REVIEW ARTICLE

Some Highlights of Animal Virus Research in 1986

By D. J. McGEOCH,1*† C. R. HOWARD2 AND U. DESSELBERGER1
1Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR and 2Department of Medical Microbiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.

In 1986 we published a review article entitled 'Some highlights of animal virus research in 1985', in which we attempted to summarize important and interesting research on animal viruses published in 1985 (McGeoch et al., 1986b). The limitations of the exercise were recognized: coverage was selective rather than systematic and, outside clearly pre-eminent subjects, was overtly influenced by personal scientific interests. In this article we now present a similar account of animal virology in 1986. Much the same ground rules have been retained. Selectivity remains of the first importance, and our treatment concerns work published in 1986 as far as possible, but with a little flexibility where necessary.

As with 1985, animal virology in 1986 was dominated by work on acquired immune deficiency syndrome (AIDS) retroviruses. Outside this area, however, we were not able to identify any small set of truly outstanding advances, although work in many virus systems clearly made excellent progress. This evaluation posed problems for a review of the type undertaken, in particular for achieving a balanced selection of topics. With this background we also found it more necessary than before to address virus groups in separate sections (representing a step back from the previous, integrative stance). We have chosen to present aspects of research on, first, the AIDS viruses, and then in succession on hepatitis viruses, herpesviruses, poxviruses, papovaviruses, picornaviruses and negative strand viruses, and to finish with a section on virion structure.

AIDS

During 1986 the extent of the public health problem posed by AIDS continued to increase worldwide. Certainly, it was in this year that the scale, the immediacy and the longer term implications of this viral disease finally became a prominent part of public consciousness in Great Britain. Prospective cohort studies in Great Britain (Weber et al., 1986) and in the U.S.A. (Goedert et al., 1986) demonstrated a high and increasing incidence of AIDS 3 to 5 years after infection with the AIDS virus. Also, in 1986, the very wide spread of infection in several Central African countries became much more clearly recognized (Mann et al., 1986a, b; Quinn et al., 1986). Basic research on the retrovirus responsible for AIDS continued to advance rapidly, but problems of treatment and prevention appear to be of a more intransigent nature. On a somewhat bathetic note, we register that the AIDS retrovirus has now been assigned the vernacular name 'human immunodeficiency virus' (HIV) (Coffin et al., 1986; Brown, 1986). Here we recount aspects of research on, first, the AIDS viruses, and then in succession on hepatitis viruses, herpesviruses, poxviruses, papovaviruses, picornaviruses and negative strand viruses, and to finish with a section on virion structure.

The wide variations in the genomes of HIV isolates were further examined. Alizon et al. (1986) determined the complete genomic sequences of two strains from Zaire, in Central Africa. These had biological properties indistinguishable from those of other HIV isolates, and their sequences showed the same layout of reading frames and other features, but sequence comparisons revealed a much greater degree of variability than previously observed, with
numerous point mutations and addition/deletion changes, including small duplications. The env gene encoding the envelope glycoproteins of the virion, was found to be particularly variable, with up to 26.4% differences in pairwise comparisons, between isolates, of the amino acid sequences of the external portion of the protein. In this part there are three major hypervariable regions, which are likely to represent superficial, antigenic structures. Outside these regions, the sequences are much more conserved. Occurrences of cysteines are essentially invariant. The picture thus emerging is of a protein whose fundamental structure is largely invariant while external features, exposed to immune pressures, vary. The gag and pol sequences were found to be more conserved, although certain of the genes peculiar to HIV (tat, 3'orf; see below) showed variability similar to that of env. Starchch et al. (1986) reported similar findings from partial sequence analysis of two additional American HIV isolates.

In another type of study, Hahn et al. (1986) obtained sequential isolates of HIV from infected persons over a period of 1 or 2 years, and carried out partial sequence analyses. In comparisons of successive isolates from one individual, they suggested that the differences between them were best accounted for by the isolates being strains evolving in parallel from a common progenitor. On this basis it was estimated that the rate of mutational change in the env gene might be at least $10^{-3}$ substitutions per nucleotide per year, with the rate for the gag gene some 10-fold lower. There was some evidence for cocirculation of genomically different subpopulations of HIV in individual hosts. Strains from one host were, however, much more similar to each other than to isolates from other individuals. There appears to be a discrepancy between the observed rapid mutation of HIV, and the existence of only a few dominant cocirculating strains in one person, and the authors suggested that this may indicate the operation of some interference mechanism to prevent multiple infection with divergent strains.

Following observations that in West African countries (far removed from the epidemic areas in Central Africa) certain AIDS patients exhibited no serological evidence of HIV infection, Clavel et al. (1986) isolated a distinct retrovirus from two such patients. The biological properties of this virus, HIV-2, were very similar to the original HIV (now termed HIV-1) but it differed substantially in the antigenicity of its envelope glycoproteins, and by nucleic acid hybridization tests was only distantly related to HIV-1. By the same criteria HIV-2 was found to be more related to, but still distinct from, simian lentivirus isolates previously described as the cause of an AIDS-like disease in captive macaques. In another study, Kanki et al. (1986) described the isolation of a similar virus in West Africa from apparently healthy humans, distinct from HIV-1 but related to the simian isolates. The relationship between these two new human viruses is presently unreported. These findings are evocative of human AIDS viruses having recently evolved from an animal reservoir.

Figure 1 depicts present understanding of the content and layout of genes in the genome of HIV. Apart from the characteristic retroviral genes gag, pol and env, there are at least four other genes. These have been variously named; here we follow the nomenclature of Chen (1986) and call them sor, tat, art/trs and 3'orf. The tat and art/trs genes each possess two polypeptide coding exons. Analysis of these HIV-specific genes has proceeded rapidly in the last 2 years; to a great extent this work has depended on the use of transfection to introduce full length, infectious DNA clones of the HIV genome (Fisher et al., 1985; Sodroski et al., 1986b, c; Seigl et al., 1986; Adachi et al., 1986). Proteins encoded by the sor, tat and 3'orf genes are all produced in natural infection (Allan et al., 1985; Goh et al., 1986, Kan et al., 1986; Lee et al., 1986).

The functions of the sor and 3'orf proteins remain unknown. These two species have been shown not to be essential for virus growth in cultured cells or for production of cytopathic effect, although sor deletion mutants do not grow as well (Sodroski et al., 1986c). The tat and art/trs proteins are, however, essential for virus replication (Dayton et al., 1986; Sodroski et al., 1986b, c; Feinberg et al., 1986; Fisher et al., 1986). In 1985 it was shown that the tat protein functioned in activating HIV gene expression. Rosen et al. (1986) and Feinberg et al. (1986) proposed that tat protein does not stimulate transcription, but is required for translation of virus mRNAs. On the other hand, Cullen (1986) did find a stimulatory effect on mRNA production. As discussed by Chen (1986), this issue remains unresolved. The art/trs gene is also involved in regulation of virus gene expression (Sodroski et al., 1986b; Feinberg et al., 1986). Again, the mechanism of
Fig. 1. Organization of genes in the genome of HIV. The upper part of the figure indicates the DNA form of the HIV genome, with long terminal repeat sequences as open boxes. The locations of the reading frames for the seven recognized genes are indicated. The tat and art/trs genes both possess two coding exons (joined by lines). Based on Chen (1986).

action is not yet well defined. One possibility is that the art/trs protein is active in regulating the extent of splicing of viral mRNA (Feinberg et al., 1986).

Studying the cytopathic action of HIV, Lifson et al. (1986b) observed that infection of a T4 lymphocyte line with HIV caused cell fusion and formation of large, multinucleate cells which soon died, and they proposed that a similar mechanism also operates in vivo. Further work showed that this effect can be obtained by expression of the env protein alone (Lifson et al., 1986a; Sodroski et al., 1986a). Uninfected cells were incorporated into such syncytia only if they carried the T4 surface antigen. Thus, the outcome of HIV infection, as well as initial susceptibility, depends on the presence of the T4 receptor. Formation of a complex between the T4 receptor and env protein was shown by McDougal et al. (1986). Maddon et al. (1986) cloned the gene for the T4 protein, and showed that its transfection into human cells and subsequent expression rendered them susceptible to HIV infection. In similarly treated mouse cells, however, the infection was blocked during internalization of virus. Levy et al. (1986) and Adachi et al. (1986) showed by transfection of complete DNA copies of the HIV genome that there was no intracellular obstacle to HIV replication in various cell types from a number of species. In this context the findings of Folks et al. (1986) are of interest. After infection of a human T cell line with HIV, a surviving fraction of less than 5% of the cells was found. These lacked T4 receptors and did not produce virus, but did contain the virus genome. Virus production could be reactivated by treatment with 5-iodo-2'-deoxyuridine as late as 90 days after infection. This observation may be developed into a model for HIV latent infections.

It is evident from the literature that many approaches to vaccination against HIV are being explored. These must cope with two fundamental attributes of HIV infection. First, as outlined above, the env gene of HIV is clearly evolving rapidly. Complementing the sequence data, serological studies have shown that HIV-1 isolates are antigenically heterogeneous (Clavel et al., 1985; Weiss et al., 1985, 1986; Matthews et al., 1986). Next, the fact that the virus directly attacks a central component of the immune system must have the most important bearing on any planned stimulation of immunity. Several papers have reported on the poor neutralizing capacity of sera containing HIV antibody (Clavel et al., 1985; Weiss et al., 1985; Matthews et al., 1986). It seems that the dynamics of HIV infection remain far from clear; particularly intriguing is the fact that in infected individuals the proportion of T4 lymphocytes actively infected (as judged by presence of HIV mRNAs) is small (Harper et al., 1986). Finally, we note that there is a paucity of animal models for HIV infection; so far, chimpanzees are the only animal species known to sustain a persistent infection with human HIV (Fultz et al., 1986).
Lasky et al. (1986) used DNA manipulative methods to produce a secreted form of the HIV glycoprotein from culture cells, and evaluated it in animal immunization experiments. Bolognesi et al. (1986) and Gallo et al. (1986) described work in progress to develop vaccines by incorporating env gene products (purified from virus particles or expressed from recombinant vectors) into glycoside lattices to generate multimeric complexes, called immunostimulatory complexes (ISCOMs; Morein et al., 1984). The ISCOM approach, using glycoproteins from feline leukaemia virus, has been found to elicit protective antibody to this virus in cats (Osterhaus et al., 1985). Another approach would involve the use of a potent anti-idiotype antibody as immunogen (Bolognesi et al., 1986; Gallo et al., 1986).

The potential use of oligopeptides was reported in several separate approaches. Antiserum to a synthetic peptide from the env protein sequence was found to react with the whole glycoprotein, and the sera of AIDS patients recognized the oligopeptide (Kennedy et al., 1986). The protective efficiency of oligopeptide antibodies remains to be tested. Another interesting possibility for prevention or attenuation of spread of HIV infection was opened by the finding that an octapeptide which is part of a neuropeptide receptor-like portion of the T4 protein blocked HIV infection of human T cells in vitro at a low concentration (0-1 nM; Pert et al., 1986). Reiher et al. (1986) reported that there is amino acid sequence similarity between an active site of interleukin 2 (IL2) and a portion of the env protein of HIV and other immunosuppressing retroviruses. This has given rise to the suggestion that this part of the env protein might act as an antagonist to IL2 either directly, or indirectly by eliciting cross-reactive antibody. Weigent et al. (1986) found that a hexapeptide corresponding to the region of homology did act antagonistically to IL2, whereas an oligopeptide representing an adjacent, conserved part of the env sequence did not. This might allow an oligopeptide vaccine approach in which antibody binds to HIV but not to IL2, so avoiding further immunosuppression.

Finally, we note that clinical improvement was observed in some AIDS patients under treatment with 3'-azido-3'-deoxythymidine (AZT), an inhibitor of HIV replication (Yarchoan et al., 1986), and that this has aroused hope that AZT will prove a potent and useful antiviral drug. The evaluation of long term treatment, however, needs further study.

Viral hepatitis

In this section we record some aspects of work on hepatitis B virus (HBV), and also on the hepatitis B-associated delta agent. The genome of HBV is an open circular DNA molecule of only 3200 residues, which is replicated by way of an RNA intermediate. A great drawback in studying HBV has been the lack of tissue culture systems which permit productive infection with the virus. Two reports in 1986 addressed this problem. First, Tuttleman et al. (1986) succeeded in growing duck HBV in primary cultures of duck hepatocytes. However, only 10% of the cells were susceptible to the virus, and even these became resistant after several days in culture. Next, Sureau et al. (1986) were able to transfect cloned, circular HBV DNA into an established human hepatoma cell line, and to observe at least a partial infectious cycle.

Expression of the HBV genome remains imperfectly understood. There are thought to be four genes, encoding the core protein (C), the virion surface protein (S) and related species (see below), the replicative RNA-dependent DNA polymerase (P), and a species (X) predicted from DNA sequence data. These are arranged unidirectionally, in the order C-S-X, with the long P reading frame overlapping the downstream end of the C reading frame, the whole of the S frame and the upstream part of the X frame. The least characterized genes are those encoding P and X. Will et al. (1986) obtained evidence for the existence of C-P fusion proteins in hepatic carcinoma cells containing HBV DNA. They suggested that synthesis of P protein might thus proceed by a ribosomal frameshift following translation of the C reading frame, with subsequent processing; i.e. by a mechanism similar to that employed by retroviruses in expression of reverse transcriptase. Two papers reported production of amino acid sequences corresponding to the predicted X protein, by chemical synthesis (Moriarty et al., 1985) and by expression in an Escherichia coli recombinant (Meyers et al., 1986). In both cases, the X-related peptides reacted with a proportion of sera from persons with markers of HBV infection, providing evidence that X is a real species produced during HBV infection.
From a number of recent papers, it is clear that expression of surface protein species from the S gene region is complex (see, for instance, Heermann et al., 1984; Neurath et al., 1984; Wong et al., 1985; Pfaff et al., 1986; Howard, 1986). The major virus surface protein S is translated from the 3' portion of an open reading frame. This is the hepatitis B s antigen (HBsAg). Two further species are also translated from this reading frame, called preS1 and preS2, which contain all of the S amino acid sequence and also, respectively, 163 and 55 additional residues at their N termini. In addition, all three polypeptides become variably glycosylated. All of these species are present in the virus particle. Neurath et al. (1986b) showed that a section of the polypeptide chain unique to preS1 could bind to a hepatocyte surface component, and proposed that this represented a part of the virion's host cell binding mechanism. At present, the molecular mechanisms underlying S gene expression remain obscure. Persing et al. (1986) examined translation of preS1 and S in transfected culture cells and in microinjected Xenopus oocytes, and found that preS1 exerted a marked and specific inhibitory effect on translation of S.

It has now been shown that a similar complexity of expression also exists for the C gene, where there are two possible in-phase initiation codons separated by a short 'pre-core' region, which encodes an N-terminal sequence resembling the hydrophobic signal peptides of membrane-bound proteins (Uy et al., 1986; Ou et al., 1986; Miyanohara et al., 1986). In HBV-infected cells, translation can start at the first initiation codon to give a version of the core protein which is translated on membrane-bound ribosomes and secreted from cells, in a distinct antigenic form as the hepatitis B e antigen (HBeAg). This is a commonly sought serum marker of active hepatitis B infection. Translation products devoid of the pre-core sequence form immature nucleocapsids around the so-called pre-genomic RNA, within which the genomic minus strand DNA is then generated by reverse transcription.

The introduction of monoclonal antibody-based assays for HBV has led to the finding of hitherto unrecognized serological variants of the virus that are not detected by conventional tests (Wands et al., 1986). The genome size of these variants is approximately the same as that of prototypic HBV, with which extensive homology has been shown by in situ hybridization. However, non-human primates immune to HBV infection remain susceptible to infection by one of the variants. This work has implications for the diagnosis of HBV infection and for the effectiveness of existing HBV vaccines. Another member of the hepadnavirus family was identified in 1986, which has tree squirrels as its natural host (Feitelson et al., 1986a, b).

This last year has seen the licensing in several countries of the first vaccines for human use to have resulted from gene cloning. These are yeast recombinant hepatitis B vaccines, and they overcome reservations concerning the use of plasma from groups at high risk of AIDS as a source of HBsAg for vaccination. It should be stressed that clinical trials showed that these products induce antibody responses similar to those against plasma-derived vaccines (Davidson & Krugman, 1986; S. E. Brown et al., 1986). The new recombinant hepatitis B vaccines are devoid of preS sequences, which several groups claim are important for conferring full immunity. For example, Neurath et al. (1986a) reported that an antiserum to a synthetic preS peptide protected chimpanzees from challenge with HBV. However, a vaccine devoid of preS antigens has proved to be effective in healthy adults for nearly 10 years. Moreover, recent work from Hellstrom et al. (1986) has suggested that antibodies to preS antigens decline rapidly after acute infection but persist in individuals with HBV-mediated chronic liver disease.

Striking progress was made in 1986 in characterization of HBV-associated delta agent. This is, in effect, a dependent virus species, which requires helper functions supplied by the HBV genome, and as a consequence is found only in close association with hepatitis B infections. Patients with existing hepatitis B infections who are exposed to the delta agent often develop a more serious, acute delta-related hepatitis superimposed on the pre-existing disease, and this may have serious long-term sequelae. The delta agent particle consists of a genomic RNA molecule associated with two delta-specific protein species of Mr 27000 and 29000, and enveloped by a lipid membrane carrying HBV S protein species (Bonino et al., 1986).

Denniston et al. (1986) described a small cDNA clone of delta genomic RNA, and its diagnostic application. Three papers reported that delta RNA is a covalently closed circle (Kos et al., 1986; Wang et al., 1986; Chen et al., 1986). The complete sequence of the RNA was
determined by Wang et al. (1986), as a 1678 nucleotide circle. The RNA has a relatively high G + C content (60%) and under non-denaturing conditions exists as a rod-like structure with extensive intramolecular base pairing. Wang et al. expressed parts of the genome in a bacterial expression vector and identified an open reading frame, on the non-virion strand, coding for protein which reacted with antiserum directed against delta agent. This reading frame could produce a protein of 215 amino acids, and this is in reasonable agreement with the analysis of delta proteins by Bonino et al. (1986). These data suggested that the delta agent possesses a negative stranded genome, but it should be emphasized that this is a preliminary characterization. Chen et al. (1986) detected complementary RNA in infected cells, and some of this was also circular. Dimeric RNA species were also observed. The genomic RNA possesses structural similarities to plant viroids, virusoids and satellite RNAs, and certain sequence features of these elements are also present. The presence of dimeric RNA species in infected cells would be consistent with a rolling circle mode of replication similar to that proposed for viroids.

The delta agent thus presents a number of unusual features as an animal pathogen. This is the first description of a circular RNA genome for an animal virus, and it is also the smallest animal virus genome known, although to regard the agent as a conventional virus may not prove justified. If the apparent similarities to viroids and similar entities are validated by further work, it will represent a wholly new class of animal infectious agent.

**Herpesviruses**

Many advances were reported in research on herpesviruses during 1986. The most interesting single item concerned a newly recognized virus of humans, termed human B-lymphotropic virus (HBLV), which is the first new human herpesvirus discovered since Epstein–Barr virus (EBV) was reported in 1964. The complete literature so far on HBLV consists of two papers, describing the isolation and biological properties of the virus (Salahuddin et al., 1986), and a preliminary analysis of its genome (Josephs et al., 1986). Isolates were made from peripheral blood cells of six patients exhibiting a range of lymphoproliferative diseases. These isolates were shown by serological tests to be closely related. All the patients possessed antibodies to HBLV, but the virus appears rare in the general population: only four out of 220 sera from healthy subjects contained antibodies to HBLV. Two of the isolates were obtained from AIDS patients, and HBLV might, therefore, in these instances have the status of an opportunistic pathogen. However, the other isolates were not associated with AIDS cases.

Replication of HBLV was achieved in vitro only in freshly isolated human B lymphocytes, and resulted first in formation of giant cells and then in cell death. The virus did not evince any growth transforming activity. No serological cross-reactions were detected between HBLV and other human and primate herpesviruses. By electron microscopical examination, virus particles possessed a typical herpesvirus morphology. The virus genome consisted of a large, double-stranded DNA molecule, not yet characterized in detail but containing more than 110 000 base pairs (that is, in a size range occupied by known herpesvirus genomes). A 9000 base pair cloned fragment of the genome did not hybridize to the genomes of other human herpesviruses, or to those of herpesviruses of other primates.

At this very early stage of investigation our picture of this new human virus is of a species which is relatively rare in the human population, unlike the previously known human herpesviruses, and which may exhibit an aetiological connection with lymphomas and lymphoproliferative disorders.

Turning to the better known herpesviruses, we next describe some aspects of genome structural analyses. A large amount of herpesvirus genome sequence information was published in 1986, only a selection of which is outlined here. The high point of this area was the publication of the complete genome sequence of varicella-zoster virus (VZV); this is the first sequence of an alphaherpesvirus genome to be completed. The genome of VZV contains approximately 125 000 base pairs of double-stranded DNA. The entire sequence was determined by Davison & Scott (1986), and analysed to give a model for the number and organization of VZV genes. This view of the gene set of VZV is based on analyses of open reading frames and codon usage, on
functions or partial functions have now been assigned to 22 of the genes by several means, leaving 45 of presently unknown function. The largest number of these assignments came from comparisons with HSV genes of known sequence or of known genomic location. From inspection of hydrophobicities of predicted amino acid sequences, it seemed likely that five genes encoded glycoproteins, and three of these have now been identified as encoding known, external virion glycoproteins (Davison et al., 1985; Ellis et al., 1985; Keller et al., 1986). A new glycoprotein nomenclature taking the sequence data into account has been proposed (Davison et al., 1986). Finally, as discussed below, two genes are thought to encode previously unrecognized enzymes.

From our present incomplete knowledge of gene organization in the otherwise better studied HSV, it is clear that these two alphaherpesviruses are mostly very similar in the layout and nature of their genes, although there are local differences. The most marked divergences are in the short regions. Davison & McGeoch (1986) showed by comparisons of predicted amino acid sequences that all of the VZV short region genes have HSV-1 counterparts, but that six of the HSV-1 genes in the short unique region have no VZV homologues. The differences in gene layout between the two genomes could be understood as resulting from a small number of partially illegitimate recombination events which had occurred since divergence of the present day viruses from a common precursor.
A number of papers describing DNA sequences of parts of the genomes of HSV-1 or HSV-2 appeared in 1986, of which two are mentioned here. First, the complete sequence of the short repeat (Rs) of HSV-1 was described (McGeoch et al., 1986a). This comprises 6600 base pairs and has a base composition of 79.5% G + C, which is the highest value for any large DNA sequence so far determined. The region contains one gene, for the transcriptional regulatory protein IE175 (or ICP4), and the coding region of this is a 3894 base region with base composition 81.5% G + C. This value is sufficiently extreme that the amino acid content of IE175 is strongly biased towards those encoded by G- and C-rich codons. The evolutionary or functional significance of these observations remains unknown. The second HSV gene sequence noted is that for the IE110 (or ICP0) protein, also a transcriptional activator. This gene lies in the long repeat (Rs) region and possesses two introns (Perry et al., 1986). It is thus the first HSV gene definitively shown to contain multiple introns.

The protein IE175 is a major transcriptional regulator of HSV infection. Studies on its mode of action have recently advanced significantly. It was shown that the protein can bind specifically to sequences in target promoters, and can directly modulate transcription in cell-free systems, inhibiting transcription of immediate early genes and stimulating later genes (Kristie & Roizman, 1986; Faber & Wilcox, 1986; Beard et al., 1986; Pizer et al., 1986). IE110 has also emerged as an important regulator of HSV transcription. Interestingly, its predicted amino acid sequence exhibits a cluster of cysteine residues similar to regions described for a number of other transcriptional regulatory proteins (Miller et al., 1985; Berg, 1986; Perry et al., 1986). These regions are postulated to be involved in the coordination of metal ions, which may participate in DNA binding. Another aspect of IE110 emerged from a study by Stow & Stow (1986), who constructed a strain of HSV-1 containing deletions of almost the whole protein coding regions in both copies of the IE110 gene. Surprisingly, this virus could still grow in normal culture cells. While growth at a high m.o.i. appeared normal, the mutant virus plaqued very inefficiently. Thus, IE110 plays an enabling, but not essential, role in getting the replicative cycle under way, and this can be dispensed with more effectively when several incoming virus particles are present in a single cell.

Weston & Barrell (1986) described the DNA sequence of a 44000 base pair section of the genome of human cytomegalovirus (HCMV), comprising the short unique region (Us), the short repeat region (Rs) and part of the long repeat region (Rl). This is by far the most extensive sequence so far published for a betaherpesvirus, and represents about 19% of the enormous genome of HCMV. The organization of genes within this DNA, and the characteristics of encoded proteins, were deduced from the sequence data. Prior to this work, nothing was known about the genetic potential of this part of the genome of HCMV. Weston & Barrell identified 38 or 39 open reading frames which probably represent the coding regions of separate genes. None of these showed any detectable amino acid sequence similarity with any published sequences for alpha- and gamma-herpesviruses, including genes in the short regions of HSV-1, VZV and EBV. Thus, despite similarities in the layout of unique and repeated units of the HCMV genome with those of HSV and VZV, it seems likely that the short regions of alpha- and beta-herpesvirus DNAs are evolutionarily unrelated.

It was found that many of the predicted HCMV polypeptides showed similarities in their amino acid sequences. Such proteins fell into five separate groups, as indicated in Fig. 3, i.e. there are five multigene families encoded in this part of the HCMV genome. Degrees of relatedness between members of a family varied widely, and in some cases additional reading frames were present which showed no sequence homology to a given family, but which were similar to the family in their distribution of amino acid types, and are therefore probably to be included in that family as particularly divergent members. Characteristics of the amino acid sequences from family 2 (see Fig. 3) suggested that these proteins may be membrane-inserted glycoprotein species. Sequences from families 3 and 5 exhibit multiple hydrophobic segments, suggesting that these proteins may be multiply inserted in membranes. It is thus clear that duplication and divergence of genes have played a major role in the evolution of this DNA. This phenomenon has no known counterpart on a comparable scale in any other virus system, and certainly goes some way towards accounting for the greater size of typical betaherpesvirus
Review: Animal virus research in 1986

Fig. 3. Organization of genes in a section of the genome of HCMV. A 44 000 base pair section of the genome of HCMV is represented on two successive lines. The short unique region (Us) is shown by a solid line, and the short repeat regions (IRs and TRs) by open boxes. Locations and orientations of open reading frames thought to correspond to genes are shown as arrows. Membership of multigene families is marked, for family 1 to family 5. Numbers in brackets indicate possible, but widely diverged, members. Drawn from the data of Weston & Barrell (1986).

genomes relative to those of other herpesviruses. Perhaps the most striking aspect of this finding is the fact that, within a given family, the individual members have often become widely diverged, but the reading frames have nonetheless remained open, suggesting that each member in the family now possesses an independently necessary role in some aspect of the virus cycle.

To close this section on genome organization, we note that some herpesvirus genes are now known to have much more complex structures than the simple models used for interpretation of the VZV and HCMV sequences just discussed. Several papers have reported characterization of cDNA clones made from mRNAs in lymphocytes latently infected with EBV, and have shown that there exists a class of transcripts which are derived from up to 100 000 residues of the EBV genome (Bodescot & Perricaudet, 1986; Bodescot et al., 1986; Sample et al., 1986). At least some of the five EBV-specified nuclear proteins (EBNAs) associated with maintenance of the genome in its latent state are expressed from this family of mRNAs (Kallin et al., 1986; Dillner et al., 1986). The primary transcripts from which these mRNAs are derived evidently run through many ‘normal’ genes and transcriptional control signals, and they become multiply spliced, with excision of up to 18 introns, to yield final transcripts of around 2000 residues. There are several different patterns of splicing, which result in different translatable reading frames.

Herpesviruses encode a number of enzymes, and interest in these has been high for many years, in particular as potential targets for antiviral agents. A number of new findings on herpesvirus enzymes were published in 1986. In the section on poxviruses, below, we mention relationships of herpesvirus DNA polymerases with other DNA polymerases. A novel observation with the HSV ribonucleotide reductase was reported by two groups. The activity of this essential enzyme was inhibited by oligopeptide species corresponding to the C-terminal region of the smaller subunit of the enzyme (Dutia et al., 1986; Cohen et al., 1986). It was proposed that the oligopeptide acted by interfering with interactions between the subunits. Although these enzyme inhibitors did not themselves display antiviral activity in tissue culture, this finding may contribute a new general rationale in searches for antiviral compounds.

Evidence for two additional enzymes encoded by herpesviruses was obtained by the rather new technique of comparing predicted amino acid sequences, deduced from DNA sequences, with protein sequence libraries. First, McGeoch & Davison (1986) showed that the products of genes US3 of HSV-1 and HSV-2, and of gene 66 of VZV, are homologous to members of the protein kinase family. The distantly related EBV does not possess a corresponding gene. This predicted enzyme species may be the new protein kinase which has been detected in cells infected with HSV-1 (Purves et al., 1986), but its function is still unknown.
The second enzyme predicted by way of DNA sequence analysis was thymidylate synthase. Like thymidine kinase and ribonucleotide reductase, this is an enzyme of nucleotide anabolism, which methylates dUMP to TMP. Genes highly homologous to known prokaryotic and eukaryotic thymidylate synthase genes were detected for herpesvirus saimiri (HVS) (Honess et al., 1986; Bodemer et al., 1986) and for VZV (Davison & Scott, 1986). In the former case, the enzyme activity was subsequently found in cells infected with HVS. However, no virus-specific activity was found with HSV, HCMV or EBV. For those herpesviruses which do encode this enzyme, it seems certain that it will become a new target in searches for antiviral compounds.

**Poxviruses**

We distinguish two aspects of work with poxviruses: research on the viruses per se, and development of their uses as vehicles for the expression of foreign genes. The latter field is becoming a prominent part of studies on surface antigens, most notably with a view to eventual vaccine use. In this section we first discuss recent results on characterization of poxvirus genes and proteins, and then review progress in the use of vaccinia virus (VV) as an expression vector and for recombinant vaccines.

Poxviruses specify their own RNA polymerase, which transcribes the virus genome within the cytoplasm of the infected cell. Poxvirus RNA polymerases are complex enzymes with many subunits, and in this respect are reminiscent of RNA polymerases of both eukaryotic and prokaryotic organisms. Broyles & Moss (1986) determined the DNA sequence of a 6000 base pair region of the VV genome containing the gene for the largest subunit of RNA polymerase and also the gene for a small subunit. They then mapped the corresponding mRNAs and deduced the amino acid sequences of the subunits. These were compared with available RNA polymerase sequences, and it emerged that the largest subunit of VV polymerase (Mr 147000) showed extensive homology with the largest subunits of *E. coli* RNA polymerase (β'), yeast RNA polymerases II and III, and a partial sequence for *Drosophila* RNA polymerase II. The similarities to the eukaryotic sequences were higher than that to the *E. coli* sequence. However, the VV sequence resembled the two yeast sequences to a near equivalent extent so that no judgement was reached on whether the VV gene had evolved from that for eukaryotic RNA polymerase II, or that for RNA polymerase III, or from some precursor of both. From this result it appears likely that other subunits of poxvirus RNA polymerase may also turn out to be homologous to subunits of the eukaryotic polymerases.

However, there remain complexities in processes of poxvirus transcription beyond the scheme just outlined. It has been known for some time that cellular RNA polymerase plays a role in poxvirus replication. For example, virus production is sensitive to the RNA polymerase II inhibitor α-amanitin, although purified poxvirus RNA polymerase is not. Transcription of the host's genome is not involved. Two papers investigated this area in 1986. Morrison & Moyer (1986) showed that antibodies against a large subunit of eukaryotic RNA polymerase II cross-reacted with rabbit poxvirus RNA polymerase, and vice versa. The rabbit poxvirus subunit apparently corresponds to the VV subunit described above. This result is, of course, consistent with the sequence data of Broyles & Moss (1986). However, it was also found that, after infection, the cellular RNA polymerase subunit was translocated from the nucleus into cytoplasmic virus factories, and was packaged into virus, from which it could be copurified with viral RNA polymerase. Wilton & Dales (1986) obtained similar results with VV. They showed that treatment of infected cells with α-amanitin had little effect on virus transcription. In addition, mRNA extracted from such cells was translated normally in a cell-free system. The basis of the involvement of cellular RNA polymerase II is still therefore unresolved. Wilton & Dales suggested that it may act in transcription of RNAs which are involved, directly or indirectly, in control of translation.

Earl et al. (1986a) determined the sequence of a 5400 base pair region of VV DNA containing the mapped locus of the virus DNA polymerase, located the polymerase mRNA's 5' terminus and identified the DNA polymerase coding region as a 938 codon open reading frame. Protein sequence library comparisons showed that the amino acid sequence of VV DNA polymerase exhibited extensive similarity to DNA polymerase sequences of EBV and HSV. In addition, a
local similarity previously observed between the herpesvirus sequences and the sequence of adenovirus 2 DNA polymerase was present also in the VV sequence. This result shows clearly that the herpesvirus and poxvirus DNA polymerase genes are evolutionarily related, and this is the first reported genetic homology between these diverse groups. It appears plausible that all of these viral DNA polymerase genes have originated from some cellular gene; the best apparent candidate is the gene for the catalytic subunit of DNA polymerase α. This finding is thus quite reasonable, and ought not to have been surprising, but for a virologist steeped in accounts of the profound differences between the structures and strategies of poxviruses and herpesviruses it provides a cogent lesson in biological unity.

Many poxviruses produce skin lesions of a deep red colour, which is caused by local haemorrhage. Pickup et al. (1986) investigated the genetic basis of this effect in cowpox virus. White pock variants, which do not cause haemorrhage, can be readily isolated. Pickup et al. mapped the genome location primarily determining pock colour, using a set of deletion mutants, and showed that the protein responsible was an abundant early species of M, about 40 kDa. They determined the DNA sequence of a 1465 base pair region of the virus genome which contained the gene, and identified a probable polypeptide coding region specifying a protein of M, 37 kDa. The amino acid sequence of this species showed extensive similarity to members of a family of serine protease inhibitors, including antithrombin III, α1-antichymotrypsin, α1-antitrypsin and heparin cofactor III. Pickup et al. speculated that the virus protein could also be a protease inhibitor, with the following mechanism of action. Following virus infection of vascular tissue, the processes of blood coagulation may be activated as part of the normal host response. There are enzymes involved in this process, such as thrombin, which are serine proteases. The virus protein may inhibit these, interfering with normal clotting and producing a local haemorrhage.

Descriptions of these three poxvirus genes have been given with some detail because we consider them to represent particularly instructive examples for the origins and functioning of these large and complex viruses. Other studies on poxvirus DNA sequences were also reported. Shida (1986) mapped the location of the haemagglutinin gene of VV and determined its sequence. The haemagglutinin of orthopoxviruses is not a virion structural species, but is a glycoprotein expressed at the cell surface late in infection. The sequence analysis shows it to be a 315 amino acid chain (including signal sequence) with a C-proximal hydrophobic transmembrane region and five potential N-glycosylation sites. Hirt et al. (1986) analysed the structure of a VV gene encoding the major envelope antigen. In natural infection, mature VV particles are enclosed in a lipid membrane, in a process involving Golgi body membranes. Details of how the mature virus is released are obscure, but its surface antigens differ from those of the virus obtained in standard experimental systems by cell disruption. Both forms are infectious. The major antigen of the envelope, a species of M, 37 kDa, is not glycosylated but contains aliphatic fatty acids. The sequence analysis did not cast much light on the mechanism by which this protein becomes membrane-associated. It may not have a signal sequence for translation on rough endoplasmic reticulum, and there is no C-proximal potential transmembrane sequence, although two hydrophobic regions are present near the centre of the amino acid chain, which Hirt et al. suggested may be membrane-spanning.

Finally, the largest poxvirus sequence reported was that for the 16000 base pair HindIII D fragment of VV (Niles et al., 1986). In addition, Weinrich & Hruby (1986) determined a sequence of 5100 base pairs overlapping with HindIII D. Niles et al. (1986) interpreted the HindIII D sequence as representing 14 genes, since corrected to 13 (Rodriguez et al., 1986). These were distributed in both orientations and possessed very closely packed reading frames. Eight of these are the loci of temperature-sensitive (ts) mutations. Most of the gene products are of unknown function, but they include guanylyl transferase (an mRNA capping enzyme), an ATPase and also the protein mediating sensitivity of virus replication to rifampicin (Tartaglia & Paolletti, 1985; Weinrich & Hruby, 1986; Niles et al., 1986; Rodriguez et al., 1986).

We turn now to discussion of the uses of VV for carrying and expressing foreign genes. Work in this area has been mostly concerned with genes for proteins which form surface structures in the virus or organism from which they originated, and are immunogenic. VV recombinants have been used for basic characterization of such genes and, more commonly, with the intent to
develop a vaccine. In the first category, for instance, Cranage *et al.* (1986) studied an HCMV gene, identified by DNA sequence analysis, which was homologous to the HSV gene encoding the virion surface glycoprotein, gB. The HCMV gene was cloned and expressed in VV, and it was shown that the product was recognized by anti-HCMV sera. Further, rabbits immunized with the recombinant VV produced HCMV-neutralizing antibodies. This work illustrated a very direct route from identification of a gene by sequence analysis to evaluation of its role and importance.

In the second category, that of vaccine development work, many surface antigen genes, of diverse origins, have now been cloned into VV. No potential vaccines for use in the human population have yet progressed beyond animal model studies. F. Brown *et al.* (1986) have enumerated procedures and standards to be observed in implementing any such live recombinant vaccine. Antigens of human viruses which in 1986 were expressed in VV recombinants and evaluated in model systems included both external glycoproteins (G and F) of respiratory syncytial virus (Olmsted *et al.*, 1986; Elango *et al.*, 1986), and the envelope glycoprotein of HIV (Chakrabarti *et al.*, 1986; Hu *et al.*, 1986). In an animal system, aimed at providing a model for immunization against a retrovirus, Earl *et al.* (1986b) cloned the envelope protein of Friend leukaemia virus in VV, and showed that immunization with this recombinant could protect against challenge with the virus. Work with recombinant VV expressing the surface glycoprotein of rabies virus progressed to testing with the intended target species, foxes and raccoons, with satisfactory results (Blancou *et al.*, 1986; Rupprecht *et al.*, 1986).

Development of poxvirus vector systems continues. One aim of DNA sequencing studies is to delineate new cloning sites in non-essential genes, other than the commonly used thymidine kinase gene. For instance, Shida (1986) suggested the use of the haemagglutinin gene. Fathi *et al.* (1986) described a procedure for efficient insertion of an unselected sequence into a VV vector. In a novel development, Fuerst *et al.* (1986) cloned and expressed in VV the RNA polymerase gene of phage T7. Test plasmids were constructed with a chloramphenicol acetyltransferase (CAT) gene flanked by phage T7 transcriptional control elements. After transfection into cells previously infected with recombinant virus, and producing phage T7 RNA polymerase, the CAT gene was expressed at a very high rate.

**Papovaviruses and transgenic mice**

In this short section we register briefly two aspects of papovavirus work: the characterization of new virus-coded activities, and the study of virus genes by introduction into the germ line of mice.

The 'large tumour antigen' or T antigen of simian virus 40 (SV40) is an intensively studied protein of Mr 94000, with several distinct activities. It is involved in cell transformation by SV40, and is also required for regulation of virus gene expression and for virus DNA replication. Stahl *et al.* (1986) have now provided evidence that this versatile molecule possesses a DNA helicase activity, that is, it can unwind DNA at a replication fork in an ATP-dependent reaction. This accounted for the previously described ATPase activity of T antigen.

In tissue culture cells, the genome of bovine papilloma virus 1 (BPV-1) persists as a circular, nuclear plasmid at a stable copy number, suggesting the operation of a regulatory mechanism. Roberts & Weintraub (1986) and Berg *et al.* (1986) produced evidence that BPV-1 encodes a trans-acting function which negatively regulates replication of the virus genome. Two cis-acting loci at origins of replication were defined, which presumably represent the sites of action of the trans-acting repressor. The trans-acting function was encoded in the 5' portion of an open reading frame (El), whose 3' part was already known to encode a positive regulatory activity, so that the control of copy number is clearly of some intricacy. Roberts & Weintraub (1986) also showed that in a composite SV40-BPV-1 plasmid, BPV-1-encoded inhibition of replication was dominant over uncontrolled replication driven by the SV40 T antigen.

It has been possible for several years to introduce foreign DNA into a fertilized mouse ovum, then to implant this in the uterus of a mouse and allow development to maturity. A proportion of such animals may contain the foreign sequences chromosomally integrated; these are termed
'transgenic'. Several papers have reported such work with the SV40 genome region encoding T antigen, under varying transcriptional control (Brinster et al., 1984; Hanahan, 1985). Mice carrying these sequences developed tissue-specific tumours or other pathological conditions. In 1986, similar studies were reported for three other papovaviruses. Small et al. (1986a, b) produced mice carrying the early gene regions of JC virus or of BK virus. These are human viruses belonging to the Polyomavirus group, and their early regions correspond to the T antigen coding sequences of SV40. Both are widespread and usually asymptomatic, although JC virus is associated with a fatal demyelinating disease in persons with impaired cellular immunity. It was found that transgenic mice carrying sequences from these viruses developed tissue-specific tumours; in neural cells for JC virus, and in epithelial cells for BK virus. JC viral sequences also caused dysmyelination in the central nervous system. Thus, the general tissue specificity observed in the natural human host was preserved.

A particularly interesting study was reported by Lacey et al. (1986) on mice carrying the genome of BPV-1. This virus induces fibropapillomas in cattle and also, experimentally, in other species. Lacey et al. introduced a terminally redundant, linear copy of the whole 8000 base pair BPV-1 genome, a form chosen to facilitate the possibility of generating excised, circular forms in transgenic animals. They succeeded in producing a mouse whose genome contained about five head-to-tail copies of the sequence, and a line of mice was then bred. The mice developed normally into adults, but then began to display various skin abnormalities and tumours. Analysis of the state of the BPV-1 genome in normal tissues showed that it remained stably integrated. Tumour cells, however, contained extrachromosomal, full length, circular BPV-1 genomes, as well as showing amplification of integrated DNA. Because the skin tumours arose only slowly, and were particularly common in skin areas affected by trauma, it was clear that factors additional to the mere presence of the genome were necessary to induce papillomas.

In summary, the use of transgenic mice is providing a new dimension in the analysis of tissue specificities in the actions of virus genomes or parts thereof. Although the primary concern of the reports described here was with viral oncogenesis, one expects the technique also to find application outside this area.

**Picornaviruses**

Here we summarize several aspects of recent developments in picornavirus research, with emphasis on poliovirus. The genome of poliovirus is a 7500 residue, positive sense RNA molecule. This contains one large open reading frame, with a long, 740 nucleotide, 5' non-coding region and a shorter, 80 nucleotide, 3' non-coding region followed by a 3' poly(A) tract. The primary translation product is processed into various subspecies by proteolytic cleavage.

The non-coding regions 5' and 3' to the single open reading frame of the poliovirus genome are highly conserved between serotypes (Toyoda et al., 1984). Several recent papers have made a start in examining the functions of these regions, by producing poliovirus strains containing defined changes in non-coding parts of the genome and evaluating their properties. These experiments all depend on the genetic manipulation of plasmids carrying poliovirus sequences, followed by transfection into culture cells, transcription of the plasmid and rescue of virus.

Semler et al. (1986) produced a hybrid virus in which residues 220 to 627 of type 1 poliovirus, within the 5' non-coding region, were replaced with an equivalent section of the coxsackievirus B3 genome. The resulting recombinant was viable but, unlike the parental strains, exhibited ts growth. This was the first demonstration of a viable recombinant between two distinct picornaviruses. Semler et al. registered their surprise at the recombinant's viability, since their attempts to make small deletions in the 5' non-coding region of poliovirus had not yielded infectious virus (B. L. Semler et al., unpublished results). Racaniello & Meriam (1986), however, reported construction of a viable type 1 poliovirus strain with a single nucleotide deleted from position 10. This mutant too was ts, and also, unlike the wild-type, was sensitive to actinomycin D. The deleted C residue was at the base of a predicted hairpin loop structure in the RNA. Non-ts revertants were isolated and shown to contain a G to U mutation at position 34, which in effect restored the potential for a hairpin of wild-type length. Racaniello & Meriam carried out experiments to exclude virion uncoating, encapsidation or particle thermolability as the cause of
the $ts$ phenotype. They showed that viral RNA synthesis was $ts$, while viral protein synthesis was slower than with wild-type, but not $ts$.

In an analogous study of the other extremity of the poliovirus genome, Sarnow et al. (1986) described a poliovirus strain carrying an eight nucleotide insert in the 3' non-coding region, which exhibited a $ts$ phenotype. However, strains with two or 10 nucleotide inserts were viable but not $ts$. All of these studies demonstrate the importance of the non-coding regions in virus function, and indicate rather delicate and precise requirements of sequence and presumably higher structure. At present, it appears most likely that the $ts$ effects are mediated through some undefined roles in RNA synthesis. Another type of information on functional effects of the 5' non-coding region was provided by Evans et al. (1985) and Minor et al. (1986), who showed that a point mutation in the 5' non-coding region of the poliovirus type 3 live vaccine strain had consistently reverted to the wild-type in virus strains from cases of vaccine-associated poliomyelitis, and that virus with the same reversion was excreted rapidly and abundantly soon after primary vaccination.

During and after translation of poliovirus RNA, the primary translation product is cleaved at some 11 or 12 sites to yield the virus structural proteins and other species, including the replicase. The cleavage sites can be classified by the nature of the flanking amino acid residues. Thus, eight or nine cleavages are between glutamine and glycine residues (Q-G in single letter code) and these are known to be catalysed by virus-coded protein 3C. Of the other cleavages, one represents the maturation of capsid protein VP0 into VP2 and VP4 at a late stage in capsid assembly, and may be autocatalytic. The remaining two occur at tyrosine-glycine (Y-G) pairs. The apparently more important of these separates the precursor of the capsid proteins (designated P1) from the non-structural proteins' precursors (P2 and P3) during translation.

Toyoda et al. (1986) examined the mechanism of this cleavage using fragments of the poliovirus reading frame expressed in $E. coli$. They found that cDNA sequences spanning the P1–P2 junction gave correctly cleaved product only when the region encoding protein 2A was present and intact. It was therefore concluded that 2A is a second processing protease encoded by poliovirus. The N terminus of 2A is the cleavage site between P1 and P2, and it seems that this site is cleaved in an intramolecular reaction by the newly synthesized 2A polypeptide. Toyoda et al. (1986) and Lloyd et al. (1986) point out that 2A and 3C contain a similar short sequence around a cysteine residue, thought in 3C to constitute the active site, so that 2A is probably also a thiol protease. In poliovirus-infected cells, yet another significant protein cleavage takes place: host protein synthesis is inhibited by a mechanism involving cleavage of a component of the cap-binding protein (CBP) complex, which is active in associating capped mRNAs with ribosomes. However, the CBP cutting activity does not appear to be either the 3C or 2A protein (Lee et al., 1985; Lloyd et al., 1985, 1986), so that this last protease has yet to be accounted for. In addition, a further complexity still lurks in the CBP protease system, since Bernstein et al. (1985) have demonstrated that mutants of 2A are defective, presumably by an indirect mechanism, in CBP cutting and inhibition of host protein synthesis.

A second protease encoded in the genome of foot-and-mouth disease virus (FMDV) was also described in 1986 (Strebel & Beck, 1986). FMDV is classified in a different genus of the family *Picornaviridae* from poliovirus. FMDV possesses a 3C protease, but the genomes of these two viruses differ in that FMDV encodes an additional polypeptide, designated L, as the proximal part of its polyprotein, and the reading frame in FMDV corresponding to the 2A species is only 16 amino acids long (Robertson et al., 1985). Strebel & Beck (1986) have now shown that the L protein has a protease activity which catalyses its own excision, and they note that the amino acid sequence of the L protein has local similarities to that of 3C. However, the function of L protein in the virus cycle, outside this self-processing action, remains obscure, since present evidence does not associate L with any other of the polyprotein processing events.

Another aspect of the enzymology of picornavirus infection also advanced in 1986. Replication of poliovirus RNA in a cell-free system requires the presence of a protein factor of cellular origin, in addition to the virus-specified RNA polymerase. Andrews et al. (1985) had suggested that this host factor could be a terminal uridylyltransferase (TUT), that is, an enzyme which adds U residues to the 3' terminus of an RNA molecule. In a further paper (Andrews &
Baltimore, 1986a) it was then shown that TUT activity purified extensively with, and was inseparable from, the host factor. Preparations of TUT were capable of adding several U residues, around five in the conditions used, to the 3' poly(A) tract of poliovirus RNA. This oligo(U) then formed a hairpin with the poly(A) for polymerase to initiate copying of the genomic RNA. In accord with this model, it has been established that the product RNA made in vitro can be found covalently linked to the template to give molecules of twice the template length (Young et al., 1985; Hey et al., 1986; Lubinski et al., 1986). It should be noted that the findings of Andrews & Baltimore (1986a) are in conflict with those of Morrow et al. (1985), who reported that the host factor was a protein kinase. The TUT mechanism in its present form is incomplete, since it does not account for the presence of a protein molecule (VPg) at the 5' terminus of poliovirus minus strand RNA extracted from cells, and it seems that some processing events must follow copying of the RNA (Young et al., 1986; Andrews & Baltimore, 1986b).

To close this section, we describe work on the structural basis of the actions of two antiviral compounds, designated WIN 51711 and WIN 52084 (Smith et al., 1986). These are members of a family of compounds known to inhibit picornavirus replication by preventing uncoating of the virus in the newly infected cell. Each has a structure consisting of a seven carbon aliphatic chain linked at one end to a heterocyclic isoxazole ring, and at the other end through a phenyl group to an oxazoline heterocycle. They differ by one methyl group. Smith et al. exposed crystals of rhinovirus 14 to these compounds to obtain complexes with a stoichiometry of approximately one molecule bound per capsid protomeric unit. Following the previous determination of the three-dimensional (3D) structure of rhinovirus 14 by X-ray diffraction (Rossmann et al., 1985), they then obtained diffraction measurements on crystals containing the WIN compounds, and were able to visualize the position of the antiviral molecule in the virion protomer. The isoxazole ring was found to lie within a hydrophobic interior pocket in VP1, with the aliphatic connector leading out of the pocket and the other rings blocking a channel which runs from the exterior to the RNA-containing interior of the virion. This suggested two possible mechanisms for the antiviral action: the compounds might act to prevent collapse of the hydrophobic pocket as a part of the uncoating process, or alternatively the blockage of the channel could prevent necessary passage of ions or small molecules. This study represents the first stereochemical analysis of an antiviral compound at this resolution, and gives fresh insight into the structural dynamics of picornaviruses.

**Negative strand RNA viruses**

Very many nucleotide sequence studies on negative strand genomes continue to be published, in particular for paramyxoviruses. Here we mention two large analyses only, on the genomes of Sendai virus and rabies virus. We note that the hepatitis delta agent described above should perhaps also be regarded formally as a negative strand RNA virus, albeit of a new type.

Completion of the genome sequence of Sendai virus, of some 15400 residues, was reported by Shioda et al. (1986). This is the first complete description of a paramyxovirus sequence. The genome possesses seven genes, two of which are overlapping. The most distal gene encodes the RNA polymerase-associated protein L, with a calculated M, of 224005. However, another report (Morgan & Rakestraw, 1986) conflicts with Shioda et al. (1986) in the detail of the reading frame arrangement in the region around the start of the L gene, and this remains to be resolved. Only very limited similarity of amino acid sequence was detectable between the Sendai virus L and the apparently corresponding polymerase of the rhabdovirus vesicular stomatitis virus (VSV), also called L, so that ideas on the evolutionary relationship between these two groups remain nebulous.

Two papers by Tordo et al. (1986a, b) reported the sequence of the 3' terminal half of the genome of rabies virus, encoding four of the five genes and stopping within the fifth and last, the L gene. Both rabies virus and VSV belong to the family Rhabdoviridae, but are assigned to separate groups, called respectively the Lyssavirus and Vesiculovirus genera. These two viruses have similar genomic organizations, but exhibit only low homologies between corresponding protein sequences. The most striking difference observed by Tordo et al. was that rabies virus...
possesses a tract of 423 nucleotides between the G and L genes. This has, at its extremities, apparently appropriate consensus sequences for the 5' and 3' termini of an mRNA, but it does not contain any open reading frame longer than 18 codons. Tordo et al. noted that this structure is intermediate between the VSV layout, where the G and L genes are closely contiguous, and the case of a fish rhabdovirus, haematopoietic necrosis virus, where an extra, functional gene lies in the corresponding genome location (Kurath et al., 1985). They thus favoured the interpretation that the rabies sequence is a non-functional remnant of a gene.

**Virion structure**

To finish this review, we describe work on the structures of three viruses, namely the present status of X-ray crystallographic analysis of the major virion protein of adenovirus 2, current ideas on polyoma virus structure, and application of cryo-electron microscopy to studying the enveloped Semliki Forest virus (SFV). In addition, an aspect of the crystallographic study of rhinovirus 14 was described in the picornavirus section.

The major event in structural studies of animal viruses in 1986 was the determination of the 3D structure of the adenovirus 2 hexon protein by X-ray crystallography (Roberts et al., 1986). The icosahedral capsid has 252 capsomers: 240 of these consist of trimers of the hexon polypeptide, while the remaining 12, located at the vertices, are composed of pentamers of another protein, from which project the characteristic fibres of adenovirus. Because of the size and complexity of the adenovirus particle, structural analysis has proceeded by determination of the 3D structure of the hexon from crystals of isolated trimer, together with electron microscopical analysis and model building to study the relationships between trimers in the capsid.

The hexon polypeptide contains 967 amino acid residues, and is the longest protein chain whose structure has been determined crystallographically. It is clear from the account of Roberts et al. that deduction of the path of the polypeptide chain from electron density maps required considerable effort, and at present there remain five internal gaps in the chain tracing. Apart from the size of the molecule, they attributed these difficulties to the complex association of chains from each subunit in the trimer, and to the relatively low content of regular secondary structure.

The hexon monomer was found to contain two so-called β-barrels, denoted P1 and P2, each consisting of eight strands of β-sheet. P1 and P2 have very similar secondary structures, but no visible similarity of primary sequence. In the trimer, the six β-barrels are arranged in parallel, with a hexagonal cross-section and a central cavity, to form the 'pedestal' of the trimer, as indicated in Fig. 4. This fascine arrangement has a height of 5.2 nm. Above it is a 'top' of height 6.4 nm, containing three pronounced 'towers', each of which is composed of polypeptide chains from all three subunits, highly interwoven. Roberts et al. drew particular attention to the complexity of this feature and the stability it must impart to the whole structure. An additional protein may be involved in the formation of the trimer, and they speculated that this may be necessary to achieve proper association of the chains in the towers. In the virion, each trimer is arranged with the pedestal perpendicular to the virion surface and the base of the pedestal buried. It is not possible to arrange the 12 trimers on each face of the icosahedral capsid to give equivalent bonding between each adjacent pair, and possible interactions between trimer capsomers are being evaluated by model building.

Several comparative aspects emerged from these analyses. First, comparison with the adenovirus 5 sequence (Kinloch et al., 1984) showed that differences are mostly confined to surface regions at the top of the trimer. Secondly, the P1 and P2 domains are so similar in 3D structure as to suggest that they may have evolved from an ancient gene duplication. Finally, Roberts et al. (1986) pointed out that the topologies of the β-barrel structures from the adenovirus hexon and from structural proteins of picornaviruses are identical, and have so far been observed only in these viruses. However, in adenovirus the barrels are arranged quite differently to the picornavirus examples, where the barrels are essentially parallel to the virion surface. The barrel structure therefore appears to be a versatile building module, and a common evolutionary precursor may be indicated.
Salunke et al. (1986) presented an electron microscopical study on the self-assembly of polyoma virus capsids from the purified major capsid protein, VP1. Polyoma has an icosahedral capsid containing a total of 72 capsomeres, 60 of which each interact with six neighbours (that is, are six-coordinated), and 12 of which, at the vertices, are five-coordinated. It was previously shown, by low-resolution X-ray diffraction of capsid crystals, that all of the capsomeres are pentamers of VP1 (Rayment et al., 1982). This was contrary to expectations from ideas on quasi-equivalent bonding of capsomeres, which had predicted that the six-coordinated capsomeres would be hexameric, and implies non-equivalent modes of bonding between capsomeres.

Salunke et al. (1986) expressed and purified VP1 from an E. coli recombinant, so that complications from normal, variable post-translational processing of VP1, and from any activities of the minor capsid proteins VP2 and VP3, were avoided. They found that VP1 preparations spontaneously assembled to give pentamers only, which then associated in high salt to give larger assemblies, including capsid-like structures. Thus, post-translational modification of VP1, and presence of the minor capsid proteins, are not essential for capsid formation. The pathway for assembly of pentamers into capsids is not known. However, it is clear that it must involve switching of bonding specificities as it proceeds, and this must depend on the previous steps in the process.

Vogel et al. (1986) investigated the structure of the alphavirus SFV using cryo-electron microscopy. In this relatively new technique, a suspension of unfixed and unstained virus is prepared in a thin, vitrified layer, and is maintained and observed at a temperature below −160 °C in a transmission electron microscope. In this system, variations in focus can be used to probe a structure to different depths (Stewart & Vigers, 1986). Images of many, variously oriented virus particles can then be combined and presented by computer methods.

Semliki Forest virus possesses an icosahedral nucleocapsid enclosed in a lipid envelope, through which are inserted glycoprotein spikes. With alphaviruses, the lipid envelope is closely drawn around the nucleocapsid, and the surface spikes form a regular array, presumably...
through interaction of their interior domains with the nucleocapsid’s exterior. Vogel et al. (1986) showed that there are 80 surface spikes, and proposed from the triangular shape of the spike heads that each of these is a trimer, containing three copies each of the three spike polypeptides. Two types of spike head differing in the detail of their shape were resolved, corresponding to 20 spikes on the threefold axes of the particle (that is, in the centres of icosahedral faces) and 60 peri-pentonal spikes (that is, surrounding the icosahedral vertices). Other morphological details were also resolved, including a spreading of the spikes just exterior to the lipid membrane, and structures joining adjacent threefold spikes. Vogel et al. pointed out that this latter bonding defines an icosahedral shell with 20 capsomeres, and they speculate that this could represent the major structure in formation of the particle, with the peri-pentalonal spikes playing a less important, filling-in role.

We are grateful to many colleagues for various forms of assistance, and in particular to A. Davison, S. Inglis, V. Mautner, C. Pringle, B. Rima, F. Rixon, D. Rowlands, V. ter Meulen and R. Weiss.

REFERENCES


Review: Animal virus research in 1986


