). Independent experiments using in situ hybridization methods have also localized *ev* 7 to the z chromosome (*Tereba* et al. 1981). Thus far, *ev* 7 is the only endogenous provirus to be localized to the z chromosome.

2.7 *ev* 9

ev 9 is one of two loci, the other being *ev* 6, that code for envelope protein produced in gs$^-$chf$^+$ cells. *ev* 9 was first identified in several embryos of this phenotype and later observed to segregate with the phenotype in backcross matings (*Astrin* 1978; *Astrin* et al. 1980b).

The proviral structure and transcriptional products of *ev* 9 are shown in Fig. 1. The provirus has no apparent deletions; however, a 34S transcript is observed (*Hayward* et al. 1980). This size is slightly smaller than the transcript of an intact provirus which is 35S. The 34S transcript contains U5, *gag*, *pol*, *env*, and U3 sequences; is present in 20–60 copies per cell; but is confined to the nuleus (*Baker* et al. 1981). A second transcript of 3 kilobases (21S) contains U5, *env*, and U3 sequences; is present in 60–150 copies per cell; and is found in both nuclei and cytoplasm. This transcript is most likely the message for the production of envelope protein characteristic of *ev*-9-containing cells.

2.8 *ev* 10, *ev* 11, and *ev* 12

These three loci each code for a distinct subgroup E virus produced by a particular line of inbred white leghorn chickens (*Astrin* et al. 1980b). *ev* 10 is present in lines C, 151_4, and 151_5 (*Tereba* and *Astrin* 1980). It has been shown to segregate with the ability to produce infectious subgroup E virus (V$^+$ phenotype) in an analysis of 48 progeny of a backcross mating of line 151_4 chickens (*Crittenden* and *Astrin* 1981). *ev* 11 is present in line 151_4 and 151_5 (*Tereba* and *Astrin* 1980), and has similarly been shown to segregate with the V+ phenotype in backcross matings (Crittenden and Astrin unpublished results). *ev* 12 is present in RPRL line 15_1 (*Tereba* and *Astrin* 1980), and has been shown to segregate with the V+ phenotype in matings of line 15_1 chickens (*Smith* and *Crittenden* 1981).

ev 10, *ev* 11, and *ev* 12, like *ev* 2, produce only very small amounts of virus spontaneously. However, if the producer cells have receptors for subgroup E virus, additional proviruses are accumulated through infection, and much larger quantities of virus are produced. It is likely that the control mechanisms regulating expression of *ev* 2 (QV) are also operating in the cases of *ev* 10, *ev* 11, and *ev* 12.

2.9 *ev* 13, *ev* 14, *ev* 15, and *ev* 16

ev 13 is a locus which was identified in gs$^-$chf$^-$ cells, and shown by in situ hybridization to be present on chromosome 1 in a unique location (*Tereba* 1981; *Tereba* and *Astrin* 1982). No information other than its chromosomal location is available.