Volume 17 Number 12 1989 Nucleic Acids Research

Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes

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

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

Received May 3, 1989; Accepted May 18, 1989

ABSTRACT

In the course of systematic analysis of protein sequences containing the purine NTP-binding motif, a new superfamily was delineated which included 25 established or putative helicases of Escherichia coli, yeast, insects, mammals, pox- and herpesviruses, a yeast mitochondrial plasmid and three groups of positive strand RNA viruses. These proteins contained 7 distinct highly conserved segments two of which corresponded to the "A" and "B" sites of the NTP-binding motif. Typical of the new superfamily was an abridged consensus for the "A" site, GxGKS/T, instead of the classical G/AxxxxGKS/T. Secondary structure predictions indicated that each of the conserved segments might constitute a separate structural unit centering at a β-turn. All previously characterized mutations impairing the function of the yeast helicase RAD3 in DNA repair mapped to one of the conserved segments. A degree of similarity was revealed between the consensus pattern of conserved amino acid residues derived for the new superfamily and that of another recently described protein superfamily including a different set of prokaryotic, eukaryotic and viral (putative) helicases.

INTRODUCTION

Molecular machineries utilized by cells and viruses for genome replication, recombination and repair, transcription and mRNA translation are replete with DNA- and RNA-dependent NTPases, at least some of which possess helicase activity, i.e. ability to promote DNA, RNA or DNA-RNA duplex unwinding (1-3). Most of these NTPases contain a common sequence motif consisting of two separate units (x, any residue; hy, hydrophobic residue): G/AxxxxGKS/T ("A site") and (3hy, 2x) D ("B" site). This motif is conserved not only in (putative) helicases but in a vast class of purine NTP-ulilizing enzymes (4-6). Where X-ray data have been reported, it was shown that each site of the NTP-binding motif comprised a distinct structural unit of the β-strand-β-turn-α-helix type (the "B" site sometimes lacking the α-helix) directly involved in NTP binding (7-10).

Recently, we, and independently Hodgman, delineated, by sequence comparison, a superfamily of (putative) DNA and RNA helicases including E.coli proteins uvrD, rep, recB and recD,

yeast helicase PIF, and proteins involved in herpesvirus DNA and positive strand RNA virus RNA replication (11-13). Subsequently, several groups described a set of rather closely related (putative) RNA helicases (14-19) which was christened 'D-E-A-D' family, after the sequence conserved in the 'B' site of the NTP-binding motif (19). It was suggested that this family might constitute a subdivision within the above superfamily (17-20).

Here, we show that the 'D-E-A-D' family is in fact a subset of another distinct superfamily of (putative) helicases which, just like the first one, includes proteins of E.coli, eukaryotes, and DNA and RNA viruses. A distant but significant relationship could be established between the two superfamilies.

METHODS

Amino acid sequences

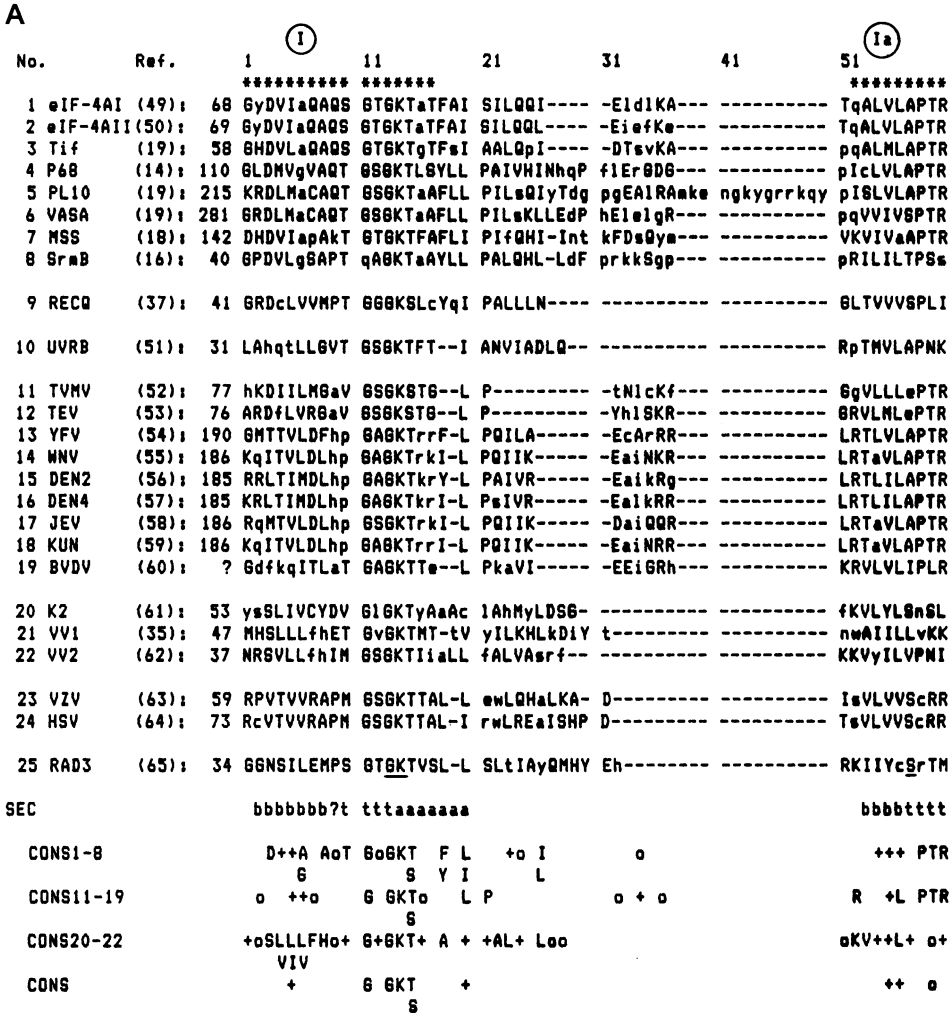
Amino acid sequences compared were those of CI proteins of potyviruses: tobacco vein mottling virus (TVMV) and tobacco etch virus (TEV); NS3 proteins of flaviviruses: yellow fever virus (YFV), West Nile virus (WNV), Dengue virus types 2 (DEN2) and 4 (DEN4), Japanese encephalitis virus (JEV), Kunjin virus (KUN); polyprotein of bovine viral diarrhea virus (BVDV, a pestivirus); NTPases I (ORF 11 of the HindIII D fragment of genomic DNA) and II (ORF 6 of the same fragment) of vaccinia virus (VV, a poxvirus); ORF 4 of Kluyveromyces lactis mitochondrial plasmid pGKL2 (K2); herpesvirus proteins: gene 51 product (gp51) of varicella-zoster virus (VZV) and UL9 protein of herpes simplex virus type 1 (HSV); murine proteins p68 and PL10; human translation initiation factor 4A (eIF-4AI,II); Drosophila melanogaster protein VASA; Escherichia coli proteins SrmB, recQ and uvrB; Saccharomyces cerevisiae proteins Tif1/Tif2 (translation initiation factor; Tif), MSS116 and RAD3. Sequences were from current literature; references are indicated in Fig.1A.

Sequence comparisons

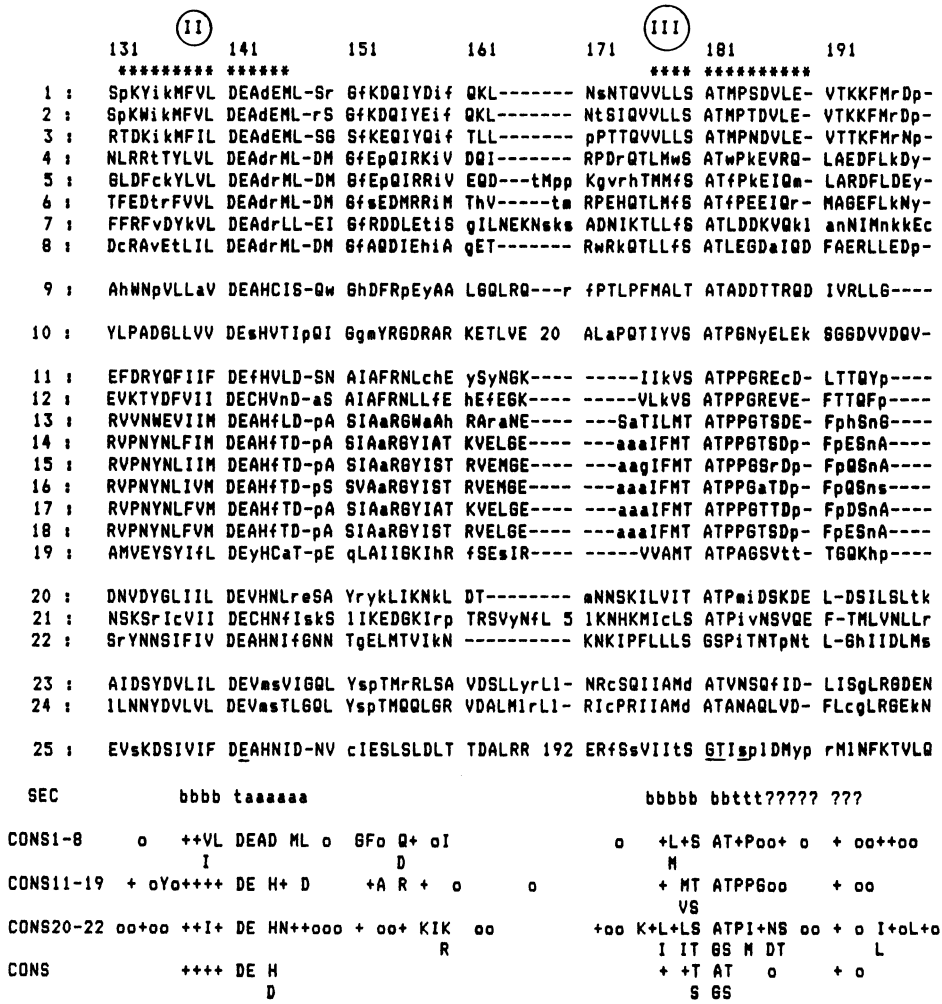
Amino acid sequences were compared by programs DIAGON (21) and OPTAL (22, 23), using the amino acid residue comparison matrix MDM78 (24). The program OPTAL, based on the Sankoff algorithm (25), performs optimal alignment of multiple amino acid sequences and its statistical evaluation by a Monte Carlo procedure. The significance of the obtained alignment is assessed by calculation of alignment score (AS): $AS=S^o-S^r/\sigma$ where S^o is the score obtained upon alignment of real sequences, S^r is the mean score for 25 random permutations of the same sequences, and σ is the standard deviation (SD). AS is expressed as the number of SD above the mean. The final alignment of 25 protein sequences was generated by combining several pairwise and group alignments, using also results of DIAGON comparisons and visual inspection. To assess the statistical significance of the alignment thus obtained, approximate probability of the observed similarity between two protein sequences being fortuitous was calculated as

$$P \approx l_1 * l_2 * \prod_{i=1}^{i=n} p_i .$$

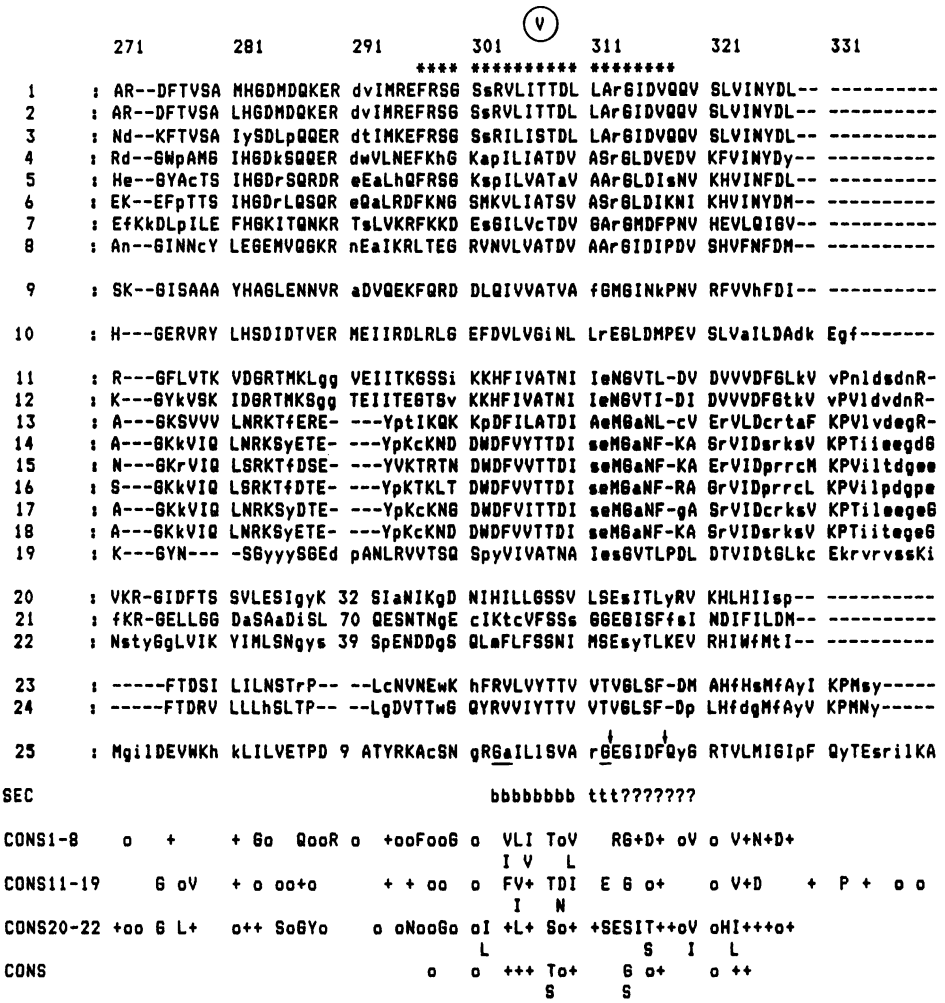
A


No. Ref. 1 11 21 31 41 51
(I) (Ia)
1 eIF-4AI (49): 68 GyDVIaQAQS GTGKTaTFAI SILQQI---- -EldlKA--- ---------- TqALVLAPTR
2 eIF-4AII(50): 69 GyDVIaQAQS GTGKTaTFAI SILQQL---- -EiefKe--- ---------- TqALVLAPTR
3 Tif (19): 58 GHDVLaQAQS GTGKTgTFsI AALQpI---- -DTsvKA--- ---------- pqALMLAPTR
4 P68 (14): 110 GLDMVgVAQT GSGKTLSYLL PAIVHINhqP flErGDG--- ---------- pIcLVLAPTR
5 PL10 (19): 215 KRDLMaCAQT GSGKTaAFLL PILsQIyTdg pgEAlRAmke ngkygrrkqy pISLVLAPTR
6 VASA (19): 281 GRDLMaCAQT GSGKTaAFLL PILsKLLEdP hElelgR--- ---------- pqVVIVSPTR
7 MSS (18): 142 DHDVIapAkT GTGKTFAFLI PIfQHI-Int kFDsQya--- ---------- VKVIVaAPTR
8 SrmB (16): 40 GPDVLgSAPT qAGKTaAYLL PALQHL-LdF prkkSgp--- ---------- pRILILTPSs
9 RECQ (37): 41 GRDcLVVMPT GSGKSLcYqI PALLLN---- ---------- ---------- GLTVVVSPLI
10 UVRB (51): 31 LAhqtLLGVT GSGKTFT--I ANVIADLQ-- ---------- ---------- RpTMVLAPNK
11 TVMV (52): 77 hKDIILMGaV GSGKSTG--L P--------- -tNlcKf--- ---------- GgVLLLePTR
12 TEV (53): 76 ARDfLVRGaV GSGKSTG--L P--------- -YhlSKR--- ---------- GRVLMLePTR
13 YFV (54): 190 GMTTVLDFhp GAGKTrrF-L PQILA----- -EcArRR--- ---------- LRTLVLAPTR
14 WNV (55): 186 KqITVLDLhp GAGKTrkI-L PQIIK----- -EaiNKR--- ---------- LRTaVLAPTR
15 DEN2 (56): 185 RRLTIMDLhp GAGKTkrY-L PAIVR----- -EaikRg--- ---------- LRTLILAPTR
16 DEN4 (57): 185 KRLTIMDLhp GAGKTkrI-L PsIVR----- -EalkRR--- ---------- LRTLILAPTR
17 JEV (58): 186 RqMTVLDLhp GSGKTrkI-L PQIIK----- -DaiQQR--- ---------- LRTaVLAPTR
18 KUN (59): 186 KqITVLDLhp GAGKTrrI-L PQIIK----- -EaiNRR--- ---------- LRTaVLAPTR
19 BVDV (60): ? GdfkqITLaT GAGKTTe--L PkaVI----- -EEiGRh--- ---------- KRVLVLIPLR
20 K2 (61): 53 ysSLIVCYDV GlGKTyAaAc lAhMyLDSG- ---------- ---------- fKVLYLSnSL
21 VV1 (35): 47 MHSLLLfhET GvGKTMT-tV yILKHLkDiY t--------- ---------- nwAIILLvKK
22 VV2 (62): 37 NRSVLLfhIM GSGKTIiaLL fALVAsrf-- ---------- ---------- KKVyILVPNI
23 VZV (63): 59 RPVTVVRAPM GSGKTTAL-L ewLQHaLKA- D--------- ---------- IsVLVVScRR
24 HSV (64): 73 RcVTVVRAPM GSGKTTAL-I rwLREaISHP D--------- ---------- TsVLVVScRR
25 RAD3 (65): 34 GGNSILEMPS GTGKTVSL-L SLtIAyQMHY Eh-------- ---------- RKIIYcSrTM
SEC bbbbbbb?t tttaaaaaaa bbbbtttt
CONS1-8 D++A AoT GoGKT F L +o I o +++ PTR
G S Y I L
CONS11-19 o ++o G GKTo L P o + o R +L PTR
S
CONS20-22 +oSLLLFHo+ G+GKT+ A + +AL+ Loo oKV++L+ o+
VIV
CONS + G GKT + ++ o
S

```
          61         71         81         91         101        111        121
          ********** *
  1 :     ELAQQIQKVV MALGDYMGaS chAc----IG gtNV---rae VqklqMeAph IIVGTPGR-V FDMLNRR-YL
  2 :     ELAQQIQKVV MALGDYMGaT chAc----IG gtNV---rNe MqklqaeAph IVVGTPGR-V FDMLNRR-YL
  3 :     ELALQIQKVV MALAFhMDIK VhAc----IG gtsF---vEd aeGl-r-DaQ IVVGTPGR-V FDNIQRR-Rf
  4 :     ELAQQVQQVa AEYcrACRLK STci----YG gApk---gpq Irdler-GVE IcIATPGR-L IDFLEcG-KT
  5 :     ELAVQIYEEa RKFsYRSRVR pCVv----YG gADI---gQq IrDler-GCh LLVATPGR-L VDMMERG-hI
  6 :     ELAIQIFNEa RKFAFESYLK IGIv----YG gtsF---rhq necitr-GCh VVIATPGR-L LDFVDRT-FI
  7 :     DLALQIEAEV KKIHDMNYgL KKYAcVSLVG gtDFraaaNk Mnkl-r--pN IVIATPGR-L IDVLEKYSnK
  8 :     rAGDaGVRSc pRTGETYASg YRHh----HR rrsL---yEp rgsvqrkSgh RRSpRPDV-c VQYIKEE-nf

  9 :     SLMKDQVDQL QANGVAAAcL NSTQ----tR eQQL----Ev MtGcrTGQir LLYIaPERLM LDNFLE--hL

 10 :     TLAAQLYGEM KEFFPENAVE YFVSyYDYYQ pEAY  201     ycSGIEN ySrFLSGRgp gEpPpTL-FD

 11 :     pLAENVTKQM RGSpFFASpT LRMrNLStFG ---------- -------Ssp ITVMTTGF-a LHFFANNV-K
 12 :     pLTDNMhKQL RsEpFNCFpT LRMrgKStFG ---------- -------Ssp ITVMTSGF-a LHhFARNI-a
 13 :     vVLsEMKEaF hGLDVKFhTQ aFSAhgSg-- ---------- -------REv IDaMchAt-L tYRMLEp--T
 14 :     vVAAEMSEaL RGLpIRYQTS aVHrEhSg-- ---------- -------NEi VDVMchAt-L tHRLMSp--h
 15 :     vVAAEMEEaL RGLpIRYQTp aIrAEhTg-- ---------- -------REi VDLMchAt-f tMRLLSp--I
 16 :     vVAAEMEEaL RGLpIRYQTp aVkSEhTg-- ---------- -------REi VDLMchAt-f tTRLLSS--T
 17 :     vVAAEMAEaL RGLpVRYQTS aVQrEhQg-- ---------- -------NEi VDVMchAt-L tHRLMSp--N
 18 :     vVAAEMAEaL RGLpIRYQTS aVArEhNg-- ---------- -------NEi VDVMchAt-L tHRLMSp--h
 19 :     AAAESVYQyM RLKHPSISfN LRIGDMKE-- ---------- ----gdmAtg ITYASYGY-f CQMPQPKLRa

 20 :     NsIDNfSNEY EKVV-LDSRL NS---LKKni ---------- ---------t IKSFS-kF-- YNcekRGe-S
 21 :     ALIEDpWMNt ILRY-ApEIT KC--DIFIny ---------- ---------D DEnFRNkF-- FTNIkT---I
 22 :     NILKifNyNM GVAMNLFNdE FIAENIFIHs ---------- ---------t tSFYSlNY-- NDNVInYngL

 23 :     SfTQTLIQrF NDAGLSGFVT YLTSEtYIMG ---------- ---------f KRLIVQLE-s LHRVSS---E
 24 :     SfTQTLAtrF AESGLVDFVT YFSStnYIMN ---------- ------dRpf hRLIVQVE-s LHRVGP---N

 25 :     SeIEKALVEL ENLMDYrTKE LGYQE-DFrG 1G1t   88     rmIsLCN IIIYSYhYLL DpKIAERVsN

SEC       aaaaaaaaaa a

CONS1-8    LA Q+ o     +                +G G o+     o  o +       + + TPGR + +D++oo

CONS11-19  +  oM o   +R   +o           o                        o  Vo M  A       + o
             V                                                      I     G
CONS20-22 N+IooFo+N+  + + L ooo o   oIFIN                          o oooFo oF   +oN+Ko o +
           L     E                  L                                Y   Y
CONS       +  o+                                                                  + o
```


II
III
131 141 151 161 171 181 191
1 : SpKYikMFVL DEAdEML-Sr GfKDQIYDif QKL------- NsNTQVVLLS ATMPSDVLE- VTKKFMrDp-
2 : SpKWikMFVL DEAdEML-rS GfKDQIYEif QKL------- NtSIQVVLLS ATMPTDVLE- VTKKFMrDp-
3 : RTDKikMFIL DEAdEML-SG SfKEQIYQif TLL------- pPTTQVVLLS ATMPNDVLE- VTTKFMrNp-
4 : NLRRtTYLVL DEAdrML-DM GfEpQIRKiV DQI------- RPDrQTLMwS ATwPkEVRQ- LAEDFLkDy-
5 : GLDFckYLVL DEAdrML-DM GfEpQIRRiV EQD---tMpp KgvrhTMMfS ATfPkEIQm- LARDFLDEy-
6 : TFEDtrFVVL DEAdrML-DM GfsEDMRRiM ThV-----ta RPEHQTLMfS ATfPEEIQr- MAGEFLkNy-
7 : FFRFvDYkVL DEAdrLL-EI GfRDDLEtiS gILNEKNsks ADNIKTLLfS ATLDDKVQkl anNIMnkkEc
8 : DcRAvEtLIL DEAdrML-DM GfAQDIEhiA qET------- RwRkQTLLfS ATLEGDaIQD FAERLLEDp-
9 : AhWNpVLLaV DEAHCIS-Qw GhDFRpEyAA LGQLRQ---r fPTLPFMALT ATADDTTRQD IVRLLG----
10 : YLPADGLLVV DEsHVTIpQI GgmYRGDRAR KETLVE 20 ALaPQTIYVS ATPGNyELEk SGGDVVDQV-
11 : EFDRYQFIIF DEfHVLD-SN AIAFRNLchE ySyNGK---- -----IIkVS ATPPGREcD- LTTQYp----
12 : EVKTYDFVII DECHVnD-aS AIAFRNLLfE hEfEGK---- -----VLkVS ATPPGREVE- FTTQFp----
13 : RVVNWEVIIM DEAHfLD-pA SIAaRGWaAh RAraNE---- ---SaTILMT ATPPGTSDE- FphSnG----
14 : RVPNYNLFIM DEAHfTD-pA SIAaRGYIAT KVELGE---- ---aaaIFMT ATPPGTSDp- FpESnA----
15 : RVPNYNLIIM DEAHfTD-pA SIAaRGYIST RVEMGE---- ---aagIFMT ATPPGSrDp- FpQSnA----
16 : RVPNYNLIVM DEAHfTD-pS SVAaRGYIST RVEMGE---- ---aaaIFMT ATPPGaTDp- FpQSns----
17 : RVPNYNLFVM DEAHfTD-pA SIAaRGYIAT KVELGE---- ---aaaIFMT ATPPGTTDp- FpDSnA----
18 : RVPNYNLFVM DEAHfTD-pA SIAaRGYIST RVELGE---- ---aaaIFMT ATPPGTSDp- FpESnA----
19 : AMVEYSYIfL DEyHCaT-pE qLAIIGKIhR fSEsIR---- -----VVAMT ATPAGSVtt- TGQKhp----
20 : DNVDYGLIIL DEVHNLreSA YrykLIKNkL DT-------- aNNSKILVIT ATPmiDSKDE L-DSILSLtk
21 : NSKSrIcVII DECHNfIskS lIKEDGKIrp TRSVyNfL 5 lKNHKMIcLS ATPivNSVQE F-TMLVNLLr
22 : SrYNNSIFIV DEAHNIfGNN TgELMTVIkN ---------- KNKIPFLLLS GSPiTNTpNt L-GhIIDLMs
23 : AIDSYDVLIL DEVmsVIGQL YspTMrRLSA VDSLLyrLl- NRcSQIIAMd ATVNSQfID- LISgLRGDEN
24 : lLNNYDVLVL DEVmsTLGQL YspTMQQLGR VDALMlrLl- RIcPRIIAMd ATANAQLVD- FLcgLRGEkN
25 : EVsKDSIVIF DEAHNID-NV cIESLSLDLT TDALRR 192 ERfSsVIItS GIIsplDMyp rMlNFKTVLQ
SEC bbbb taaaaaa bbbbb bbttt????? ???
CONS1-8 o ++VL DEAD ML o GFo Q+ oI o +L+S AT+Poo+ o + oo++oo
I D M
CONS11-19 + oYo++++ DE H+ D +A R + o o + MT ATPPGoo + oo
VS
CONS20-22 oo+oo ++I+ DE HN++ooo + oo+ KIK oo +oo K+L+LS ATPI+NS oo + o I+oL+o
R I IT GS M DT L
CONS ++++ DE H + +T AT o + o
D S GS

```
                                                                 (IV)
          201        211        221        231        241        251        261
                                                               ** ********** **********
 1 :      ---IRILvKK EELTLEgIRQ FYINVER-EE WKLDTLCDLY ---EtLTiTQ AVIFINTRRK VDWLTEKM-h
 2 :      ---IRILvKK EELTLEgIKQ FYINVEr-EE WKLDTLCDLY ---EtLTiTQ AVIFLNTRRK VDWLTEKM-h
 3 :      ---VRILvKK DELTLEgIKQ FYVNVEE-EE YKYEcLtDLY ---DsISvTQ AVIFCNTRRK VEELTTKL-R
 4 :      ---IhINiGA LELSANhNIL QIVDVch-DV EKDEKLIrLM EeIasEKENK TIVFVETKRR cDELTRKh-R
 5 :      ---IfLAvGR VGSTSEnITQ KVVwVEE-AD KRsFLLDILN ---atkGDSL ILVFVETKKG ADSLeDFL-Y
 6 :      ---VfVAiGI VGGAcsdVKQ TIyEVNk-yA KRsKLIEiLs -----EQADG TIVFVETKRG ADFLASFL-S
 7 :      LfLDtVDkNE PEAhERIDQS VVISEkfANS IFaaVEhiKk QikErDSNyK AIIFaPTVKF TsFLcSILKN
 8 :      ---VEVSANp STRERKKIhQ WYyRADD-LE HKtaLLVhLL ---kQpEATR SIVFVrnRLE AVcMSWqTGc

 9 :      --------LN DPLIQISSFD RRNIrym-LM EKFKPLDQLM RyVQEQREKS GIIYCNSRaK VEDTAAAL-Q

10 :      --------VR PTGLLDpIIE VRpVATQ-VD DLLSEIRQRA -----AINER VLVtTITKRM AEDLTEYLEE

11 :      --------VE LLIEEQLSLR DFVDAQGTDA HADVVKKG-- --------DN ILVYVaSYNE VDQLSKMLNE
12 :      --------VK LKIEEALSFQ EFVSLQGTGA NADVISCG-- --------DN ILVYVaSYND VDSLGKLLvQ
13 :      --------EI EDVQTDIpSE pW--NTGhDw ILADK----- --------RP TaWFLPSIRA ANVMAAsLRK
14 :      --------pI SDMQTEIpDR aW--NTGyEw ItEYV----- --------gK TVWFVPSVKM GNEIALcLQR
15 :      --------pI MDEEREIpER SW--NSGhEw VtDFK----- --------gK TVWFVPSIKT GNDIAAcLRK
16 :      --------pI EDIEREIpER SW--DTGfDw ItDYQ----- --------gK TVWFVPSIKA GNDIANcLRK
17 :      --------pI hDLQDEIpDR aW--sSGyEw ItEYA----- --------gK TVWFVaSVKM GNEIAMcLQR
18 :      --------pI SDLQTEIpDR aW--NSGyEw ItEYI----- --------gK TVWFVPSVKM GNEIALcLQR
19 :      --IEEFIApE VMKGEDLg-S QFLDIAGLkI pVDEMK---- --------gN MLVFVPTRNM AVEVAKKLKa

20 :      eTSrIIfs-- -ENkIDIkIS YVgQEINGET LFLSEMKGqQ 1 36 EQsSK InaFINSIKE GELTVLFSfY
21 :      pgSLQhQsLf -ENkRLVDEK EVsKLGGLcS YIVNNEfSIF D 69 EIATL yndFkNSLRD rEFSkSALDT
22 :      eeTIDFGeII SRGkKVIQTL LNERGVNVLK DLLKGRISyY E 98 NIsSK fkYFINrIQT LNgkhFIYfS

23 :      IhTIVcTyAG VGfSGRTcTI LRDMGIDTLV RVIKRSPEhE D 19 QcGhN IcIFSsTLsF SELVAQFcAi
24 :      VhVVVgEyAM PGfSARRcLf LprLGTELLQ AALRPgPpsG p 22 GGGDN IcIFSsTVsF AEIVARFcRQ

25 :      kSyaMtLAKK SfLpMIITKg SDQVAIs-SR FeIRNDPSIV R 11 ITpDG MVVFfPSYLY MESIVSMWQT

SEC                                                            b bbbbtttaaa aaaaaaaaaa

CONS1-8     +  +          o   o  ++o o      o   +  L       o o o  +IF oToo   o L o
                                                                   V
CONS11-19               oooI  o  W   oG o    o                 o  ++FV S+o   oo+   L
                           L     F                                  Y
CONS20-22 EoS+o+ o++  oNKo+Iooo +Voo +N+ o ++Looo+S++ o    EISSK +o FINSIoo  E+o +++Fo
            T              V                 V             N  T         L    N
CONS                          o  +                                  F  S o   o
                                                                    Y  T
```


V
271
281
291
301
311
321
331
1 : AR--DFTVSA MHGDMDQKER dvIMREFRSG SsRVLITTDL LArGIDVQQV SLVINYDL-- ----------
2 : AR--DFTVSA LHGDMDQKER dvIMREFRSG SsRVLITTDL LArGIDVQQV SLVINYDL-- ----------
3 : Nd--KFTVSA IySDLpQQER dtIMKEFRSG SsRILISTDL LArGIDVQQV SLVINYDL-- ----------
4 : Rd--GWpAMG IHGDkSQQER dwVLNEFKhG KapILIATDV ASrGLDVEDV KFVINYDy-- ----------
5 : He--GYAcTS IHGDrSQRDR eEaLhQFRSG KspILVATaV AArGLDIsNV KHVINFDL-- ----------
6 : EK--EFpTTS IHGDrLQSQR eQaLRDFKNG SMKVLIATSV ASrGLDIKNI KHVINYDM-- ----------
7 : EfKkDLpILE FHGKITQNKR TsLVKRFKKD EsGILVcTDV GArGMDFPNV HEVLQIGV-- ----------
8 : An--GINNcY LEGEMVQGKR nEaIKRLTEG RVNVLVATDV AArGIDIPDV SHVFNFDM-- ----------
9 : SK--GISAAA YHAGLENNVR aDVQEKFQRD DLQIVVATVA fGMGINkPNV RFVVhFDI-- ----------
10 : H---GERVRY LHSDIDTVER MEIIRDLRLG EFDVLVGiNL LrEGLDMPEV SLVaILDAdk Egf-------
11 : R---GFLVTK VDGRTMKLgg VEIITKGSSi KKHFIVATNI IeNGVTL-DV DVVVDFGLkV vPnldsdnR-
12 : K---GYkVSK IDGRTMKSgg TEIITEGTSv KKHFIVATNI IeNGVTI-DI DVVVDFGtkV vPVldvdnR-
13 : A---GKSVVV LNRKTfERE- ---YptIKQK KpDFILATDI AeMGaNL-cV ErVLDcrtaF KPVlvdegR-
14 : A---GKkVIQ LNRKSyETE- ---YpKcKND DWDFVYTTDI seMGaNF-KA SrVIDsrksV KPTiieegdG
15 : N---GKrVIQ LSRKTfDSE- ---YVKTRTN DWDFVVTTDI seMGaNF-KA ErVIDprrcM KPViltdgee
16 : S---GKkVIQ LSRKTfDTE- ---YpKTKLT DWDFVVTTDI seMGaNF-RA GrVIDprrcL KPVilpdgpe
17 : A---GKkVIQ LNRKSyDTE- ---YpKcKNG DWDFVITTDI seMGaNF-gA SrVIDcrksV KPTileegeG
18 : A---GKkVIQ LNRKSyETE- ---YpKcKND DWDFVVTTDI seMGaNF-KA SrVIDsrksV KPTiitegeG
19 : K---GYN--- -SGyyySGEd pANLRVVTSQ SpyVIVATNA IesGVTLPDL DTVIDtGLkc Ekrvrvss Ki
20 : VKR-GIDFTS SVLESIgyK 32 SIaNIKgD NIHILLGSSV LSEsITLyRV KHLHIIsp-- ----------
21 : fKR-GELLGG DaSAaDiSL 70 QESNTNgE cIKtcVFSSs GGEGISFfsI NDIFILDM-- ----------
22 : NstyGgLVIK YIMLSNgys 39 SpENDDgS QLaFLFSSNI MSEsyTLKEV RHIWfMtI-- ----------
23 : -----FTDSI LILNSTrP-- --LcNVNEwK hFRVLVYTTV VTVGLSF-DM AHfHsMfAyI KPMsy-----
24 : -----FTDRV LLLhSLTP-- --LgDVTTwG QYRVVIYTTV VTVGLSF-Dp LHfdgMfAyV KPMNy-----
25 : MgilDEVWKh kLILVETPD 9 ATYRKAcSN gRGaIL1SVA rGEGIDFQyG RTVLMIGIpF QyTEsrilKA
SEC bbbbbbbb ttt???????
CONS1-8 o + + Go QooR o +ooFooG o VLI ToV RG+D+ oV o V+N+D+
I V L
CONS11-19 G oV + o oo+o + + oo o FV+ TDI E G o+ o V+D + P + o o
I N
CONS20-22 +oo G L+ o++ SoGYo o oNooGo oI +L+ So+ +SESIT++oV oHI+++o+
L S I L
CONS o o +++ To+ G o+ o ++
S S

```
                        (VI)
      341        351        361        371        381        391        401        411
                 ********** ********
  1 : ---------- PtNrENYIhR IGRGGRfGRK G---VAINMV TEEDkRTLRD -IETFYNTSI EEMpLNVaDL I   0
  2 : ---------- PtNrENYIhR IGRGGRfGRK G---VAINMV TEEDkRTLRD -IETFYNTTV EEMpMNVaDL I   0
  3 : ---------- PtNrENYIhR IGRGGRfGRK G---VAINMV TEEDkRTLRE -LEKFYSTQI EELpsDIaTL L   1
  4 : ---------- PNSSEDYIhR IGRTARstKt G---tAYtFF TPNNIKQVSD -LISVLREAN QAINpkLLQL V 139
  5 : ---------- PsDIEEYVhR IGRTGRVGNl G---LATsFF NERNINITKD -----LLDLL VEAKQEVPSw L  84
  6 : ---------- PskIDDYVhR IGRTGCVGNN G---RATsFF DPEKDRAIAA ----DLVKIL EGsGQTVPDf L  41
  7 : ---------- PsELANYIhR IGRTARsGKE G---ssVLFI cKDELpFVRE -LEDAKNIVI AKQEKYEpsE e 154
  8 : ---------- PRSGDTYLhR IGRTARAGRK G---tAIsLV EAHDHLLLGK ----VGRyIE EpIKarVIDE L  65

  9 : ---------- PRNIESYYQe tGRAGRDGLP A---EAMLFY DPADMAWLRR cLEEKpQGQL QDIErHKLNA M 237

 10 : ---------- LRSERGLIQT IGRAARN-VN G---KAILYG DKiTpSMaKA -IGEtERRRE KQQKYNEEhG I  89

 11 : ---lVsycki PiSLGErIQR fGRVGRNK-- ----PGVaLr iGETIKGLVE -IPSMIATEa AF--LcfVyG L 241
 12 : ---aVqynkt VvSYGErIQK LGRVGRhK-- ----EGVaLr iGQTNKTLVE -IPEMVATEa AF--LcfMyN L 240
 13 : K--vaikgpl riSASSAAQR rGRIGRNpNR D---GDSyYY SEpTSENnAh -hVcWLEASM LLDNMEVrgG M 115
 14 : R--vIlgeps AiTAASAAQR rGRIGRNpSQ V---GDEycY gGHTNEDdSN -fAhWtEARI MLDNINMPNG L 114
 15 : R--vIlagpm PvThSSAAQR rGRIGRNpRN E---NDQyIY mGEpLENdeD -cAhWkEAKM LLDNINTPEG I 111
 16 : R--vIlagpi PvTpASAAQR rGRIGRNpaQ E---DDQyVF SGDpLKNdeD -hAhWtEAKM LLDNIyTPEG I 114
 17 : R--vIlgnps PiTSASAAQR rGRVGRNpNQ V---GDEyhY gGATSEDdSN -lAhWtEAKI MLDNIHMPNG L 114
 18 : R--vIlgeps AvTAASAAQR rGRTGRNpSQ A---GDEycY gGHTNEDdSN -cAhWtEARI MLDNINMPNG L 114
 19 : pfivtglkrm AvTVGEqAQR rGRVGRVK-- ----PGRyYr SQETaTGSKD -yhyDLLQAQ RY---GIEDG I   ?

 20 : ---------- FwNYGQIkQS IGRAIRIGSh E--gLEDksM kvylHAAhYD -kEgKDIDIw KI-AYDKNkD I 159
 21 : ---------- twNEASLrQi VGRAIRLnSh VltpPERrYV NvHFIMARLG ----NGMpTV DE---DLFEI I 114
 22 : ---------- PDTFSQYnQi LGRSIRkfSy A---DISEpV NvylLAAVYS -dfNDEVTSL ND--YTQDEL I 141

 23 : ---------- GpDMvSVYQS LGRVrlLlLN E--vLMYVdG SRtRcGpLFS pMLLNFTiAN KFQwFpThTQ I 423
 24 : ---------- GpDMvSVYQS LGRVrtLRKg E--lLIYMdG SGARSEpVFT pMLLNhVvSs cGQwpAQFSQ V 418

 25 : RlefMre ll FDaMRhAAQc LGRVLRGKDD y---GVMVLA DRRFSRkRSQ -LPKWIAQGL sDADLNLsTD M  66

SEC                 aaaaaaa atttt?????

CONS1-8          Poo ooY+HR IGR GR Goo G    A o++ o oo   + o                o+ o   +
                                A
CONS11-19         VT  o QR   GR+GR                     o   o     + o     + +     G +
                  IS
CONS20-22         Wo+ o+oQI +GRAIR+ SH      Eoo V oVY++AA +o  o oo+oo+  oo  Yoooo+ I
                                                M
CONS                o  o  Q   GR  R                         o                       +
                          H
```

B

BVDV	Ia	RVLVLIPlRAAA
DEN2	Ia	RTLILaPTRVVA
recQ	Ia	lTVVVSPlIsLM
BVDV	IV	NMLVfVPTRnMA
WNV	IV	KTVWfVPSVkMG
RAD3	IV	qMVVffPSvLVM
TVMV	III	KIIkVSATPpGr
BVDV	III	RVVaMTATPAGS
uvrB	III	QTIYVSATPGnv
TVMV	V	KkhFIVATnIIe
BVDV	V	spyVIVATnAIe
p68	V	KapILIATdVAS

Fig.1. (A) Alignment of conserved regions of (putative) helicases of the new superfamily.
Abbreviations of viruses stand for respective proteins (see Methods), and VV1 and VV2 for NTPases I and II of VV, respectively. CONS1-8,11-19,20-22 are consensus amino acid residue patterns for the 'D-E-A-D' family (entries 1-8), the family of RNA viral proteins (entries 11-19), and that of DNA viral/plasmid proteins (entries 20-22). CONS is the joint consensus derived as the overlap of the three patterns. +, hydrophobic residues (I,L,V,M,F,Y,W); o, charged or polar residues (S,T,D,E,N,Q,K,R). Where single symbols are indicated, one exception was allowed. For positions where two residues were observed, only pairs of similar residues were included in the consensus patterns. Residues belonging to one of the following groups were counted as similar: L,V,I,M; G,A; S,T; K,R; D,E,N,Q; F,Y,W. Residues having no identical or similar matches in sequences of other families or individual proteins outside the families are shown in lower case. Dashes designate gaps introduced for optimal alignment. The numbers of amino acid residues in terminal regions of all proteins and in inserts available in some of the proteins are indicated. Question marks indicate that precise distances to the protein termini are unknown. For BVDV, polyprotein fragment from residue 1898 to 2223 is shown. The alignment of the 'D-E-A-D' proteins was from (19), with minor modifications. The residue numbering above the alignment is arbitrary, beginning from the first aligned residue. Conserved segments are numbered I to VI. Asterisks denote residues used for statistical analysis. Where gaps were introduced into conserved segments, those segments of the respective sequences were omitted from calculations. Secondary structure prediction: a, α-helix; b, β-strand; t, β-turn; ?, prediction ambigous. Sites of amino acid substitutions in RAD3 (see text) are underlined. Arrows indicate insertions of 3 and 2 residues in segment V of RAD3. Source references are in parentheses preceding each of the aligned sequences.
(B) Alignment of selected sequence stretches from different conserved segments of the proteins of the new superfamily. Amino acid residues having identical or similar counterparts in 'heterologous' segments are shown in upper case.

Here, l_1 and l_2 are the lengths of the two compared sequences, and p_i are double matching probabilities calculated for each of n conserved segments aligned without gaps, using the algorithm of McLachlan (26). To obtain the upper limit estimate for P, it was accepted $l_1=l_2=1000$ which is sowewhat above the maximal length of the compared proteins, and the spacing of the conserved segments was not taken into account. The program DIAGON was written in the C programming language and run on a WicatS150 computer. The program OPTAL was written in FORTRAN 77 and run on IBM PC AT. Secondary structure prediction was by the Chou and Fasman method (27).

RESULTS AND DISCUSSION

Formation of a new superfamily of putative helicases

Sequence comparison of NTP-binding motif containing proteins revealed several distinct families [(5,6,12,13,23,28), and manuscript in preparation]. For two of such families, one including putative NTP-binding domains of replicative proteins of three groups of positive strand RNA viruses, and the other NTPases of vaccinia virus and a yeast mitochondrial plasmid-encoded protein, consensus patterns of conserved amino acid residues resembling that of the 'D-E-A-D' family were derived. The sequences of the three families were aligned so as to maximize the overlap between these patterns. This allowed delineation of 7 conserved segments (Fig.1A). Most striking was the similarity between the 'D-E-A-D' family and RNA viral proteins confirmed by pairwise alignments some of which yielded high AS values (e.g. app. 7 SD for CI protein of TEV vs. eIF-4AI). All available sequences of other NTP-binding motif-containing proteins (6, and manuscript in preparation) were screened for complete or partial correspondence to the joint consensus pattern derived upon comparison of the three families. As the result, a set of 25 proteins was delineated (Fig.1A). E.coli protein recQ displayed an unexpectedly high similarity to the 'D-E-A-D' family, with AS of app. 11 SD for comparisons with eIF-4A and p68 sequences. High local similarities were also observed between this family and uvrB, despite two insertions in the latter protein. For two herpesvirus proteins and yeast helicase RAD3, more modest segmental similarities were observed, the spacer lengths between the 7 conserved segments varying significantly (Fig.1A).

The significance of the final alignment was assessed by calculating the probability of simultaneous chance occurence of all 7 segments for each pair of sequences as described under Methods. These calculations showed that all aligned proteins were linked into a single network by highly significant matches, with the possible exception of RAD3 (Fig.2). However, numerous data on mutagenesis of this protein are available (see below), corroborating our identification of segments important for its function.

Characterization of the conserved segments

The final alignment contained 6 invariant amino acid residues distributed among 7 conserved segments (Fig.1A). Of these residues, 2 were observed in segment I, 2 in segment II,

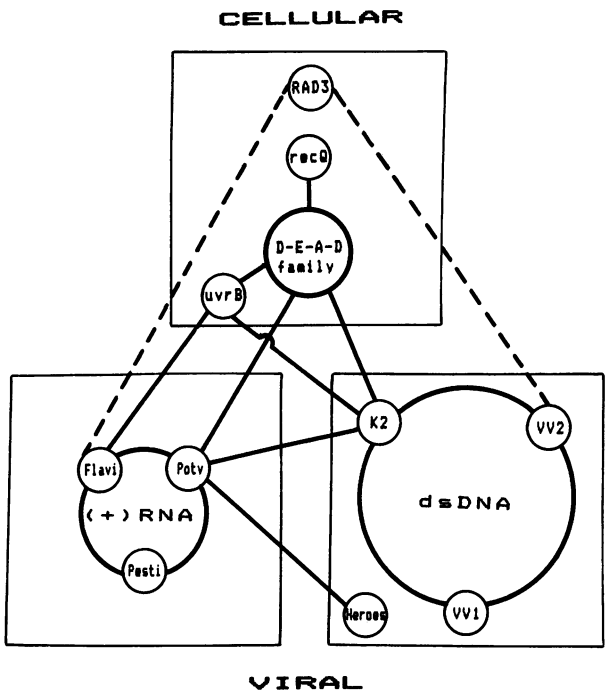
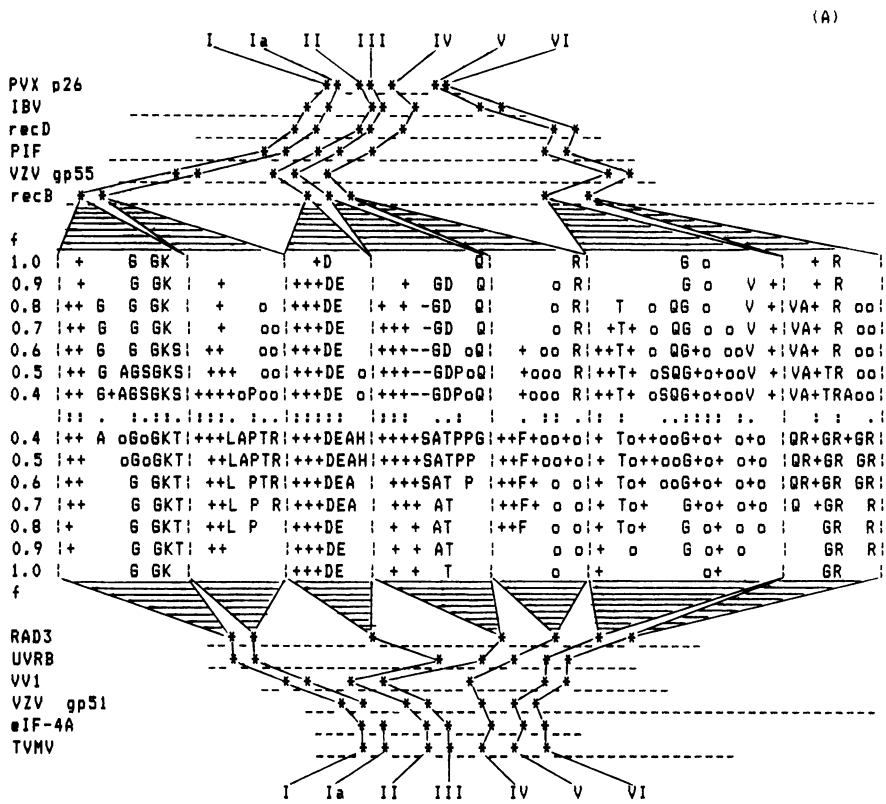
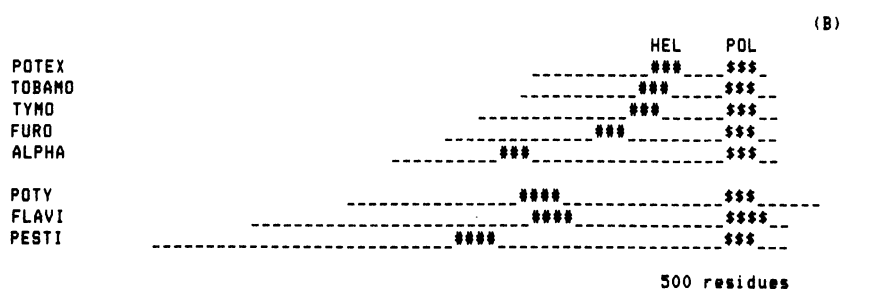


Fig.2. A schematic representation of the relationships between the members of the new superfamily of (putative) helicases. Squares enclose proteins of similar origin (cellular, RNA viral and DNA viral). Names of virus groups (flavi, poty, pesti and herpes) enclosed into small circles stand for respective sets of closely related proteins. The large circles link proteins constituting groups delineated by sequence comparison (probability of chance similarity $P<10^{-9}$). The diameters of these circles are in approximate reverse proportion to the degree of similarity between the members of each group. Solid straight lines indicate significant connections between the groups ($P<10^{-7}$). Of the 'D-E-A-D' family, only the sequences of eIF-4AI and p68 were used for calculations. For any two groups only one best connection is shown. Dashed lines correspond to $10^{-3}<P<10^{-5}$.

and 2 in segment VI. Segment I corresponded to the "A" site of the NTP-binding motif. The N-terminal G/A fixed in the "A" consensus was replaced by a bulky residue in 12 proteins of the new superfamily (position 8 in Fig.1A). Another G residue was conserved in position 10, presumbly maintaining the flexible loop conformation typical of this site (7-10). Segment II corresponded to the "B" site of the NTP-binding motif thought to interact with the Mg^{2+} cation of Mg-NTP through the conserved D residue (7-10). Segment VI, the 3rd most conserved


(A)
I Ia II III IV V VI
PVX p26
IBV
recD
PIF
VZV gp55
recB
f
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.4
0.5
0.6
0.7
0.8
0.9
1.0
f
RAD3
UVRB
VV1
VZV gp51
eIF-4A
TVMV
I Ia II III IV V VI
200 residues


(B)
HEL
POL
POTEX
TOBAMO
TYMO
FURO
ALPHA
POTY
FLAVI
PESTI
500 residues

segment in the proteins of the new superfamily, might be a special kind of nucleic acid binding site, provided the abundance of positively charged residues. A similar motif has recently been implicated in RNA binding in several nuclear proteins (29). A correlation between the conserved patterns of segments II and VI might be of interest. In segment II, most proteins of the superfamily outside the 'DEAD' family had the signature 'DExH'. In segment VI, the signature of the 'DEAD' family was 'HxxGRxxR', and that of other proteins 'QxxGRxxR', suggesting a sort of compensation. Sequence motifs revealed in segments Ia and III to V were less strictly conserved, and only degenerate forms of some of them could be identified in certain proteins (Fig.1A). A degree of similarity could be revealed between different segments, suggesting they might be considered imperfect repeats (Fig.1B).

Secondary structure predictions indicated that each of the conserved segments centered at a β-turn usually flanked from

Fig.3. Comparison of the proteins constituting the two (putative) helicase superfamilies.
(A) Correspondence between the conserved amino acid residue patterns of SF1 (upper) and SF2 (lower). Additional abbreviations: PVX, potato virus X (a potexvirus); IBV, infectious bronchitis virus (a coronavirus); PIF, a yeast mitochondrial helicase. For SF1, the data are from an updated version of the published alignment (13), and for SF2 from the alignment shown in Fig.1A. Asterisks designate conserved segments numbered as in Fig.1. Their positioning in the proteins designated by dashed horizontal lines is shown to scale. For each superfamily, a representative sampling was generated including proteins of different origin (i.e. RNA viral, DNA viral, prokaryotic and eukaryotic) to show the entire length span of the spacers separating the conserved segments. The boundaries of the IBV protein were predicted from the analysis of putative cleavage sites (A.E.G. et al., submitted). f, approximate frequency of the consensus residues. The designation system for the consensus patterns is from (66), with modifications. Colons highlight complete correspondence between the two consensus patterns, and dots partial correspondence. Other designations are as in Fig.1A.
(B) Location of the putative helicase domains of the two superfamilies in multidomain proteins of positive strand RNA viruses.
Multidomain proteins (dashed lines) and the conserved regions of the putative helicases (HEL) and of the RNA polymerases (POL) are shown to scale. For tobamo-, alpha- and potyviruses, more detailed schemes have been published (23). For potexviruses, the data are from (67,68), for tymoviruses from (69), for flaviviruses from (54-59), and for pestiviruses from (70). For potex- and furoviruses having each two putative helicases, only those embodied in multidomain proteins are shown.

the N-side by a β-strand, and only in segment VI by an α-helix (Fig.1A).

Implications for protein functions

The sequence, and presumably structural, similarity between the proteins of the new superfamily suggests they should be similar to some extent also functionally. The best guess is that their common activity might be that of a nucleic acid-dependent NTPase, possibly a helicase. This had been documented for only a few proteins, but what is known of the functions of the others supports to some extent, or at least does not contradict this proposal. RNA helicase activity has been revealed in p68 (14), SrmB (16) and eIF-4A (30). RAD3 is a DNA helicase involved in yeast DNA repair, and possibly replication (31,32). UvrB is a subunit of uvrAB helicase (33) displaying, under certain conditions, ATPase activity (34). DNA-dependent ATPase activity was described for the two vaccinia proteins (35,36). RecQ is a component of the recF recombination pathway in E.coli whose specific activity is unknown (37). UL9 protein of HSV specifically binds to the virus DNA replication origin (38). RNA viral proteins are poorly studied but for flavivirus NS3 involvement in RNA replication is strongly suggested (39). A survey of spontaneous and artificial mutants of RAD3 (32,40-42) showed that all the numerous mutations impairing its activity in excision DNA repair and/or its essential function fell exactly within the conserved segments I to V identified here (Fig.1A). This lends strong support for the involvement of these segments in the helicase function of RAD3 and, by implication, of other proteins of the new superfamily.

Comparison of the two helicase superfamilies

It was of interest to compare the pattern of conserved structural elements of the putative helicase superfamily described here (hereafter SF2) with that of the superfamily identified previously (SF1). Proteins of both groups have 7 conserved segments of which most are probably similar at the level of secondary structure (cf.13 and Fig.1A). Superposition of these segments revealed a number of coincidences beyond the NTP-binding motif proper, particularly in segments I, II, V and VI. For other segments which were more variable within each superfamily, the similarity was not that obvious (Fig.3A). The lengths of spacers separating the conserved segments in the proteins of the two superfamilies overlapped in each case (Fig.3A). Interestingly, the putative NTPases of both superfamilies occupied similar locations in multidomain proteins of positive strand RNA viruses relative to conserved RNA polymerase domains (Fig.3B). Taken together, it could be concluded that the two superfamilies were distinct but distantly related.

Previously, the correspondence between segments I,Ia (18,43), II, V and VI (18) has already been established for some of the proteins now included into SF1 and SF2. In other works, superpositions which are now to be regarded as partially erratic have been presented (17,20,44). Presumably, this could be due to scant representation of SF2.

CONCLUDING REMARKS

Hopefully, identification of the two (putative) helicase superfamilies and demonstration of a distant relationship between them may initiate formation of a conceptual framework for further studies of these important enzymes. There are several well characterized helicases which could not be included neither in SF1 nor in SF2 (unpublished observations). These include SV40 T antigen (45) whose sequence is related to those of NS1 proteins of parvoviruses (28), E.coli proteins recA (46), dnaB (47) and rho (48), and some others. Thus, conservation of the sequence motifs typical of SF1 and/or SF2 is not obligatory for a helicase. Revelation of functional constraints leading to this conservation is a tantalizing goal for future studies.

ACKNOWLEDGEMENTS

The authors are grateful to Professor V. I. Agol for constant interest and useful criticisms, to Dr. K. M. Chumakov for help with computer programming, and to Dr. M. N. Rozanov for helpful suggestions. We thank Dr. M. Ashburner for sending a preprint of his work.

*To whom correspondence should be addressed.

REFERENCES

1. Muskavitch,K.M.T. and Linn,S. (1981) in Boyer,P.D. (ed.) The Enzymes, vol.14, pp.233-250, Academic Press, New York.
2. Geider,K. and Hoffman-Berling,H. (1981) Annu.Rev.Biochem. 50, 233-260.
3. Merrick,W.C., Abramson,R.D., Anthony,D.D., Dever,T.E. and Caliendo,A.M. (1987) In Ilan,J. (ed.) Translational Regulation of Gene Expression, pp.265-286.
4. Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) EMBO J. 1, 945-951.
5. 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.
6. Doolittle,R.F. (1986) In Inouye,M. (ed.) Protein Engineering, pp.15-27, Plenum, New York.
7. La Cour,T.F.M., Nyborg,J., Thirup,S. and Clark,B.F.C. (1985) EMBO J. 4, 2385-2388.
8. Fry,D.C., Kuby,S.A. and Mildvan,A.S. (1986) Proc.Nat.Acad.Sci.USA 83, 907-911.
9. 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.
10. Jurnak,F. (1988) Trends Biochem.Sci. 13, 195-198.
11. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov, V.M. (1988) Nature 333, 22.
12. Hodgman,T.C. (1988) Nature 333, 22-23.

13. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov, V.M. (1988) FEBS Lett. 235, 16-24.
14. Ford,M.J., Anton,I.A. and Lane,D.P. (1988) Nature 332, 736-738.
15. Lasko,P.F. and Asnburner,M. (1988) Nature 335, 611-617.
16. Nishi,K., Morel-Deville,F., Hershey,J.W.B., Leighton,I. and Schnier,J. (1988) Nature 336, 496-498.
17. Hay,B., Jan,L.Y. and Jan,Y.N. (1988) Cell 55, 577-587.
18. Seraphin,B., Simon,M. and Boulet,A. (1989) Nature 337, 84-87.
19. Linder,P., Lasko,P.F., Asnburner,M., Leroy,P., Nielsen,P.J., Nishi,K., Schnier,J. and Slonimski,P.P. (1989) Nature 337, 121-122.
20. Lane,D.P. (1988) Nature 334, 478.
21. Staden,R. (1982) Nucleic Acids Res. 10, 2951-2961.
22. Pozdnyakov,V.I. and Pankov,Yu.A. (1981) Int.J.Peptide Prot. Res. 17, 284-291.
23. Gorbalenya,A.E., Blinov,V.M., Donchenko,A.P. and Koonin,E.V. (1989) J.Mol.Evol. 28, in press.
24. Dayhoff,M.O., Barker,W.C. and Hunt,L.T. (1983) Meth.Enzymol. 91, 524-545.
25. Sankoff,D. (1972) Proc.Nat.Acad.Sci.USA 69, 1-3.
26. McLachlan,A.D. (1971) J.Mol.Biol. 61, 409-424.
27. Chou,P.Y. and Fasman,G.D. (1978) Adv.Enzymol. 47, 45-148.
28. Astell,C.R., Mol,C.D. and Anderson,W.F. (1987) J.Gen.Virol. 68, 885-893.
29. Christensen,M.E. and Fuxa,K.P. (1988) Biochem.Biophys.Res.Commun. 155, 1278-1283.
30. Ray,B.K., Lawson,T.G., Kramer,J.C., Cladaras,M.H., Grifo, J.A., Abramson,R.D., Merrick,W.C. and Thach,R.E. (1985) J.Biol.Chem. 260, 7651-7658.
31. Sung,P., Prakash,L., Matson,S.W. and Prakash,S. (1987) Proc.Nat.Acad.Sci.USA 84, 8951-8955.
32. Sung,P., Higgins,D., Prakash,L. and Prakash,S. (1988) EMBO J. 7, 3263-3269.
33. Oh,E.Y. and Grossman,L. (1987) Proc.Nat.Acad.Sci.USA 84, 3638-3642.
34. Friedberg,E.C. (1988) Microbiol.Rev. 52, 70-102.
35. Rodriguez,J.F., Kahn,J.S. and Esteban,M. (1986) Proc.Nat.Acad.Sci.USA 83, 9566-9570.
36. Broyles,S.S. and Moss,B. (1987) J.Virol. 61, 1738-1742.
37. Irino,M., Nakayama,K. and Nakayama,H. (1986) Mol.Gen. Genet. 205, 298-304.
38. Olivo,P.D., Nelson,N.J. and Challberg,M.D. (1988) Proc.Nat.Acad.Sci.USA 85, 5414-5418.
39. Grun,J.B. and Brinton,M.A. (1987) J.Virol. 61, 3641-3644.
40. Naumovski,L., Chu,G., Berg,P. and Friedberg,E.C. (1985) Mol.Cell.Biol. 5, 17-26.
41. Naumovski,L. and Friedberg,E.C. (1986) Mol.Cell.Biol. 6, 1318-1327.
42. Naumovski,L. and Friedberg,E.C. (1987) Mol.Gen.Genet. 209, 458-466.

43. Gorbalenya,A.E. and Koonin,E.V. (1988) Nucleic Acids Res. 16, 7734.
44. Foury,F. and Lahaye,A. (1987) EMBO J. 6, 945-951.
45. Wiekowski,M., Schwartz,M.W. and Stahl,H. (1988) J.Biol.Chem. 263, 436-442.
46. Kowalczykowski,S.C. (1987) Trends Biochem.Sci. 12, 141-145.
47. Dombrosky,A.J. and Platt,T. (1988) Proc.Nat.Acad.Sci.USA 85, 2538-2542.
48. Bear,D.G. and Peabody,D.S. (1988) Trends Biochem.Sci. 13, 343-347.
49. Nielsen,P.J., McMaster,G.K. and Trachsel,H. (1985) Nucleic Acids Res. 13, 6867-6880.
50. Nielsen,P.J. and Trachsel,H. (1988) EMBO J. 7, 2097-2105.
51. Backendorf,C., Spaink,H., Barbeiro,A.P. and van de Putte, P. (1986) Nucleic Acids Res. 14, 2877-2890.
52. Domier,L., Franklin,K.M., Shahabuddin,M., Hellmann,G.M., Overmeyer,J.H., Hiremath,S.T., Siaw,M.F.E., Lomonosoff,G.P., Shaw,J.G. and Rhoads,R.E. (1986) Nucleic Acids Res. 14, 5417-5430.
53. Allison,R., Johnston,R.E. and Dougherty,W.G. (1986) Virology 154, 9-20.
54. Rice,C.M., Lenches,E.M., Eddy,S.R., Shin,S.J., Sheets,R.L., Strauss,J.H. (1985) Science 229, 726-733.
55. Castle,E., Nowäk,T., Leidner,U., Wengler,G. and Wengler,G. (1986) Virology 145, 227-236.
56. Hahn,Y.S., Galler,R., Hunkapiller,T., Dalrymple,J.M., Strauss,J.H. and Strauss,E.G. (1988) Virology 162, 167-180.
57. Mackow,E., Makino,Y., Zhao,B., Zhang,Y.M., Markoff,L., Buckler-White,A., Guiler,M., Channock,R. and Lai,C.J. (1987) Virology 159, 217-228.
58. Sumiyoshi,H., Mori,C., Fuke,I., Morita,K., Kuhara,S., Kondou,J., Kikuchi,Y., Nagamatu,H. and Igarashi,A. (1987) Virology 161, 497-510.
59. Coia,G., Parker,M.D., Speight,G., Byrne,M.E. and Westaway, E.G. (1988) J.Gen.Virol. 69, 1-21.
60. Collett,M.S., Larson,R., Gold,C., Strick,D., Anderson,D.K. and Purchio,A.F. (1988) Virology 165, 191-199.
61. Tommasino,M., Ricci,S. and Galeotti,C.L. (1988) Nucleic Acids Res. 16, 5863-5878.
62. Niles,E.G., Condit,R.C., Caro,P., Davidson,D., Matusick,L. and Seto,J. (1986) Virology 153, 1728-1742.
63. Davison,A.J. and Scott,J.E. (1986) J.Gen.Virol. 67, 1759-1816.
64. 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.
65. Reynolds,P., Higgins,D.R., Prakash,L. and Prakash,S. (1985) Nucleic Acids Res. 13, 2357-2372.
66. Patthy,L. (1987) J.Mol.Biol. 198, 567-577.
67. Forster,R.L.S., Bevan,M.W., Harbison,S.-A. and Gardner,R.C. (1988) Nucl.Acids Res. 16, 293-303.

68. 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 300, 711-716.
69. Morche,M.-D., Boyer,J.-C. and Haenni,A.L. (1988) Nucl.Acids Res. 16, 6157-6173.
70. Collett,M.S., Anderson,D.K. and Retzel,E. (1988) J.Gen.Virol. 69, 2637-2643.