Some Highlights of Virus Research in 1988

By ROGER HULL 1* AND DUNCAN J. McGEOCH 2

1John Innes Institute, AFRC Institute of Plant Science Research, Colney Lane,
Norwich NR4 7UH and 2MRC Virology Unit, Institute of Virology, University of Glasgow,
Church Street, Glasgow G11 5JR, U.K.

Our review of the highlights in virus research in 1988 differs from the previous highlight reviews (McGeoch et al., 1986, 1987, 1988c) in that it includes work on plant viruses as well as animal viruses. This change is an attempt to make these reviews more general and to indicate areas of similarity and contact between the viruses of different kingdoms. Nonetheless it remains a highly selective view of progress in a wide and expanding scientific field.

We start by describing the current work on three specific groups of viruses, herpesviruses, hepadnaviruses and hepatitis delta virus (HDV). This leads into a discussion of mechanisms by which viral genes are replicated and expressed and then into some recent findings on viral genome organization, virus structure, the use of transformation in the study of plant viruses, and we conclude with an account of new disease agents. For the first time we do not have a section devoted to AIDS and human immunodeficiency virus (HIV); this reflects both the selective nature of our account, and also an evaluation that HIV research has now emerged from its initial phase of rapid discovery, so that truly notable advances are becoming less frequent, despite the very large efforts being made.

Research on herpesviruses

In 1988 determination of the DNA sequence of herpes simplex virus type 1 (HSV-1) was completed (McGeoch et al., 1988a; Perry & McGeoch, 1988). The sequence contains 152260 residues of overall base composition 68.3% G + C. This is the third herpesvirus sequence to be published, after those for Epstein–Barr virus (EBV) and varicella-zoster virus (VZV). The HSV-1 sequence appears to contain at least 72 genes. At present there is no significant information on function for approximately half of these.

During 1987 and 1988, results were published concerning transcription of HSV DNA in latently infected neurons. It has been found, for both experimental and natural infections, that in these neurons transcripts originate largely from the long repeat element of the genome (Rock et al., 1987; Stevens et al., 1987; Wagner et al., 1988a, b; Spivack & Fraser, 1987, 1988; Steiner et al., 1988; Wechsler et al., 1988a, b; Krause et al., 1988). Their 5' termini are from a region without any assigned protein-coding function (the largest such region in the genome) and their 3' portions overlap partially, in the opposing sense, with one of the immediate early (IE) genes, which encodes a transcriptional regulator (IE110 or ICP0). These latency-associated transcripts (LATs) are up to about 2 kb, include spliced species and are not polyadenylated. While there is great interest in LATs as a possible mechanism in establishment or maintenance of the latent state, it has to be said that their significance is still obscure. The LAT region may not be protein-coding (Perry & McGeoch, 1988), and no clear function has yet been assigned to the transcripts. One possibility is that they inhibit expression of the IE110 gene, by annealing to its mRNA. Use of a deletion mutant of HSV has indicated that LAT expression is not required to establish latent infection in mice (Javier et al., 1988).

A significant advance in the study of herpesvirus DNA replication was reported in 1988, with the enumeration and characterization of a set of seven HSV-1 genes minimally required for DNA synthesis (genes UL5, UL8, UL9, UL29, UL30, UL42 and UL52; Wu et al., 1988; McGeoch et al., 1988b). This made use of an assay in which fragments of HSV-1 DNA bearing
such genes were introduced into cells and their effect on amplification of a test plasmid carrying an HSV-1 origin of DNA replication was measured. The set includes genes for the previously known DNA polymerase (UL30), the single-stranded DNA binding protein (UL29) and a second DNA-binding species (UL42). Gene UL9 has subsequently been shown to encode the origin-binding protein (Olivo et al., 1988). Of the remaining three genes, UL5 may encode the DNA helicase found in infected cells (Crute et al., 1988; Zhu & Weller, 1988) and the functions of UL8 and UL52 remain unknown. Interestingly, the set did not include the gene for the deoxyribonuclease, which is highly active in infected cells, and no role is yet defined for this enzyme. In a related area, the UL2 gene of HSV-1 and the corresponding HSV-2 gene were identified as encoding the DNA repair enzyme uracil-DNA glycosylase (Worrad & Caradonna, 1988).

Another advance in analysis of herpesvirus DNA replication reported in 1988 was the identification by Hammerschmidt & Sugden (1988) of a region in the genome of EBV which, when cloned in a plasmid, caused amplification of the plasmid in cells supporting lytic growth of EBV. This element, termed oriLyt, is thus thought to be the long sought EBV origin of lytic phase DNA synthesis. OriLyt is distinct from the previously characterized oriP, which acts in maintenance replication of the EBV episome in latently infected cells. The DNA products obtained in the two cases also differ: oriP-dependent replication yields circular plasmids, whereas oriLyt gives concatemeric molecules, the form expected for packaging into virions. In most EBV strains there are two copies of oriLyt, present in the widely separated repeat elements DL and DR. OriLyt is complex, comprising two regions of 321 and 374 bp separated by 263 bp of sequence not required for origin function, and it contains multiple repeats. It does not resemble the previously described origins of HSV and of VZV, which appear to be considerably more simple.

An aspect of transcriptional regulation in HSV-infected cells concerns the role of one of the virus's structural proteins in activating expression of the set of IE genes. Vmw65 (or VP 16) is the major protein species in the tegument layer of the virion, and several years ago it was found that this polypeptide acted in the newly infected cell to stimulate transcription of the IE genes, by interaction with upstream sequence elements. It has now been demonstrated that Vmw65 does not itself bind DNA, but that specific binding can be observed in vitro when both Vmw65 and an extract of cellular proteins are supplied (Marsden et al., 1987; McKnight et al., 1987; O'Hare & Goding, 1988; Preston et al., 1988). The active cellular species is probably the previously known nuclear factor III or octamer binding protein. It has also been shown that Vmw65 possesses a functional 'acid tail' (an acidic C-terminal region) as found for a number of other transcriptional activators (Friedman et al., 1988; Sadowski et al., 1988). This protein thus has two distinct roles, virion structural and regulatory, and the latter function appears to be closely tied to cellular transcriptional control systems. It has now been demonstrated by mutational analysis of plasmid-borne copies of the Vmw65 gene that portions of the protein crucial for transcriptional activation and for the structural function can be distinguished (Ace et al., 1988).

An interesting result of 1988 in the human cytomegalovirus (HCMV) system emerged from sequence analysis of a previously unknown HCMV gene, the encoded protein of which displayed sequence similarity to class I HLA proteins (Beck & Barrell, 1988). It had previously been shown that preparations of this virus bind $\beta_2$ microglobulin, which is normally associated with class I HLA antigens on cell surfaces, and that this interaction could enhance virus infectivity by enabling binding to cell surface HLA antigens (Grundy et al., 1987a, b). In addition, HCMV from urine was found to carry $\beta_2$ microglobulin, unlike virus grown in tissue culture (McKeating et al., 1987). Thus, the system may represent a device for evading the host's immune system. It was thus suggested that the HCMV-encoded HLA-like protein is the virion surface species responsible for binding of $\beta_2$ microglobulin, although direct evidence of this is still lacking.

The DNA sequence data for herpesviruses which have emerged in the last several years have, in general, indicated relationships between herpesviruses largely concordant with the existing taxonomic scheme for the family, which was based largely on biological properties. However, an interesting study by Buckmaster et al. (1988) has now provided new insights into herpesvirus
lineages. These workers examined the genomes of Marek's disease virus (MDV) and of herpesvirus of turkeys (HVT), two lymphotropic avian herpesviruses, by cloning one or two hundred randomly generated fragments of the virus DNAs into an M13 vector and determining their sequences. Possible encoded amino acid sequences were then compared with the complete sets of VZV- and EBV-encoded sequences, and many fragments of MDV and HVT genes thus identified. Further work located these on the restriction maps of their genomes. It emerged that the genomes of the two avian viruses were much more closely related to VZV DNA than to that of EBV, as judged by the number of homologues found, the strengths of the similarities and the arrangements of genes. Insofar as MDV and HVT are, like EBV, lymphotropic whereas VZV is neurotropic, this result was not expected. Clearly, then, more than one major lineage of lymphotropic herpesviruses exists.

**Research on hepadnaviruses**

The use of systems for expressing viral gene products has led to several experiments to further the understanding of the functioning of hepadnavirus gene products. Other aspects of this virus group were reviewed in 1987 (McGeoch et al., 1987). This group of DNA viruses which replicate by reverse transcription have genomes with four open reading frames (ORFs) encoding the core or nucleocapsid protein (C protein), the polymerase with reverse transcriptase, DNA polymerase and ribonuclease H activity (P protein), the surface antigen (S protein), and the X protein. The mature C protein is produced from a precursor protein, the pre-C protein, by removal of the N-terminal part. Three proteins, pre-S1, pre-S2 and S, are produced from different transcripts of the surface antigen region which give ORFs starting at each of the first three ATGs.

The significance of the pre-C region, which is associated with the production of the e antigen, was analysed using an *in vitro* system for cotranslational processing at microsomes of transcripts from cloned DNA (Bruss & Gerlich, 1988). The P25 pre-C is cleaved cotranslationally as a signal peptide to give the P23 which was translocated to the lumen of the microsomes. An arginine-rich C-terminal domain was suggested to act as a novel type of translocation stop signal as well as acting as a binding region for the RNA template for reverse transcription. A model (see section on Virus structure for more details) for the structure of the C protein, as well as predicting a basic structure (an eight-stranded β-barrel) similar to that of some RNA virus capsid proteins, indicates that this C-terminal region is unstructured (Argos & Fuller, 1988). DNA-binding studies suggested that, in addition to this C-terminal basic region, there was another region close to the precore/core boundary which also had DNA-binding capacity (Matsuda et al., 1988). Thus it is interesting to note that this might be analogous to the ‘cys’ sequences found in retrovirus core proteins and that, like retroviruses, hepadnaviruses have two nucleic acid-binding domains.

Although hepadnavirus particles had for some time been known to contain nucleic acid polymerase activity and a candidate gene for reverse transcriptase had been suggested, there was no direct evidence that this gene was expressed *in vivo*. This gene product has now been found in sera from humans infected with hepatitis B virus (HBV) (Stemler et al., 1988) and has been identified in HBV particles (Bavand & Laub, 1988; Mack et al., 1988). Two proteins with reverse transcriptase activity were found in activity gels, that of 90K being the size expected from the sequence and one of 70K which was presumed to be product of processing (Bavand & Laub, 1988); immunoblotting detected only a product of 65K (Mack et al., 1988). It is possible that processing is achieved by a protease domain identified by sequence homology with retrovirus gene product sequences (Miller, 1988). Radziwill et al. (1988) developed a technique for relaxing the structure of duck HBV cores without destroying enzymic activity. They showed that, unlike in retroviruses, the viral polymerase is tightly associated with the core structure and can not switch templates. They suggest that this is the mechanism to sequester the reverse transcriptase so that it does not cause untoward damage to the host cell. However, Offensperger et al. (1988) reported that they could use the HBV polymerase activity on exogenous templates including poly(rA)-poly(dT)12 and poly(rA)-poly(dT)12-18 which indicated that it had both RNA-dependent and DNA-dependent DNA polymerase activities. It is possible that the differences
in the ability to reverse transcribe exogenous templates were due to differences in procedures for extraction or for permeabilizing the cores. The reverse transcribing and polymerase activities were inhibited by suramin, chloroquine, phosphonoformate and ara-ATP (Offensperger et al., 1988).

The 5' end of HBV minus-strand DNA has a covalently attached protein. Using antisera specific to the P gene it has now been shown that this 5' protein is derived from the N-terminal domain of the P gene product (Bosch et al., 1988; Bartenschlager & Schaller, 1988) and that it is attached to the minus-strand DNA by a phosphodiester link to a tyrosine residue.

S protein production has been analysed in an adenovirus-based expression system (Molnar-Kimber et al., 1988). When expressed alone pre-S1 is not secreted whereas pre-S2 and S are. The secreted pre-S2 and S proteins assemble into 22 nm particles both in the adenovirus-based system and in one using bovine papillomavirus (Molnar-Kimber et al., 1988; Yoneyama et al., 1988). When the three proteins are expressed together the secretion of pre-S2 and S is inhibited more than 95% by the presence of pre-S-1 (Molnar-Kimber et al., 1988). This is taken as suggesting differential biosynthesis, transport and processing of the three proteins during HBV infection and could lead to more efficient assembly and secretion of virions. In a study of protein expression after infection of ducks with duck HBV, Yokosuka et al. (1988) showed that C protein could be detected as early as 24 h after infection; pre-S1 and -S2 were first observed after 3 days. They suggested that C protein was used for the initial packaging of the nucleic acid and that pre-S1 and -S2 were used for the production of virions near the end of the replication cycle. Within the S gene transcript, but 3' of the coding region, there is an enhancer region. Removal of this enhancer gives a significant reduction in the level of S gene expression but does not affect the specificity of transcript initiation or the stability of the mRNA (Bulla & Siddiqui, 1988). The enhancer is cell-specific (Factor et al., 1988).

Some information has accrued on the expression and function of the X protein. It is probably translated from a heterogeneous population of non-polyadenylated transcripts of about 0.65 kb which are found in the nuclei of liver cells (Kaneko & Miller, 1988); these transcripts are present in very low abundance, representing less than 0.1% of the virus-specific RNA. The X protein appears to trans-activate the enhancer noted above (Spandau & Lee, 1988). It will also trans-activate the simian virus 40 (SV40) and Rous sarcoma virus (RSV) enhancers in CV-1 cells but not the SV40 enhancer in COS-1 cells (Spandau & Lee, 1988). However, the X gene product is expressed in COS-1 cells (Koike et al., 1988).

Research on hepatitis delta virus and RNA self-cleavage

Considerable progress was made in 1988 in the understanding of HDV; thus although this unusual virus was reviewed 2 years ago (McGeoch et al., 1987) its similarities to plant systems warrant discussion here. HDV requires hepadnavirus helper function(s) for propagation in human, chimpanzee or woodchuck hepatocytes. The 36 nm HDV particles contain a single-stranded circular RNA genome of 1.7 kb, encapsidated in lipid and HBV S protein; within the particles are HDV-specific proteins termed hepatitis delta antigen (HDAg). HDAg has now been shown to be encoded by the largest ORF in the antigenomic RNA (Weiner et al., 1988; Kuo et al., 1988a). HDAg is usually found as two protein species of 24K to 25K and 26K to 29K but the relationship between these two molecules is not yet clear. At least one of the proteins has been shown to be phosphorylated at serine residues and to bind to HDV genomic RNA (Chang et al., 1988).

The RNA of HDV has several properties in common with plant-infecting viroids, virusoids and satellite RNAs. These include being single-stranded circular RNA, having considerable base-pairing giving an unbranched rod structure and there being multimeric lengths of genomic and antigenomic RNAs (see Sharmeen et al., 1988 for summary of similarities). These similarities now extend further with the demonstration that HDV RNA undergoes self-cleavage both in vivo and in vitro (Kuo et al., 1988b; Sharmeen et al., 1988). The self-cleavage occurs at a single specific location in both genomic and antigenomic RNA. At least in antigenomic RNA the cleavage gives junction fragments consistent with 5'-hydroxyl and cyclic 2',3'-monophosphate ends (Sharmeen et al., 1988), similar to those found in certain plant viral satellites and
viroids. This cleavage is adjacent to the 3' end of the HDAG ORF and thus could process the circular RNA to give an mRNA. The minimum length of contiguous sequence needed for self-cleavage is 30 bases 5' and 74 bases 3' of the cleavage site. However, unlike some other self-cleaving RNAs this sequence could not be folded into the 'hammerhead' structure (Kuo et al., 1988b) predicted for the cleavage sites of the plant pathogens.

This leads us directly to a discussion of the recent work on the self-cleaving RNA molecules involved in plant infections. Certain naturally occurring RNA molecules, including a viroid (avocado sunblotch viroid), and the satellite RNAs of tobacco ringspot virus and lucerne transient streak virus, possess this property of self-cleavage. The sites of the cleavage are specific and are associated with domains of conserved sequence and secondary structure ('hammerhead' structures). Haseloff & Gerlach (1988) separated the substrate and cleavage activity of these sequences and showed that the substrate contained none of the conserved secondary structures whereas the fragment with cleavage activity did. This led them to design the 'ribozyme', an RNA molecule which could cleave substrate RNA at a specific site. The ribozyme consists of two basic domain structures. The flanking sequence domain comprises nucleotides which will base pair with the substrate on either side of the cleavage site, which in the example they showed (the satellite of tobacco ringspot virus) was GUC. This domain positions the conserved sequence domain that has the catalytic activity. Synthetic ribozymes cleaved the mRNA of chloramphenicol acetyltransferase in vitro. Haseloff & Gerlach point out various potential applications of ribozymes including the major one of inactivating gene transcripts in vivo, which could have implications in approaches to non-conventional resistance.

Sequences and genome organizations

In this section we describe information on genome organization derived from the determination of the sequences of members of virus groups for which full sequence information was not previously available. There were, at least, five such sequences reported in 1988, those of the animal-infecting pestiviruses and spumaretroviruses and of three groups of plant-infecting viruses.

Three viruses are classified in the genus Pestivirus: bovine viral diarrhoea virus (BVDV), hog cholera virus and border disease virus of sheep. These are enveloped viruses with monomolecular single-stranded RNA genomes, and are at present classified as a genus in the family Togaviridae. Although economically important, they have not been well characterized, owing in part to poor growth in cell culture. In a series of three papers, Collett et al. (1988a, b, c) changed this situation profoundly for BVDV by analysing the sequence of its genome, using the data to identify virus-specified protein species and locate the genome regions encoding them, and to re-examine relationships with other virus groups.

A library of cDNA clones was constructed from genomic BVDV RNA, ordered and used to determine the nucleotide sequence (Collett et al., 1988 c). The final nucleotide sequence obtained had 12573 residues. Since the genomic termini were not analysed directly, the determined sequence may be incomplete at the extremities. The genome RNA does not contain a Y poly(A) tract. The sequence contained one large ORF in the genomic (plus-strand) orientation, representing 3988 codons (Fig. 1). Thus it seems that the genome is expressed as a polyprotein, in the manner of picornaviruses or flaviviruses, and this correlates with the absence in infected cells of subgenomic RNA species.

A series of bacterial plasmids each expressing a portion of the BVDV polyprotein fused to β-galactosidase was then made, and antisera were raised against these polypeptides (Collett et al., 1988b). The antisera were used to detect authentic viral protein species by precipitation from extracts of infected cells. In this way eight polypeptides, as well as a number of putative precursors of higher Mr, were found and mapped on the genome; all are presumed to arise by proteolytic processing of the primary polypeptide translation product. Several of the polypeptides, from a region near the 5' end of the genome, were glycosylated. The strategy failed for two regions predicted to encode particularly hydrophobic amino acid sequences. Proteins representing 83% of the genomic coding capacity were detected, to give the first map of pestivirus genome function.
This description of pestivirus genome organization differs significantly from that of the alphaviruses, which constitute the major genus of the Togaviridae, indicating that the present classification is not valid. Recently another group, the flaviviruses, were removed from the Togaviridae to form their own family (Flaviviridae). Collett et al. (1988a) pointed out a compelling set of similarities between flavivirus and pestivirus genome organizations. Both have single-strand plus-sense RNA genomes without poly(A) tails. Both exhibit a single ORF occupying almost all of the genome. However, there is no extensive RNA sequence similarity, and the BVDV sequence is some 2 kb longer than those of flaviviruses. Nonetheless, when the hydrophobicity profiles of the polyproteins of BVDV and of yellow fever virus are compared, they were seen to be convincingly similar, and the proposed proteins of BVDV correlate well with those of yellow fever virus. Collett et al. (1988a) therefore proposed that these two groups are evolutionarily related and that the Pestivirus genus should be reassigned to the Flaviviridae family.

1988 saw genome sequences for several plant viruses including members of three virus groups which had not been previously sequenced, the potexvirus, luteovirus and tymovirus groups. Each of these viruses has a positive-sense, single-stranded RNA genome. The genome organizations of these viruses are shown in Fig. 1. Five ORFs are found in the genomes of the
two potexviruses, potato virus X (PVX) and white clover mosaic virus (WCIMV) (Huisman et al., 1988; Forster et al., 1988) and in the genome of the luteovirus, barley yellow dwarf virus (BYDV) (Miller et al., 1988), and four in the tymovirus, turnip yellow mosaic virus (TYMV) (Morch et al., 1988). For each virus the product of the largest ORF shows homology with RNA polymerases; that of BYDV shows a striking similarity to the putative polymerase of carnation mottle virus (CarMV) (Millet et al., 1988). The polymerase and the coat protein are the only genes to which functions can be ascribed with any degree of confidence. The functions of the other gene products await elucidation. However, some intriguing sequence homologies have been found between (potential) gene products from these viruses and those previously sequenced. The coat protein of BYDV shows a distant relationship to those of some other icosahedral viruses, e.g. tomato bushy stunt virus, southern bean mosaic virus and CarMV (Miller et al., 1988), perhaps reflecting similarities in protein structure. These authors also noted that the 50K readthrough protein from the BYDV coat protein (CP28) has similarities to the readthrough protein from the coat protein of beet necrotic yellow vein virus (BNYVV), a rod-shaped virus. The 12K protein of PVX shows appreciable homologies at the amino acid level with proteins of similar sizes from BNYVV, barley stripe mosaic virus and WCIMV (Huisman et al., 1988). These sequence homologies are providing further evidence for the 'modular' evolution of RNA viruses, a point made by Miller et al. (1988) in discussing the data on BYDV.

The sequence data throw further light on the methods of expression of these viruses. Miller et al. (1988) suggest that there is a frameshift between the two 5' ORFs of BYDV which would give a product of the size found by in vitro translation. They also suggest the possibility of readthrough from the coat protein to the 3' ORF though they were unable to ascribe a function to the readthrough product.

The first full sequence of a member of the third subfamily of retroviruses, the spumaretroviruses, has been completed with the report of the sequence from the 5' end of the genome of human spumaretrovirus (HSRV) (Maurer et al., 1988) adding to the previous report of the 3' sequence (Flugel et al., 1987). The genome of more than 12 kb is larger than those of other retroviruses and has seven ORFs. The 5' ORFs, gag, pol and env have various features in common with those of oncoviruses. There is an overlap of 22 nucleotides between gag and pol, and the seven nucleotides just 5' of two consecutive stop codons in the gag sequence are identical to those in the oncovirus RSV. This is suggestive of the HSRV pol gene being expressed as a polyprotein by frameshift from the gag gene as has been shown for RSV. Sequence homology analysis showed that the HSRV pol gene had the reverse transcriptase, RNase H and integrase domains characteristic of retroviruses and that the reverse transcriptase and integrase domains had a high degree of homology with those of the oncovirus murine leukaemia virus. One difference from other retroviruses is that the gag gene lacked the nucleic acid-binding 'cys' sequence and had in its place a strongly basic region. The organization at the 3' end of the HSRV genome resembles that of lentiviruses, with several small ORFs. Two of the bel genes have limited protein similarity (28 and 20% respectively) to the tat and 3'orf gene products of HIV-2 (Maurer et al., 1988). Thus this spumavirus has the basic genome organization of retroviruses with specific similarities to both the oncovirus and lentivirus subfamilies.

Variations in mechanisms of gene expression

In recent years the detailed study of mechanisms in expression of viral genes has revealed a number of unanticipated complexities. These include the transcriptional processing phenomena of splicing, the occurrence of translational frameshifting and suppression and also mechanisms for processing polyproteins. In this and the subsequent three sections we describe some further elaborations in transcription and translation, a more detailed understanding of the transcription in negative-strand viruses, a system in which changes in expression result in changes in pathogenesis and some recent work on proteolytic processing.

Thomas et al. (1988) analysed the structure and expression of the P gene of the paramyxovirus simian virus 5 (SV5). It was previously known that this encoded two structurally related proteins P (Mr 44000) and V (Mr 24000). P is a component of the virus transcriptase, but the function of V is unknown. Sequence analysis of the gene by Thomas et al. now showed that two overlapping
ORFs, of 222 and 250 codons, are present. The use of specific antibodies and manipulation of plasmid-borne versions of the gene then revealed that V was encoded by the upstream ORF; P contained sequences specified by both ORFs.

At this stage, then, the system looked like another instance of ribosomal frameshifting. However, this was then discounted, since P could not be translated from transcripts copied in vitro from the cloned gene. Sequence analysis of a number of independent cDNA clones of mRNAs then showed that there were two populations which differed in sequence in the region of overlap of the two ORFs, in that one set possessed a tract of four G residues, while the other had six Gs at the equivalent location. The genomic RNA had four Cs at the corresponding site. The two additional G residues converted the two separate ORFs into a single ORF of 392 codons, and P protein could be translated in vitro from a ‘six G’ template. It appears, then, that the addition of two extra G residues is a device to enable expression of two related proteins from one coding unit. In principle, the two extra residues could be added during transcription, or post-transcriptionally. The latter would be comparable to the phenomenon of ‘editing’ of transcripts described recently for several non-viral systems (Feagin et al., 1988a, b; Shaw et al., 1988). However, a transcriptional mechanism seems more likely, in that the virus RNA-dependent RNA polymerase is already known to direct synthesis of non-templated 3’-terminal poly(A).

Details of the expression of the P gene region appear to be particularly variable among members of the Paramyxoviridae group. For some other members including Sendai virus, one mRNA species is produced, which yields two or more proteins, it is suggested via multiple translational starts. In 1988 two groups showed that the 5’-proximal start codon for the Sendai virus mRNA is ACG (Curran & Kolakofsky, 1988; Gupta & Patwardhan, 1988), rather than the standard AUG. In all, this mRNA is thought to exhibit four initiation codons; relatively inefficient initiation at each in turn then allows translation of four polypeptides. One suspects, however, that the complexities of this coding region are not yet exhausted. In their paper on the SV5 system, Thomas et al. (1988) pointed out that the C-terminal region of the V protein (that part not found in the P protein) has a cluster of seven cysteine residues, and that equivalents of these and their surrounding sequence can be found, strongly conserved, in alternative ORFs of other paramyxovirus and morbillivirus P genes.

Turning to a quite different genetic system, bacteriophage T4, we note another ‘bizarre’ variant of translation. Huang et al. (1988) showed that in T4 gene 60 (encoding a topoisomerase subunit) there is, within the coding region, a 50 nucleotide untranslated sequence. Unlike intron sequences, this remains part of the mature mRNA: it appears that the translating ribosomes ‘skip’ it, perhaps because of an unusual folded structure.

These diverse and idiosyncratic variants in expression mechanisms emphasize that, although we possess a considerable formal knowledge of genetic coding, we are not fluent ‘colloquially’; cautionary tales for interpreters of nucleic acid sequences.

**Transcription in negative-strand viruses**

Analysis of the RNA-dependent transcription processes of negative-strand viruses has long been hindered by the nature of the templates (which are nucleoprotein complexes rather than free RNAs), the large size of proteins involved, and by tight coupling of RNA synthesis and processing. Nonetheless, advances are being made and, in this section, we describe results on characterization of genes involved, on reconstruction of nucleoproteins, and on cell-free replicative (as opposed to transcriptive) RNA synthesis.

In the non-segmented negative-strand Rhabdoviridae family the major component of the RNA-dependent RNA polymerase is the large protein (M, around 240000) designated L. The L gene sequence of vesicular stomatitis virus (VSV) Indiana serotype was reported some time ago (Schubert et al., 1984), and in 1988 the L gene sequences were reported for VSV New Jersey (Feldhaus & Lesnaw, 1988) and for rabies virus (so completing determination of the rabies virus genome sequence; Tordo et al., 1988). The L protein sequences of the two VSV serotypes were found to be closely related, while that of rabies virus is more like the VSV Indiana sequence than are the other corresponding protein sequences of the two viruses.

In the other group of non-segmented negative-strand viruses, the Paramyxoviridae, there
is a corresponding protein, also termed L. In 1988 L gene sequences were published for parainfluenza virus 3 (to complete the sequence of this genome; Galinski et al., 1988), and measles virus (Blumberg et al., 1988) supplementing the previously known Sendai virus and Newcastle disease virus (NDV) sequences (Shioda et al., 1986; Morgan & Rakestraw, 1986; Yusoff et al., 1987). Both of the new L protein sequences were strongly similar to that for Sendai virus, and showed a lesser degree of similarity to the NDV sequence.

With the numbers of rhabdovirus and paramyxovirus sequences now available, it is clear (as discussed by several of the above papers) that there are low but convincing similarities between the L protein sequences of the two families. This provides direct evidence for a previously surmised evolutionary relationship between these virus groups. The resulting definition of conserved elements in the L proteins should assist in their functional dissection.

Several papers have appeared recently on assembly of nucleocapsids in negative-strand virus systems. Mirakhur & Peluso (1988) used extracts from cells infected with wild-type VSV to make nucleocapsids with VSV defective interfering (DI) RNAs. The structures obtained were indistinguishable by biophysical criteria from nucleocapsids obtained from cells during DI particle replication and, impressively, they were replication-competent when introduced into cells with a helper virus. This particular example was directed towards development of a genetic manipulation system rather than to studying RNA polymerase function.

Kingsbury et al. (1987) reported the assembly of ribonucleoprotein complexes from influenza A virus nucleocapsid protein (NP) produced in Escherichia coli and single-stranded RNAs. Again these had properties similar to authentic complexes. Szewczyk et al. (1988) described experiments in which the proteins from influenza viral nucleoprotein were separated by electrophoresis in polyacrylamide gels containing SDS. NP and the three polymerase proteins (PA, PB1 and PB2) were then recovered from the gel and renatured in the presence of thioredoxin. It was possible to reconstitute nucleoprotein from the four proteins plus genomic RNA, and this was transcriptionally active in vitro, giving mRNA-sized plus-strand product. This notable achievement should certainly open the way for further analysis of influenza virus transcription.

Another significant step in study of influenza virus RNA synthesis was the development by Shapiro & Krug (1988) of a cell-free system to study replicative synthesis of full-length plus and minus strands (as opposed to transcription of mRNAs, which are incomplete plus-strands and experimentally more accessible). It was shown that nuclear extracts from infected cells could synthesize full-length plus and minus strands. When extracts were depleted of free NP molecules, both activities were inhibited.

**Genes of encephalitic measles virus**

Evidence has been accumulating for several years that measles virus strains causing progressive encephalitis are partially defective in genes for viral coat components. Early work concentrated on defective expression of the matrix protein (M), but it is now clear that defects in expression of the surface glycoproteins can also be relevant. Molecular examination of such virus strains by the most direct route, cloning viral mRNA species directly from infected brain tissue, was reported by Cattaneo et al. (1986), and in a more extensive paper by Cattaneo et al. (1988). The authors of the latter paper were able impressively to clone at least one complete transcript of each virus gene except the L gene, from two cases of subacute sclerosing panencephalitis (SSPE) and from one case of measles inclusion body encephalitis (MIBE). Extensive sequence analysis was carried out.

Considerable variability in nucleotide sequences was found. A problem in quantifying such comparisons is that the progenitor strains from which the encephalitic viruses developed are not accessible. However, with respect to a derived consensus reference sequence, it was estimated that the SSPE and MIBE strains had accumulated mutations in 2% of their nucleotides, and 35% of these had given rise to amino acid changes. These values are much higher than those observed in lytic strains. Mutations resulting in defects in expression or structure of the fusion glycoprotein and of the M protein were observed. In the MIBE case, the changes in the M gene were unprecedented: 132 (or 50%) of the uracil residues in the gene were replaced by cytosines,
resulting in gross changes in the encoded protein. Further analysis showed that these biased mutations were limited to the M gene plus a small adjacent region, and had probably been generated in a single event. Mutagenesis of this nature and scale represents a highly novel and interesting phenomenon, presumably reflecting some aspect of the functioning of the viral RNA replicative machinery. (Note that in the section on Variations in mechanisms of gene expression above, another unusual mechanism in paramyxovirus RNA synthesis was outlined.)

**Proteolytic processing**

One common way by which viruses overcome the restrictions on eukaryotic ribosomes in translating non-5' cistrons is to express all or large sections of their genetic information as a polyprotein which is subsequently cleaved to functional proteins. In most cases the protease(s) involved is (are) virus-encoded and thus present attractive possible targets for virus control by the use of inhibitors or by other means. Before such methods can be devised it is necessary to have detailed information on the mechanisms of action of these proteases.

The proteases of at least three groups of viruses, the animal picornaviruses and the plant-infecting comoviruses and potyviruses fall into the same protease group with cysteine active centres (Cys centre). Bazan & Fletterick (1988) pointed out that these proteases have homology to a family of trypsin-like serine proteases. They suggest that the active site nucleophile of the catalytic triad, which in trypsin is Ser 195, is the Cys residue found in all these viral proteases; the other two residues of the trypsin triad, His 57 and Asp 102, are conserved in all the viral proteases. The 2A and 3C proteases of picornaviruses are homologous to the small and large subclasses of trypsin-like serine proteases respectively. This enabled Bazan & Fletterick (1988) to map residues from the viral proteases onto tertiary structures and to identify precisely the amino acids which determine specificity for proteases 2A and 3C.

The action of picornavirus protease 2A appears to be on the host cap-binding protein complex, thus preventing cellular mRNAs from binding to ribosomes (Lloyd et al., 1988). By sequence comparisons picornaviruses can be divided into two groups, those which have the 18 amino acid Cys centre in the C-terminal region of the 2A protease and those which do not. Enteroviruses and rhinoviruses which have the Cys centre cleave the p220 component of the cap-binding protein. The two cardioviruses, encephalomyocarditis virus and Theiler's encephalomyelitis virus, which lack the Cys centre, do not cleave p220. However, foot-and-mouth disease virus (FMDV) which has a 2A protein of only 16 amino acids, lacking the Cys centre, can process the p220. This processing has been shown to be by the FMDV leader protein (Devaney et al., 1988). This constitutes the first report of a function for this protein other than autolytic cleavage of the FMDV polyprotein (Strebel & Beck, 1986).

The in vitro translation of RNA transcripts from mutagenized cDNA clones of picornaviruses has shown that a precursor of the 3C protease, 3CD, processes the coat protein precursor P1 (Nicklin et al., 1988; Ypma-Wong et al., 1988). The subsequent processing of the Gln-Gly sites in the P2 and P3 precursors is by 3CD and 3C together (Ypma-Wong et al., 1988).

The proteases of the comovirus cowpea mosaic virus (CPMV) and the potyviruses tobacco etch virus (TEV) and tobacco vein mottling virus are in an analogous position in the genome to, and show sequence homology to, the picornavirus 3C protease (see Goldbach & Wellink, 1988). The CPMV protease, a 24K polypeptide, is encoded by the viral bottom (B) component RNA and is responsible for all the cleavages of the polyproteins from both B and middle (M) component RNAs. Vos et al. (1988) reported that the CPMV 32K protein, encoded by B component, is essential for the cleavage of the Glu-Met site in the M polyprotein even though it does not itself have proteolytic activity. It is interesting to note that the 32K protein is in an analogous position to the picornavirus 2A protease. The TEV polyprotein is cleaved by the virus-encoded 49K protein at Gln-Ser sites surrounded by a conserved sequence (Carrington et al., 1988; Dougherty et al., 1988). Carrington & Dougherty (1988) introduced a 25 amino acid segment containing a putative cleavage site into the capsid protein sequence and showed that it was cleaved by the 49K protease. Even the insertion of a synthetic linker comprising the conserved seven amino acid segment led to protein cleavage. The authors pointed out the possibilities of engineering proteins to contain cleavage sites, which would be activated by the addition of the TEV protease.
Retrovirus proteases have been shown to be evolutionarily related to the cellular Asp protease family (Pearl & Taylor, 1987). With the obvious interest of the potential of controlling virus disease by interfering with the protease action, the protease of HIV-1 has been purified and studied. There are three basic approaches to obtaining the enzyme for study, namely isolation from infected cells, expression in a vector system and chemical synthesis. Using the first two approaches the protease has been identified as a 10-7K protein (Debouck et al., 1987; Krausslich et al., 1988; Lillehoj et al., 1988; Mous et al., 1988) in which deletion of the C-terminal 17 amino acids abolishes activity (Krausslich et al., 1988). The enzyme has been chemically synthesized as a 99 residue protein which specifically cleaved HIV-1 gag protein to the expected fragments (Schneider & Kent, 1988). Two cleavage sites other than the previously reported Phe/Tyr-Pro site were determined as being Met-Met and Leu-Ala.

Virus structure

In this section we discuss two topics: binding of influenza virus haemagglutinin (HA) with its receptor and molecular modelling of virus proteins.

For virology, the major crystallographic study of 1988 was determination of the structure of influenza A virus HA complexed with analogues of its cell surface receptor (Weis et al., 1988). It is well known that HA binding involves sialic acid, as part of a cell surface glycoprotein. In effect, sialic acid consists of a pyranose ring with carboxylate, acetamido and glyceryl side chains, and it is found in either \( \alpha(2,3) \) or \( \alpha(2,6) \) linkage to a neighbouring sugar residue. From previous, indirect data it was thought likely that the binding site was a depression or pocket on the membrane-distal surface of the HA molecule. The amino acid residues forming this pocket are refractory to evolutionary change, and it is surrounded by antibody binding sites.

Weis et al. (1988) determined the structure of an effectively wild-type HA complexed with the trisaccharide \( \alpha(2,3) \) sialyl-lactose, and of a mutant HA complexed with \( \alpha(2,3) \) sialyl-lactose. In each case, the sialic acid moiety is seen bound within the pocket and filling it, while the other sugar residues are not specifically bound. Each of the specific side chains of the sialic acid interacts with the protein, but the atoms of the pyranose ring do not. No conformational change in HA was seen to result from binding of sialic acid. In the case of the mutant HA studied, which has increased affinity for \( \alpha(2,3) \)-linked sialic acid, the mutated residue does not itself interact directly with the sialic acid, but apparently gives rise to a set of small conformational changes in the head of HA, which result in an altered affinity.

Thus, sialic acid is the one cell surface component involved in the initial binding of influenza A viruses. This work has provided the first detailed model of molecular interactions between a virus particle and its cell surface receptor. It will clarify ideas on the mechanisms of virus neutralization by antibodies, and it should facilitate design of antiviral agents.

Molecular modelling of proteins seeks to use computer methods to produce a valid three-dimensional structure, usually using sequence alignments between the protein under study and a related protein, plus X-ray crystallographic data for the related species. This is a rather new activity in virological research, but it is becoming increasingly used, and in 1988 several interesting studies in this area were published, including modelling of the trypsin-like serine proteases, of picornaviruses and of HBV core protein. The first of these has already been mentioned in the section on proteases.

Luo et al. (1988) derived three-dimensional models for FMDV and for hepatitis A virus (HAV). The amino acid sequences for FMDV and HAV structural proteins were first aligned with these for human rhinovirus 14 and Mengo virus (whose structures had been determined previously by X-ray crystallography). Models of the structures of individual proteins were then built, and hence models of the virions. The resulting FMDV model was consistent with data on FMDV epitopes accessible to neutralizing antibodies. In both models, the most notable feature was that the surfaces differed greatly from those of other picornaviruses, particularly in the structure of the depressions thought to contain binding sites for cellular receptors. The value of this approach will depend on the extent to which it succeeds in guiding experimentation. In the case of FMDV, direct analysis of the structure is in progress (Fox et al., 1987) and will provide a direct test of the model's validity. [Postscript to this section. Since the above was written, the structure of the FMDV particle as derived by X-ray crystallography has been published]
(Acharya et al., 1989). The judgement of Acharya et al. is that the sequence alignment derived by Luo et al. (1988) which was used in their model building exercise was partially incorrect, and that although the model correctly predicted basic, common features, it did not succeed in identifying distinctive FMDV structures.]

The other model-building exercise described here represents, in our view, a considerably more speculative undertaking. Argos & Fuller (1988) carried out sensitive tests for similarities between the sequences of HBV core protein and those of known capsid proteins of other viruses (see section on Research on hepadnaviruses). They found convincing similarities with the VP3 proteins of picornaviruses, and proceeded to build a model for the HBV core protein molecule as an eight-stranded $\beta$-barrel. Details of the model were consistent with antigenic and other data.

The eight-stranded $\beta$-barrel architecture was first observed with the structural proteins of small RNA viruses, and similar elements were subsequently reported for the adenovirus hexon protein (see reviews by McGeoch et al., 1986, 1987). Fuller & Argos (1987) reported that alphaviruses may also make use of this building block. Rossmann (1988) has proposed that the $\beta$-barrel structure may be extremely widespread among viruses; specifically, he suggested that the $gag$ protein p24 of HIV may belong to this group. While the proposed generality may be startling, the possibilities raised are certainly intriguing and could be very productive.

Coat protein protection in plants

Since the report by Powell Abel et al. (1986) that plants transgenic in the coat protein of tobacco mosaic virus (TMV) were resistant to infection by that virus considerable attention has been paid to this form of non-conventional resistance. Basically the research has taken two approaches, i.e. determining the extent of this phenomenon amongst plant viruses and studying its mechanism. Up to 1988 coat protein protection has been demonstrated for five further viruses: alfalfa mosaic virus (AlMV) (Loesch-Fries et al., 1987; Tumer et al., 1987; van Dun et al., 1988), cucumber mosaic virus (Cuozzo et al., 1988), PVX (Hemenway et al., 1988), tobacco rattle virus (TRV) (van Dun & Bol, 1988) and tobacco streak virus (TSV) (van Dun et al., 1988). The protection against viruses not contributing the coat protein is related to the similarity between the coat proteins of the infecting and protecting viruses or strains. Plants transgenic in TRV strain TCM coat protein show considerable resistance to the Dutch isolate of pea early browning virus, the coat protein gene of which cross-hybridizes extensively with that of TRV-TCM. On the other hand there is no protection against the PLB strain of TSV the coat protein of which has only 39\% sequence identity with that of TRV-TCM (van Dun & Bol, 1988). It has been known for some time that the genome of AlMV requires the addition of AlMV coat protein for replication; this activation can also be performed by the coat protein of the related ilarvirus TSV. van Dun et al. (1988) showed that this activation also takes place in plants transgenic in TSV coat protein; these plants do not protect against AlMV infection. In most of the virus–transgenic plant systems the protection is overcome by inoculation with viral RNA (see Loesch-Fries et al., 1987; Nelson et al., 1987). However, plants transgenic in PVX coat protein protect against infection with the viral RNA (Hemenway et al., 1987).

Mesophyll protoplasts from tobacco plants transgenic in TMV coat protein are resistant to infection with TMV (Register & Beachy, 1988) but, as with whole plants, are not protected with viral RNA or with pH 8-treated virus which has lost some of its 5' coat protein subunits. This suggests that the protection results from interference with an early stage in the infection cycle. Register & Beachy discuss two models to explain this protection. They suggest that the expressed coat protein either blocks the sites where TMV uncoating takes place or shifts the kinetics of virus disassembly.

New animal disease agents

Two novel disease agents, that of bovine spongiform encephalopathy and the seal morbillivirus featured prominently in the U.K. press in 1988 and were, and still are, a cause of public concern. This section briefly describes these agents.

Wells et al. (1987) described a novel spongiform encephalopathy in cattle in the U.K. A strong resemblance of this disease to scrapie and similar conditions was noted. In one case, abnormal
fibrils were demonstrated, similar to the characteristic fibrils which can be extracted from scrapie-infected brain, and are composed of the neuronal membrane protein PrP. The disease is now termed bovine spongiform encephalopathy (BSE). Morgan (1988), considering the possible relationship of BSE to scrapie, pointed out that some cattle are fed a dietary supplement derived from sheep offal, and that changes to lower temperature methods of rendering offal have been made in the last decade. Thus scrapie agent from affected sheep could survive current processing techniques. Epidemiological evidence supports this aetiology (Fraser et al., 1988).

Hope et al. (1988) purified BSE-associated fibrils and showed by a number of criteria, including protein sequence, that they were indeed composed of the bovine equivalent of PrP. Fraser et al. (1988) reported producing a scrapie-like condition in mice by inoculation with preparations of brain from BSE-infected cattle.

BSE is thus now regarded as the bovine counterpart of the sheep disease scrapie. It appears to have arisen recently. Other properties remain unknown, including whether affected cattle could be infectious, or whether transmission to humans might also be a possibility.

1988 saw a severe epidemic in common or harbour seals, first in the Baltic and then in the North Sea. Mortality rates were high, with many dead seals being washed-up on coasts. There was also a report of the disease extending into porpoises (Kennedy et al., 1988a) and a suggestion that it had spread to the common seal from the harp seal (Goodhart, 1988). At first the causal agent was identified serologically as being canine distemper virus (CDV) or a closely related morbillivirus (Osterhaus & Vedder, 1988). This appeared to be confirmed by infection of dogs with the seal virus, in which host the presence of nucleocapsids characteristic of paramyxoviruses was observed (Osterhaus et al., 1988), and by its histopathology (Kennedy et al., 1988b). However, more detailed analysis by nucleic acid hybridization and the serological relationships of the various viral proteins showed that the seal virus differed from CDV and was most likely a previously unknown morbillivirus now tentatively named phocine distemper virus (Cosby et al., 1988; Mahy et al., 1988).

We thank Dr T. M. A. Wilson for his helpful comments on this review and M. Harvey for the graphics in the figure.

REFERENCES


