

Two early genes of bacteriophage T5 encode proteins containing an NTP-binding sequence motif and probably involved in DNA replication, recombination and repair

V.M. Blinov, E.V. Koonin*, A.E. Gorbalenya, A.V. Kaliman⁺ and V.M. Kryukov⁺

*Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, Moscow Region, *Institute of Microbiology, USSR Academy of Sciences, Moscow and ⁺Institute of Biochemistry and Physiology of Microorganisms, Pustchino, Moscow Region, USSR*

Received 29 May 1989

It is demonstrated, by computer-assisted analysis, that T5 bacteriophage early genes D10 and D13 encode proteins containing the purine NTP-binding sequence motif. The D10 gene product is shown to be a member of a recently characterized superfamily of (putative) DNA and RNA helicases. The D13 gene product is related, at a statistically significant level, to the gene 46 product of bacteriophage T4 which is a component of an exonuclease involved in phage DNA replication, recombination and repair. A lower but also significant degree of sequence similarity was detected between the gene D12 product of T5 and the gene 47 product of T4, the second component of the same nuclease. It is hypothesized that both D10 and D13 gene products of T5 might be NTPases, possibly DNA-dependent, mediating NTP-consuming steps during phage DNA replication, recombination and/or repair.

NTP-binding sequence motif; Helicase; Exonuclease; DNA replication; DNA recombination; DNA repair; (Bacteriophage T5, Bacteriophage T4)

1. INTRODUCTION

T5 is a large coliphage with a typical complex-shaped virion and an approx. 121 kb dsDNA genome with direct terminal repeats [1,2]. Unlike T7 and T4 phages, whose replication machineries have been extensively studied both structurally and functionally, very limited information is available on the molecular genetics of T5, especially at the

level of gene sequences. Several T5 gene products have been implicated in DNA replication but only DNA polymerase [3], DNA-binding protein [4], 5'-exonuclease [5] and dihydrofolate reductase [6] functions have been assigned to specific genes. Recently, two of us reported the sequence of approx. 10 kb of T5 DNA encompassing several early genes [7]. Here, we present results of computer analysis of the sequences of the proteins encoded by two of these genes, D10 and D13. Both proteins contain a purine NTP-binding motif and might possess NTPase activity. A helicase activity is proposed for D10 protein whose sequence was related to those of a superfamily of (putative) helicases. The sequence of D13 is highly similar to that of gp46 of phage T4. A degree of sequence similarity was observed also between D12 protein of T5 and gp47 of T4. Together, these observations suggested that, like gp46 and gp47, T5 proteins D13 and D12 might be subunits of an exonuclease involved in

Correspondence address: A.E. Gorbalenya, Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, Post Office Institute of Poliomyelitis, 142782 Moscow Region, USSR

Abbreviations: gp, gene product; TEV, tobacco etch virus; WNV, West Nile encephalitis virus; BVDV, bovine viral diarrhoea virus; K2, yeast mitochondrial plasmid pGK12; VV, vaccinia virus; VZV, varicella-zoster virus; S.c., *Saccharomyces cerevisiae*; M.l., *Micrococcus luteus*

I

Man eIF-4A : 68 gydvlaQAqs GtSKTatfAI sllqq 7 tQalVLapT rela-qQIqk vvMal
 E.coli RECQ : 41 grdCLVvapt GgGKSLcyQI palll 1 GltvVVspl iSlMkdQVdq lQang
 M.l. UVRB : 40 EkdvVLsBat GtSKSat--t Awlve 3 rptlVMVqn kTla-aQLan efreI
 S.c. RAD3 : 34 ggnslEmps GtSKTVsl-L sLtia 8 rkliycsrI aSeI-ekalv elenI

TEV CI : 76 ardfLVrBav GsSKStg--L p---- 7 GrvMLepT rplT-dNMhk qlrae
 MNV NS3 : 186 kQitVLDlhP GsSKTrki-L pqiik 7 lrtavLapT rvva-aEMse Alrgl
 BVDV p125 : ? gDfkqItlat GsSKTte--L pkAvi 7 krviVLlpl raaa-esVYq yarlK

K2 P4 : 53 ysSlIVcydv GtSKTyA-a cLAhm 5 fkvlyLsnS lnsI-dNfsn eyeKv
 VV NTPaseI : 47 mhSlLLfhet GvGKTMT--t vyilK 7 nwaillLvK kali-eDpWm ntIlR
 VV NTPaseII : 37 NrSVLLfhia GsSKTIIA-L lfAlv 4 kkvyILVpn iniLkifnYn agVaa
 VZV gp51 : 59 rpvtVVRApM GsSKTtal-L ewlqh 5 isvlVVsCr rSft-qtLiQ rfnda
 CONS ++ B SKT + + ++L P ++ + +

S

T5 D10 : 100 DDTClINGkP GfSKTILA-L ALAYk 1 GQKtLVICT nTSl-remWa AEVRK

II

Man eIF-4A : 47 lspKyI kafVLDEade mLSRgF 12 sn tqvVILSATM psdvle-Vtk Kf 45
 E.coli recQ : 45 lahwnp vILaVDEaHc isqugh 16 pt lpfMALTAta ddttrqDIVr lI 40
 M.l. uvrB : 245 Dyfpdd fllVVDEsHv tlpqig 36 ri gqtVyLSATp gayElgQadg yv 37
 S.c. RAD3 : 134 Nevskd siVIfDEaHn idnvci 206 rf ssvlitSGTI splDmyprMI nf 59

TEV CI : 36 aevKtY dfVIIDEChv ndASai 12 gk --vLkVSATp pgREve-fft qf 33
 MNV NS3 : 34 hrvpny nlfIMDEaHf tdpaSi 12 ge aaalfMTATp pgtedp-fpe sN 29
 BVDV p125 : 39 aamveY syIfLDEyHc atpeql 12 ir --vVAMTATp agsvtt-tgq Kh 34

K2 P4 : 29 sdnvdy GIIILDEVHn lreSaY 12 nn skilVITATp midskdEL-d si 82
 VV NTPaseI : 29 inSKsr icVIIDEChn fIsKSl 23 kn hkmIcLSATp ivnsqvEf-t ml 120
 VV NTPaseII : 34 lsrynn siFIVDEaHn ifSntt 10 nk ipfLILSGSp iitntptl-g hi 147
 VZV gp51 : 33 EaideY dvLILDEVms vlgqly 19 rc sqilAMdATV nsqfid-LIs gl 68
 CONS + +++DE+H ++++TAT +

S

T5 D10 : 33 NISKvF GtVIVDEVHh cVATTf 7 ca rykIGLSGTL krKDglQVMF KD 16

IV

Man eIF-4A : t qavIfInTrR kvDwLtekMh Ardftvsamh gdMd 6 iarefRsBss rVlittdLla
 E.coli recQ : k sgilycnBra kvEdtaaaLQ skgisAaayh agLe 6 vqEkfqrddI qIvvAtvaFg
 M.l. uvrB : e rvlVtLTkR maEdLtdyLl eagvKveyih sdVd 6 llrelRkGtf dVlvGinLlr
 S.c. RAD3 : d GmvVfppSyl yaEsivSMwQ tMgildevwk hKLi 14 tyrkAcnngR gaillevarg

TEV CI : d nilVyVaSyn dvDslgkLLv qkgyKvskid grtm 6 iiteBtsvkk hfivAtnIie
 MNV NS3 : g ktvwfVpSVK mgNeIalcLQ ragkKvIqln rksy 6 -ypkcKnddw dfvyttidIse
 BVDV p125 : g nmlVfVpTrn maveVakkLk Akgyn---s gyyy 6 NlrVvtsqsp yVivAtnaie

K2 P4 : s kinafInSiK egELtvlfsf yVkr-GIdft ssVl 38 sianiKGdni hIllGsSVIS
 VV NTPaseI : t lyndfknSlR drEfskSaLD tfkr-Gellg gdas 76 QesntnGeci ktcvfsSsgg
 VV NTPaseII : s kfKyfInriQ TlNgkhfIyf snStyGglvi kyIm 45 spEnddGsqI MflfssnImS
 VZV gp51 : h nicIfsSTIs fsELVaafca ifTDsiLiln strp 0 lcnVnewkhf rVlvytTVvT
 CONS + f S + + + +++ T +

T

T5 D10 : m GhKVlIVSdR T-ELIqTILE ALTQRGVttY fIlg 11 E-DIAKG6pc VLaaAqSIFS

				VI			
Man σ IF-4A	:	rGIdVQQVSI VInydL	0	pt NrenyihriG RggRfgrkg-	--VainMVte	edK	26 [28]
E.coli recQ	:	mGInkpNVrf VVhfdI	0	pr NiesyyQetG RagRdqlpa-	--eamLfyDp	adm	264 [29]
M.l. uvrB	:	EGIdMpDLpe VslvaI	8	Lr stTsLiQtIG RaaRn-vsg-	--evhMyagn	Vtd	142 [30]
S.c. RAD3	:	EGIdfQygrrt VLMigI	30	fD aarhaaQcIG RVIRgkDdy-	--gvaVLaDr	rfs	92 [31]
TEV CI	:	NGVTI-DIdv VVdfgt	18	Vv sygeriQkIG RVgRhk----	--egvaLrig	qtn	264 [32]
MNV NS3	:	mGanf-kaSr VIdark	20	ai taasaaQrrG RIgRnpsqv-	--gDeycygg	htn	140 [33]
BVDV p125	:	sGVTLPdLdt VIdtgL	22	av tvgeqaQrrG RVgRVk----	--pgryyrsq	eta	? [34]
K2 P4	:	ESITLyrVkh LhliSp	0	fw NygqIkQsiG RaiRigshe-	-gLEdksMkv	yLh	184 [35]
VV NTPaseI	:	EGISffsInd IfildM	0	tw NEasLrQivG RaiRLnshvl	tppeErryVNV	hfi	133 [36]
VV NTPaseII	:	ESyTLkEVrh IwfATI	0	pD tfSgyNQIIG RsiRkfaya-	--disepVNV	yLi	165 [37]
VZV gp51	:	vGLSf-DMah fhsmfa	7	gp DmvsVyQsIG RVrILLine-	-vLmyVdgar	trc	450 [38]
CONS		G+ + + V+				Q G R+ R	
T5 D10	:	EGISLNELSc LIMgSL	0	IN NESIIEQLaG RVqRIVEgk-	--LDPIVVDL	IMK	43 [7]

Fig.1. Alignment of the amino acid sequence of T5 bacteriophage D10 gene product with conserved segments of the helicase superfamily. Only selected sequences from the superfamily including more than 20 members (A.E.G. et al., submitted) are shown. Source references for the sequences are given in parentheses. Conserved segments are numbered above the alignment. Segments I and II correspond to the A and B sites of the NTP-binding motif, respectively. Numbers of amino acid residues in terminal regions and in spacers separating the conserved regions are indicated. Residues identical or similar to the respective residues of D10 are shown in capitals. Residues belonging to one of the following groups were considered similar: A,G; S,T; L,I,V,M; F,Y,W; D,E,N,Q; R,K. CONS denotes the consensus pattern derived for 20 proteins of the superfamily. (+) Hydrophobic residues. Residues substituted in mutants of uvrB and RAD3 with impaired function in excision repair and/or helicase activity [30,39,40] are indicated in italics. Arrows denote two insertions of two and three amino acid residues in RAD3.

phage DNA replication, recombination and/or repair.

2. METHODS

2.1. Protein sequences

Sequences of T5 proteins were translated from the previously published DNA sequence [7]. Other protein sequences were from the references cited in the figures.

2.2. Protein sequence comparison

Initial searching of protein sequences for the NTP-binding motif was by visual inspection, or by the pattern-searching program SITE. Segmental multiple sequence alignment was performed by manually fitting a new sequence into a previously generated alignment, and the statistical significance of the result was assessed by the program SCORE [8]. This program calculated the difference between the score obtained upon comparison of a query sequence with an alignment and the mean and maximal scores obtained upon 300 simulations of such a comparison with randomly scrambled versions of the query sequence. This difference was expressed in standard deviation (SD) units. Scores were computed using the MDM78 amino acid residue comparison matrix [9]. Preliminary pairwise sequence comparison was by the program DOTHELIX generating the full map of local similarity between two sequences (A.E.G. et al., in preparation). Pairwise alignments were generated by the program OPTAL which is an implementation of the Sankoff algorithm of sequence alignment [10] allowing optimal align-

ment of amino acid sequences and its statistical evaluation in SD units [11,12].

3. RESULTS AND DISCUSSION

3.1. Putative NTPases of T5

Inspection of the open reading frames available in the sequenced portion of T5 genome revealed, in putative proteins D10 and D13, the so-called 'A' site of the NTP-binding sequence motif G/Axx(G)xGKS/T typical of numerous ATP- and GTP-utilizing enzymes [12-15]. We compared the sequences of D10 and D13 to those of other NTP-binding motif-containing proteins. It was shown that the D10 sequence contained the 7 sequence motifs (the 'A' and 'B' sites of the NTP-binding motif included) conserved in the members of a recently characterized superfamily of (putative) DNA and RNA helicases (fig.1; A.E.G. et al., submitted). The functional importance of 6 of the conserved segments was confirmed by the results of mutational analysis of RAD3 and uvrB proteins (cf. fig.1). Quantitative evaluation of the D10 sequence alignment with the conserved segments of 20 proteins of this superfamily revealed that the alignment score exceeded the mean random score

by 11.3 SD and the maximum of 300 randomizations by 8.1 SD. This demonstrated definite evidence for the relatedness of the D10 protein to the helicase superfamily.

For D13 protein, highly significant similarity (approx. 15.2 SD above the mean) was detected to the gene 46 product of T4 bacteriophage through

the entire lengths of both proteins. Importantly, the most prominent conservation was observed in the vicinity of the A and B sites of the NTP-binding motif (fig.2). An unusual feature of both proteins was the very long distance separating these sites, the B sites being adjacent to the C-termini (cf. figs 1,2; cf. [12-15]).

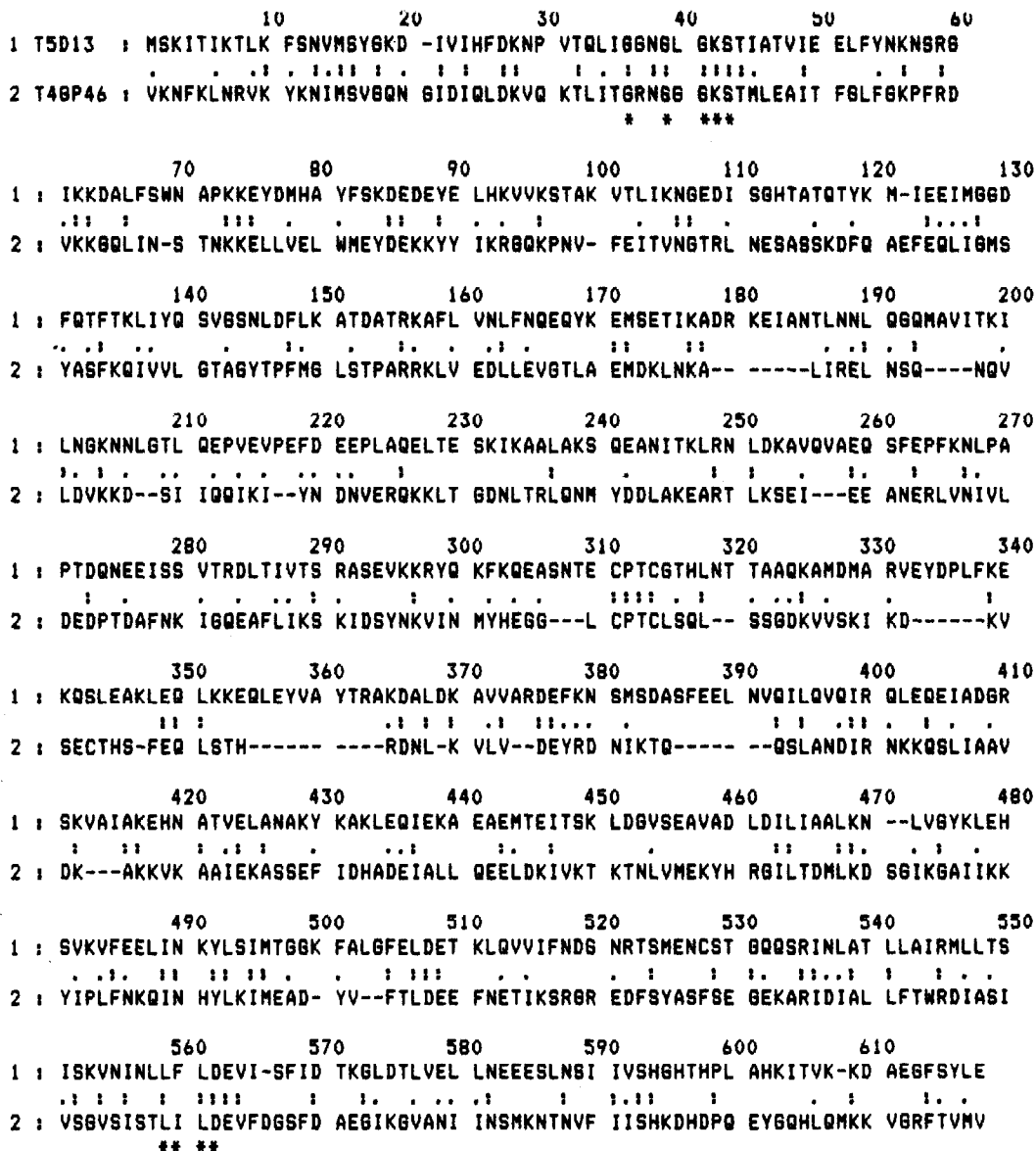


Fig.2. Alignment of amino acid sequences of D13 protein of T5 and gp46 of T4. The alignment was generated by the program OPTAL written in FORTRAN 77 and run on an IBM PC AT. Colons denote identical residues, and dots similar residues (defined as in fig.1). Asterisks indicate conserved residues of the A and B sites of the NTP-binding motif. The gp46 sequence was from [41].

In T4, gp46, as a complex with gp47, the product of the neighboring gene, constitutes an exonuclease involved in phage DNA recombination, replication and repair [16–18]. Comparison of the amino acid sequences of gp47 and the D12 gene product of T5 revealed similarity at the level of approx. 5 SD (not shown). No comparable similarity could be revealed between the sequences of other proteins encoded in the respective genome regions of the two phages. Nevertheless, as the actual percentage identity between D12 and gp47 was low (<15%), and no data pertaining to possible functional sites in these proteins are available, this relationship could not be established with certainty.

3.2. Implications for phage replication

Proteins containing the NTP-binding motif are encoded by genomes of many viruses belonging to highly diverse groups, including bacteriophages T7 and T4, parvo-, papova-, herpes- and poxviruses as well as a number of groups of RNA viruses ([12,19–23] and A.E.G. et al., in preparation). Most of these proteins are involved in DNA or RNA replication and/or transcription; one of their main functions appears to be that of a DNA(RNA) helicase ([24,25] and A.E.G. et al., in preparation). A helicase function is also plausible for the D10 gene product of T5, as demonstrated by the observation that this protein belongs to a superfamily of (putative) helicases. The gene D13 product has been implicated in phage DNA replication [2]. It is tempting to speculate that products of T5 genes D13 and D12, like their probable T4 counterparts gp46 and gp47, may form a complex with an exonuclease activity. This assignment is in agreement with the results of very recent experiments demonstrating that a plasmid expressing D12 and D13 complemented mutants in genes 46 and 47 when introduced into bacteria infected with mutant T4 (A.V.K. and V.M.K., unpublished). In the (putative) phage exonucleases gp46 and D13 protein are probably NTPase subunits, whereas gp47 and D12 protein might confer the nuclease activity. In this respect, the phage enzyme complexes seem analogous to multifunctional nucleases/helicases involved in *E. coli* DNA recombination and repair such as uvrABC and recBCD [26,27].

It seems a common feature of large DNA viruses to encode two or more proteins containing the purine NTP-binding motif predominantly involved

in genome replication. This is the case for T4, poxviruses, at least some of the herpesviruses (A.E.G. et al., in preparation), and T5 (this paper). It would be no surprise if further exploration of the T5 genome revealed additional proteins of this class.

Acknowledgements: The authors are grateful to Drs L.I. Brodsky, A.P. Donchenko, K.M. Chumakov and A.M. Leontovich for providing computer programs.

REFERENCES

- [1] McCorquodale, D.J. (1975) *CRC Crit. Rev. Biochem.* 4, 101–159.
- [2] McCorquodale, D.J. and Warner, H.R. (1988) in: *Viruses*, vol.1, Bacteriophages (Calendar, R. ed.) pp.439–476, Plenum, New York.
- [3] Fujimura, R.K., Tavtigian, S.V., Choy, T.L. and Roop, B.C. (1985) *J. Virol.* 53, 495–500.
- [4] McCorquodale, D.J., Gossling, J., Benzinger, R., Chesney, R., Lawhorne, L. and Moyer, R. (1979) *J. Virol.* 29, 322–327.
- [5] Kaliman, A.V., Krutilina, A.I., Kryukov, V.M. and Bayev, A.A. (1986) *FEBS Lett.* 195, 61–64.
- [6] Hendrickson, H.E. and McCorquodale, D.J. (1972) *J. Virol.* 9, 981–989.
- [7] Kaliman, A.A., Kryukov, V.M. and Bayev, A.A. (1988) *Nucleic Acids Res.* 16, 10353–10354.
- [8] Koonin, E.V., Gorbalenya, A.E. and Chumakov, K.M. (1989) *FEBS Lett.* 252, 42–46.
- [9] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) *Methods Enzymol.* 91, 524–545.
- [10] Sankoff, D. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1–4.
- [11] Pozdnyakov, V.I. and Pankov, Yu.A. (1981) *Int. J. Peptide Protein Res.* 17, 284–291.
- [12] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1989) *J. Mol. Evol.* 28, 256–268.
- [13] Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.* 1, 945–951.
- [14] Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., Buckel, S.D., Bell, A.W. and Hermodson, M.A. (1986) *Nature* 323, 448–450.
- [15] Dever, T.E., Glynnias, M.J. and Merrick, W.C. (1987) *Proc. Natl. Acad. Sci. USA* 84, 1814–1818.
- [16] Cunningham, R.P. and Berger, H. (1977) *Virology* 80, 67–82.
- [17] Luder, A. and Mosig, G. (1982) *Proc. Natl. Acad. Sci. USA* 79, 1101–1105.
- [18] Wakem, L.P. and Ebisuzaki, K. (1981) *Virology* 112, 472–479.
- [19] Seif, I. (1984) *Virology* 138, 347–352.
- [20] Gorbalenya, A.E., Blinov, V.M. and Koonin, E.V. (1985) *Mol. Genet. (Moscow)* 11, 30–36.
- [21] Doolittle, R.F. (1986) in: *Protein Engineering* (Inouye, M. ed.) pp.15–27, Plenum, New York.

- [22] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1987) *Mol. Biol. (Moscow)* 21, 1566–1572.
- [23] Astell, C.R., Mol, C.D. and Anderson, W.F. (1987) *J. Gen. Virol.* 68, 885–893.
- [24] Hodgman, T.C. (1988) *Nature* 333, 22–23.
- [25] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1988) *FEBS Lett.* 235, 16–24.
- [26] Oh, E.Y. and Grossman, L. (1989) *J. Biol. Chem.* 264, 1336–1343.
- [27] Finch, P.W., Storey, A., Brown, K., Hickson, I.D. and Emmerson, P.T. (1986) *Nucleic Acids Res.* 14, 8583–8594.
- [28] Nielsen, P.J., McMaster, G.K. and Trachsel, H. (1985) *Nucleic Acids Res.* 13, 6867–6880.
- [29] Irino, N., Nakayama, K. and Nakayama, H. (1986) *Mol. Gen. Genet.* 205, 298–304.
- [30] Shiota, S. and Nakayama, H. (1988) *Mol. Gen. Genet.* 213, 21–29.
- [31] Reynolds, P., Higgins, D.R., Prakash, L. and Prakash, S. (1985) *Nucleic Acids Res.* 13, 2357–2372.
- [32] Allison, R., Johnston, R.E. and Dougherty, W.G. (1986) *Virology* 154, 9–20.
- [33] Castle, E., Nowak, T., Leidner, U., Wengler, G. and Wengler, G. (1986) *Virology* 145, 227–236.
- [34] Collett, M.S., Larson, R., Gold, C., Strick, D., Anderson, D.K. and Purchio, A.F. (1988) *Virology* 165, 191–199.
- [35] Tommasino, M., Ricci, S. and Galeotti, C.L. (1988) *Nucleic Acids Res.* 16, 5863–5878.
- [36] Rodriguez, J.F., Kahn, J.S. and Esteban, M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9566–9570.
- [37] Niles, E.G., Condit, R.C., Caro, P., Davidson, D., Matusick, L. and Seto, J. (1986) *Virology* 153, 1728–1742.
- [38] Davidson, A.J. and Scott, J.E. (1986) *J. Gen. Virol.* 67, 1759–1816.
- [39] Friedberg, E.C. (1988) *Microbiol. Rev.* 52, 70–102.
- [40] Sung, P., Higgins, D., Prakash, L. and Prakash, S. (1988) *EMBO J.* 7, 3263–3269.
- [41] Gram, H. and Ruger, W. (1985) *EMBO J.* 4, 257–264.

NOTE ADDED IN PROOF

Since the acceptance of this paper, the manuscript on the helicase superfamily mentioned on page 49 as 'A.E.G. et al., submitted' has been accepted for publication. It will appear as Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. in volume 17 of *Nucleic Acids Research*.