Alphaherpesviruses possess a gene homologous to the protein kinase gene family of eukaryotes and retroviruses

Duncan J. McGeoch and Andrew J. Davison

MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

Received 2 January 1986; Accepted 21 January 1986

ABSTRACT
The US3 genes of herpes simplex virus serotypes 1 and 2, and the corresponding gene of varicella-zoster virus, encode proteins whose sequences are clearly homologous to members of the protein kinase family of eukaryotes and retroviruses. Similarity is most characteristic, and strongest, in an 80 residue region comprising part of the catalytic structure of the kinases. In this region the herpesvirus proteins are most like a yeast cell division control protein, and least like the retrovirus protein-tyrosine kinases. We consider that the herpesvirus proteins are probably involved in modulation of cellular processes during lytic infection, although other roles are also possible, for example in latent infection.

INTRODUCTION
The genome of the alphaherpesvirus herpes simplex virus (HSV) comprises about 155,000 base pairs of double stranded DNA (1), and contains approximately 80 genes, most of which are presently of unknown function. We are engaged on large scale sequence analysis of the DNA of HSV-1, and also of the distantly related varicella-zoster virus (VZV), and have determined the structures of many previously uncharacterized genes (see, for instance, ref. 2). One way to attempt to advance the understanding of such unknowns is to compare their predicted amino acid sequences with collections of known protein sequences, using computer methods. From the work of others and of ourselves, it has emerged that some alphaherpesvirus proteins are related to proteins of the gammaherpesvirus, Epstein-Barr virus (EBV), and also, more strikingly, that some are related to non-herpesvirus proteins (3,4,5; also, unpublished data). In the latter class, homologues have been found with ribonucleotide reductase (for HSV and VZV) and thymidylate synthetase (VZV only).
of eukaryotes and prokaryotes (ref. 4; Davison, unpublished data). Both of these examples are of enzymes involved in nucleotide metabolism. In this paper we report a new class of homology between an alphaherpesvirus protein and a non-herpesvirus protein: we have found that the protein encoded by HSV-1 gene US3, termed 53K, is clearly homologous to members of the protein kinase family of eukaryotes and of oncogenic retroviruses (6). In addition, related genes exist in HSV-2 and in VZV. This is a particularly interesting and intriguing finding, since protein kinases are known to be involved in regulation or modulation of many normal cellular processes, and in addition are central to one mechanism by which retroviruses cause oncogenic transformation of cells (7,8). With this background, there is a clear implication that the homologous herpesvirus proteins could play an important regulatory role in some aspect of alphaherpesvirus activity. In this paper we present and evaluate the evidence that the HSV and VZV genes do encode protein kinase related proteins, and, from what is known at present, discuss their possible functions.

The organization and surroundings of HSV-1 gene US3 are shown in Figure 1 (2,9). The gene has two separate promoters. In both resulting mRNA species, the 5' proximal AUG opens the 481 codon 53K coding sequence. The right neighbour of gene US3 is US4, which is transcribed in the same orientation. US3 mRNAs extend through gene US4 to form a 3' coterminal set with US4 mRNA. The left neighbour of gene US3 is US2, which is transcribed in the opposite orientation. The 5' terminal regions of US2 and US3 mRNAs possess substantial opposing strand overlap, and US2 mRNA is initiated within the predicted US3 protein coding region. Our studies in mapping US3 mRNAs showed that mRNA species from both promoters are present early after infection (at 4 h), and have similarly declined in abundance by 7 h post infection (9). From interpretation of the DNA sequence, we think that HSV-1 gene US3 encodes a 481 amino acid protein of Mr 52,831 ("53K"). Recently, we have analysed the corresponding region in the genome of HSV-2 and have deduced that it encodes a protein exhibiting 75% amino acid sequence homology with HSV-1 53K (D.J.
Figure 1. Genomic location and organization of HSV-1 gene US3. The upper part of the figure shows a conventional representation of the HSV-1 genome, with unique sequences as lines and major repeat elements as open boxes. The lower part expands a 3 kb section of the short unique (Ug) region containing the US3 gene, with numbering for the Ug sequence (2). The positions and orientations of mRNAs for genes US2, US3 and US4 are shown by arrows (9). Filled circles indicate 5' termini of mRNAs (two for gene US3) and open boxes the locations of predicted protein coding regions.

McGeoch, H.W.M. Rixon and D. McNab, unpublished data). In addition, 53K has homology with a species encoded by the distantly related alphaherpesvirus VZV, termed 44K (10,11). 44K lacks residues corresponding to the N terminal 98 amino acids of the HSV proteins, but in the remainder of the sequence HSV-1 53K and 44K show 45% identity. None of these predicted gene products has yet been recognized as a protein species in infected cells.

METHODS

Protein sequences were extracted from the National Biomedical Research Foundation sequence database, version 5. Protein sequence homology searches used the WORDSEARCH program of the University of Wisconsin Genetics Computer Group (12). Graphical presentation of similarity between two sequences used the CINTHOM program of Pustell and Kafatos (13), with modification by our colleague P. Taylor to allow scoring for similar amino acids with the replacement probability table of Dayhoff et al. (14), and output on a graphics matrix printer. Optimal alignments of pairs of sequences were obtained with the HOMOL program (15).
Figure 2. Comparison of the amino acid sequences of HSV-1 53K and yeast CDC28. (a) shows relations between 53K and CDC28 protein sequences (16), as presented by a matrix comparison. For CDC28 protein, on the x-axis, the whole 298 amino acid sequence was used. For 53K, on the y-axis, the data for residues 161 to 481 (C terminus) are shown. Axes are marked at 50 residue intervals. The sequences were compared using a window of 25 residues. Similarity scores of 22 to 48 were plotted as thin lines, and higher scores as heavy lines. (b) shows alignment of the sequences. 53K is on the upper line, and identical residues are marked by asterisks. 53K residues 182 to 481 (C terminus) and CDC28 protein residues 2 to 298 (that is, excluding only the initiating methionine) were used. The overlined section is the region of strongest and characteristic similarity (see text).
RESULTS

We describe relations between the herpesvirus proteins and the protein kinase family in several stages. First, as a low resolution overview, Figure 2(a) gives a graphical comparison of the amino acid sequences of HSV-1 53K and the protein encoded by the CDC28 gene of Saccharomyces cerevisiae, which is involved in the regulation of cell division (16). This graph scores for occurrence of similar amino acids as well as identical residues in each sequence. Extended regions of sequence similarity are evident. The yeast protein has no sequence counterpart to approximately 170 amino acids at the N terminus of 53K, but the positions of the C termini are closely equivalent. In Figure 2(b) the sequence similarity between 53K and the CDC28 protein is shown explicitly by aligning the sequences. Much of the sequences, particularly in the region overlined, show convincing similarity.

We have compared 53K with available sequences of the protein kinase family, and here present the results with seven representative members. In preliminary comparisons we established that homologies were most characteristic, and strongest, in a region of approximately 80 amino acids, corresponding in the case of HSV-1 53K and CDC28 protein to the overlined region in Figure 2(b). This was the case for HSV/VZV comparisons, for comparisons between herpesvirus proteins and recognized members of the protein kinase family, and for comparisons between kinase family members. Since describing relations between sets of clearly similar, but not necessarily highly conserved, sequences poses problems for concise presentation, we restricted the analysis presented here to this region, and proceeded by constructing separate consensus sequences for the herpesvirus proteins and for the kinase family members. In Figure 3(a), the three herpesvirus sequences are aligned and a consensus of conserved residues extracted. This consensus also registers positions which were not conserved, but which were occupied only by hydrophobic residues, since a number of such occurred. In Figure 3(b) the corresponding sections from two cellular, cyclic nucleotide dependent kinases, the CDC28
protein and four retrovirus onc gene proteins are presented, with minimal gapping introduced to obtain an overall alignment. With this set, only 11 positions are completely conserved, so the consensus was extended to include residues occurring in at least 5 of the 7 sequences at a given position. We note that this is still a rather stringent measure and excludes many partially conserved features. In Figure 3(c) the two consensus sequences are compared, with occurrences of identical residues and also
Table 1. Pairwise comparisons of best conserved regions of polypeptide chains. Optimal alignments were computed for each pair of sequences shown in Figure 3, and the results presented as number of identical residues for each aligned pair. The main diagonal indicates the length of each sequence used.

<table>
<thead>
<tr>
<th></th>
<th>HSV-1</th>
<th>HSV-2</th>
<th>VZV</th>
<th>cAMP PK</th>
<th>cGMP PK</th>
<th>CDC 28</th>
<th>v-mos</th>
<th>v-src</th>
<th>v-erbB</th>
<th>v-abl</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV-1</td>
<td>82</td>
<td>72</td>
<td>52</td>
<td>29</td>
<td>27</td>
<td>33</td>
<td>33</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>HSV-2</td>
<td>-</td>
<td>82</td>
<td>50</td>
<td>28</td>
<td>28</td>
<td>31</td>
<td>29</td>
<td>22</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>VZV</td>
<td>-</td>
<td>-</td>
<td>81</td>
<td>30</td>
<td>26</td>
<td>33</td>
<td>27</td>
<td>23</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>cAMP PK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>77</td>
<td>47</td>
<td>34</td>
<td>25</td>
<td>21</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>cGMP PK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>79</td>
<td>32</td>
<td>31</td>
<td>25</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>CDC 28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81</td>
<td>29</td>
<td>25</td>
<td>27</td>
<td>25</td>
</tr>
<tr>
<td>v-mos</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>83</td>
<td>31</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>v-src</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81</td>
<td>39</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>v-erbB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>82</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>v-abl</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81</td>
</tr>
</tbody>
</table>

partially conserved and similar residues being marked at each position. This comparison demonstrates that almost all of the non-herpesvirus conserved residues are also represented in the more highly conserved herpesvirus set, that most of the herpesvirus conserved residues have at least a partial correspondence in the non-herpesvirus set, and that the non-conserved regions in the herpesvirus set are also not conserved in the non-herpesvirus set. This homology is in a region considered to be involved in catalysis of phosphorylation (18). In addition, there exists in members of the protein kinase family a sequence N proximal to the region discussed so far, which is involved in ATP binding and exhibits several conserved residues (18). The herpesvirus proteins show only minimal conservation of this element with respect to the protein kinase family. Nonetheless, the similarities of the 80 residue regions are compelling, and we conclude that the herpesvirus proteins are
Figure 4. Positions of the conserved region within polypeptide chains. For the 10 proteins listed in Figure 3, the whole of each sequence is represented as a solid line proportional to chain length, with N termini at the left. The sequences are aligned by the positions of the best conserved approximately 80 residues (Figure 3); this region is boxed. The positions of the N and C termini of cAMP dependent protein kinase are indicated by dashed lines. The scale bar shows a span of 100 amino acids.

most likely protein kinases, or, possibly, some other type of phosphotransferase.

Relative similarities between pairs of sequences were evaluated by computing optimal alignments for all pairs of the 80 residue conserved regions. Results are presented in Table 1 as numbers of conserved positions in each comparison. This demonstrates clearly that the three herpesvirus sequences resemble most the yeast CDC28 protein, followed by the mos protein and the CAMP and cGMP dependent protein kinases, while the src, erbB and abl products are distinctly less similar. The protein kinase family contains two types of enzymes: kinases which phosphorylate serine or threonine residues, and those which act on tyrosine residues. The latter include the products of several onc genes (src and abl, but not mos; 23, 24, 25). We have no definitive criterion to establish the class to which the herpesvirus proteins belong, but we note that their sequences least resemble the tyrosine kinases in the 80 residue conserved region. In addition, several tyrosine kinases are found to be phosphorylated at a tyrosine residue within a partially conserved sequence (25), and this site is not present in the herpesvirus sequences.
Examination of the location of the 80 residue conserved region within each of the ten polypeptide sequences shows that it is not present at a constant position with respect to either terminus (Figure 4). Thus, aligning with respect to the conserved region, we see that the three herpesvirus proteins, cAMP and cGMP dependent kinases, CDC28 protein, and the mos and src proteins all have their C termini within a 50 residue bracket, but that the erbB and abl proteins have large C terminal extensions. At the N termini, all but one of the proteins are extended with respect to the cAMP dependent kinase. Thus, the herpesvirus proteins are unexceptional in the positions of their termini. In the better studied proteins, some of the reasons for these "extra" sequences are clear: they constitute functional entities additional to the basic, catalytic structures of the protein kinase enzyme molecules. Thus, for the cGMP dependent kinase, the N proximal sequence is regulatory, with cGMP binding activity (18). (This is in contrast to the case of cAMP dependent kinase, where regulation is effected by a separate polypeptide chain (7)). In the case of the onc gene products, it is known, at least for the src and erbB proteins, that the extra sequences direct cellular location of the protein (8,21,26). Thus, the N terminal extensions of the herpesvirus proteins could represent separate functions, and might accordingly have homology to other protein sequences. We therefore performed separate searches of the protein sequence database for species most similar to the N terminal regions of HSV-1 53K and VZV 44K, but no significant additional homologies were found. There are two aspects of gene organization which could reduce the evaluated importance of the N terminal region of 53K. First, VZV 44K is 98 residues shorter than 53K at the N terminus. Second, as shown in Figure 1, US2 mRNA is initiated, on the opposite strand, within the 53K coding sequences. If the US2 promoter is typical of other HSV promoters, which occupy roughly 80 to 100 base pairs upstream of the transcription initiation site (27,28), then the DNA encoding 53K residues 28 to about 60 (that is, a significant part of the N terminal extension) must also be subject to requirements of US2 promoter functionality, suggesting possibly a non-critical role for this part of 53K.
We have searched all recognized open reading frames predicted from the complete DNA sequence of the gammaherpesvirus EBV, strain B95-8 (3) for a protein kinase candidate, using standard similarity search methods. In addition, totally conserved elements of the consensus sequences in Figure 3 were used as query sequences: all occurrences were listed and their surroundings examined for other elements of the consensus. These procedures constituted a thorough search process, but no candidate sequence was found, so it seems very likely that EBV does not possess such a gene. Finally, Davis and Huang (29) recently described the sequence of a gene of the betaherpesvirus, human cytomegalovirus, thought to encode a protein kinase. However, this shows no homology with the genes discussed in this paper.

DISCUSSION

The presence of a gene for a polypeptide related to the protein kinase family in equivalent genomic positions in both serotypes of HSV and in the substantially divergent VZV indicates that the gene was acquired by a progenitor of today's alphaherpesviruses at a distant stage of their evolution. We suppose that originally the gene might have been obtained either directly from a cellular genome, or indirectly; for instance, by way of a retrovirus. Since the HSV-1 and VZV coding sequences differ in size (by 88 codons) and in base composition (by 21.3 percentage points), and since 55% of their corresponding amino acid sequences are non-identical, these genes have clearly undergone extensive changes since incorporation into a herpesvirus genome. We thus think it likely that the present day versions are quite distinct from any cellular gene, unlike the retrovirus onc genes, which have close counterparts in cell genomes (30).

A careful search of open reading frames demonstrated that EBV, strain B95-8, does not encode a recognizable kinase related protein gene. At the moment, therefore, presence of this gene appears to be characteristic of the Alphaherpesvirinae sub-family, rather than of the whole Herpesviridae family (31). In fact, none of the genes in the completely sequenced short
regions (short unique plus short repeat) of both HSV-1 and VZV have detectable homologues in EBV, although the long regions exhibit numerous homologies (2,10,32; also, unpublished data). We speculate that loss or gain of the short region, including the kinase related protein gene, could have been a significant event in the divergent evolution of the Alpha- and Gamma-herpesvirinae.

At present we do not have a function defined for an alphaherpesvirus protein kinase. This ignorance is a close consequence of the genome structure approach which led to the discovery of these genes. However, it seems very likely that we have uncovered an element of a regulatory system important in some aspect of alphaherpesvirus activity. Possibly novel protein kinase activities have been detected in herpesvirus infected cells (33,34), but the relation between such observations and the genes discussed here is unclear. Since HSV-1 US3 mRNAs are expressed early in lytic infection, we tentatively favour the idea that the protein might act on some process during the infectious cycle, as opposed to being made for package into virions, in the manner of the HSV virion kinase described by Lemaster and Roizman (35). A protein kinase could act to regulate or modify processes of the host cell, or on virus encoded proteins. For the latter case, we note that many HSV proteins do become phosphorylated in infected cells (36). We think that redirection of host cell processes is the most likely role for this kinase. This could be negative (for instance, turn off of host protein synthesis) or positive (for instance, enabling some aspect of HSV DNA synthesis in non-dividing cells, by analogy with the CDC28 protein (16)). It also remains possible that the in vivo role of a protein kinase could be in a part of alphaherpesvirus capability other than lytic infection, as the establishment, maintenance or reactivation of latent infection in neurones.

In recent years it has become clear that many retrovirus oncogenes encode protein kinases, which belong to the family of proteins discussed in this paper (8,25). In view of a possible connection of HSV-2 with uterine cervical carcinoma, and from experimental demonstration of culture cell transformation by HSV (37,38,39), the possibility that HSV could carry a transforming
kinase is certainly intriguing and provocative. However, at present we think this rather unlikely, since such an activity would probably have come to light during the extensive work performed on cell transformation by HSV, or by HSV DNA or fragments thereof (40,41).

Biochemical characterization of protein phosphorylation in infected cells and in virions has been pursued with many virus systems, but (putting aside the acute transforming retroviruses) in most cases no clear functional synthesis has emerged. The potential of this present case differs in that a probable protein kinase gene has now been characterized. Given the power of modern methods, it should realistically be possible to identify the encoded protein and to study its behaviour and properties. Since the gene can be manipulated, it should also be possible to evaluate the effects of the protein on cellular and viral processes. However, judging from the experience of retrovirus onc gene studies (8), the further stage, of understanding in detail how such effects are obtained, may prove more refractory.

ACKNOWLEDGEMENTS

We thank P. Taylor for computer support, D. Leader and W. Stevely for discussion, and M. Dalrymple, M. Frame and J. Macnab for criticism of the manuscript.

REFERENCES
