VIRUS GENES 7:1, 89–94, 1993

Gene A32 Product of Vaccinia Virus may be an ATPase Involved in Viral DNA Packaging as Indicated by Sequence Comparisons with Other Putative Viral ATPases

EUGENE V. KOONIN,[1] TATJANA G. SENKEVICH,[2] AND VICTOR I. CHERNOS[2]
[1]*Institute of Microbiology, Academy of Sciences and*
[2]*Institute of Viral Preparations, Academy of Medical Sciences, Moscow, Russia*

Received November 12, 1991
Revised and accepted February 11, 1992

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

Key words: amino acid sequence comparison, ATPase, vaccinia virus, virion morphogenesis

Abstract

Statistically significant sequence similarity was revealed between the gene A32 product of vaccinia virus (VV), gene I products (gpI) of filamentous single-stranded DNA bacteriophages, and IVa2 gene products of adenoviruses. Four conserved sequence motifs were delineated, the two N-proximal of which correspond to the A and B motifs of the purine NTP-binding pattern. Based on the role of gpI and IVa2 proteins in virion morphogenesis, and on the conservation of the NTP-binding pattern in these proteins, we hypothesize that the A32 gene product might be involved in an ATP-consuming function in VV virion formation, e.g., packaging of the DNA in the virus particle.

Introduction

All large DNA viruses, for which complete genome sequences have been reported to date, encode proteins containing the so-called purine NTP-binding sequence pattern. Participation of these proteins in various ATP(GTP)-consuming processes in virus reproduction, e.g., DNA duplex unwinding during replication and/or transcription, DNA packaging, or DNA precursor biosynthesis, has been either directly demonstrated or postulated (reviewed in ref. 1). In particular, vaccinia

virus (VV) encodes eight proteins of this class. These include four putative DNA and/or RNA helicases encoded by genes D6, D11, I8, and A18 (2,3); thymidine kinase encoded by gene J2 (4,5); thymidylate kinase encoded by gene A48 (5,6); a putative ATPase encoded by gene D5 and performing some not yet understood function in virus DNA replication (1,7); and finally, a gene A32 product whose function remains completely unknown (8). Previous computer-assisted analyses of the (putative) NTPase of VV have led to the potentially important prediction of the helicase activity for four proteins, a hypothesis that is compatible with the available experimental data for at least two of them, D6 and D11 (9,10).

In an attempt to shed some light on the possible functions of the two not yet characterized putative VV NTPases, D5 and A32, we compared their amino acid sequences with the database created from the Swissprot bank (release 18) by selection of all proteins containing either of the two forms of the so-called A motif of the NTP-binding pattern, G/Ax_4 GKS/T, or GxGKS/T (x = any amino acid residue). The search was performed using the program SMART (11), which includes a prescreening step analogous to the FASTP program of Lipman and Pearson (12), followed by detailed local similarity analysis by the DotHelix algorithm (13). No sequences scored high enough to be immediately considered homologs of the putative NTPases of VV. Thus we proceeded to further scrutiny of the sequences showing marginally significant scores. This analysis failed to reveal any interesting relationships for D5 protein. On the other hand, moderate but meaningful similarity was found to exist between A32 protein, gene I products (gpI) of small filamentous single-strand DNA bacteriophages, and IVa2 proteins of adenoviruses. Using the multiple alignment program OPTAL based on the Sankoff algorithm (14), the alignments of the A32 sequence with those of gpI and IVa2 scored 6.5 and 6.8 standard deviation (SD) above the mean of 25 random simulations, respectively, which is a strong indication that the observed relationship is not fortuitous (15). Although inspection of the resulting alignment showed that only very few amino acid residues were strictly conserved in all three groups of proteins (regarding A32 as a separate group), four conserved motifs could be delineated (Fig. 1). Motifs I and II corresponded to the A and B motifs, which together constitute the purine NTP-binding pattern and have been directly implicated in the interaction of several enzymes with ATP (reviewed in ref. 1). Motifs III and IV have not been described previously. Motif III showed the organization that is typical of many conserved motifs in purine NTP-binding proteins, particularly in DNA(RNA) helicases (1,2). It consists of a run of five bulky hydrophobic amino acid residues, presumably folding in a beta-strand and terminated by a partially conserved polar residue (Gln or His, for this particular case; Fig. 1). On the other hand, motif IV had a different structure, with the nearly invariant His residue followed by a putative hydrophobic beta-strand (Fig. 1).

The similarity between the amino acid sequences of gpI, A32, and IVa2 suggested that these proteins are evolutionarily, and probably functionally related, whereas the presence of the two motifs of the purine NTP-binding pattern among the best conserved sequence stretches indicated that they probably possess

NTPase activity. Moreover, it seems likely that these proteins are ATPases, not GTPases, as the latter enzymes encompass additional highly conserved signature sequences (16) that could not be found in the new protein family described here. Unfortunately, the data pertaining to the specific functions of phage gpI and adenovirus IVa2 proteins are rather scarce. GpI is required for bacteriophage assembly, although it is not a component of the phage particle (17,18). GpI is an integral membrane protein anchored in the bacterial inner membrane by its C-terminal domain and is thought to mediate the formation of the so-called export pore, i.e., the site of phage extrusion from the bacterial cell (19,20). In addition, a function in DNA packaging has been proposed for this protein based on the fact that certain mutations in gene I suppressed mutations in the packaging signal, the region in phage DNA required for assembly initiation (21). The presence of the NTP-binding pattern in the amino acid sequence of gpI suggests that this function might be ATP consuming. Adenovirus IVa2 protein is also thought to be involved in the maturation of the virus particle. This protein has been detected in virus-like particles that are intermediates in the assembly process but not in mature virions (22). These data allow us to hypothesize that A32 protein may be an ATPase involved in VV virion morphogenesis. As A32 lacks a counterpart to the C-terminal membrane-spanning domain of phage gpI, a direct function in VV DNA packaging seems most appropriate for this protein. It remains to be elucidated experimentally what common aspect may exist in the mechanisms of packaging of small phage circular ssDNA and large dsDNA of adenoviruses and poxviruses.

A further intriguing parallel may be envisaged between the (putative) functions of A32 protein and bacteriophage gpI in the morphogenesis of the respective viruses. GpI is involved in an interaction with bacterial thioredoxin that is essential for phage assembly (reviewed in ref. 18). Recently it has been shown that VV gene O2 product is homologous to glutaredoxins, small proteins mediating electron transfer in deoxyribonucleotide biosynthesis and functionally analogous to thioredoxins (8). Although glutaredoxins and thioredoxins do not show significant similarity, they are of nearly identical sizes, possess similar patterns of active Cys residues, and may adopt analogous spatial structures. Thus it is tempting to speculate that VV-encoded O2 protein may interact with A32 and function in virus morphogenesis in a manner analogous to the participation of thioredoxin in filamentous phage assembly.

In addition to adenoviruses and filamentous ssDNA bacteriophages, participation of proteins containing the NTP-binding pattern in DNA packaging has also been demonstrated for bacteriophages T7 and phi29 (1,23). Moreover, direct measurement of the DNA-dependent ATPase activity of phi29 packaging protein gp16 has been reported (23). However, detailed sequence comparisons that we performed failed to reveal any significant similarity (beyond the NTP-binding motif as such) of either this protein or the putative packaging ATPase of T7, gp19, with the protein family described above.

The observed significant similarity between protein sequences of such appar-

```
                              I("A")

gp1 Pf3    (2-214)  ITLITAVPGSGKTLYAIGLIEAALSEGRP-VFTNIS----------------
gp1 Cf1c   (8-213)  ISLLTGLPGSGKSLRIIQAIRYLMDKGAH-VY--V-----------------
gp2 SpV1  (77-287)  LSVIIGKLGTGKTLTLTYLSQTMKLLTDE-IYSNY-----------------
gp1 I2     (3-244)  VYVVTGKLGAGKTLVAVSRIQRTLAKGGI-VATNLNL--------KLHHFP
gp1 Ike    (3-244)  VYVVTGKLGAGKTLVAVSRIQRTLAKGGI-VATNLNL--------KLHHFP
gp1 M13    (3-244)  VYFVTGKLGSGKTLVSVGKIQDKIVAGCK-IATNLDL--------RLQNLP
                      ::**  * ***:  :      :
A32 VV     (50-286) RMVLTGGSGSGKTIYLLSLF-STLVKKYK-HIFLFT---------PVYN-P
           (20-256)   *: * :* **:  *  *  * *:      :*            :
IVA2 Ade2(166-432)  IGVIYGPTGCGKSQLLRNLLSSQLISPTPETVFFIAPQVDMIPPSELKAWE
IVA2 Ade7(166-432)  IGVIYGPTGCGKSQLLRNLLSSQLITPAPETVFFIAPQVDMIPPSELKAWE
consensus           zxZZxgxxGxGKtxxxxxxzxxxzxxxxxxxxxxZxxxxxxxxxxxxxxxx
                        a      s

gp1 Pf3    ------GLVKDK----FSNPHLLS--------D----APD-DWRDTPEGS---------
gp1 Cf1c   ------CNIDGI---SVPGTTPWA--------D-----PH-KWQDLPAGS---------
gp2 SpV1   ------PLEDDKVKVLTFKNLDFT--------D------R-TKPVPPDDS---------
gp1 i2     -QV--GRYAKQCRVMRIADKPTLE--------DLEIG-GN-LSYDESKNG---------
gp1 Ike    -QV--GRYAKQCRVMRIADKPTLE--------DLEIG-GN-LTYDESKNG---------
gp1 M13    -QV--GRFAKTPRVLRIPDKPSIS--------DLLIG-GN-DSYDENKNG---------
            :     .   .  .  ..          :.    .
A32 VV     -DY--DGYIWPNHINFVSSQESLE-------YNLIRTKSNIEKCIAVAQNHKKSAHF--
            :     *       :    :*         *: :    * :
IVA2 Ade2 MQICEGNYAPGPDGTIIPQSGTLRPRFVKMAYDDLILEHNYDVSDPRNIFAQAAARGPI
IVA2 Ade7 MQICEGNYAPGPQGTFIPQSGTLRPKFIKMAYDDLTQEHNYDVSDPRNVFARAAAHGPI
consensus xxxxxxxzxxxxxxxxZxxxxxZxxxxxxxxxdxxxxxxxxxxxxxxxxxxxxxxxxxx
                                          n

            II("B")                                    III

gp1 Pf3    LVVYDEAQQAHLYPSNAQRGPVTDERLTAMETH-RH-TGHDLVFITQAPT----FVHHH
gp1 Cf1c   ILFVDEAQ--HFFP--ARRGGDPVETIKAMSTI-RH-DGVRLVLATQQPN----YLDTY
gp2 SpV1   VILFDESYLYIDGTSPHDEKKVHSGKIPWIVLA-RH-FGNRALFTAQREG----MIWNN
gp1 i2     LIVLDECG--TWFNSRNWSDKSRQPVIDWFLHA-RK-LGWDVIFIIQDIS----LMDKQ
gp1 Ike    LLVLDECG--TWFNSRNWSDKSRQPVIDWCLHA-RK-LGWDIIFIIQDIS----LMDKQ
gp1 M13    LLVLDECG--TWFNTRSWNDKERQPIIDWFLHA-RK-LGWDIIFLVQDLS----IVDKQ
           ::::*: .           .  . .*..   . *  .   :: . *        :
A32 VV     LLIFDDVG--------DKLSKC-NTLIEFLNFG-RH-LNTSIILLCQTYR----HVPIL
            :: *:            :**        *          :::::            :
IVA2 Ade2 AIIMDECM--ENLGGHKGVSKFFHAFPSKLHDKFPKCTGYTVLVVLHNMNPRRDMAGNI
IVA2 Ade7 AIIMDECM--ENLGGHKGVSKFFHAFPSKLHDKFPKCTGYTVLVVLHNMNPRRDLGGNI
consensus xZZZDexxxxxxxxxxxxxxxxxxxxxxxzxxxxxkxxxxxzZZzxqxxxxxxxxxxxx
               d                                  h         h
```

Fig. 1. Alignment of the amino acid sequences of vaccinia virus A32 protein, putative morphogenetic ATPases of filamentous bacteriophages, and adenovirus IVa2 proteins. The conserved motifs are designated I–IV. Motifs I and II correspond to the A and B motifs of the NTP-binding pattern, respectively. Asterisks indicate identical amino acid residues in A32 and the bacteriophage protein sequences (with one possible exception in the latter), or in A32 and the two adenovirus IVa2 sequences; colons indicate similar residues in the same groups of sequences; periods indicate identical or similar residues in A32 and the three coliphage protein sequences. The grouping of the amino acid residues by physico-chemical similarity was as follows: G,A; S,T; D,E,N,Q; K,R; I,L,V,M; F,Y,W. The consensus line included strictly invariant residues (upper case) and residues conserved in all except one sequence (lower case); Z(z) indicates a bulky hydrophobic residue (I, L, V, M, F, Y, or W). The positions of the aligned sequences in the proteins are indicated by numbers in parentheses. The data for A32 proteins of two VV strains, Copenhagen (upper row) and WR (lower row), are presented; apparently, the Copenhagen protein has a 30 amino acid residue N-terminal extension (24). The amino acid sequences were from the Swissprot bank (Release 20).

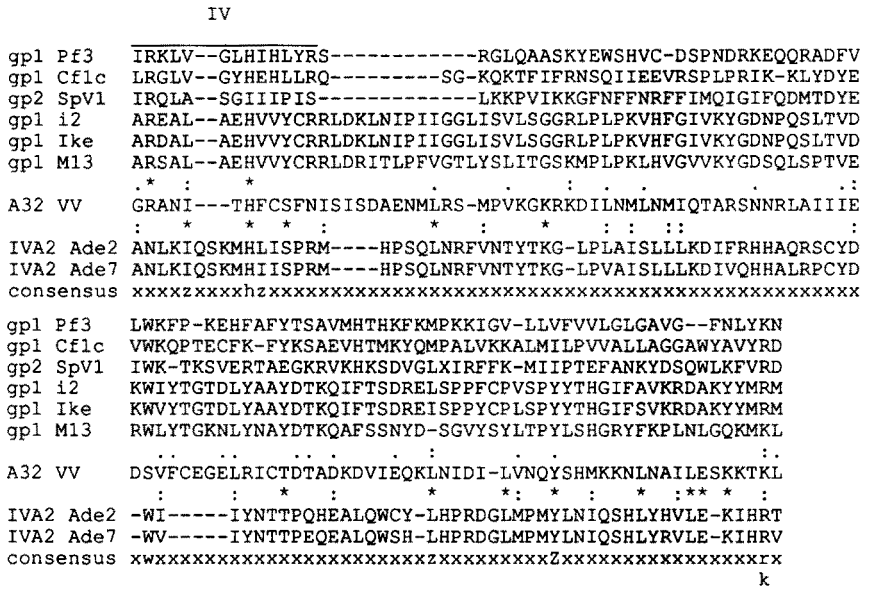

Fig. 1. Continued

ently unrelated viruses as VV and small filamentous ssDNA bacteriophages is unexpected. Although moderate in itself, this similarity appears striking, given the extent of sequence divergence within the group of the phage morphogenetic proteins. In the present study we expanded this family, showing that it includes not only gene I proteins of filamentous coliphages but also the distantly related proteins of two *Pseudomonas* (*Xanthomonas*) phages (Pf3 and Cf1c), and Spiroplasma virus 1 (SpV1). Inspection of the alignment in Fig. 1 showed that these proteins are actually not much more closely related to those of coliphages than the VV A32 protein is. From general considerations, a sequence similarity could be more readily expected to exist between proteins performing related functions in VV and such a large dsDNA bacteriophage as T7, which is, however, not the case. These findings emphasize once more the complexity of the pathways of virus evolution and the possibility of mechanistic parallels between molecular mechanisms involved in the expression of very different viruses. One possible evolutionary explanation for these findings might be independent acquisition of related cellular genes by otherwise distant viruses.

References

1. Gorbalenya A.E. and Koonin E.V., Nucleic Acids Res. *17*, 8413–8440, 1989.
2. Gorbalenya A.E., Koonin E.V., Donchenko A.P., and Blinov V.M., Nucleic Acids Res *17*, 4713–4730, 1989.

3. Koonin E.V. and Senkevich T.G., J Gen Virol *73*, 989–993, 1992.
4. Black M.E. and Hruby D.E., J Biol Chem *265*, 17584–17592, 1990.
5. Koonin E.V. and Senkevich T.G., Virus Genes *6*, 187–196, 1990.
6. Smith G.R., De Carlos A., and Chan Y.S., Nucleic Acids Res *17*, 7581–7590, 1989.
7. Roseman D.A. and Hruby D.E., J Virol *61*, 1398–1406, 1987.
8. Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P., and Paoletti E., Virology *179*, 247–266, 1990.
9. Kahn J.S. and Esteban M., Virology *174*, 459–471, 1990.
10. Broyles S.S. and Fesler B.J., Virology *64*, 1523–1529, 1990.
11. Brodsky L.I., Drachev A.L., Tatuzov R.L., and Chumakov K.M., Biopolimery i Kletka *7*, 10–14, 1991 (in Russian).
12. Lipman D.J. and Pearson W.R., Science *227*, 1435–1441, 1985.
13. Leontovich A.M., Gorbalenya A.E., and Brodsky L.I., Biopolimery i Kletka *6*, 12–19, 1990 (in Russian).
14. Gorbalenya A.E., Blinov V.M., Donchenko A.P., and Koonin E.V., J Mol Evol *28*, 256–268, 1989.
15. Doolittle R.F. *Of URFs and ORFs*. University Science Books, Mill Valley, CA, 1986.
16. Dever T.E., Glynias M.J., and Merrick W.C., Proc Natl Acad Sci USA *84*, 1814–1818, 1987.
17. Pratt D., Tzagoloff H., and Erdahl W.S., Virology *30*, 397–410, 1966.
18. Russel M., Mol Microbiol *5*, 1607–1613, 1991.
19. Lopez J. and Webster R.E.J., Bacteriology *163*, 1270–1274, 1985.
20. Horabin J.I. and Webster R.E., J Biol Chem *263*, 11575–11583, 1988.
21. Russel M. and Model P., J Virol *63*, 3284–3295, 1989.
22. Persson H., Mathisen B., Philipson L., and Pettersson U., Virology *93*, 198–208, 1980.
23. Guo P., Peterson C., and Anderson D., J Mol Biol *197*, 229–236, 1987.
24. Smith G.L., Chan Y.S., and Howard S.T., J Gen Virol *72*, 1349–1376, 1991.