Volume 17 Number 21 1989 Nucleic Acids Research

Viral proteins containing the purine NTP-binding sequence pattern

Alexander E.Gorbalenya* and Eugene V.Koonin[1]

Institute of Poliomyelitis and Viral Encephalitides, The USSR Academy of Medical Sciences, 142782 Moscow region and [1]Institute of Microbiology, The USSR Academy of Sciences, 117819 Moscow, USSR

Received June 30, 1989; Revised and Accepted September 20, 1989

ABSTRACT

A compilation is presented of viral proteins containing the NTP-binding sequence pattern, and criteria are suggested for assessment of the functional significance of the occurence of this pattern in protein sequences. It is shown that the distribution of NTP-binding pattern-containing proteins through the viral kingdom is strongly non-random. Sequence comparisons led to delineation of several families of these proteins, some of which could be brought together into superfamilies including also cellular proteins. The available biochemical evidence is compatible with the proposal that viral proteins in which the NTP-binding pattern is evolutionarily conserved might all be NTPases involved in: i) duplex unwinding during DNA and RNA replication, transcription, recombination and repair, and possibly mRNA translation; ii) DNA packaging, and iii) dNTP generation.

INTRODUCTION

Essentially all the main biochemical processes including DNA replication, recombination, repair and transcription, protein synthesis, membrane transport, signal transduction, and so on are coupled to nucleoside triphosphate (NTP) hydrolysis. Numerous, though not all, NTPases possess a common sequence pattern first recognized by Walker and co-workers (1). The pattern consisted of two separate motifs, the N-terminal 'A' site, and the C-terminal 'B' site. Subsequent studies led to a great increase in the number of proteins containing sequences related to this pattern and, not unexpectedly, to curtailment of the motif formulae themselves (2- 7). The latest analysis suggests that the motifs can be best presented as follows

<hydrophobic stretch>(G/A)xx(G)xGKS/T ('A'), and

<hydrophobic stretch>D(E/D) ('B').

The hydrophobic stretches contain at least 2 hydrophobic residues (as defined in reference 8) out of 5 ('A'), or 3 out of 5 ('B'). The residues shown in parentheses occur in most, but not all, sequences containing the pattern. However, in the 'A' motif, either G/A in the 1st position, or G in the 4th are necessarily present, thus generating the extended ('classical') and the short versions of this motif (reference 7 and discussion below). Where X-ray data have been obtained, it has been shown that both motifs usually folded in Rossmann-type beta-turn-alpha structural units,

©IRL Press

the N-terminal hydrophobic beta-strands being best conserved (9-11).The same studies revealed that the central flexible Gly-rich loop of the 'A' site bound the phosphoryl moiety of ATP or GTP, whereas the invariant D of the 'B' site chelated Mg^{2+} of Mg-NTP. Similar structural organization, and by implication, similar functional layout was suggested by secondary structure prediction of many, though not all, proteins containing the NTP-binding pattern (here after NTPp proteins; cf. references 7,12,13).

Previously we and others noticed the presence of the NTP-binding pattern in a number of proteins of RNA and DNA viruses (4,6,7,14-20). Here, a systematic survey of viral NTPp proteins is presented. Viral proteins are attractive for such a treatise for several reasons: (i) though numerous, viral protein sequences are still relatively easily handable; (ii) many complete viral genomic sequences are known; thus it may be possible to assess the actual status of NTPp proteins in viral reproduction; (iii) many small viruses have no non-essential proteins; hence, it is almost certain that highly conserved sequence patterns are of vital importance (iv) preliminary observations indicated that the distribution of NTPp proteins among virus groups was non-random in that they are nearly ubiquitous in some virus classes, and totally absent in others. There is reasonable hope that the results of such analysis might have applications outside virology.

APPROACH

To compile the complete list of viral NTPp proteins, the NBRF Protein Identification Resource database (release 15.0) and subsequently published viral protein sequences were searched for either of the two forms (extended or short) of the 'A' motif of the NTP-binding pattern. Sequences thus extracted were screened for the 'B' motif in the appropriate location, i. e. between the 'A' motif and the protein C-terminus. Clearly, however, due to the looseness of both motifs their presence in a protein sequence is in itself but a tentative indication for the query protein having an NTP-consuming function, and the NTP-binding pattern being crucial for this function. The problem of the relationship between the NTP-binding pattern and NTP-binding capacity can be depicted by a simple Venn diagram shown in Fig. 1. Three partially overlapping sets of proteins with known amino acid sequences are to be considered: (i) NTPp proteins, (ii) NTP-binding proteins, and (iii) proteins constituting distinct families delineated by sequence similarity. There is every reason to believe that in the proteins constituting the overlap between all three sets (designated I in Fig. 1) the residues of the NTP-binding pattern, provided it is conserved throughout a family, actually interact with NTP. The same can be predicted for those proteins within the overlap between the sets (i) and (iii) in which the NTP-binding pattern is evolutionarily conserved (II in Fig.1). The premise underlying this notion, which is very important, provided the prevalence of sequence over functional information, is that the residues conserved in a protein family are functionally indispensable. If these conserved residues fit the NTP-binding pattern formula, this function is very likely to be an NTP-consuming one (cf. also references 4-6). Conversely, wherever the NTP-binding pattern is not conserved

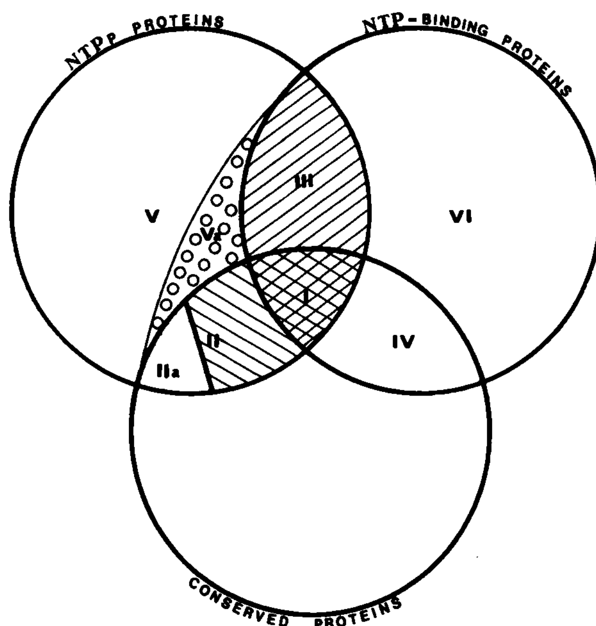

Fig. 1. A Venn diagram depicting the relationship between the sets of NTPp proteins, NTP-binding proteins, and proteins constituting distinct families delineated by amino acid sequence conservation ('conserved proteins'). The overlaps between the three sets are highlighted by hatching. Different groups of proteins are designated by Roman numbers as discussed in the text.

within a protein family, its presence in the sequences of some of the members of such a family is almost for sure due to chance (IIa). The presence of the NTP-binding pattern is likely to be significant also in the proteins falling within the overlap between the sets (i) and (ii), i.e. NTPp proteins with no known related sequences (III in Fig. 1), but with purported NTP-binding capacity. On the other hand, the overlap between sets (ii) and (iii) is likely to serve as the source of new significant patterns some of which might be related to the present one, thus leading to its relaxation (IV). Of particular interest will be direct identification of deviant forms of the NTP-binding pattern by analysis of statistically significant alignments of sequences of NTP-binding proteins lacking the presently accepted form of this pattern with those of NTP-binding NTPp proteins. The remaining parts of the sets (i) and (ii) are zones of ambiguity. For these zones, the significance of the presence of the NTP-binding pattern in protein sequences remains uncertain (V), whereas identification of functional sites in NTP-binding proteins is hardly feasible (VI). Scrutiny of the initial set of viral NTPp proteins allowed to place most of them into one category, or another, and to correspondingly reveal whether the

presence of the NTP-binding pattern was due to chance, or not. Inevitably, however, several proteins remained in the ambiguity zone. Thus the alignment of short sequence stretches surrounding the 'A' and 'B' motifs in those viral NTPp proteins where the presence of the pattern could be shown to be significant was analyzed to derive a two new consensus patterns (one for the 'extended', and the other for the 'short' types of 'A' motif') including additional partially conserved residues. The new patterns were used to search the sequences of the remaining viral NTPp proteins. Those with good correspondence to the consensus were tentatively included in the final list of putative NTP-binding NTPp proteins (Va in Fig. 1).

RESULTS AND DISCUSSION

The NTP-binding pattern was identified in over 100 viral proteins. Those of these proteins in which the presence of this pattern could be considered significant by the criteria outlined under APPROACH are characterized in Tables 1 to 3. Hereafter, we mean only this set, when referring simply to NTPp proteins. The proteins where the occurence of the pattern appeared likely to be fortuitous as well as viral NTP-binding proteins not containing the pattern are briefly discussed in the respective section below.

Distribution of NTPp proteins among virus groups

NTPp proteins are non-randomly distributed among virus classes (Table 1). Specifically, they are typical in dsDNA viruses, being found in all groups for which complete genome sequences have been reported, and in positive strand RNA viruses where most groups have them. In the latter class, a curious correlation between the genome size and the presence of NTPp proteins is observed. Among viruses sequenced so far, all those whose genome size exceeds app. 5.8 kb possess NTPp protein(s), and those with smaller genomes lack them. This correlation appeared to hold also for dsRNA viruses, though here the sampling of groups with known genomic structures is too small to draw definite conclusions. Among ssDNA viruses, some groups have NTPp proteins, whereas others do not, but no obvious correlation with genome size, or with any other trait could be discerned. Genomes of negative strand RNA viruses and of retroid viruses sequenced so far do not encode NTPp proteins, except for ras oncoproteins of certain oncogenic retroviruses whose genes are clearly of recent cellular origin (21). As complete sequences are available for members of three of the five known families of negative strand RNA viruses, and for all three families of retroid viruses, it seems reasonable to generalize on these observations and to hypothesize that NTPp proteins are generally not typical of the viruses of these classes.

Interestingly, not only dsDNA viruses with large genomes (such as T4, T5, herpes- and poxviruses), but also some of the positive strand RNA viruses have more than one NTPp protein. In this respect, no correlation with the genome size appears to exist, and potexvirus genomes lying exactly on the aforementioned threshold still encode two NTPp proteins (cf. Table 1 and reference 19). It remains a challenge for future studies to unravel the evolutionary and/or functional grounds for the non-random distribution of NTPp proteins among viruses.

TABLE 1
DISTRIBUTION OF NTP-BINDING PATTERN CONTAINING PROTEINS AMONG VIRUS GROUPS

No	Group (family) of viruses and prototype	Host range	Number of genomes sequenced	Genome size, kb	Number of NTPp proteins encoded per genome
	dsDNA VIRUSES				
1	Papovaviridae (SV40)	Mammals, birds	N	5.0-7.9	1
2	Corticoviridae (PM2)	Eubacteria	0	9	u
3	Plasmaviridae (MV-L2)	Mycoplasma	0	11	u
4	Phage P4 group	Eubacteria	p	11.5	1
5	Tectiviridae (PRD1)	Eubacteria	p	13	u
6	Phage ϕ29 group	Eubacteria	2	19.2	1
7	Adenoviridae (human adenoviruses)	Mammals, birds	1,p	28-45	1
8	Phage Mu group	Eubacteria	p	38	1
9	Podoviridae (T7)	Eubacteria	1,p	40	2
10	Siphoviridae (λ , P22)	Eubacteria	1,p	49	1
11	Phage N4 group	Eubacteria	0	72	u
12	Baculoviridae (AcNPV)	Insects	p	80-160	u
13	Phage T5 group	Eubacteria	p	121	2
14	Herpesviridae (HSV)	Vertebrates, invertebrates	3,p	125-172	2-3
15	Iridoviridae (FV3)	Vertebrates	p	150-400	u
16	Myoviridae (T4)	Eubacteria	p	165	7
17	Poxviridae (VV)	Vertebrates, insects	p	200	4
18	Polydnaviridae	Insects	p	?[i]	u
	ssDNA VIRUSES				
19	Circovirus (PCV)	Mammals	0	1.9	u
20	Geminiviridae (CLV)	Plants	7	2.5-5.5	1
21	SpV4 group	Mycoplasma	1	4.4	0
22	Parvoviridae (MVM)	Mammals, insects	5,p	5.1-5.5	1
23	Microviridae (ϕX174)	Eubacteria	3	5.4-5.7	0
24	Inoviridae (M13)	Eubacteria	4,p	6.4-6.9	1

		dsRNA VIRUSES			
25	Mycoviruses	Fungi	1,p	4.6; 3-9	0;u[ii]
26	Cryptic viruses	Plants	0	5.5	u
27	Birnaviridae (IBDV)	Animals	1,p	5.8-5.9	0
28	Cystoviridae (φ6)	Eubacteria	1	13.4	1
29	Reoviridae	Vertebrates, insects, plants	1,p	18.6-23.5	1
		POSITIVE STRAND ssRNA VIRUSES			
30	Leviviridae (MS2)	Eubacteria	4	3.5-4.0	0
31	Sobemovirus (SBMV)	Plants	1	4.0	0
32	Carmovirus (CarMV)	Plants	3	4.0	0
33	Tombusvirus (TBSV)	Plants	p	4.0	u
34	Tobacco necrosis virus group	Plants	p	4.0	u
35	Nodaviridae (BBV)	Insects	1	4.5	0
36	Nudaurelia β virus group	Insects	0	5.5	u
37	Luteovirus (BWYV, BYDV)	Plants	3	5.6-5.8	0
38	Potexvirus (PVX)	Plants	3	5.8-6.3	2
39	Dianthovirus (CarRSV)	Plants	p	6	u
40	Tymovirus (TYMV)	Plants	1,p	6.3	1
41	Tobamovirus (TMV)	Plants	1,p	6.4	1
42	Carlavirus (CarLV)	Plants	p	7-8	1(2)[ii]
43	Picornaviridae (PV)	Mammals, insects[iv]	N	7.5-8.5	1
44	Caliciviridae (SVEV)	Mammals	0	8	u
45	Pea enation virus group	Plants	0	8	u
46	Tobravirus (TRV)	Plants	1,p	8.3-10.4	1
47	Tricornaviridae (BMV)	Plants	3	9.3	1
48	Comovirus (CPMV)	Plants	1,p	9.4	1
49	Potyvirus (TEV)	Plants	3,p	9.5	1

50	Hordeivirus (BSMV)	Plants	1	10	2
51	Maize chlorotic dwarf virus group	Plants	0	10	u
52	Flaviviridae (YFV)	Mammals, arthropods	6,p	10.5-10.7	1
53	Furoviridae (BNYVV)	Plants	1	11.3	2
54	Alphaviridae (SNBV)	Mammals, insects	3,p	11.7	1
55	Nepovirus (TBRV)	Plants	1	12.0	1
56	Pestivirus (BVDV)	Mammals	1	12.6	1
57	Closterovirus (BYV)	Plants	0	13	u
58	Toroviridae (VB)	Mammals	0	20	u
59	Coronaviridae (IBV)	Mammals, birds	1,p	27.6	1
		NEGATIVE STRAND ssRNA VIRUSES			
60	Arenaviridae (LCMV)	Mammals, insects	p	10.5	u
61	Rhabdoviridae (VSV)	Animals, plants	2,p	11.2	0
62	Bunyaviridae (SSHV)	Mammals, insects	p	14-16	u
63	Orthomyxoviridae (InfV)	Mammals, birds	N	15	0
64	Paramyxoviridae (SV)	Mammals	3,p	15.4	0
65	Philoviridae (MarV)	Mammals	p	20	u
		RETROID VIRUSES			
66	Hepadnaviridae (dsDNA) (HBV)	Mammals, birds	4	3.2	0
67	Caulimoviridae (dsRNA) (CaMV)	Plants	3	7.7-7.9	0
68	Retroviridae (ssRNA) (RSV)	Vertebrates	N	9-10	0

SUMMARY OF TABLE 1

Genome type	Number of groups	Number of groups for complete sequences are available	Number of groups having NTPp proteins	Number of groups lacking NTPp proteins
dsDNA	18	6	11	0
ssDNA	6	5	3	2

dsRNA	5	4	2	2
(+)ssRNA	30	20	16	5
(-)ssRNA	6	3	0	3
Retroid	3	3	0	3
Total	68	41	32	15

The classification of viruses was from references 49 and 50, with modifications and additions based on subsequently published data. The list of dsDNA viral groups is incomplete: a number of eubacterial, archebacterial and algal viruses which have not been properly classified are not mentioned; for some of these viruses partial sequences have been reported but no NTPp containing proteins could be identified so far. Where appropriate, prototype viruses of respective groups are indicated. The values of genome size were from the references containing respective sequences (see Table 2 and the list of source references at the end of the paper), or, where sequence information was not available, from references 49 and 50. In the latter case, the values are to be regarded as approximate. Abbreviations: PM2, MV-L2, P4, ϕ29, PRD1, T7, λ, P22, ϕX174, M13, ϕ6, MS2, respective phages; AcNPV, Autographa californica nuclear polyhedrosis virus; HSV, herpes simplex virus; FV3, frog virus 3; VV, vaccinia virus; PCV, porcine circovirus; SpV4, Spiroplasma virus 4; CLV, cassava latent virus; MVM, minute virus of mice; IBDV, infectious bursa disease virus; SBMV, southern bean mosaic virus; CarMV, carnation mottle virus; TBSV, tomato bushy stunt virus; BBV, black beatle virus; BWYV, beet western yellows virus; BYDV, barley yellow dwarf virus; PVX, potato virus X; CarRSV, carnation ringspot virus; TYMV, turnip yellow mosaic virus; TMV, tobacco mosaic virus; CarLV, carnation latent virus; PV, poliovirus; SVEV, swine vesicular exanthema virus; TRV, tobacco rattle virus; BMV, brome mosaic virus; CPMV, cowpea mosaic virus; TEV, tobacco etch virus; BSMV, barley stripe mosaic virus; YFV, yellow fever virus; BNYVV, beet necrotic yellow vein virus; SNBV, Sindbis virus; TBRV, tomato black ring virus; BVDV, bovine viral diarrhoea virus; BYV, beet yellows virus; VB, virus Berne; IBV, infectious bronchitis virus; LCMV, lymphocytic choriomeningitis virus; VSV, vesicular stomatitis virus; SSHV, showshoe hare virus; InfV, influenza virus; SV, Sendai virus; MarV, Marburg virus; HBV, hepatitis B virus; CaMV, cauliflower mosaic virus; RSV, Rous sarcoma virus.
Designations: NTPp proteins, proteins in which the presence of the NTP-binding pattern could be considered significant according to the criteria delineated under Approach; p, only partial sequences available; u, situation with NTPp proteins uncertain due to lack of complete sequences; N, numerous sequences available and precise calculation is hampered by lack of a good definition of a species.
Footnotes:
[i]polydnaviruses have polydisperse genomes, the exact genome size necessary for reproduction unknown;
[ii]there are at least 6 distinct groups differing by genome size; the complete 4.6 kb genome sequence has been determined only for yeast L-A virus;

Clues to classification of NTPp proteins based on sequence similarity and its relationship to their structure and functions

Alignment of short sequence stretches of viral proteins encompassing the 'A' and 'B' motifs of the NTP-binding pattern is presented in Table 2. As indicated under APPROACH, inspection of this alignment and detailed sequence comparisons (see below) allowed refinement of the NTP-binding pattern itself (the 'consensus' rows in Table 2). First, it became obvious that the extended and the short forms of the 'A' motif are indeed related, and appear to be of equal predictive value. Moreover, further deterioration of the 'A' motif seems to be forthcoming as manifested by the striking substitution of Phe for the otherwise invariant Gly of the GKT/S in the T4 bacteriophage protein UvsX. One can speculate that the Lys and Thr(Ser) residues conserved in the 'A' motif are directly involved in NTP binding, whereas the role of the Gly residues is structural, i.e. maintaining the flexible loop conformation. Usually, this requires the sequence GxxGxG but in some cases two, or even one Gly (UvsX) suffice (Table 2), perhaps through some yet obscure compensatory changes. Second, preference for two negatively charged residues was typical of the 'B' motif, with the doublet frequencies ranged as follows: DE>DD>EE>ED. On the other hand, the second negatively charged residue was substituted in many instances, and in one notable case (that of bacteriophage P4 primase) substitution of the first one was observed. Thus, ultimately, the only strict requirement for the 'B' site might be a negatively charged residue (though isolated Glu has not yet been observed) flanked, from the N-terminal side, by a hydrophobic stretch with a beta-strand forming propensity. Third, the preference for hydrophobic residues was largely confined to specific positions in the putative N-terminal beta-strands of both sites, and in the putative C-terminal alpha-helix of the 'A' site (see the 'consensus' rows). Finally, an extremely wide range of variation of the distances separating the 'A' and 'B' motifs was observed (Table 2). As the two sites are supposed to be juxtaposed in the NTP-binding center, this implies that large domains can be looped out in the proteins of this class. These conclusions are corroborated both by experimental studies of several viral NTPp proteins, and by phylogenetic analysis as discussed below.

Sequence comparisons allowed delineation of 4 vast superfamilies, and 5 smaller groups of NTPp proteins, of which all but one included both viral and cellular proteins (Table 3). The criteria for delineation of these groups were that, within each group, the proteins should constitute a contigous network linked by statistically significant sequence similarity and

[iii]the sequenced portion of the carlavirus genome encodes one NTPp protein; however, the similarity of carlavirus genome organization to that of potexviruses suggests that the former also contains two such genes;
[iv]the partial sequence of CrPV, an insect picornavirus, differs very significantly from those of mammalian picornaviruses; hence, inclusion of this virus and its relatives in Picornaviridae is conditional.

contain distinct highly conserved sequence stretches, in addition to the 'A' and 'B' motifs of the NTP-binding pattern.

Superfamilies 1 and 2 each included proteins involved in genome replication of positive strand RNA viruses and of large DNA viruses, along with cellular helicases such as rep and uvrD (superfamily 1), or RAD3 (superfamily 2). DNA$^-$ ts mutants of HSV1 UL5 (superfamily 1) protein mapped near the 'B' motif, supporting the notion of its functional importance (22). Still more convincing evidence of this sort have been reported for cellular helicases RAD3 and uvrB belonging to superfamily 2 (cf. 7,23). Proteins of each superfamily contain 7 conserved sequence stretches together spanning, in some proteins at least, most of the polypeptide chain (7,19,20,23). Conceivably, some of these conserved segments should be responsible for the helicase activity, and our hypothesis is that viral proteins included in these families might be DNA or RNA helicases. Importantly, superfamily 2 brought together proteins with the extended and the short forms of the 'A' motif, confirming the conclusion of their equal significance. A careful sequence comparison of these two (putative) helicase superfamilies suggested they might be distantly related, constituting a higher rank group (7).

Superfamily 3 is, in a sense, even more diverse, including proteins implicated in genome replication of positive strand RNA viruses, and both ssDNA and dsDNA-containing small DNA viruses (24). Here too, the functionality of the 'A' and 'B' sites was validated by localization of mutations of poliovirus 2C protein conferring resistance to, or dependence on, guanidine, an inhibitor of viral RNA replication (25). SV40 T antigen is a well studied DNA and RNA helicase involved in viral DNA replication, as well as in transcription, and possibly mRNA translation regulation (26,27,27a). Antibody-inhibition as well as genetic studies confirmed that the NTP-binding pattern-containing domain of T antigen is indeed indispensable for the NTPase and helicase activities (12,28,29). The helicase hypothesis is compatible also with the established or proposed functions of the other proteins of this superfamily (Table 3). Still, in this case it is somewhat less warranted than for superfamilies 1 and 2 as the convincing sequence similarity is confined to a relatively small region, including only one highly conserved segment in, addition to the 'A' and 'B' sites, and probably comprising the NTPase domain proper.

Superfamily 4 differs from the other three in that it includes cellular proteins of very different functions, among them NTPase components of numerous membrane transport systems, eubacterial proteins uvrA which is a repair endonuclease/helicase subunit, mutS and hexA also involved in DNA repair, recN and recF, both implicated in recombination and postreplicative repair, and yeast repair protein RAD50 (30-32). Viral proteins belonging to this group are most closely related to UvrA, and at least two of them are ATP-dependent endonucleases or subunits thereof (Table 3).

A group of proteins whose sequences are related to that of E.coli dnaB protein encompasses bacteriophage primase-associated helicases involved in duplex unwinding during DNA replication initiation and, at least in T7, in elongation too (33,33a,34). In the sequences of these proteins, elimination of the second negatively charged residue of the 'B' motif should be noted.

TABLE 2
CONSERVED SEQUENCES OF VIRAL NTPp pPROTEINS

Group No	Source reference	Virus/	Protein	N-	'A' site	spacer length	'B' site	-C
					1 10 20		1 10	
					* * ***		**	
1	1	SV40	T	420	YWLFKGPIDSGKTTLAAALLE	26	FLVVFEDVKGTG	229
	2	BKV	T	422	YWLFKGGPISGKTTLAAGLLD	26	YMVVFEDVKGTG	198
	3	PYV	T	567	NILFRGPVNSGKTGLAAALIS	26	FVVCFEDVKGQI	159
	4	BFDV	T	347	YYIFKGPVNTGKTTVAAAILA	26	FMVLFEDVKGTP	148
	5	HPV1a	E1	434	CLLIFGPPNTGKSMFCTSLLK	24	KIGLLDDATKPC	117
	6	BPV1	E1	427	CLAFIGPPNTGKSMLCNSLIH	24	RAALVDDATHAC	124
4	7	P4	alpha protein	495	FLEVTGPGGSGKSIMAEIATL	27	SLIRLpDQEKWS	222
6	8	∅29, PZA	gp16	18	LNFVIGARGIGKSYAMKVYPI	73	?STIVFDEFIREK	208
7	9	ADE2	(IVA2)	165	IGVIYGPTGCGKSQLLRNLLS	57	?VKMAYDDLILEH	194
						88	?IAIIMDECMENL	163
8	10	Mu	B protein	94	IAVVCGNPGVGKTEAAREYRR	52	?GLVIIDEADHLG	133
9	11	T7	(gp19)	52	KFILQAFRGIGKSFITCAFVV	81	?DIIIADDVEIPS	420
						435	?GALAHDDRLDAL	66
10	12	λ	EA59	189	IHAFIGRNGCGKTTILNGMIG	175	SLVLFDEPEVHL	128
10	13	P22	gp12	194	LVIIAARPGMGKTELALKIAE	97	SLIMADYLGLIE	139
13	14	T5	D10	102	TCIINGKPGFGKTILALALAY	64	GTVIVDEVHHCV	251
	15	T5	D13	29	VTQLIGGNGLGKSTIATVIEE	502	NLLFLDEVISFI	48
14	16	EBV	BBLF4	66	AYVITGTAGAGKSTSVSCLHH	109	NVIVVDEAGTLS	613
	17	VZV	gp55	84	VVLISGNAGSGKSTCIQTLNE	132	NVIVIDEAGLLG	643
	18	HSV1	UL5	91	VYLITGNAGSGKSICVQTINE	131	NVIVIDEAGLLG	667
	19	VZV	gp51	61	VTVVRAPMGSGKTTALLEWLQ	69	DVLILDEVMSVI	673
	20	HSV1	UL9	75	VTVVRAPMGSGKTTALIRWLR	73	DVLVLDEVMSTL	670
	21	EBV	TK	286	SLFLEGAPGVGKTTMLNHLKA	80	CWILMDRHLLSA	209
	22	VZV	TK	14	RIYLDGAYGIGKTTAAEEFLH	90	KIMLSDRHPIAS	206
	23	HSV1	TK	51	RVYIDGPHGMGKTTTTQLLVA	85	LTLIFDRHPIAA	208
	24	MarHV	TK	12	RVYLDGPHGVGKSTTAEALVA	92	FVLVVDRHAVGP	240
	25	EHV	TK	28	RIYLDGVYGIGKSTTGRVMAS	85	LTVVFDRHPVAS	173
16	26	T4	(p64.0)	84	TLRLDGGRQSGKSIAVTNFAA	65	ILYIIDEPVKSP	145
	27	T4	gp41	191	LNVLMAGVNVGKSLGLCSLAA	93	TIIIVDYLGIVK	121
	28	T4	gp46	30	KTLITGRNGGGKSTMLEAITF	467	STLILDEVFDGS	50
	29	T4	UvsX	54	LLILAGPSKSfKSNFGLTMVS	62	VVVFIDSLGNLA	241
	30	T4	TK	3	LIFTYAAMNAGKSASLLIAAH	57	HCVFVDEAQFLK	98
	31	T4	PNK	2	KILTIGCPGSGKSTWAREFIA	136	?KAVIFDVDGTLA	129
						247	?VKLAIDDRTQVV	18
17	32	VV	ts17 protein	496	LTFFFGETATGKSTTKRLLKS	87	?YKPVFDRIDNAL	158
						70	?ATIIIDTNYKPV	175
						128	?KVKLLDEGLDGK	117
						205	?GYILYDNVVTLP	51
	33	VV	TK	5	IQLIIGPMFSGKSTELIRRVR	50	SVIGIDEGQFFP	89

					* * ***		**	
	34	FPV	TK	5	IHVITGPMFSGKTSELVRRIK	56	QVIGIDEAQFFL	89
	35	ShFV	TK	5	IHLIIGPMFAGKSTELIRLVR	50	EVIGIDEGQFFP	88
20	36	MSV	AL1''	16	SLYIVGPTRTGKSTWARSLGV	18	IYNIVDDIPFKF	88
	37	WDV	AL1''	?14	SIYICGPTRTGKTSWARSLGT	18	KYNIIDDIPFKF	84
	38	CSMV	AL1''	? 7	SLYICGPTRTGKTSWARSLGT	18	QFNIIDDIPFKF	78
	39	BCTV	AL1	212	SIILEGDSRTGKTMWARSLGA	20	EYNLIDDLDPTY	93
	40	CLV	AL1	214	SIVIEGDSRTGKTIWARSLGP	20	WYNVIDDVDPHY	92
	41	TGMV	AL1	217	SIIIEGDSRTGKTMWARSLGP	20	EYNVIDDVTPQY	82
22	42	AAV	NS1	328	TIWLFGPATTGKTNIAEAIAH	22	MVIWWEEGKMTA	152
	43	B19	NS1	318	TLWFYGPPSTGKTNLAMAIAK	22	SLVVWDEGIIKS	289
	44	MVM	NS1	393	TVLFHGPASTGKSIIAQAIAQ	22	NLIWVEEAGNFG	223
	45	H1	NS1	393	TVLFHGPASTGKSIIAQAIAQ	22	NLIWVEEAGNFG	226
	46	BPV	NS1	307	STLFYGPASTGKTNLAKAICH	22	MILWWEECIMTT	366
	47	ADV	NS1	516	CIWFYGPGGTGKTLLASLICK	22	NIIWAEECGNFG	108
	48	MDV	NS1	878	GMVLEGITNAGKSLILDNLLA	25	GSILFEEPMITP	244
24	49	M13	gp1	2	VYFVTGKLGSGKTLVSVGKIQ	59	?GLLVLDECGTWF	288
						50	?GNSDYDENKNGL	297
	50	IKe	gp1	2	VYVVTGKLGAGKTLVAVSRIQ	59	?GLLVLDECGTWF	271
						50	?GNLTYDESKNGL	280
	51	PF3	p301	1	ITLITAVPGSGKTLYAIGLIE	40	?SLVVYDEAQQAH	227
						59	?RGPVTDERLTAM	208
28	52	φ6	P4	121	VTALMGATGSGKSITLNEKLR	9	?VAEAYDELDTAV	169
						40	?FNVAVDSVRPLL	138
						102	?AILCADGNVSRT	76
						128	?APLAADTHMPSM	50
29	53	REO	λ2	887	CMLSLGAAAAGKSMTFDAAFQ	206	?GSFVVDSPDVDI	163
38	54	WClMV	p147	564	MSVIHGAGGSGKSHAIQKALR	46	KIIVFDDYSKLP	651
	55	WClMV	p26	22	PIVVHAIAGSGKSTVIRKILS	32	TLDILDEYGQLP	149
	55	PVX	p180	729	ACVIHGAGGSGKSHAIQKALR	45	SIVIFDDYSKLP	650
	57	PVX	p25	23	PLVVHAVAGAGKSTALRKLIL	32	GFAILDEYTLDN	138
	58	NMV	p186	862	GTIIHGCGGSGKSFAIQEWMR	47	PILVFDDYTKLP	697
	59	NMV	p26	24	PIIIHGVAGCGKSTIIQKIAL	30	PLDILDEYLGGP	150
40	60	TYMV	p206	970	VVHFAGFAGCGKTYPIQQLLK	45	SILVIDEIYKMP	792
41	61	TMV	p126	827	VVLVDGVPGCGKTKEILSRVN	52	KRLFIDECLMLH	184
46	62	TRV	p134	898	FELVDGVPGCGKSTMIVNSAN	51	DVLHFDEALMAH	205
47	63	BMV	RNA 1pr	679	ISMVDGVAGCGKTTAIKDAFR	49	HRLLVDEAGLLH	179
	64	CMV	RNA 1pr	707	ISQVDGVAGCGKTTAIKSMFN	48	SRVLVDEVVLLH	180
	65	AMV	RNA 1pr	832	VTIVDGVAGCGKTTNIKQIAR	50	QRLIFDECFLQH	190
50	66	BSMV	p130	832	FELIDGVPGCGKSTMILNSCD	54	DRFHFDEALKVH	221
	67	BSMV	p58	263	TGIISGVPGSGKSTIVRTLLK	39	DLLIIDEYTLAE	195
53	68	BNYVV	p237	936	LEYVKGGPGTGKSFLIRSLAD	43	QIIFVDEFTAYD	1097
	69	BNYVV	p43	127	VGIVLGAPGVGKSTSIKNLLD	44	NTMLVDEVTRVH	188
54	70	SNBV	nsP2	180	TIGVIGTPGSGKSAIIKSTVT	45	EVLYVDEAFACH	549
	71	SFV	nsP2	180	VVGVFGVPGSGKSAIIKSLVT	45	DILYVDEAFACH	540
	72	RRV	nsP2	180	VIGVFGVPGSGKSAIIKSVVT	45	ENLYVDEAFACH	540
60	73	IBV	'HEL'	276	RTTVQGPPGSGKSHFAIGLAV	71	DILLVDEVSMLT	221
43	74	PV1	2C	118	CLLVHGSPGTGKSVATNLIAR	26	GVVIMDDLNQNP	147

```
                                *  * ***                   **
    75 CVB3 2C         123 CLLLHGSPGAGKSVATNLIGR  26  AVVIMDDLCHNP 147
    76 BEV  2C         123 CALIHGSPGTGKSLATMIVGR  26  AVVVMDDLIQNP 148
    77 HRV2 2C         114 AIVIHGPPGAGKSITTNFLAK  24  SVVIMDDIMQNP 147
    78 HRV89 2C        118 AVLIHGSPGTGKSLATSVLAR  24  SVVIMDDIMQNP 146
    79 HRV14 2C        119 CVLIHGTPGSGKSLTTSIVGR  26  EVVIMDDLNQNP 147
    80 EMCV 2C         109 VIVLRGDAGQGKSLSSQVIAQ  28  FAAIMDDLGQNP 150
    81 TMEV 2C         110 VVVLRGAAGQGKSVTSQIIAQ  28  FSVIMDDLGQNP 150
    82 FMDV 2C          99 VVCLRGKSGQGKSFLANVLAQ  29  TVVVMDDLGQNP 152
    83 HAV  2C         132 VCYLYGKRGGGKSLTSIALAT  30  LVCIIDDIGQNT 144
48  84 CPMV p58        162 TIFFQGKSRTGKSLLMSQVTK  29  PFVLMDDFAAVV 369
    85 TBRV p72        209 WIYLFGQRHCGKSNFMATLDN  28  TFFHVDDLSSVK 377

49  86 TEV  CI          78 DFLVRGAVGSGKSTGLPYHLS  68  DFVIIDECHVND 527
    87 TVMV CI          79 DIILMGAVGSGKSTGLPTNLC  68  QFIIFDEFHVLD 526
    88 PPV  CI          79 DILIRGAVGSGKSTGLPFHLS  68  KCIIFDECHVHD 506

CONSENSUS 1                -!!!-g--g-gKs-----!--18-502--!!!de-----
                                a      t                   ed

9   89 T7   gp4        243 VIMVTSGSGMGKSTFVRQQAL  91  DVIILDHISIVV 136
16  90 T4   gp44        44 IILHSPSPGTGKTTVAKALCH  36  KVIVIDEFDRSG 206
17  91 VV   NTPaseI     49 SLLLFHETGVGKTMTTVYILK  56  ICVIIDECHNFI 483
    92 VV   NTPaseII    39 SVLLFHIMGSGKTIIALLFAL  57  SIFIVDEAHNIF 501

22  93 SDV  NS1         68 TFQIVSPPSAGKNFFIETVLA  25  RVNYWDEPNFEP 761

42  94 PVM  p25         22 PIVVHCVPGAGKSSLIRELLE  34  KFVVLDEYTLLT 137
52  95 YFV  NS3        191 MTTVLFHPGAGKTRRFLPQIL  70  EVIIMDEAHFLD 329
    96 WNV  NS3        187 QITVLLHPGAGKTRKILPQII  70  NLFIMDEAHFTD 328
    97 DEN2 NS3        187 LTIMDLHPGAGKTKRYLPAIV  70  NLIIMDEAHFTD 328
    98 KUN  NS3        187 ITVLDLHPGAGKTRRILPQII  70  NLFVMDEAHFTD 328
    99 JEV  NS3        188 MTVLDLHPGSGKTRKILPQII  70  NLFVMDEAHFTD 328
   100 TBEV NS3        193 ITVLDMHPGSGKTHRVLPELI  70  EVAIMDEAHWTD 325
56 101 BVDV p125      1899 FKQITLATGAGKTTE-LPKAV  74  SYIFLDEYHCAT 1982

CONSENSUS 2                --!!----G-GKt--!---!- 25-91-!!!!De-----
                                       s
```

<u>Designations</u>. 'CONSENSUS 1' and 'CONSENSUS 2', patterns of (partially) conserved amino acid residues derived for the proteins with the 'extended' and the 'short' types of the 'A' motif, respectively; invariant residues are shown in upper case (but see note viii); (!) hydrophobic residues; (*) residues fixed in the NTP-binding pattern formula; unique substitutions of P for D in the 'B' motif of phage P4 alpha protein, and of F for G in phage T4 UvsX protein are shown in lower case.

<u>Notes</u>.

(i) Where the identification of the 'B' motif could not be verified by phylogenetic analysis (see text), the respective sequences are preceded by '?'. In these cases, the sites best conforming to the derived consensus were included. Where ambiguity could not be avoided, more than one sequence is shown.

(ii) The names of proteins which were included based only on their good correspondence to the consensus (Va in Fig. 1) are bracketed.

(iii) The numbers in the first column correspond to the numbers of virus groups in Table 1.
(iv) Source references for each sequence are given at the end of the paper under the same numbers as in the second column of the table.
(v) Generally, the sequences are arranged as in Table 1, i.e. by genome size increase. However, sequences of positive strand RNA viral proteins are grouped into three families revealed upon sequence comparison (separated by blank rows in the table).
(vi) For VV ts17 protein, the 'B' motif sequence shown in the upper line is suggested by the alignment with herpesvirus thymidine kinases and yeast thymidylate kinase (unpublished). However, as the statistical significance of this alignment was relatively low, other putative 'B' sequences with closer resemblance of the consensus are also shown.
(vii) In two monopartite geminiviruses, WDV and CSMV, the open reading frames encoding the AL1'' proteins lack start codons. The authentic mRNAs for these proteins are thought to be generated via splicing (50a), but the exact sizes of the proteins have not been determined (marked by '?' in the table).
(viii) Of the two insect parvovirus (densovirus) protein sequences, one (SDV, entry 93) shows striking deviations from the "A" site consensus, whereas the other (MDV, entry 48) fully conforms to it. It cannot be excluded that the SDV sequence was determined erroneously.
(ix) For gpI of filamentous ssDNA phages, the 'B' sequences shown in the upper lines fully conformed to the consensus; however, direct alignment of the closely related protein sequences of M13 and Ike with that of PF3 (44) failed to match these sequences and suggested alternative candidate sites which are also shown, despite the fact they did no satisfy the 'B' formula.
(x) The boundaries of the putative helicase of IBV were determined from analysis of putative cleavage sites in the viral polyprotein (51).
(xi) For BVDV, the polyprotein boundaries are indicated as the exact positions of cleavage sites are unknown.
(xii) The exhaustive search of the literature was terminated on April 1, 1989; however, several subsequently published or yet unpublished sequences also were included. Where very similar sequences of several strains, or serotypes of the same virus species were available, only one was included (e.g. picornaviruses and flaviviruses); also, certain sequences of papova-, pox-, gemini- and parvoviruses very similar to those incorporated in the Table are not shown.
<u>Abbreviations</u> (not used in Table 1):
PZA, IKe, PF3, respective bacteriophages; BKV, BK virus, PYV, human polyoma virus, BFDV, budgerigar fledgling disease virus; HPV1a, human papilloma virus type 1a, BPV1, bovine papilloma virus (papovaviruses); ADE2, human adenovirus type 2; EBV, Epstein-Barr virus, VZV, varicella-zoster virus, HSV1, herpes simplex virus type 1, MarHV, marmoset herpes virus (herpesviruses); VV, vaccinia virus, FPV, fowlpox virus; ShFV, Shope fibroma virus (poxviruses); WDV, wheat dwarf virus; CSMV, cassava stripe mosaic virus; BCTV, beet curly top virus; TGMV, tomato golden mosaic virus, CLV, cassava latent virus (geminiviruses); AAV, adeno-associated virus, of mice, BPV, bovine parvovirus; ADV, Aleutian disease virus; MDV, mosquitoe

Another small group unites phage T4 UvsX protein and bacterial recA proteins, strand-exchanging helicases involved in DNA recombination and repair, and, in the case of UvsX, also in replication (35-37). The unique substitution of the consensus Gly residue in the putative 'A' motif of UvsX has already been mentioned. Importantly, the direct relationship between the sequence of UvsX shown in Table 2 and the classical 'A' motif could be confirmed by highly significant alignment with recA sequences (unpublished observations).

Together, these 6 groups encompass the majority of viral NTPp proteins which all appear to be involved in genome replication and/or transcription, in most, if not all, cases mediating NTP-dependent DNA or RNA duplex unwinding.

Two other small groups include two types of thymidine kinases of which one (poxviruses and T4) is closely related to cellular thymidine kinases (38-40), and the other (herpesviruses) display a more remote similarity to cellular thymidilate kinase (41). A vaccinia virus protein with yet unknown activity involved in virus DNA replication also could be tentatively included in the latter group (unpublished observations). Spontaneous and artificial mutants of HSV thymidine kinase have been extensively explored, and mutations modifying activity have been mapped directly in, or in the close proximity of 'A' and 'B' sites (42,43).

Finally, the NTPp proteins of filamentous ssDNA phages involved in virion morphogenesis, whereas (for the while) having no cellular homologs, display sequences divergence sufficient to warrant their description as a distinct group (44).

This brief discussion, together with the data of Table 3, shows that a general correlation exists between the tentative classification of NTPp proteins based on sequence similarity and their functional grouping. For several groups of NTPp proteins of cellular origin, e.g. H^+-ATPases and related transcription termination factor rho, and the vast GTPase family (cf. 5), no viral counterparts could (yet) be identified. It is of obvious interest if this gap could be eventually filled, perhaps upon accumulation of sequences related to those of viral NTPp proteins which, at this stage, could not be classified (cf. Table 3).

densonucleosis virus; SDV, silkworm densonucleosis virus (parvoviruses); WClMV, white clover mosaic virus, NMV, narcissus mosaic virus (potexviruses); PVM, potato virus M (carlavirus); CMV, cucumber mosaic virus, AMV, alfalfa mosaic virus (tricornaviruses), SFV, Semliki forest virus, RRV, Ross river virus (alphaviruses); PV1, poliovirus type 1, CV, coxsackie virus type B3, HRV2,89,14 human rhinoviruses of respective serotypes; EMCV, encephalomyocarditis virus, TMEV, Theiler murine encephalomyelitis virus, FMDV, foot-and-mouth disease virus, type A10, HAV, hepatitis A virus (picornaviruses), TVMV, tobacco vein mottling virus, PPV, plum pox virus (potyviruses); WNV, West Nile virus, DEN2, Dengue virus type 2; KUN, Kunjin virus; JEV, Japanese encephalitis virus (flaviviruses); gp, gene product; pr, product; TK, thymidine kinase; PNK, polynucleotide kinase; 'HEL', putative helicase.

TABLE 3

GROUPS OF VIRAL NTPp PROTEINS DELINEATED BY SEQUENCE COMPARISON AND THEIR FUNCTIONAL CHARACTERISTICS

Viral group, or organism	NTPp protein (references)	Properties/ Activity	Function(s) in viral reproduction or in the cell
	PUTATIVE HELICASE SUPERFAMILY 1 (19,20)		
	Viral proteins		
Potexvirus	p147,p180,p186 p26,p25	unknown unknown	unknown unknown
Carlavirus	p26	unknown	unknown
Tymovirus	p206	unknown	unknown
Tobamovirus	p126 (52,53)	NTP-binding protein	RNA synthesis
Tricorna-viridae	RNA 1 product (54)	unknown	RNA synthesis
Tobravirus	p130	unknown	unknown
Hordeivirus	p58 p130	unknown unknown	unknown unknown
Furovirus	p43 p237	unknown unknown	unknown unknown
Alphaviridae	nsP2 (55)	unknown	RNA synthesis, particularly that of subgenomic RNA
Coronaviridae	'HEL' (51)	unknown	unknown
Herpesviridae	BBLF4, gp55, UL5 (22,56)	unknown	DNA replication
	Cellular proteins		
E.coli	uvrD (57)	DNA helicase	DNA replication
E.coli	rep (58)	DNA helicase	DNA replication
E.coli	helicase IV (59)	DNA helicase	unknown
E.coli	recB (60)	DNA-dependent ATPase, subunit of recBCD endo-nuclease	DNA recombination and repair

E.coli	recD (61,62)	unknown, subunit of recBCD	DNA recombination and repair
yeast	PIF (63)	DNA helicase	Mitochondrial DNA replication
	PUTATIVE HELICASE SUPERFAMILY 2 (7,23)		
		Viral proteins	
Potyvirus	CI (64)	RNA-dependent ATPase	unknown
Flaviviridae	NS3 (65)	unknown	RNA synthesis
Pestivirus	p125	unknown	unknown
Herpesviridae	gp51, UL9 (66)	DNA replication origin-binding protein	DNA replication, probably initiation
Poxviridae	NTPases I,II (67,68)	DNA-dependent NTPases	DNA transcription
T5	D10	unknown	unknown
		Plasmid protein	
Yeast mitochondrial plasmid pGKL2	P4	unknown	unknown
		Cellular proteins	
E.coli, slime mold, yeast, Drosophila, mammals	eIF-4A and related proteins ('DEAD' family) [(69,70), and references therein]	(putative) RNA helicases	translation initiation, ribosome assembly, transcription regulation (?)
E.coli	recQ (71)	unknown	DNA recombination and repair
E.coli M.luteus	uvrB (72)	DNA-dependent ATPase, subunit of uvrABC endonuclease-helicase	DNA repair
Yeast	RAD3 (73)	DNA helicase	DNA repair and replication
	PUTATIVE HELICASE SUPERFAMILY 3 (18,24)		
		Viral proteins	
Picornaviridae	2C (74,75)	unknown	RNA synthesis

Comovirus	p58 (76)	unknown	RNA synthesis
Nepovirus	p72	unknown	unknown
Papova-viridae	T antigen, E1 (26,27)	DNA-dependent ATPase, DNA and RNA helicase	Initiation and elongation of DNA replication, transcription and translation (?) regulation
Parvo-viridae	NS1 (77,78)	unknown	DNA replication, transcription regulation
Gemini-viridae	AL1 (79,80)	unknown	DNA replication
P4	alpha protein (81)	DNA primase	Initiation of DNA replication
		Cellular protein	
E.coli	La (82)	ATP-dependent DNA-stimulated protease	Abnormal polypeptide degradation
		SUPERFAMILY 4 (30-32)	
		Viral proteins	
Siphoviridae (λ)	EA59 (83)	DNA-dependent ATPase, ATP-dependent endonuclease	Unknown
Myoviridae (T4)	gp46 (84)	exonuclease subunit	DNA replication
T5	D13 (85)	unknown	DNA replication
		Cellular proteins	
E.coli, M.luteus	uvrA (86,87)	ATPase, GTPase, subunit of uvrABC endonuclease-helicase	DNA repair
E.coli	recN (88)	unknown	DNA recombination and repair
E.coli	recF (89)	unknown	DNA recombination and repair
Salmonella typhimurium	mutS (90)	ATPase	DNA repair

Streptococcus pneumoniae	HexA (90)	unknown	DNA repair
Yeast	RAD50 (90a)	unknown	DNA repair and recombination
Eubacteria, slime mold, plants, insects, mammals	'active transport proteins' [(31), and references therein]	ATPases or ATP-binding proteins	Active transport of various compounds, cell division, nodulation
PRIMASE-ASSOCIATED HELICASE FAMILY (33,33a,34)			
		Viral proteins	
Siphoviridae (P22)	gp12 (91)	DNA-dependent ATPase	Initiation of DNA replication
Podoviridae (T7)	gp4 (92)	helicase, DNA primase	Initiation and elongation of DNA replication
Myoviridae (T4)	gp41 (93)	helicase	Initiation of DNA replication
		Plasmid protein	
Cryptic plasmid of Chlamydia	dnaB-like protein (33)	unknown	DNA replication
		Cellular protein	
E.coli	dnaB (94)	helicase	Initiation of DNA replication
RecA-RELATED HELICASE GROUP (35, and unpublished observations)			
		Viral protein	
Myoviridae (T4)	UvsX (36)	strand-exchanging helicase	DNA replication, recombination and repair
		Cellular proteins	
Eubacteria	RecA (37)	strand-exchanging helicase	DNA recombination and repair
THYMIDINE KINASE FAMILY 1 (38-40)			
		Viral proteins	
Myoviridae (T4)	TK (95)	thymidine kinase	DNA precursor synthesis

Poxviridae	TK (96)	thymidine kinase	DNA precursor synthesis
		Cellular proteins	
Mammals, birds	TK (96)	thymidine kinase	DNA precuror syntheais
		THYMIDINE KINASE FAMILY 2 (41)	
		Viral proteins	
Herpes-viridae	TK (96)	thymidine and thymidilate kinase	DNA precursor synthesis
(Poxviridae	ts17 (97) protein	unknown	DNA replication)
		Cellular proteins	
Yeast	CDC8 (98)	thymidilate kinase	DNA precursor synthesis
		PHAGE MORPHOGENETIC PROTEIN GROUP (44)	
Inoviridae (M13,f1,Pf3)	gpI (99)	unknown	Virion morphogenesis
		UNCLASSIFIED VIRAL PROTEINS	
Cystoviridae (ϕ6)	P4 (100)	ATPase	RNA synthesis
Reoviridae (human reo-virus)	λ 2 (101)	mRNA:guanylyl transferase	RNA synthesis (mRNA maturation)
ϕ29, PZA	gp16 (46)	DNA- and prohead-dependent ATPase	Virion morphogenesis
Mu	B protein (102)	DNA-dependent ATPase	DNA replicative transposition
Myoviridae (T4)	gp44 (103)	helicase component	DNA replication
	PNK (104)	polynucleotide kinase	DNA replication
	orf64.0	unknown	unknown
Podoviridae (T7)	gp19 (105)	unknown	Virion morphogenesis

Adenoviridae	IVa2 (106)	unknown	Virion morphogenesis

Notes:
(i) references indicated after the group names pertain to the respective group characterization by sequence comparison, and those at the protein designations to biochemical properties of respective proteins and their function(s);
(ii) only preliminary data have been reported on ATPase activity of potyvirus and o6 phage proteins
(iii) in the thymidine kinase family 2, VV ts17 protein is shown in brackets to emphasize that its inclusion in this family is preliminary.

__Fortuitous occurence of the NTP-binding pattern in viral proteins and viral NTP-binding proteins lacking the pattern__

Viral proteins in which the presence of the NTP-binding pattern appeared to be due to chance included adenovirus DNA-binding protein, Epstein-Barr virus DNA polymerase, alfalfa mosaic virus RNA polymerase, influenza virus hemagglutinin, Sendai virus hemagglutinin-neuraminidase, and some other. In most of these cases the pattern was not conserved in closely related protein sequences. Notably, these proteins either most likely lack NTP-binding capacity (e.g. negative strand RNA virus hemagglutinins), or are thought to possess NTP-binding centers of a different type (e.g. DNA and RNA polymerases). Several viral proteins with reported NTPase activity lack the NTP-binding pattern conforming to the formulae accepted here. These included F protein of human immunodeficiency virus, gpA (terminase) of phage, gp17 and dNMP kinase of T4 phage, and conceivably this list will be expanded. For the first three proteins, potential NTP-binding sites have been suggested (45,46), but our analysis failed to confirm the relationship between these sequences and the NTP-binding pattern. At best, these proteins encompass remote deviants of this pattern. Deviant forms of the NTP-binding pattern appeared to be conserved in certain viral proteins for which NTP-binding capacity has not yet been reported, e.g. homologous proteins encoded by spliced transcripts of herpesviruses [(47) and unpublished observations). Thus, in future further loosening of the pattern can be anticipated. This will require application of novel database search strategies, e.g screening for several partially overlapping sequence motifs conserved in distinct protein families, like those described above.

__Evolutionary implications__

The degree of sequence similarity within each of the groups delineated above suggested evolutionary relatedness between the respective proteins. Clearly, this applied to the NTP-binding pattern which, in each case, was among the best conserved sequences. A tempting further speculation might be that this pattern, apparently the only one common to all groups, arose in evolution only once, perhaps in the common ancestor of all extant forms of life.

The distribution of viral NTPp proteins among groups arising from sequence comparison generally did not follow the

classification of viruses by genome type, not to speak of families and smaller groups. Interestingly, NTPp proteins of positive strand RNA viruses are interspersed among the three large superfamilies of (putative) helicases. On the other hand, monophyletic origin seems a plausible possibility for these proteins as they are related not only by sequence similarity but also by similar localization in viral multidomain proteins (6,17). By contrast, NTPp proteins of eukaryotic small DNA viruses all belong to the same superfamily. These observations may suggest a very ancient origin for positive strand RNA viruses which could subsequently give rise to several groups of small DNA viruses and dsRNA viruses (cf.24,48). Large dsDNA viruses are represented in each group of NTPp proteins, suggesting, perhaps not unexpectedly, a complex evolutionary network. Generally, evolutionary interpretation of the grouping of NTPp proteins based on sequence similarity awaits further careful consideration in the context of comparison of complete genomes.

CONCLUSIONS

(i) Using the suggested criteria, about 100 virus proteins were identified in which the presence of the NTP-binding pattern could be considered significant. These proteins are encoded by genomes of viruses of 32 groups. Of 41 virus groups for which complete sequences have been reported, NTPp proteins were detected in 26, making the NTP-binding pattern one of the best (if not the very best) conserved sequences in viral proteins. Of the 6 main virus classes defined by genome type, two (negative strand RNA viruses and retroid viruses) appear to lack NTPp proteins. All dsDNA viruses for which complete sequences are available, and most groups of positive strand RNA viruses and of ssDNA viruses have such proteins. For positive strand RNA viruses, and apparently for dsRNA viruses, a strong correlation was observed between genome size and occurence of NTPp proteins.

(ii) Based on sequence comparisons, viral NTPp proteins could be classified into distinct families some of which could be brought together into superfamilies. The latter included both viral and cellular NTPp proteins. In all these groups the NTP-binding pattern was among the best conserved sequences.

(iii) What is known of functions of viral and related cellular NTPp proteins, either directly supports, or at least does not contradict the proposal about their NTPase activity. The NTP-consuming functions of these proteins include: a) duplex unwinding during DNA and RNA replication, transcription, recombination, and repair, and perhaps also mRNA translation; b) DNA packaging, and c) dTTP generation.

(iv) Although there is no one-to-one relationship between the NTP-binding pattern and ATP- or GTP-binding capacity, the present analysis showed that this pattern is the best available predictor of such capacity. Expansion of the set of NTPp proteins led to partial deterioration of this pattern formula. Thus, we suggest that, to extract the sequences of putative NTP-binding proteins from databases both comprehensively and selectively, the latter should be screened sequentially with patterns conserved in different groups of NTPp proteins.

ACKNOWLEDGEMENTS
The authors are greatful to Professor V. I. Agol for interest and constant encouragement, to Drs. V. M. Blinov and A. P. Donchenko, and Mr. Yu. I. Wolf for collaboration in the original work incorporated in this review, to Drs. B. N. Afanasiev, C. Fritsch, G. Gustafson, Yu. V. Kozlov, S. Yu. Morozov and A. G. Pletnev for communication of sequence data prior to publication, and to Drs. K. M. Chumakov, S. Yu. Morozov and M. N. Rozanov for useful discussions.

*To whom correspondence should be addressed

REFERENCES

1. Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) EMBO J. 1, 945-951.
2. Halliday,K.R. (1984) J Cyclic Nucleotides Protein Phosphorylation Res. 9, 435-448.
3. Moller,W. and Amons,R. (1985) FEBS Lett. 186, 1-7.
4. Doolittle,R.F. (1986) In Inouye,M. (ed.) Protein Engineering, pp.15-27, Plenum, New York.
5. Hodgman,T.C. (1989) CABIOS 5, 1-13.
6. Gorbalenya,A.E., Blinov,V.M., Donchenko,A.P. and Koonin,E.V. (1989) J.Mol.Evol. 28, 256-268.
7. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1989) Nucleic Acids Res. 17, 4713-4730.
8. Biou,V., Gibrat,F., Levin,J.M., Robson,B. and Garnier,J. (1988) Protein Engineering 2, 185-191.
9. La Cour,T.F.M., Nyborg,J., Thirup,S. and Clark,B.F.C. (1985) EMBO J. 4, 2385-2388.
10. Fry,D.C., Kuby,S.A. and Mildvan,A.S. (1986) Proc.Nat.Acad.Sci.USA 83, 907-911.
11. De Vos,A.M., Tong,L.M., Milburn,M.V., Matias,P.M., Jancarik, J., Noguchi,S., Nishimura,S., Miura,K., Ohtsuka, E. and Kim, S.-H. (1988) Science 239, 888-893.
12. Bradley,M.K., Smith,T.F., Lathrop,R.H., Livingston,D.M. and Webster,T.A. (1987) Proc.Nat.Acad.Sci.USA 84, 4026-4030.
13. Chin,D.T., Goff,S.A., Webster,T.A., Smith,T.F. and Goldberg, A.L. (1988) J.Biol.Chem. 263, 11718-11728.
14. Seif,I. (1984) Virology 138, 347-352.
15. Gorbalenya,A.E., Blinov,V.M. and Koonin,E.V. (1985) Molek.Genetika (Moscow) No.11, 30-36.
16. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov, V.M. (1987) Molek.Biol. (Moscow) 21, 1566-1572.
17. Domier,L.L., Shaw,J.G. and Rhoads,R.E. (1987) Virology 158, 20-27.
18. Astell,C.R., Mol,C.D. and Anderson,W.F. (1987) J.Gen.Virol. 68, 885-893.
19. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov, V.M. (1988) FEBS Lett. 235, 16-24.
20. Hodgman,T.C. (1988) Nature 333, 22-23.
21. Barbacid,M. (1987) Ann.Rev. Biochem. 56, 779-827.
22. Zhu,L. and Weller,S. (1988) Virology 166, 366-378.
23. Blinov,V.M., Koonin,E.V., Gorbalenya,A.E., Kaliman,A.V. and Kryukov,V.M. (1989) FEBS Lett. 252, 47-52.
24. Gorbalenya,A.E., Koonin,E.V. and Wolf,Yu.A., submitted.

25. Pincus,S.E., Rohl,H. and Wimmer,E. (1987) Virology 157, 83-88.
26. Wold,M.S., Li,J.J. and Kelly,T.J. (1987) Proc.Nat.Acad.Sci.USA 84, 3643-3647.
27. Wiekowski,M., Schwartz,M.W. and Stahl,H. (1988) J.Biol.Chem. 263, 436-442.
27a. Scheffner,M., Knippers,R. and Stahl,H. (1989) Cell 57, 955-963.
28. Cole,C.N., Tornow,J., Clark,R. and Tjian,R. (1986) J.Virol. 57, 539-546.
29. Mole,S.E., Gannon,J.V., Ford,M.J. and Lane,D.P. (1987) Phil.Trans.R.Soc.Lond. B317, 455-469.
30. Doolittle,R.F., Johnson,M.S., Husain,I., Van Houten,B., Thomas,D.C. and Sancar,A. (1986) Nature 323, 451-453.
31. Higgins,C.F., Gallagher,M.P., Mimmack,M.L. and Pearse,S.R. (1988) BioEssays 8, 111-116.
32. Gorbalenya,A.E. and Koonin,E.V., in preparation.
33. Hatt,C., Ward,M.E. and Clarke,I.N. (1988) Nucleic Acids Res. 16, 4053-4067.
33a. Wong,A., Kean,L. and Maurer,R. (1988) J.Bacteriol. 170, 2668-2675.
34. Koonin,E.V. and Gorbalenya,A.E., in preparation.
35. Fujisawa,H., Yonesaki,T. and Minagawa,T. (1985) Nucleic Acids Res. 13, 7473-7481.
36. Formosa,T and Alberts,B. (1986) J.Biol.Chem. 261, 6107-6118.
37. Smith,K.C. and Wang,T.V. (1989) BioEssays 10, 12-16.
38. Boyle,D.B., Coupar,B.E.H., Gibbs,A.J., Seigman,L.J. and Both,G.W. (1987) Virology 156, 355-367.
39. Gibbs,A.J. (1987) J.Cell.Sci.Suppl. 7,319-337.
40. Koonin,E.V. et al., in preparation.
41. Robertson,G.R. and Whalley,J.M. (1988) Nucleic Acids Res. 16, 11303-11317.
42. Liu,Q. and Summers,W.C. (1988) Virology 163, 638-642.
43. Sawyer,M.H., Inchauspe,G., Biron,K.K., Waters,D.J., Straus, S.E. and Ostrove,J.M. (1988) J.Gen.Virol. 69, 2585-2593.
44. Koonin,E.V. and Gorbalenya,A.E., in preparation.
45. Guy,B., Kieny,M.P., Riviere,Y., Le Peuch,C., Dott,K., Girard,M., Montagnier,L. and Lecocq,J.-P. (1987) Nature 330, 266-269.
46. Guo,P., Peterson,C. and Anderson,D. (1987) J.Mol.Biol. 197, 229-236.
47. Gosta,R.H., Draper,K.G., Kelly,T.J. and Wagner,E.K. (1985) J.Virol. 54, 317-328.
48. Koonin,E.V., Gorbalenya,A.E. and Chumakov,K.M. (1989) FEBS Lett. 252, 42-46.
49. Matthews,R.E.F. (1982) Intervirology 17, 1-199.
50. Ackermann,H.-W. (1987) Microbiol.Sci. 4, 214-218.
50a. Schalk,H.-J., Matzeit,V., Schiller,B., Schell,J. and Gronenborn,B. (1989) EMBO J. 8, 359-364.
51. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1989) Nucleic Acids Res. 17, 4847-4862.
52. Evans,R.K., Haley,B.E. and Roth,D.A. (1985) J.Biol.Chem. 260, 7800-7804.
53. Jaspars,E.M.J., Gill,D.S. and Symons,R.H. (1985) Virology 144, 410-425.

54. Sarachu,A.N., Huisman,M.J., Van Vloten-Doting,L. and Bol,J.F. (1985) Virology 141, 14-22.
55. Strauss,E.G. and Strauss,J.H. (1986) In Schlesinger,S. and Schlesinger,M.J., ed. The Togaviridae and Flaviviridae, Plenum, NY, pp.35-90.
56. McGeoch,D.J., Dalrymple,M.A., Dolan,A., McNab,D., Perry, L.J., Taylor,P. and Challberg,M.D. (1988) J.Virol. 62, 444-453.
57. Matson,S.W. (1986) J.Biol.Chem. 261, 10169-10175.
58. Gefter,M.L. (1981) In Boyer,P.D., ed. The Enzymes, Vol.14, pp.367-373, Acad.Press, Orlando, FL.
59. Wood,E.R. and Matson,S.W. (1989) J.Biol.Chem. 264, 8297-8303.
60. Roman,L.J. and Kowalczykowski,S.C. (1989) Biochemistry 28, 2873-2880.
61. Finch,P.W., Storey,A., Brown,K., Hickson,I.D. and Emmerson, P.T. (1986) Nucleic Acids Res. 14, 8583-8594.
62. Julin,D. and Lehman,I.R. (1987) J.Biol.Chem. 262, 9044-9051.
63. Foury,F. and Van Dyck,E. (1985) EMBO J. 4, 3525-3530.
64. Lain,S., Riechmann,J.L., Martin,M.T., Lopez-Otin,C. and Garcia,J.A. (1989) In Abstracts of the 2nd Int. Symp. Positive Strand RNA Viruses, Vienna, p.25.
65. Grun,J. and Brinton,M. (1987) J.Virol. 61, 3641-3644.
66. Olivo,P.D., Nelson,N.J. and Challberg,M.D. (1988) Proc.Nat.Acad.Sci.USA 85, 5414-5418.
67. Shuman,S., Spencer,E., Furneaux,H. and Hurwitz,J. (1980) J.Biol.Chem. 255, 5396-5403.
68. Broyles,S.S. and Moss,B. (1987) Molec.Cell.Biol. 7, 7-14.
69. Seraphin,B., Simon,M. and Boulet,A. (1989) Nature 337, 84-87.
70. Linder,P., Lasko,P.F., Ashburner,M., Leroy,P., Nielsen,P.J., Nishi,K., Schnier,J. and Slonimski,P.P. (1989) Nature 337, 121-122.
71. Nakayama,H., Nakayama,K., Nakayama,R., Irino,N., Nakayama,Y. and Hanawalt,P.C. (1984) Mol.Gen.Genet. 195, 474-480.
72. Oh,E.Y. and Grossman,L. (1989) J.Biol.Chem. 264, 1336-1343.
73. Sung,P., Prakash,L., Matson,W. and Prakash,S. (1987) Proc.Nat.Acad.Sci.USA 84, 8951-8955.
74. Kuhn,R.J. and Wimmer,E. (1987) In Rowlands,D.J., Mahy,B.W.J. and Mayo,M., ed. The Molecular Biology of Positive Strand RNA Viruses, Acad.Press, London, pp.17-51.
75. Bienz,K., Egger,D. and Pasamontes,L. (1987) Virology 160, 220-226.
76. Goldbach,R.W. and Van Kammen,A. (1985) In Davies,J.W., ed. Moleculer Plant Virology, Vol.2, Boca Raton, pp.83-120.
77. Cotmore,S. and Tattersall,P. (1987) Adv.Virus.Res. 33, 91-174.
78. Tullis,G.E., Labienic-Pintel,L., Clemens,K.E. and Pintel,D. (1988) J.Virol. 62, 2736-2744.
79. Elmer,J.S., Brand,L., Sunter,G., Gardiner,W.E., Bisaro,D.M. and Rogers,S.G. (1988) Nucleic Acids Res. 16, 7043-7060.
80. Lazarowitz,S., Pinder,A.J., Damsteegt,V.D. and Rogers,S.G. (1989) EMBO J. 8, 1023-1032.

81. Krevolin,M. and Kalendar,R. (1985) J.Mol.Biol. 182, 509-517.
82. Hershko,A. and Ciechanover,A. (1982) Ann.Rev.Biochem. 51, 335-364.
83. Count,D. and Oppenheim,A.B. (1983) In Hendrix,R.W., ed. Lambda II, Acad.Press, NY, pp.251-277.
84. Luder,A. and Mosig,G. (1982) Proc.Nat.Acad.Sci.USA 79, 1101-1105.
85. McCorquodale,D.J. and Warner,R.H. (1988) In Calendar,R., ed. Viruses, Vol.1. Bacteriophages, Plenum, NY, pp.439-476.
86. Oh,E.Y., Claassen,L., Thiagalingam,S., Mazur,S., and Grossman,L. (1989) Nucleic Acids Res. 17, 4145-4159.
87. Shiota,S. and Nakayama,H. (1989) Molec.Gen.Genet. 217, 332-340.
88. Pickslev, S.M., Attfield,P.V. and Lloyd,R.G. (1984) Mol.Gen.Genet. 195, 267-274.
89. Rothman,R.H. and Clark,A.J. (1977) Mol.Gen.Genet. 155, 279-286.
90. Modrich,P. (1987) Ann.Rev.Biochem. 56, 435-466.
90a. Alani,E., Subbiah,S. and Kleckner,N. (1989) Genetics 122, 47-57.
91. Wickner,S. (1984) J.Biol.Chem. 259, 14038-14043.
92. Nakai,H. and Richardson,C.C. (1988) J.Biol.Chem. 263, 9818-9830.
93. Venkatesan,M., Silver,L.L. and Nossal,N. (1982) J.Biol.Chem. 257, 12426-12434.
94. Lebowitz,J.A. and McMacken,R. (1986) J.Biol.Chem. 261, 4738-4748.
95. Chace,K.V. and Hall,D.H. (1973) J.Virol. 12, 343-348.
96. Kit,S. (1985) Microbiol.Sci. 2, 369-375.
97. Roseman,N.A. and Hruby,D.E. (1987) J.Virol. 61, 1398-1406.
98. Sclafani,R.A. and Fangman,W. (1984) Proc.Nat.Acad.Sci.USA 81, 5821-5825.
99. Model,P. and Russel,M. (1988) In Calendar,R.,ed. Viruses, Vol.2. Bacteriophages, Plenum, NY, pp.379-456.
100. Mindich,L., Nemhauser,I., Gottlieb,P., Romantschuk,M., Carton,J., Frucht,S., Strassman,J., Bamford,D.H. and Kalkkinen,N. (1988) J.Virol. 62, 1180-1185.
101. Cleveland,D.R., Zarbl,H. and Millward,S. (1986) J.Virol. 60, 307-311.
102. Teplow,D.B., Nakayama,C., Leung,P.C. and Harshey,R.M. (1988) J.Biol.Chem. 263, 10851-10857.
103. Mace,D.C. and Alberts,B. (1984) J. Mol.Biol. 177, 279-312.
104. Runnels,J.M., Soltis,D., Hey,T. and Snyder,L. (1982) J.Mol.Biol. 154, 273-286.
105. Roeder,G.S. and Sadowski,P.D. (1977) Virology 76, 263-285.
106. Persson,H., Mathisen,B., Philipson,L. and Pettersson,U. (1979) Virology 93, 198-208.

SOURCE REFERENCES for table 2, in condensed form (first author, year,journal, first page).

1.Fiers,1978,Nature,273,113.
2.Seif,1979,Cell,18,963.

3.Soeda,1980,Nature,283,445.
4.Rott,1988,Virology,165,74.
5.Danos,1982,EMBO J.,1,231.
6.Chen,1982,Nature,299,529.
7.Flensburg,1987,J.Mol.Biol.,195,439.
8.Garvey,1985,Gene,40,311; Paces,1986,Gene,44,107.
9.Gingeras,1982,J.Biol.Chem.,257,13475.
10.Miller,1984,Nucleic Acids Res.,12,8627.
11.Dunn,1983,J.Mol.Biol.,166,477.
12.Sanger,1982,J.Mol.Biol.,162,729.
13.Backhaus,1984,Gene,32,289.
14,15.Kaliman,1988,Nucleic Acids Res.,16,10353.
16.Baer,1984,Nature,310,207.
17.Davison,1986,J.Gen.Virol.,67,1759.
18.McGeoch,1988,J.Gen.Virol.,69,1531.
19.Davison,1986,J.Gen.Virol.,67,1759.
20.McGeoch,1988,J.Gen.Virol.,69,1531.
21.Baer,1984,Nature,310,207.
22.Davison,1986,J.Gen.Virol.,67,1759.
23.McGeoch,1988,J.Gen.Virol.,69,1531.
24.Otsuka,1984,Virology,135,316.
25.Robertson,1988,Nucleic Acids Res.,16,11303.
26.Valerie,1986,Nucleic Acids Res.,14,8637.
27.Nakanishi,1984,Genbank 44.0,Z4BPT4.
28.Gram,1985,EMBO J.,4,257.
29.Fujisawa,1985,Nucleic Acids Res.,13,7473.
30.Valerie,1986,Nucleic Acids Res.,14,8637.
31.Midgley,1985,EMBO J.,4,2695.
32.Roseman,1987,J.Virol.,61,1398.
33.Weir,1986,J.Virol.,60,920.
34.Boyle,1987,Virology,156,355.
35.Upton,1986,J.Virol.,60,920.
36.Mullineaux,1984,EMBO J.,3,3063.
37.MacDowell,1985,EMBO J.,4,2173.
38.Andersen,1988,Virology,164,443.
39.Stanley,1986,EMBO J.,5,1761.
40.Stanley,1983,Nature,301,20.
41.Hamilton,1984,EMBO J.,3,2197.
42.Srivastava,1983,J.Virol,45,555.
43.Shade,1986,J.Virol.,58,921.
44.Astell,1983,Nucleic Acids Res.,11,999.
45.Rhode,1983,J.Virol.,45,173.
46.Chen,1988,J.Virol.,60,1085.
47.Bloom,1988,J.Virol.,62,2903.
48.Golyov,1989,Dokl.Akad.Nauk SSSR,307,996.
49.Beck,1978,Nucleic Acids Res.,5,4495.
50.Peeters,1985,J.Mol.Biol.,181,27.
51.Luiten,1985,J.Virol.,56,268.
52.Mindich,1988,J.Virol.,62,1180.
53.Seliger,1987,J.Biol.Chem.,262,16289.
54,55.Forster,1988,Nucleic Acids Res.,16,293.
56,57.Krayev,1988,Dokl.Akad.Nauk SSSR,300,711.
58,59.Zuidema,1989,J.Gen.Virol.,70,267.
60.Morch,1988,Nucleic Acids Res.,16,6157.
61.Goelet,1982,Proc.Nat.Acad.Sci.USA,79,5818.
62.Hamilton,1987,J.Gen.Virol.,68,2563.
63.Ahlquist,1984,J.Mol.Biol.,172,369.

64.Rezaian,1985,Eur.J.Biochem.,150,331.
65.Cornelissen,1983,Nucleic Acids Res.,11,1252.
66.Gustafson,1989,Virology,in press.
67.Gustafson,1986,Nucleic Acids Res.,14,3895.
68.Bouzoubaa,1987,J.Gen.Virol.,68,615.
69.Bouzoubaa,1986,J.Gen.Virol.,67,1689.
70.Strauss,1984,Virology,133,92.
71.Takkinen,1986,Nucleic Acids Res.,14,5667.
72.Faragher,1988,Virology,163,509.
73.Boursnell,1987,J.Gen.Virol.,68,57.
74.Racaniello,1981,Proc.Nat.Acad.Sci.,78,4887.
75.Lindberg,1987,Virology,156,50.
76.Earle,1988,J.Gen.Virol,69,253.
77.Skern,1985,Nucleic Acids Res.,13,2111.
78.Duechler,1987,Proc.Nat.Acad.Sci.USA,84,2605.
79.Stanway,1984,Nucleic Acids Res.,12,7859.
80.Palmenberg,1984,Nucleic Acids Res.,12,2969.
81.Pevear,1987,J.Virol.,61,1507.
82.Carroll,1984,Nucleic Acids Res.,12,2461.
83.Najarian,1985,Proc.Nat.Acad.Sci.USA,82,2627.
84.Lomonosoff,1983,EMBO J.,2,2253.
85.Greif,1988,J.Gen.Virol.,69,1517.
86.Allison,1986,Virology,154,9.
87.Domier,1986,Nucleic Acids Res.,14,5417.
88.Maiss,1989,J.Gen.Virol.,70,513-524.
89.Dunn,1983,J.Mol.Biol.,166,477.
90.Spicer,1984,J.Biol.Chem.,259,15425.
91.Rodriguez,1986,Proc.Nat.Acad.Sci.USA,83,9566.
92.Niles,1986,Virology,153,96.
93.Bando,1987,J.Virol.,61,553.
94.Rupasov,1989,J.Gen.Virol.70,1861.
95.Rice,1985,Science,229,726.
96.Castle,1986,Virology,149,10.
97.Hahn,1988,Virology,162,167.
98.Coia,1988,J.Gen.Virol.,69,23.
99.Sumiyoshi,1987,Virology,161,497.
100.Pletnev,1989,Bioorg.Khim.,in press.
101.Collett,1988,Virology,165,191.