Volume 235, number 1,2, 16–24 FEB 06091 August 1988

Discussion Letter

A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination

Alexander E. Gorbalenya, Eugene V. Koonin, Alexei P. Donchenko and Vladimir M. Blinov

Institute of Poliomyelitis and Viral Encephalitides of the USSR Academy of Medical Sciences, 142782 Moscow Region, USSR

Received 23 March 1988; revised version received 23 May 1988

A statistically significant similarity was demonstrated between the amino acid sequences of 4 *Escherichia coli* helicases and helicase subunits, a family of non-structural proteins of eukaryotic positive-strand RNA viruses and 2 herpesvirus proteins all of which contain an NTP-binding sequence motif. Based on sequence analysis and secondary structure predictions, a generalized structural model for the ATP-binding core is proposed. It is suggested that all these proteins constitute a superfamily of helicases (or helicase subunits) involved in NTP-dependent duplex unwinding during DNA and RNA replication and recombination.

ATP binding; Eukaryotic virus; Helicase; Amino acid sequence; Sequence comparison; Nucleoside triphosphate-binding motif; Protein evolution

1. INTRODUCTION

Enzymes catalysing the hydrolysis of the β,γ-phosphodiester bond of NTP (primarily ATP or GTP) play a central role in the energy coupling and control of essentially all biochemical processes. Specifically, DNA(RNA)-dependent ATPases are involved in DNA replication, transcription, recombination and repair. For many of these enzymes, an ATP-dependent dsDNA (or DNA-RNA hybrid) unwinding (helicase) activity has been observed (reviews [1–4]). In bacteria (and probably even more so in eukaryotes) helicases are surprisingly numerous, the rationale for such a multitude being far from clear.

In the course of the present-day rapid accumulation of nucleic acid and deduced protein sequences, those of several helicases have been determined. They all contain the so-called NTP motif, a constellation of conserved sequence elements characteristic of the catalytic sites of a vast class of NTP-utilizing enzymes [5–8]. These elements are: (i) a flexible loop involved in the binding of the pyrophosphate moiety of NTP, with the consensus sequence G/AXXXXGKS/T (the

A preliminary account of a part of this work has been reported at the Conference on 'Molecular Biology of Eukaryotic Genes' (Institute of Molecular Biology of the USSR Academy of Sciences, Moscow, October 1987)

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

Abbreviations: CMV, cucumber mosaic virus (tricornavirus); BMV, brome mosaic virus (tricornavirus); AlMV, alfalfa mosaic virus (tricornavirus); TMV, tobacco mosaic virus (tobamovirus); TRV, tobacco rattle virus (tobravirus); BSMV, barley stripe mosaic virus (hordeivirus); SNBV, Sindbis virus (alphavirus); SFV, Semliki forest virus (alphavirus); BNYVV, beet necrotic yellow vein virus (furovirus); PVX, potato virus X (polexvirus); WClMV, white clover mosaic virus (potexvirus); IBV, avian infectious bronchitis virus (coronavirus); VZV, varicella zoster virus (herpesvirus); EBV, Epstein-Barr virus (herpesvirus)

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

so-called 'A' site), and (ii) a β-strand ending with an invariant Asp residue interacting with the Mg^{2+} coordinated with the same phosphates, with the consensus (Z)(Z)(Z)ZZD where Z is a hydrophobic residue (the so-called 'B' site). Apart from this consensus pattern, however, no significant similarities have been detected between different helicase sequences (strictly speaking, it was the A consensus only which was identified in most helicases; the B site usually was not recognized), with the exception of *E. coli* rep and uvrD proteins [9].

Here we delineate a distinct family of homologous proteins consisting of *E. coli* helicases (or helicase subunits) rep, uvrD, recB and recD. Unexpectedly, we discovered that the sequences of the conserved domains of these proteins are similar at a statistically significant level to those of a pair of herpesvirus proteins of unknown function and to a family of proteins involved in eukaryotic positive strand RNA viral replication. We suggest that the proteins of the latter two groups are involved in duplex unwinding during replication and recombination of viral DNA and RNA, respectively.

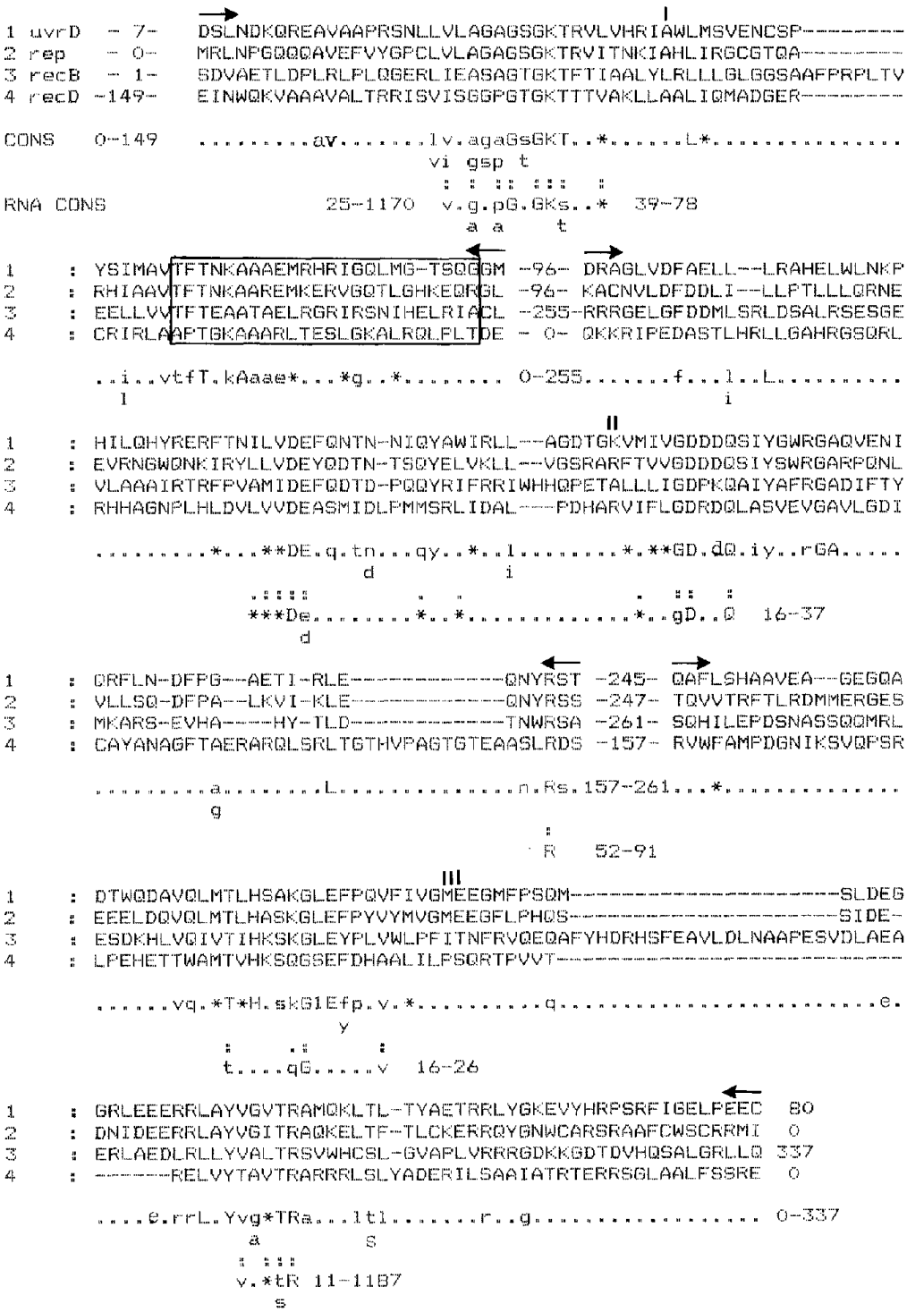
2. rep, uvrD, recB AND recD CONSTITUTE A DISTINCT PROTEIN FAMILY

Proteins rep and uvrD have about 40% identical residues [9]. The latter are mainly distributed between 3 long regions designated I–III in fig.1. The A and B sites of the NTP motif are located in regions I and II, respectively. We sought to determine whether other helicases had counterparts similar to the conserved regions of rep/uvrD. This was done by program MULDI (MULtiple DIagon) which is a version of standard DIAGON [10] adapted to compare multiple pre-aligned sequences. The program is similar to that recently described by Argos [11].

Regions of significant similarity to the 3 conserved regions of rep/uvrD were detected in 2 subunits of recBCD helicase, recB and recD. The conserved segments of the 4 proteins were aligned by the program OPTAL (OPTimal ALignment) which performs optimal alignment of multiple amino acid sequences and its statistical evaluation by a Monte Carlo procedure [12]. The resulting alignment is shown in fig.1. For regions I and III, the alignment was highly significant, with the alignment score exceeding the mean score for randomized sequences by more than 5 standard deviations (SD). In region II, the similarity of recB to the sequences of rep and uvrD was also significant, but that of recD to the 3 other sequences was less pronounced (about 3 SD). Still recD contained more than 1/2 of the residues conserved in this region in rep, uvrD and recB. Comparison of the conserved blocks of the 4 protein alignments shown in fig.1 with amino acid sequences of other helicases such as *E. coli* proteins recA, recF, uvrA, uvrB, dnaB, traY and rho, bacteriophage T_4 uvsY protein, bacteriophage T_7 gene 4 product, and polyomavirus T antigens did not reveal significant similarity. Thus, the 4 proteins characterized above constitute a distinct subset of bacterial helicases with a high degree of sequence conservation.

3. THE NEWLY ESTABLISHED FAMILY OF BACTERIAL HELICASES IS RELATED TO 2 FAMILIES OF PROTEINS PROBABLY INVOLVED IN EUKARYOTIC RNA AND DNA VIRUS REPLICATION

Our bank of NTP motif containing proteins includes more than 100 species. Many of these are grouped into distinct families. We compared the consensus derived for the family of 4 *E. coli* helicases with those of other families. Unexpectedly, a striking resemblance has been noticed to the pattern of 1 of the 3 families of NTP motif containing proteins involved in the genome replication of eukaryotic positive strand RNA viruses [13–15]. This family includes proteins of α- and coronaviruses infecting animals and those of several plant virus groups. In fact, the RNA viral consensus has been found to constitute a subset of the helicase one, the former protein family being more variable than the latter (fig.1). The conserved residues of the virus protein family are, like those of the helicases, distributed between 3 regions separated by divergent spacers of varying lengths. The sequences of the conserved domains of RNA viral proteins were aligned by OPTAL with those of recD as the differences in spacer lengths were minimal in this case as compared to the 3 other *E. coli* proteins. Two separate alignments were generated, one between the N-terminal conserved



subdomain of RNA viral proteins [14,15] and the 2 N-terminal conserved regions of recD, and the other between the C-terminal subdomain and the 3rd conserved region. Both alignments were statistically significant, at levels of approx. 8 and 7 SD, respectively. Interestingly, the similarity between the sequences of the helicases and those of RNA viral proteins discussed here is generally higher than that between the 3 families of NTP motif containing proteins of positive strand RNA viruses (the 2nd and 3rd families include proteins of picorna-/como- and potyviruses, respectively [15]). The joint consensus pattern for the 2 families includes 20 conserved residues. Further search of the sequences of NTP motif containing proteins for this pattern picked up 2 very similar herpesvirus proteins, BBLF4 protein of EBV and gene product 55 of VZV. Significant similarity (>5 SD) has been revealed upon alignment of region I of bacterial helicases and of regions II and III of RNA viral proteins with respective portions of the herpesvirus proteins.

The resultant alignment of selected fragments of proteins of 3 families is shown in fig.2. It includes 6 conserved segments of varying lengths (3–21 residues) totalling 90 aligned residues. The 1st (N-terminal) segment was extracted from the conserved region I of the helicase alignment, the 2nd–4th segments were from region II, and the 5th and 6th from region III. The conserved stretches are separated by spacers, whose lengths vary to a much greater extent, the most variable being those between the 1st and 2nd and between the 4th and 5th segments, constituting the junctions between the 3 large conserved regions (see above). As a result, the total lengths of the compared sequences differ by more than 500 residues, comprising from about 200 residues in p26 of potexviruses to approx. 760 residues in recB and the VZV protein. The 20 residues constituting the consensus pattern of the helicases and RNA viral proteins are also highly conserved in the final alignment; 8 are invariant. In addition, 17 positions in the alignment are occupied predominantly by hydrophobic residues.

4. STRUCTURAL PREDICTIONS

The striking sequence similarity between *E. coli* helicases and viral proteins suggests some degree of similarity at higher structural levels. Secondary structure predictions performed by program ALBEAT, based on the Finkelstein and Ptitsyn algorithm [16,17], indicate that all the proteins (domains) discussed here belong to the mixed $\alpha\beta$ structural type [18,19]. This is compatible with the alternating $\beta\alpha\beta\alpha$ structure ('Rossmann fold') known to be characteristic of NTP-binding domains [20,21]. A tentative structural model of *E. coli* helicases and related proteins was generated (fig.3). It includes a core formed by 6 α/β units, the β-strands constituting a pleated sheet(s). Four of the α/β units encompassing the 4 N-terminal conserved sequence segments (fig.2) have the classical β-turn–α configuration, while the 2 C-terminal units probably constitute less usual α-turn–β-folds. Conserved amino acid residues are mainly located within, or in close proximity to, β-turns. We suggest that these residues are juxtaposed, constituting the catalytic center. The conserved residues of the 2 N-terminal segments (fig.2) constituting the NTP motif proper are supposed to interact directly with NTP. Specific functions of the other conserved residues juxtaposed in our model remain to be elucidated. To maintain such juxtaposition, opposite orientations should be postulated for the 4 N-terminal and 2 C-terminal β-strands (fig.3). The putative core is surrounded by 4 additional domains, of which 2 are inserts and 2 are N- and C-terminal extensions. Only some of

Fig.1. Alignment of conserved regions of 4 *E. coli* helicase subunits. The putative DNA-binding domain is highlighted. The derived consensus pattern (CONS) is shown below the alignment. A residue was included in the consensus if it occupied the given position in 3/4 or more of the sequences, with the invariant residues capitalized. Where a position can be occupied by 2 functionally related residues, both are indicated. Asterisks denote hydrophobic residues. Below the helicase consensus, the consensus for the RNA viral proteins (RNA CONS) is shown. This was derived under the same rules, but only stretches centering at invariant residues are shown, while some partially conserved residues observed between the latter have been omitted. Colons denote coincidences between the 2 consensus patterns, and dots those positions where the consensus residue of RNA viral proteins is observed in 1 or 2 of the helicase sequences. Lengths of spacers between conserved regions and of terminal extensions are shown by numbers. Sequences from: [38], uvrD; [9], rep; [31], recB; [33], recD.

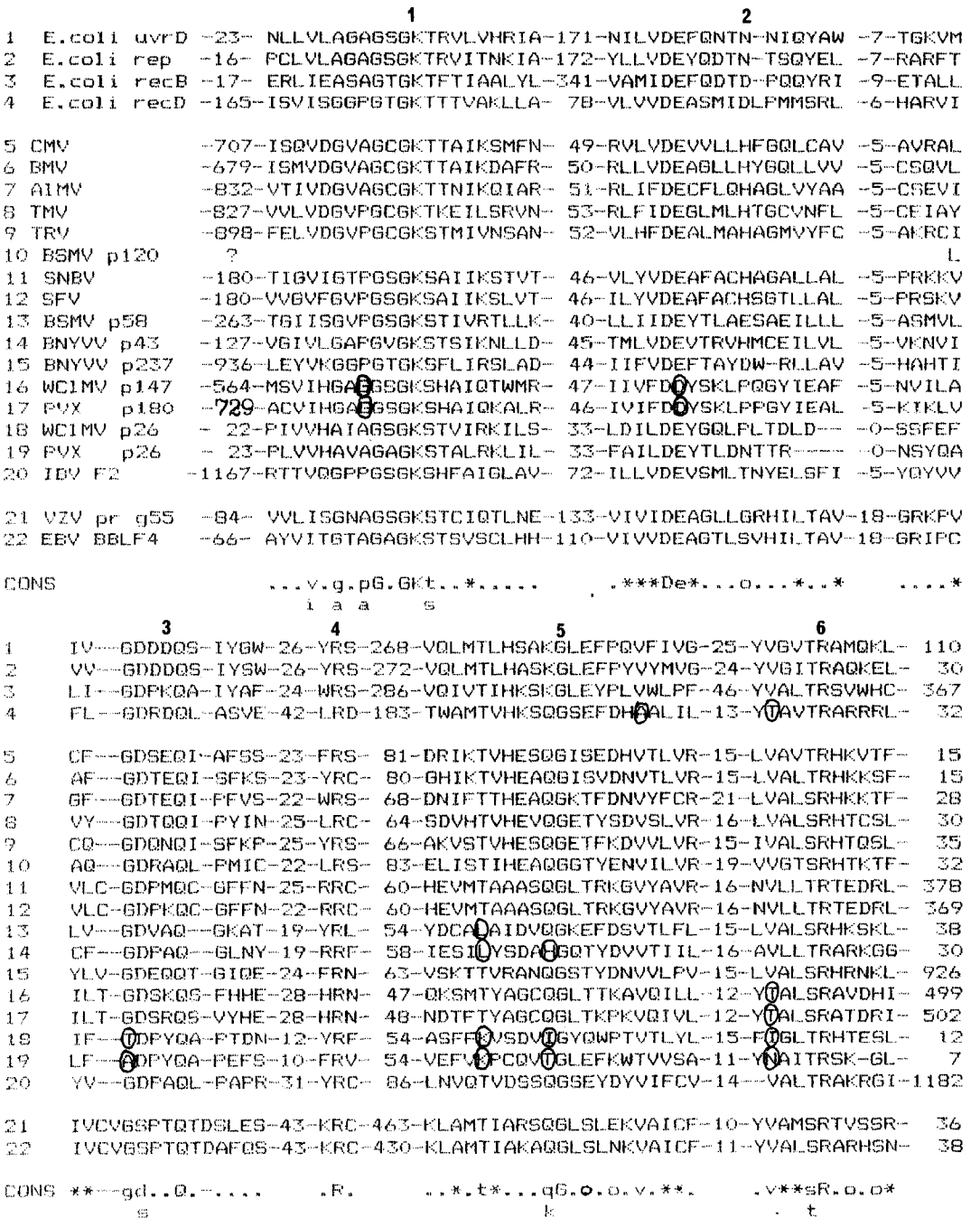


Fig.2. Alignment of highly conserved segments of bacterial helicases, RNA virus proteins and herpesvirus proteins. The aligned stretches are numbered 1–6 from the N- to C-termini of the proteins. Under the alignment the consensus pattern (CONS) is shown. The rules for consensus derivation and designations are as in fig.1 except that in those positions where different consensus residues are observed in different protein families, all are indicated and 'o' designates a hydrophilic residue. Encircled are amino acid residues deviating from the consensus in proteins presumably constituting heterodimers. Sequences from: [39], CMV; [40], BMV; [41], AlMV; [42], TMV; [43], TRV; [44], SNBV; [45], SFV; [46], p58 of BSMV; [47] and personal communication of Yu.V. Kozlov, p120 of BSMV (partial sequence); [48], p43 of BNYVV; [49], p237 of BNYVV; [50], WClMV; [51], PVX; [52], IBV; [53], VZV; the EBV sequence was from GenBank.

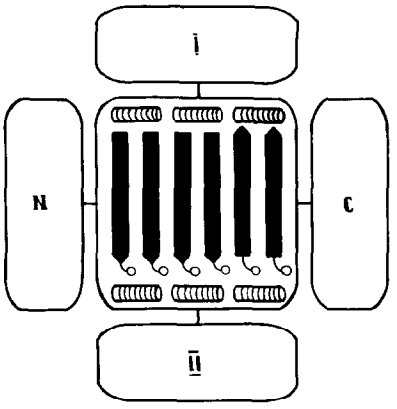


Fig.3. A schematic model depicting a possible spatial organization of the proteins of the helicase superfamily. The proposed core domain and 4 dispensable domains are shown of which one (I) is located between the 1st and 2nd conserved segments, another (II) between the 4th and 5th segments (see text and fig.2), and the remaining 2 are N- and C-terminal extensions. β-Strands are indicated by arrowheaded rectangles; α-helices by cylinders and β-turns by small circles. To obtain a consensus secondary structure prediction, α, β and turn potentials were averaged for each position of the 6 aligned conserved segments (fig.2) and surrounding stretches (the latter were chosen so as to obtain a stretch of at least 30 residues for each segment) of each of the 3 protein families, and then the average of these 3 values was calculated. The 2 β-strands shown to the right are those corresponding to the 5th and 6th conserved segments. Otherwise the β-strands and the α-helices are not specified and the connections between them are not shown as the available data are insufficient to determine their precise localizations (cf. [54]). Generally, an approximately equal number of helices is supposed to lie below and above the β-sheet.

these domains are present in each individual protein; the smallest ones, p26 of WClMV and PVX, have only short N- and C-terminal extensions (see fig.2).

5. FUNCTIONAL IMPLICATIONS

The high degree of structural similarity between the viral proteins and *E. coli* helicases suggests that the former, like the latter, should be true helicases, or at least helicase subunits. For the herpesvirus proteins, such a proposal seems natural, since the replication of the dsDNA of these viruses requires a helicase(s). Compatible with this proposal, proteins described here are among the most conserved between EBV and VZV [22], probably being vital for herpesvirus reproduction. As for RNA viruses, the need for a helicase seems less obvious as their genomes are ssRNA. However, it has been demonstrated in several systems that replication complexes isolated from cells infected with positive strand RNA viruses are capable of in vitro synthesis of only double-stranded replicative forms, and not of genomic ssRNA [23–25]. In one case, that of TMV [25], this has been shown to correlate with the absence in such preparations of p126 which possesses NTP-binding properties [26] and, according to our hypothesis, may be a helicase. Thus one can hypothesize that the function of RNA viral proteins (domains) described here is NTP (probably ATP)-dependent unwinding of double-stranded template molecules in viral RNA replication (fig.4). It seems likely that they may also be involved in RNA recombination which readily occurs in plant viruses and in coronaviruses [27,28]. The necessity for an energy-dependent unwinding function has already been postulated for another group of positive strand RNA viruses (the picornaviruses), based on some in vitro experiments with replication complexes [29,30].

In an attempt to relate the proposed helicase function with the structural features outlined above, it is possible to suggest that the core domain may be responsible for NTP binding and hydrolysis coupled to duplex unwinding. The additional variable domains may be involved in DNA (RNA) recognition and in interaction with other components of the replication machinery. A potential DNA-binding domain of the classical helix-turn-helix type has indeed been identified in the domain of recB separating the 1st and 2nd conserved segments [31]; we were able to demonstrate that this domain is conserved in similar locations in the other 3 *E. coli* proteins (fig.1) and in the herpesvirus proteins (not shown), i.e. in all the DNA-binding proteins included in our set.

Two of the *E. coli* proteins discussed here, recB and recD, are subunits of a single helicase, recBCD (exonuclease V). Of these proteins, only recB has been shown to possess an intrinsic helicase activity [32,33]. recD, on the other hand, has been shown to enhance greatly the helicase activity of recBC [34]. Thus, within the holoenzyme, the functions of the subunits are probably specialized, recB being the helicase proper, and recD performing some ATP-dependent accessory function. A rather similar situation exists in 4 RNA viruses,

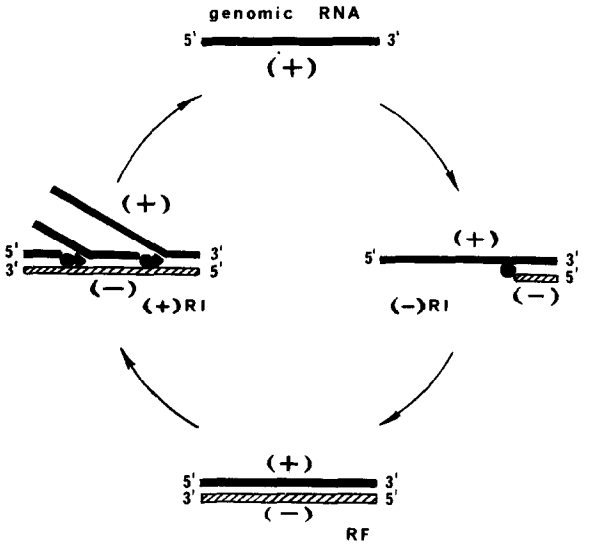


Fig.4. A scheme demonstrating the possible involvement of an RNA helicase in viral RNA replication. Two stages of replication are shown: I, synthesis of negative strands resulting in formation of double-stranded molecules; II, synthesis of daughter positive strands using the negative strand of RF as the template. Circles designate viral RNA-dependent RNA polymerase, and triangles the proposed helicase. The scheme is purposely oversimplified in that real replication complexes probably contain other components of viral and cellular origin; they are, however, poorly characterized.

BNYVV, 2 potexviruses and, most probably, BSMV, which possess 2 putative NTP-binding proteins each. It may be proposed that, analogous to recBCD, the putative helicases of RNA viruses function as oligomers containing, in most cases, identical subunits, but in the 4 viruses mentioned above heterologous ones. Interestingly, the putative NTP-binding proteins of these viruses as well as recD contain substitutions of certain otherwise conserved amino acid residues (fig.2), in line with the idea of functional diversification. Also compatible with this idea is the smallest, among all the proteins of the discussed set, size of one member of each pair of these proteins, especially p26 of potexviruses.

6. FOUR *E. coli* HELICASES AND TWO FAMILIES OF VIRAL PROTEINS MAY CONSTITUTE A MONOPHYLETIC SUPERFAMILY: EVOLUTIONARY IMPLICATIONS

It is very likely that the *E. coli* proteins rep, uvrD, recB and recD which display highly significant sequence similarity and perform similar functions constitute a monophyletic family. Significantly, the gene for recD, the protein most distantly related to others, is contained within one operon with that for recB [33], making gene duplication a realistic possibility for their origin. The same appears certain for 2 closely related herpesvirus proteins and for RNA viral proteins as argued elsewhere [15]. The possibility of the 3 families constituting a single monophyletic superfamily is not as easily acceptable. Nevertheless, in our opinion, the arguments for this are rather compelling. Although statistically significant similarity could be established only for some pairs of conserved segments of the 3 protein families, they do form a contiguous 'network' in which each sequence block of a family is related to the respective block of at least one of the 2 other families at a meaningful level. Obviously, the fortuitous simultaneous appearance of all the 6 conserved blocks in the same order in the proteins of the 3 families is most unlikely, though it is not easy to estimate the exact probability due to highly variable spacer lengths. The complete consensus pattern of 20 conserved residues could not be found in any protein outside the delineated set as shown by screening of the translated version of Genbank (Rel. 38.0). These observations demonstrate that the group of proteins described here constitutes a distinct cluster among other NTP motif containing proteins.

The relationships within the postulated helicase superfamily are non-trivial (fig.5). Strikingly, the 3 protein families overlap, i.e. numerous cases are observed when a sequence of one of the families is more closely similar to certain sequences of one or both of the other 2 families than to some members of its own family. Most surprising is the high degree of similarity between recD and certain RNA viral proteins such as those of BSMV, CMV, and especially BNYVV (see also [35]).

Finally, we should like to note that the range of organisms possessing the proteins of the proposed superfamily is rather peculiar, bringing together eubacteria, large eukaryotic DNA viruses, and a subset of positive strand RNA viruses also infecting eukaryotes. It remains uncertain as to whether additional members can be identified in eukaryotes and archebacteria, but analysis of protein se-

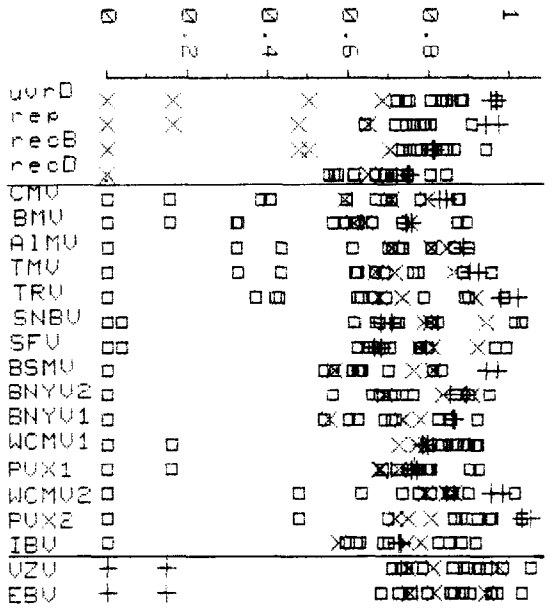


Fig.5. A graphic representation of the relationships between protein sequences within the proposed helicase superfamily. Ordinate: relative distances between the protein segments aligned in fig.2. The distances were calculated by program COMP using the following formula: $D = -\ln(S^{obs} - S^{r}/S^{m} - S^{r})$ [55] where S^{obs} is the alignment score for a given pair of sequences calculated by use of the MDM78 amino acid residue comparison matrix [10], S^{r} the mean score for the same pair of sequences upon random jumbling, and S^{m} the maximal possible score for the same pair (i.e. the average of the scores obtained upon alignment of each sequence with itself). Crosses designate sequences of *E. coli* proteins, squares those of RNA viral proteins, and plus signs those of herpesvirus proteins. BSMV = p58 of BSMV, BNYV2 = p43 of BNYVV, BNYVV1 = p237 of BNYVV, WCMV1 = p147 of WClMV, PVX1 = p180 of PVX, WCMV2 = p26 of WClMV, PVX2 = p26 of PVX; other designations are as throughout the text. The figure was generated by use of the GIM (Graphic Interactive Management) system designed by A.L. Drachev and run on a WicatS150 computer.

quences of negative strand RNA viruses, retroviruses and retrovirus-like genetic elements and small DNA viruses did not reveal any (unpublished). Nevertheless, one may suggest that the principal construction of the core domain of the (putative) helicases is a very ancient, if not a primordial, one.

ADDENDUM

During the final stage of the preparation of this manuscript we learned that T.C. Hodgman had independently reached very similar conclusions [36] and additionally included in his treatise a yeast helicase, whose sequence has been demonstrated very recently to be related to that of uvrD [37]. Interestingly, this helicase is less closely related to RNA viral proteins than recD.

Acknowledgements: The authors are grateful to Professor V.I. Agol for constant interest and encouragement, Dr A.V. Finkelstein for secondary structure predictions, Dr K.M. Chumakov for providing the program COMP, and A.L. Drachev for providing the GIM system. Thanks are also due to Drs D. Zimmern, Yu.V. Kozlov and S.Yu. Morozov for communicating results prior to publication.

REFERENCES

[1] Abdel-Monem, M. and Hoffmann-Berling, H. (1980) Trends Biochem. Sci. 5, 128–130.
[2] Muskavitch, K.M.T. and Linn, S. (1981) in: The Enzymes (Boyer, P.D. ed.) vol.14, pp.233–250, Academic Press, New York.
[3] Kornberg, A. (1980) DNA Replication, Freeman, San Francisco.
[4] Kornberg, A. (1982) Supplement to DNA Replication, Freeman, San Francisco.
[5] Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 945–951.
[6] Halliday, K.R. (1984) J. Cyclic Nucleotide Protein Phosphorylation Res. 9, 435–448.
[7] Möller, W. and Amons, R. (1985) FEBS Lett. 186, 1–7.
[8] Doolittle, R.F. (1986) in: Protein Engineering (Inouye, M. ed.) pp.15–27, Plenum, New York.
[9] Gilchrist, C. and Denhardt, D. (1987) Nucleic Acids Res. 15, 465–475.
[10] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[11] Argos, P. (1987) J. Mol. Biol. 193, 385–396.
[12] Pozdnyakov, V.I. and Pankov, Yu.A. (1981) Int. J. Peptide Protein Res. 17, 284–291.
[13] Gorbalenya, A.E., Blinov, V.M. and Koonin, E.V. (1985) Mol. Genet. (Moscow) 11, 30–36.
[14] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1987) Mol. Biol. (Moscow) 21, 1566–1572.
[15] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1988) J. Mol. Evol., in press.
[16] Finkelstein, A.V. (1975) Dokl. Akad. Nauk SSSR 223, 744–747.
[17] Ptitsyn, O.B. and Finkelstein, A.V. (1983) Biopolymers 22, 15–25.
[18] Richardson, J.S. (1977) Nature 268, 495–500.
[19] Richardson, J.S. (1981) Adv. Protein Chem. 34, 167–339.
[20] Rossmann, M.G., Moras, D. and Olsen, K.W. (1974) Nature 250, 194–199.
[21] Rossmann, M.G., Liljas, A., Branden, C.-I. and Banaszak, L.J. (1975) in: The Enzymes (Boyer, P.D. ed.) vol.11, pp.61–102, Academic Press, New York.
[22] Davison, A.J. and Taylor, P. (1987) J. Gen. Virol. 68, 1067–1079.

[23] Bove, J.M., Bove, C. and Mocquot, B. (1968) Biochem. Biophys. Res. Commun. 32, 480–486.
[24] Zaitlin, M., Duda, C.T. and Petti, M.A. (1973) Virology 53, 300–311.
[25] Jaspars, E.M.J., Gill, D.S. and Symona, R.H. (1985) Virology 144, 410–425.
[26] Evans, R.K., Haley, B.E. and Roth, D.A. (1985) J. Biol. Chem. 260, 7800–7804.
[27] Bujarski, J.J. and Kaesberg, P. (1986) Nature 321, 528–531.
[28] Lai, M.M.C., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L. and Stohlman, S.A. (1985) J. Virol. 56, 449–456.
[29] Dmitrieva, T.M., Eremeeva, T.P. and Agol, V.I. (1980) FEBS Lett. 115, 19–22.
[30] Agol, V.I., Chumakov, K.M., Dmitrieva, T.M. and Svitkin, Yu.V. (1980) in: Soviet Sci. Rev. Sect. D: Biol. Rev. (Skulachev, V.P. ed.) vol.1, pp.319–370.
[31] Finch, P.W., Storey, A., Chapman, K.E., Brown, K., Hickson, I.D. and Emmerson, P.T. (1986) Nucleic Acids Res. 14, 8573–8582.
[32] Lieberman, R.P. and Oishi, M. (1974) Proc. Natl. Acad. Sci. USA 71, 4816–4820.
[33] Finch, P.W., Storey, A., Brown, K., Hickson, I.D. and Emmerson, P.T. (1986) Nucleic Acids Res. 14, 8583–8594.
[34] Hickson, I.D., Robson, C.R., Atkinson, K.E., Hutton, L. and Emmerson, P.T. (1985) J. Biol. Chem. 260, 1224–1229.
[35] Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1988) Nature, in press.
[36] Hodgman, T.C. (1988) Nature, in press.
[37] Foury, F. and Lahaye, A. (1987) EMBO J. 6, 945–951.
[38] Finch, P.W. and Emmerson, P.T. (1984) Nucleic Acids Res. 12, 5789–5799.
[39] Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1985) Eur. J. Biochem. 150, 331–339.
[40] Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984) J. Mol. Biol. 172, 369–383.
[41] Cornelissen, B.J.C., Brederode, F.T., Moormann, R.J. and Bol, J.F. (1983) Nucleic Acids Res. 11, 1252–1265.
[42] Goelet, P., Lomonosoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5818–5822.
[43] Hamilton, W.D.O., Boccara, M., Robinson, D.J. and Baulcombe, D.C. (1987) J. Gen. Virol. 68, 2563–2575.
[44] Strauss, E.G., Rice, C.M. and Strauss, J.H. (1984) Virology 133, 92–110.
[45] Takkinen, K. (1986) Nucleic Acids Res. 14, 5667–5682.
[46] Gustafson, G. and Armour, S. (1986) Nucleic Acids Res. 14, 3895–3909.
[47] Rupasov, V.V., Afanasiev, B.N., Adyshev, D.M. and Kozlov, Yu.V. (1986) Dokl. Akad. Nauk SSSR 288, 1237–1241.
[48] Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K. and Jonard, G. (1986) J. Gen. Virol. 67, 1689–1700.
[49] Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G. and Richards, K. (1987) J. Gen. Virol. 8, 615–626.
[50] Forster, R.L.S., Bevan, M.W., Harbison, S.-A. and Gardner, R.C. (1988) Nucleic Acids Res. 16, 293–303.
[51] Krayev, A.S., Morozov, S.Yu., Lukasheva, L.I., Rosanov, M.N., Chernov, B.K., Simonova, M.L., Golova, Yu.B., Belzhelarskaya, S.N., Pozmogova, G.E., Skryabin, K.G. and Atabekov, J.G. (1988) Dokl. Akad. Nauk SSSR, in press.
[52] Boursnell, M.E.G., Brown, T.D.K., Foulds, I.J., Green, P.F., Tomley, F. and Binns, M.M. (1987) J. Gen. Virol. 68, 57–77.
[53] Davison, A.J. and Scott, J.E. (1986) J. Gen. Virol. 67, 1759–1816.
[54] Bradley, M.K., Smith, T.F., Lathrop, R.H. and Livingston, D.M. (1987) Proc. Natl. Acad. Sci. USA 84, 4026–4030.
[55] Feng, D.F., Johnson, M.S. and Doolittle, R.F. (1985) J Mol. Evol. 21, 112–125.