REVIEW ARTICLE

Influenza Virus Transcription

By J. J. SKEHEl AND A. J. HAY

Division of Virology, National Institute for Medical Research, Mill Hill, London NW7

For some time it has been known that influenza viruses contain a number of distinct single-stranded RNA molecules and in the last few years conclusive evidence has been obtained to indicate that both type A and type B viruses contain eight unique polynucleotide chains. This conclusion in the case of type A viruses is based on evidence provided by the results of direct chemical analysis involving RNase T1-oligonucleotide mapping (McGeoch et al. 1976) and nucleotide sequence analyses (J. J. Skehel & A. J. Hay, unpublished data) and by the results of genetic investigations involving analyses of the RNA components of recombinant viruses (Scholtissek et al. 1976; Hay et al. 1977a; Palese, 1977). The results of the former experiments have in addition allowed the identification of other small RNA molecules, detected in some strains of virus as 5'-terminal fragments of one of the largest unique genome segments, and have, therefore, resolved to some extent the ambiguity concerning the number of RNAs per genome. The genetic studies in turn have also led to the assignation of eight virus-specific polypeptides detected in infected cells to their corresponding RNA genes. This was achieved initially simply by comparisons of the sizes of the genes and their products (Skehel, 1972; McGeoch et al. 1976) and subsequently established precisely by comparative hybridization and polyacrylamide gel electrophoretic analyses of the RNA components of parental and recombinant viruses carrying defined antigenic or temperature sensitive genetic markers (Ritchey et al. 1976; Scholtissek et al. 1976, 1977; Hay et al. 1977a; Inglis et al. 1977; Palese et al. 1977; Ritchey & Palese, 1977; Rohde et al. 1977). As a result, the products of two of the three largest RNA molecules, of mol. wt. approx. 10^6, are proposed to be involved in transcription and the third polypeptide of similar size, to be involved in replicase activity; three genes of mol. wt. between 6.0 and 8.0 x 10^5 contain the information for the synthesis of the haemagglutinin and neuraminidase glycoproteins of the virus membrane and the nucleocapsid polypeptide, and the syntheses of the two smallest polypeptides, a non-virion component of unknown function and the virus matrix polypeptide, both of mol. wt. about 25 x 10^3, are directed by the two smallest RNAs of approximate weight 3 x 10^5 daltons.

The influenza virus genome then is made up of eight single-stranded RNA genes of mol. wt. between 10^6 and 3 x 10^5 which contain the information necessary for the synthesis of eight polypeptide products of corresponding sizes.

The expression of this information in infected cells is mediated by RNA transcripts of complementary sequence to genome RNAs (Etkind & Krug, 1974; Glass et al. 1975) and the remainder of this article contains a description of the nature of these molecules, of the processes involved in their synthesis and a discussion of their functions in virus replication. Many of the recently obtained experimental results to be mentioned come from studies of the infections of chick embryo fibroblasts and of L cells by fowl plague virus. The former is a permissive system and the latter abortive with respect to the production of infectious virus. The method used to analyse the products of transcription involves the detection by
fluorography of electrophoretically separated double-stranded hybrid molecules formed by annealing radioactive RNA extracted from infected cells or polymerase reaction mixtures in the presence of an excess of unlabelled virion RNA.

Analyses of the transcript composition of infected cell extracts using this analytical procedure led to the observation that there were two distinct classes of RNA complementary in sequence (cRNA) to virion RNA in infected cells (Hay et al. 1977b). Subsequent experiments showed that the members of only one of these classes contained polyadenylic acid sequences and furthermore that the S1-nuclease-resistant hybrids obtained using polyadenylated cRNAs, prepared by affinity chromatography, were in fact smaller than their counterparts formed by the annealing of non-polyadenylated cRNAs and virion RNA. This distinctive feature was investigated further by comparative RNase T1 oligonucleotide mapping of radioactive polyadenylated and non-polyadenylated cRNAs and also of radioactive virion RNA molecules (vRNA) obtained from the double-stranded RNA portions of hybrids formed between labelled vRNAs and either polyadenylated or non-polyadenylated cRNAs. The results obtained indicated that, whereas the non-polyadenylated cRNAs appeared to represent complete transcripts of the genome RNAs, the polyadenylated cRNAs did not (Hay et al. 1977c). Moreover, observations on the nuclease susceptibility of terminally labelled vRNA in the differently derived hybrids also suggested that all polyadenylated transcripts lacked sequences complementary to the 5'-termini of vRNAs. In turn, this conclusion was supported by the results of sequence analyses of the unique RNase T1 oligonucleotides of vRNAs 7 and 8 which were not rendered S1-nuclease-resistant following hybridization with polyadenylated cRNA. The findings indicated that oligonucleotides of the sequences 5'UUUUUUAAC(U,C)CAG3' and 5'UUUUUUAU(AUU, AUC, AAAU)AAG3', which were known to be near the 5'-termini of RNAs 7 and 8 respectively (Smith et al. 1977), were removed by nuclease S1 digestion from the virion RNAs in hybrids with polyadenylated cRNAs and more recent sequence analyses have in fact located these oligonucleotides to positions 17 to 29 and 17 to 36 from the 5'-termini of the respective molecules (Hay et al. 1977d).

In the course of these studies of vRNA nucleotide sequence an interesting feature of the genome was observed. It has been known for some time that the 5'-terminus of all the vRNAs is pppAp (Young & Content, 1970 and also that all contain uridine at their 3' ends (Lewandowski et al. 1971). The similarities were extended by the findings that the first 23 nucleotides of all the vRNAs were almost identical in being

```
U A G
G U G A
5'AGUAGAAAUUAGG UUUUUUU
G U A
A G A
```

(J. J. Skehel & A. J. Hay, unpublished data). The results of the nuclease S1 protection experiments would, therefore, suggest that the complementary sequence to this common region was not present in polyadenylated cRNAs. Additional analyses are required to establish precisely the position at which complementarity between vRNA and polyadenylated cRNA stops.

Taken together, therefore, these observations indicate that there are two classes of cRNA transcript in influenza infected cells, that molecules of only one type contain poly A and that, unlike their non-polyadenylated counterparts, these molecules do not contain sequences complementary to a conserved region at the 5'-termini of the corresponding virion RNAs.
The initial transcription of the infecting virus genome is catalysed by the action of an enzyme which is an integral component of infectious virus particles and is referred to as primary transcription. Secondary transcription includes all other synthetic activities in which cRNA is produced. In vitro, following detergent disruption of the virus membrane, the virion enzyme transcribes all of the RNAs and has two properties which are of particular interest. Firstly, polymerization is stimulated with some strains of virus by as much as 100-fold in the presence of dinucleotide primers such as GpG and ApG which are incorporated into the 5'-termini of the transcripts (McGeoch & Kitron, 1975; Plotch & Krug, 1977) and secondly, the products of the reaction are incomplete transcripts of the virion RNAs (Hay et al. 1977d). They are polyadenylated and seem to resemble polyadenylated transcripts extracted from infected cells exactly. Both of these features will be mentioned again.

In vivo, with the information described above on the nature of the two classes of complementary RNA in fowl plague virus infected cells, it was possible to determine separately the kinetics of their synthesis and their subcellular location in attempts to ascertain their function in replication (Hay et al. 1977b). Virus-specific RNA synthesis in fowl plague virus infected cells occurs at maximum rate between 1½ and 2½ h after infection and in agreement with this determination approx. 90% of the vRNA which is eventually incorporated into progeny virus particles is synthesized in the first 4 h of infection. Additionally it can be shown that at these times after infection all of the vRNAs are synthesized coordinately at the same rates, and similar observations were made concerning the synthesis of the non-polyadenylated cRNAs in several infected cell systems. However, in clear contrast with these results, the relative rates of formation of the different polyadenylated cRNAs varied considerably at different times after infection. In infected chick embryo fibroblasts the initial products of transcription include approximately equal amounts of each transcript. Subsequently the pattern of synthesis changed so that each polyadenylated cRNA was synthesized at maximum rate in the order 8, 1, 2, 3 and 5; and finally 4, 6, and 7. In addition, the largest transcripts, those of genes 1, 2 and 3, following the initial stages of infection, were always produced in much smaller amounts than any of the others. Transcription in fowl plague virus infected L cells followed a somewhat different course. In this infection the transcripts of the three largest genes were the first to achieve their maximum rate of synthesis followed in order by transcripts 8, 5 and 4. Transcription of genes 6 and 7 were severely curtailed at all stages of infection (Bosch et al. 1977). This difference in the pattern of transcription of the same virus in different infected cells will be mentioned again later. It serves here to emphasize the differences in the kinetics of synthesis of polyadenylated and non-polyadenylated transcripts.

Differences between the cRNAs of the two classes were again registered when their cellular locations were determined. Although all were predominantly located in the cytoplasm, polyadenylated cRNAs were exclusively associated with polysomes. Consequently, although in cell-free protein synthesizing systems both polyadenylated and non-polyadenylated transcripts appear to be efficient as messenger RNAs (Stephenson et al. 1977), it seems that in vivo this function is reserved for the former molecules. Concerning the site of synthesis, there is conflicting evidence which may again be a reflection of the different virus-cell systems used. However, the observations of Krug et al. (1976) that influenza messenger RNAs, in addition to 5'-terminal 7-methyl guanosine in ‘cap’ structures, contain internal N6-methyladenosine is at present the clearest indication that they are produced in the nuclei during replication.

Summarizing then, two classes of cRNA can be distinguished in influenza virus infected cells on the basis of their primary structure, their cellular locations and the differing times
after infection and relative rates at which they are synthesized. Taken together these distinctive properties suggest that polyadenylated cRNAs are messenger RNAs and that non-polyadenylated cRNAs may function as templates in genome replication, a function which their incomplete polyadenylated counterparts could not undertake.

The results of analyses of protein synthesis in virus-infected cells which provide indices of messenger RNA abundance at different times after infection are in accordance with the suggestion that the polyadenylated molecules serve as messenger RNAs. Variations in both the amounts of individual virus-specific proteins produced and the time after infection at which they are synthesized at maximum rate have been clearly recognized in several influenza virus infections (Skehel, 1972; Meier-Ewert & Compans, 1974; Inglis et al. 1976).

In fowl plague virus-infected chick cells the results indicate that the non-virion polypeptide product of RNA8 is synthesized at maximum rate before all other virus products; that the nucleocapsid polypeptide can also be detected at the earliest times after infection but is actually synthesized at maximum rate at a later time than the non-virion polypeptide; and that the haemagglutinin, the neuraminidase and the matrix protein are not detected early in infection and only achieve their maximal rates of synthesis at later times than the nucleocapsid polypeptide. Comparable data on the time of synthesis of the three largest polypeptides has not been obtained since, although they can now be resolved from each other, their resolution from host components of similar mol. wt. in samples taken early in infection has not improved and they are the least abundant of the virus specified polypeptides. However, from the available data it is possible to deduce that the rate of synthesis of the third largest polypeptide declines at earlier times than those of the other two large gene products.

These results, at least for the non-glycosylated polypeptides, are fairly accurate reflections of the polyadenylated cRNA populations of cells particularly after the initial stages of infection and are quite consistent with the proposition that molecules of this class of cRNA function as messenger RNAs. At the same time they also imply that the controlled expression of the virus genetic information as indicated by the production of virus specific proteins is dictated at the level of genome transcription.

How is this achieved? At present most of the suggestions concerning the mechanism of the control of replication come from the results of analyses of the action of inhibitors of RNA and protein synthesis on virus replication. In the presence of cycloheximide, transcription is restricted to the action of the polymerase of the infecting virus particle which, as mentioned above, leads to the production of messenger RNA and is referred to as primary transcription. An initial suggestion from the results of analyses of the polypeptides synthesized immediately after removal of the inhibitor of protein synthesis in the fowl plague virus-chick cell system was that this primary transcription process was selective (Skehel, 1973). This may be so in some cases since the results of subsequent experiments have indicated, as described above, that the initial products of transcription in this system are different from those of the same virus infection in L cells and this is also the case in the absence of protein synthesis (Bosch et al. 1977). Similar variations, particularly with regard to the transcription of RNA 7, occur in infections of other cells by fowl plague virus, of chick cells by influenza viruses isolated from humans, and are also indicated indirectly in comparisons of WSN infections of MDBK and CHO cells (Lamb & Choppin, 1976, 1977). Restricted primary transcription appears, therefore, not to be an exclusive property of the infecting virus but rather to be a reflection of an undefined interaction between virus and cell.

The production of RNA which can serve as template in genome replication must also,
of course, be one of the functions of the infecting genome-transcriptase complex but at present the mechanism of the interchange between messenger and template RNA synthesis is not known. It is certainly clear from direct analyses of the RNA synthesized in the presence of cycloheximide that only messenger RNAs are produced (Hay et al. 1977b) and it is, therefore, reasonable to conclude that newly synthesized proteins specified either by host or by virus are required to allow production of complete transcripts. In fact it is of particular interest to note that the continued production of template RNA (Hay et al. 1977d) and probably, as a consequence, of vRNA (Scholtissek & Rott, 1970; Pons, 1973) is dependent upon protein synthesis. The continued functioning of the input transcriptase complex in generating the template RNA for genome production may, therefore, require the provision in equivalent stoichiometric amounts of newly synthesized protein and the suggestion that virus-specified proteins are involved (Hay et al. 1977d) is supported by the observations that ts mutants of two complementation groups, ts 29 and 131 (Ghendon et al. 1975), which can undertake primary transcription, are unable at the restrictive temperature to synthesize template RNA (A. J. Hay, unpublished results). In this connection also the results of a number of experiments in which transcription was analysed following reversal of the cycloheximide block using cycloheximide-free medium containing the amino acid analogue p-fluorophenylalanine may be mentioned. Under these conditions complete transcripts are not produced even though polypeptide synthesis is restored. However, the pattern of RNA transcription is changed. Instead of that of the primary transcription characteristic of this system in which similar amounts of all eight messenger RNAs are produced, a pattern of transcription similar to that which occurs late in a normal infection is observed. It appears, therefore, that transcription by the input polymerase can be modified by newly synthesized proteins, that the components involved in the regulation of messenger RNA production are active under these conditions and also that they may be distinct from those required for the synthesis of template RNA. Whether the modification involves the genome or the enzyme itself is not known nor is it clear whether, in a normal replicative cycle, these components of the infecting particle are involved in secondary transcription. Regarding this point, however, the levels of messenger RNA synthesis achieved in primary transcription and of template RNA produced in a normal infection are similar and the possibility exists that following the initial rounds of messenger RNA synthesis the input polymerase complex is modified to produce template RNA and does not in fact participate in secondary transcription. In this case the latter process would be an exclusive function of newly synthesized genome RNA and newly synthesized polymerase molecules. These possibilities will be mentioned again later.

The description given above of primary and secondary transcription placed some emphasis on the influence of the host cell on the relative abundance of the messenger RNAs produced. This is only one of a number of consequences of the interplay between virus and host during virus replication. The results of experiments on infected cells incubated in the presence of actinomycin D and α-amanitin, for example, in which all transcription of vRNA is inhibited (Pons, 1977) indicate that influenza viruses are unique in requiring DNA transcription for their replication without employing virus-specific DNA as a replicative intermediate (Barry et al. 1962; Rott & Scholtissek, 1970). Again, the exact function involved is not known but from the results of experiments using cycloheximide, protein synthesis is clearly not required (Scholtissek & Rott, 1970; Pons, 1973; Etkind et al. 1977; Hay et al. 1977b). In this regard the previously mentioned stimulation of transcription in vitro by primers could reflect a requirement for similar molecules by virus-specific enzymes in vivo and this has been suggested (Plotch & Krug, 1977). Comparative
analyses of the 5'-terminal sequences of cRNAs and the 3'-termini of vRNAs may clarify this point.

In conclusion, therefore, the actual components involved in controlling virus transcription are unknown. There is a continuous requirement for host nucleoplasmic RNA synthesis even for primary transcription to occur and this has been shown clearly using α-amanitin-resistant mutant cells (Lamb & Choppin, 1977; Spooner & Barry, 1977); the continuous synthesis of virus and possibly host specified proteins is additionally required to allow template RNA synthesis, and different proteins are involved in regulating the synthesis of messenger RNA.

In considerations of the mechanisms which might be involved in controlling transcription, therefore, several different observations must be accommodated. On the one hand there are those which imply the control of messenger RNA synthesis and on the other there are those which indicate a differentiation between messenger and template RNA synthesis. Regarding the former, little information is available. It might be assumed that in most cases the initiation of transcription is the process that is controlled and certainly the possibility that the inhibition by actinomycin D is the result of an inhibition of primer production would be in accord with this assumption. The action of the p-fluorophenylalanine insensitive virus or host product in selectively altering the proportions of the messenger RNAs produced is not known but modification of transcriptase specificity in initiation is certainly a possibility. All of the other documented observations involve the 3'-termini of template and messenger RNAs since this region of the transcripts distinguishes them from each other and several mechanisms can be proposed which would give rise to incomplete transcript formation. All transcription, for example, might result in the formation of complete transcripts some of which could be subsequently nucleolytically processed. Alternatively, synthesis might terminate before transcription of the 5'-terminal regions of the genome segments. It might also be suggested that messenger RNAs are not contiguous transcripts of genome RNAs. From the results of determinations of the nuclease resistant and sensitive sequences in vRNA-messenger RNA hybrids, the last alternative appears unlikely but conclusive experimental evidence is not available to allow distinction between the first two possibilities. However, the observations that *in vitro* and *in vivo*, in the absence of protein synthesis, only incomplete polyadenylated cRNAs are produced suggest that premature termination is the most likely mechanism. If this proves to be the case then the newly synthesized proteins required for the production of complete transcripts would be involved in allowing transcription to proceed through the messenger RNA termination site rather than in inhibiting the processing of complete transcripts. The precise position of messenger RNA termination has not been determined but it is probably in the region of residue 17. The possibility exists, therefore, that the U₆ sequence between positions 17 and 22 represents a termination signal and in this regard similar sequences have been implicated in termination of prokaryotic transcripts (e.g. Rosenberg *et al.* 1976; Landy & Ross, 1977; Sanger *et al.* 1977). In this connection it is also possible that the nucleotides at the 3'-termini of complete transcripts function as recognition sequences in replicase binding and are even involved in the process of selection of a genetic complement of vRNA molecules in replication and recombination. The observations of conservation of this sequence in the virion RNAs of all strains of influenza are consistent with its involvement in such functions.
REFERENCES


(Received 25 November 1977)