VIRUS GENES 7:3, 289–295, 1993

Fowlpox Virus Encodes a Protein Related to Human Deoxycytidine Kinase: Further Evidence for Independent Acquisition of Genes for Enzymes of Nucleotide Metabolism by Different Viruses

EUGENE V. KOONIN[1] AND TATJANA G. SENKEVICH[2]
[1]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894*
[2]*Institute for Viral Preparations, Russian Academy of Medical Sciences, Moscow, Russia*

Received May 16, 1992
Accepted June 29, 1992

Requests for reprints should be addressed to Eugene V. Koonin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bldg. 38A, 8600 Rockville Pike, Bethesda, MD 20894.

Key words: fowlpox virus, vaccinia virus, deoxycytidine kinase, herpesviruses, virus evolution

Abstract

It is demonstrated that fowlpox virus (FPV) protein FP26 located in the *Hin*dIII D fragment of the genome is related to the human deoxycytidine kinase (dCK) and probably possesses the same enzymatic activity. A homologous protein is not encoded by vaccinia virus. A multiple alignment of the amino acid sequences of the human and FPV dCKs, the thymidine kinases (TK) of herpesviruses, and cellular and vaccinia virus thymidylate kinases (ThyK) was generated and the conserved motifs, at least two of which are implicated in ATP binding, were characterized. An apparent duplication of ATP-binding motif B in the dCKs was revealed, leading to the reassignment of one of the catalytic residues. Phylogenetic analysis based on the multiple alignment suggested that the putative dCK of FPV probably has diverged from the common ancestor with the human dCK at a later stage of evolution than the herpesvirus TKs, with the ThyKs being

E.V. Koonin is on leave from Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia.
T.G. Senkevich's present address is Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892.

peripheral members of the family. These results are compatible with the hypothesis that genes for enzymes of nucleotide metabolism could be acquired independently by different DNA viruses (Koonin, E.V. and Senkevich, T.G., Virus Genes 6:187–196, 1992).

Recent analysis of the relationships between the amino acid sequences of thymidine kinases (TK) and thymidylate kinase (ThyK) of large DNA viruses and the related cellular enzymes has led to the hypothesis that the genes for these enzymes of nucleotide metabolism could be captured independently not only by poxviruses and herpesviruses but also by different subdivisions of poxviruses themselves (1). Two families of TKs and ThyKs have been described based on the results of the comparative analysis of amino acid sequences. One of these included the cellular and poxvirus TKs, and the other the cellular and vaccinia virus ThyKs, together with the herpesvirus kinases, most of which exert both activities (1–4). Recently significant sequence similarity has been shown to exist between the herpesvirus TKs and human deoxycytidine kinase (dCK), and it has been hypothesized that dCK might be the direct ancestor of the herpesvirus kinases (5). In addition, the sequence of the human ThyK has been reported recently (6), allowing a more informative comparison of the sequences of cellular and viral ThyKs. With these data, it was of interest to reinvestigate the relationships between the sequences of the kinases of this family with the aim of assessing the possible pathways of evolution and the validity of the idea of the independent capture of the genes for enzymes of nucleotide metabolism by various viruses.

Unexpectedly, when the nonredundant amino acid sequence database (National Center for Biotechnology Information) was screened for similarity with the human dCK sequence (7) using the BLAST program (8), the closest relationship was observed with FP26 protein of fowlpox virus (FPV), which is encoded in the *Hin*dIII D fragment of the viral genome (9). The probability of the observed similarity being fortuitous was computed to be below 10^{-17}. When the two sequences were aligned using the OPTAL program (10), the convincing alignment score of 18.5 standard deviations (SD) above the random expectation was obtained, with the alignment containing 36.1% identical amino acid residues and 53% identical or similar residues (Fig. 1). The A and B motifs implicated in ATP binding (11–13) were conserved in the dCK and FP26 sequences, suggesting that the latter protein is most likely an active dCK, or at least a deoxypirimidine kinase. This is in agreement with the observation that FP26 is actively expressed during FPV reproduction (9).

Vaccinia virus (VV), whose complete genome sequence has been reported (14), does not encode a homologue of FP26 (9). We have explored the possibility that such a protein might be encoded by a pseudogene in VV using the option of BLAST, comparing an amino acid sequence query to the conceptual translation of a nucleotide sequence database in all six reading frames. This procedure has not revealed any significant similarities, indicating that a counterpart to FP26 is lacking altogether in VV. VV genome codes for three nucleoside/dNMP kinases,

```
                                        A motif - P loop

FP26 FPV  MD-----SINEFTS-------KKLSIEGNISSGKTDVLNILRN-------
          *      *   *:         **:******  **:  :***::
dCK human MATPPKRSCPSFSASSEGTRIKKISIEGNIAAGKSTFVNILKQLCEDWEV


FP26 FPV  INNVV-SFHDV---EDRYTPIEK------ELIRKFHENPSRWSYALQTHY
          : : *  : :*   :* :  :         :::    * * ***:  **
dCK human VPEPVARWCNVQSTQDEFEELTMSQKNGGNVLQMMYEKPERWSFTFQTYA


                                      B motif - Mg binding

FP26 FPV  CMKRVRMHLECF------VPSRVNILERSIFSDRYVFAEAATALGYMDDP
          *: *:   *                *   ***::****:**         *::
dCK human CLSRIRAQLASLNGKLKDAEKPVLFFERSVYSDRYIFASNLYESECMNET


FP26 FPV  EWALYCKQHDWYTDKL--EIQFDGIIYLRTIPESCKERINEKSITEKNYP
          ** :*   ***  :     :: ******   ** *  **  :   * :
dCK human EWTIYQDWHDWMNNQFGQSLELDGIIYLQATPETCLHRIYLRGRNEEQ--


FP26 FPV  NISIDYLKTLHEKHELWLTQC---------KKVPVLIIDGEEDFIFDPCA
          * ::**  ** *** **              **:* :* :***  *
dCK human GIPLEYLEKLHYKHESWLLHRTLKTNFDYLQEVPILTLDVNEDFK-DKY-


FP26 FPV  KKKLINEVTEFINSI
             *:: * **: ::
dCK human -ESLVEKVKEFLSTL
```

Fig. 1. Alignment of the amino acid sequences of the FP26 protein of FPV and human deoxycytidine kinase. Identical amino acid residues are designated by asterisks, and similar residues (A,G; D,E,N,Q; S,T; K,R; I,L,V,M; F,Y,W) by colons. A and B motifs constituting the ATP-binding pattern are denoted.

namely, TK, ThyK, and guanylate kinase, with the latter encoded by a split reading frame and probably inactive (14–16). Two enzymes of this type have been identified so far in FPV, the TK (1, and references therein) and the putative dCK, as outlined above.

The TK of FPV shows a different genomic location and a surprisingly low sequence similarity with the TKs of VV and other orthopoxviruses (1, and references therein). The putative dCK lacks a ortholog in VV, as stated above. Thus, in a striking fashion and in agreement with the ideas forwarded previously (1), both known nucleoside/dNMP kinases of FPV seem to have an evolutionary descent distinct from that of the enzymes of this type encoded by VV. It will be of considerable interest to find out if FPV encodes a ThyK and/or a guanylate kinase, and if it does, whether they will be orthologous to the VV kinases. Of course, at this point it cannot be excluded that the gene for dCK has been lost in the evolutionary lineage leading to VV. The assessment of this possibility should await the complete sequencing of additional poxvirus genomes. It does not seem very likely, however, as once acquired the dCK activity would probably provide the virus with a selective advantage, as demonstrated by the drop of

virulence and of the efficiency of reproduction in cell culture observed in the case of tk^- mutants of VV (17).

A multiple alignment of the amino acid sequences of the (putative) human and FPV dCKs, selected TKs of herpesviruses, and cellular and VV-encoded TKs was generated using the OPTAL program. The alignment scores were 13.3 SD for the comparison of the two dCKs with eight herpesvirus kinases, indicative of a relatively close similarity, and 6.6 SD for the comparison of the resulting alignment with the ThyKs, suggesting a considerably more distant but still significant relationship. Partial conservation of the four motifs delineated previously in the herpesvirus kinases and the human dCK (5,18) was observed, although considerable deviations were found to exist both in the putative dCK of FPV and in the ThyKs in motifs 2 and 4 (Fig. 2). This alignment highlighted an interesting peculiarity of motif 3 in the dCKs. This motif is equivalent to the so-called purine NTP-binding motif B, which contains a conserved negatively charged amino acid residue (most frequently Asp) interacting with Mg^{2+} in the phosphate binding site of the NTPase (kinase) active center (19,20). Previously this function has been assigned to an Asp residue of the human dCK, which has been aligned with the conserved Asp residue of the herpesvirus TKs; this residue is conserved also in the putative dCK of FPV (5; Fig. 2).

On the other hand, this residue was substituted by a Glu in the TK of channel catfish virus (CCV). Our present analysis showed that superposition of a Glu residue, which is conserved in the dCKs and is located upstream from the aforementioned Asp, with the critical Asp of the herpesvirus TKs produced a better alignment (Fig. 2). On the other hand, when the comparison was performed between the sequences of the dCKs and the ThyKs, the alignment of the conserved Asp residues was obtained (not shown). Apparently a duplication of the B motif exists in the dCKs (Fig. 2), leaving the assignment of the negatively charged residue involved in catalysis somewhat uncertain. Perhaps some preference should be given to the Glu residue suggested by the alignment in Fig. 2, as the negatively charged residue in the B motif of numerous NTPases and kinases is almost invariably preceded by a bulky hydrophobic residue (12; E.V.K., unpublished observations). As a matter of speculation, the possibility of a switch of the catalytic residue in the evolution of the dCKs may be envisaged.

The multiple alignment of the amino acid sequences of the dCKs, herpesvirus TKs, and the ThyKs was subjected to phylogenetic analysis using either a cluster algorithm (21) or the maximal topological similarity algorithm, which has been reported to be independent of variations of the evolutionary rate in different lineages (22). In agreement with the observed alignment scores, the ThyKs constituted the outgroup in this family, with the second branching separating the putative common ancestor of the human and FPV dCKs, and the herpesvirus kinase subfamily (Fig. 3). Not unexpectedly, these results suggested an independent acquisition of the genes for the dCK homologues by FPV and an ancestor herpesvirus. The sequence of the CCV TK has been shown to be only rather distantly related to those of the other herpesvirus TKs (5). The position of the CCV TK

```
                                motif 1                 motif 2
                                 _________               _______
consensus               xUxxUEGxxxxGKTxxxxxU      UxExuxxUxxxxxxxxx
                             D        S           N
                                                  Q
 TK   HSV1  ( 51-226)   RVY-IDGPHGMGKTTTTQLL -11- VPEPMTYWRVLGASETI -59-
 TK   FHV   ( 22-198)   RIY-IDGAYGIGKSLTAKYL -13- FPEPMLYWRSLFETDVV -58-
 TK   PRV   (  5-172)   RIY-LDGAYGTGKSTTARVM - 7- VPEPMAYWRTLFDTDTV -55-
 TK   EHV4  ( 27-202)   RIY-LDGVYGIGKSTTGRVM -12- FPEPMAYWRTLFETDVI -58-
 TK   MDV   ( 21-195)   RVY-LDGSMGIGKTSMLNEI -12- F-EPMKYWRYYF-TDLV -57-
 TK   EBV   (286-455)   SLF-LEGAPGVGKTTMLNHL -10- VPEPMRYWTHVY-ENAI -56-
 TK   HVS   (211-378)   FIF-LEGSIGVGKTTLLKSM -11- FHEPIAYWTDVF-SNSL -53-
 TK   CCV   ( 17-167)   LVFCVEGNIGCGKSTLVKAL -11- VEEPVDQWVNHNGKNYL -33-

dCK  HUMAN  (23-198)   KIS-IEGNIAAGKSTFVNIL - 9- VPEPVARWCNV---QST -58-
dCK? FPV    (11-160)   KLS-IEGNISSGKTDVLNIL - 2- INNVV-SF------HDV -40-

ThyK HUMAN  ( 7-149)   ALIVLEGVDRAGKSTQSRKL -14- FPERSTEIGKLL-SSYL -31-
ThyK YEAST  ( 6-150)   KLILIEGLDRTGKTTQCNIL -10- FPERSTRIGGLI-NEYL -35-
ThyK VV     ( 5-143)   ALIVFEGLDKSGKTTQCMNI -13- FPQRSTVTGKMI-DDYL -31-


                    motif 3   motif 4                   motif 5
                        ___     ___                      ________
consensus           xxxUUDRxxUSxxxxFxxxx      xxxUUUxxlxxxxxxuxruxxrxxxxEx
                              A    Y                             k
 TK   HSV1          LTLIFDRHPIAALLCYPAAR -23- GTNIVLGALPED-RHIDRLAKRQRPGER
 TK   FHV           VTLIIDRHPLASLVCFPLAR -23- GGNLVVTTLNIE-EHLKRLRGRSRTGEQ
 TK   PRV           MTVVFDRHPVAATVCFPLAR -23- GGNLVVASLDPD-EHLRRLRARARAGEH
 TK   EHV4          LTLVFDRHPVASTVCFPAAR -23- GGNIVVTTLNVE-EHIRRLRTRARIGEQ
 TK   MDV           LILILDRHPISATVCFPIAR -23- GCNLVIVDLHDEKEHVSRLSSRNRTGEK
 TK   EBV           CWILHDRHLLSASVVFPLML -22- GDTIVWMKLNVE-ENMRRLKKRGRKHES
 TK   HVS           MWVMFDRHPLSATVVFPYMH -22- GDNIILLNLNSQ-ENLKRVKKRNRKEEK
 TK   CCV           ---IMERSPMSATRVFCAVN -27- RPVFVYLELPPE-ECLRRMRRRDRTGEA
                      ______
dCK  HUMAN          V-LFFERSVYSDRYIFASNL -30- LDGIIYLQATPE-TCLHRIYLRGRNEEQ
dCK? FPV            V-NILERSIFSDRYVFAEAA -28- FDGIIYLRTIPE-SCKERINEKSIT-EK
                            ------
ThyK HUMAN          VTLVVDRYAFSG-VAFTGAK -17- PD-LVLF-LQLQ-LA-DAAKRGAFGHER
ThyK YEAST          KNIVMDRYVYSG-VAYSAAK -19- PD-LTLF-LSTQDVD-NNAEKSGFGDER
ThyK VV             ITLIVDRYAFSG-VAYAAAK -16- PD-LVIF-LESG--S-KEINR-NVGEEI
```

Fig. 2. Multiple alignment of the amino acid sequences of the human and FPV deoxycytidine kinases, selected herpesvirus thymidine kinases, and viral and cellular thymidylate kinases. The four blocks shown in the figure are excerpts from a longer alignment generated by the OPTAL program (see text). The boundaries of the aligned protein regions are indicated in parentheses, and the distances between the blocks are shown. The conserved motifs are designated as in refs. 5 and 18. The "consensus" line shows the amino acid residues conserved in all three groups of proteins (upper case), or in the herpesvirus kinases and the dCKs (lower case); U = a bulky hydrophobic residue (I,L,V,M,F,Y,W); x = any residue. Only a selection of the herpesvirus TK sequences that showed the greatest divergence is shown, but all the sequences of a larger set analyzed in ref. 5 retained the same consensus pattern of conserved residues. The apparent two copies of the B motif in the dCKs are denoted by underlining and overlining, respectively (see text). Herpesviruses: HSV1 = herpes simplex virus type 1; FHV = feline herpesvirus; PRV = pseudorabies virus; EHV4 = equine herpesvirus type 4; MDV = Marek disease virus; EBV = Epstein-Barr virus; HVS = herpesvirus saimiri; CCV = channel catfish virus. The sequences were from the PIR bank or were translated from the respective Genbank entries.

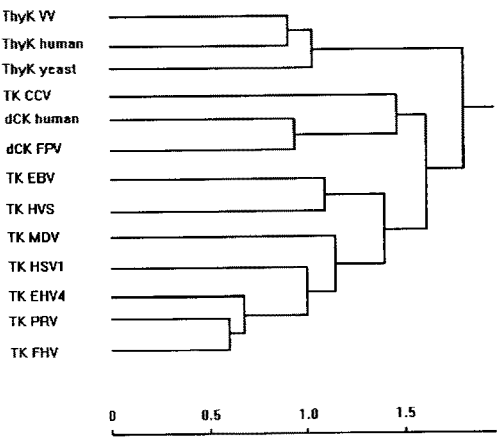

Fig. 3. A cluster dendrogram showing the relationships between the sequences of the dCKs, herpesvirus TKs, and the ThyKs. The dendrogram was generated by analysis of 141 amino acid residues aligned without gaps using the UPGMA algorithm (21) and the MDM78 matrix for amino acid residue comparison (23). The scale of distances calculated according to ref. 24 (with modifications) is shown at the bottom. The rate-independent maximum topological similarity algorithm (22) produced an identical tree (not shown).

in the tentative phylogenetic tree was uncertain, depending on the scoring matrix used for amino acid residue comparison (Fig. 3). The present analysis does not rule out the possibility of the independent capture of the nucleoside/dNMP kinase gene by CCV and the common ancestor of the other herpesviruses. Another point of uncertainty involved the relationship between the ThyK of VV and those of humans and yeast. The viral sequence was approximately equally similar to each of the two cellular sequences. It remains unclear whether this is due to the rapid evolution of the VV gene, to its origin from the putative cellular ancestor at an early stage of evolution, or whether both factors are contributing.

The observations presented in this communication are compatible with the hypothesis of independent histories of acquisition of the cellular genes for nucleoside/dNMP kinases by different large DNA viruses. Conceivably, in each case the maintenance of such genes in the viral genome was driven by the selective advantage provided to the virus by the respective enzymatic activities. We believe that this situation may serve as the model to conceptualize the process of the capture of cellular genes by these viruses in general.

References

1. Koonin E.V. and Senkevich T.G., Virus Genes *6*, 187–196, 1992.
2. Gentry G.A., Rana S., Hutchinson M., and Starr P., Intervirology *29*, 277–280, 1988.
3. Robertson G.R. and Whalley J.M., Nucleic Acids Res *16*, 11303–11317, 1988.
4. Smith G.R., de Carlos A., and Chan Y.S., Nucleic Acids Res *17*, 7581–7590, 1989.

5. Harrison P.T., Thompson R., and Davison A.J., J Gen Virol *72*, 2583–2586, 1991.
6. Su J.-Y. and Sclafani R. A., Nucleic Acids Res *19*, 823–827, 1991.
7. Chottiner E.G., Shewach D.S., Datta N.S., Ashcraft E., Gribbin D., Ginsburg D., Fox I.H., and Mitchell, B.S., Proc Natl Acad Sci USA *88*, 1531–1535.
8. Altshul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J., J Mol Biol *215*, 403–410, 1990.
9. Tartaglia J., Winslow J., Goebel J., Johnson G.P., Taylor J., and Paoletti E.J., Gen Virol *71*, 1517–1524, 1990.
10. Gorbalenya A.E., Blinov V.M., Donchenko A.P., and Koonin E.V., J Mol Evol *28*, 256–268, 1989.
11. Walker J.E., Saraste M., Runswick M.J., and Gay N.J., EMBO J *1*, 945–951, 1982.
12. Gorbalenya A.E. and Koonin E.V., Nucleic Acids Res *17*, 8413–8440, 1989.
13. Saraste M., Sibbald P.R., and Wittinghofer A., Trends Biochem Sci *15*, 430–434, 1990.
14. Goebel S.J., Johnson G.P., Percus M.E., Davis S.W., Winslow J.P., and Paoletti E., Virology *179*, 247–266, 1990.
15. Smith G.L., Chan Y.S., and Howard, S.T., J Gen Virol *72*, 1349–1376, 1991.
16. Koonin E.V. and Senkevich T.G., J Gen Virol *73*, 989–993, 1992.
17. Buller R.M., Smith G.L., Cremer K., Notkins A.L., and Moss B., Nature (Lond) *317*, 813–815, 1985.
18. Balasubramaniam N.K., Veerisetty V., and Gentry G., J Gen Virol *71*, 2979–2987, 1990.
19. Yan H. and Tsai M.-D., Biochemistry *30*, 5539–5546, 1991.
20. Story R.M. and Steitz, T.A., Nature (Lond) *355*, 374–376, 1992.
21. Sneath P. and Sokal R., *Principles of Numerical Taxonomy*. San Francisco, 1973.
22. Chumakov K.M. and Yushmanov S.Y., Molek. Genetika *3*, 3–9, 1988 (in Russian).
23. Dayhoff M.O., Barker W.C., and Hunt L.T., Meth Enzymol *91*, 524–549, 1983.
24. Feng D.F., Johnson M.S., and Doolittle R.F., J Mol Evol *21*, 112–125, 1985.