Volume 291, number 2, 277–281 FEBS 10271 October 1991

ADONIS 001457939100980N

Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain

Alexander E. Gorbalenya[1] and Eugene V. Koonin[2]

[1]*Institute of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, 142782 Moscow Region, USSR* and [2]*Institute of Microbiology, USSR Academy of Sciences, 117811 Moscow, USSR*

Received 29 July 1991

A statistically significant amino acid sequence similarity is demonstrated between the endonuclease (R) subunit of *Eco*K restriction-modification (R-M) enzyme, and RNA and DNA helicases of the so-called 'DEAD' family It is further shown that all three known sequences of R subunits of type-I and type-III R-M enzymes contain the conserved amino acid sequence motifs typical of the previously described helicase superfamily II [(1989) Nucleic Acids Res 17, 4713–4730] A hypothesis is proposed that these enzymes may exert helicase activity possibly required for local unwinding of DNA in the cleavage sites

Amino acid comparison, Sequence pattern, ATP-binding site Helicase, Restrictase

1 INTRODUCTION

Type-I and type-III restriction-modification (R-M) enzymes are three- or two-subunit complexes, respectively, requiring ATP for DNA cleavage [1,2] Type-I enzymes consist of S (Specificity), M (Modification) and R (Restriction) subunits, and type-III enzymes of S-M and R subunits The role for ATP in the action of these enzymes is apparently dual, first inducing a conformational change allowing discrimination between modified and unmodified recognition sites on DNA, and then providing the energy thought to be required for DNA translocation through the enzyme (type-I), or for enzyme release from DNA (type-III) At the first step ATP is replaceable by its non-hydrolyzable analogs, whereas the second step requires ATP hydrolysis, ATP hydrolysis is thus necessary for DNA cleavage but not for modification (for type-III enzymes, however, some cleavage is observed in the absence of ATP hydrolysis) The ATP binding and ATPase activities have been assigned to the R subunits [1] Sequence determination of *hsd* genes encoding the type-I and type-III enzyme subunits revealed little conservation at the amino acid level, especially among the R subunit sequences [3–5] Importantly, however, it has been noticed [3] that the *Eco*K R subunit sequence contained the so-called 'A' motif of the purine NTP-binding pattern typical of a wide range of ATP- and GTP-utilizing enzymes [6,7]

Here, we show that R subunits of the type-I and type-III restrictases share a distant but reliable sequence similarity and contain the sequence motifs characteristic for superfamily II of DNA and RNA helicases [8] It is speculated that R subunits of type-I and type-III restrictases posses helicase activity which may be involved in local unwinding of DNA in the cleavage sites.

Abbreviations Pl, plasmid, BP, bacteriophage, VV, vaccinia virus, VZV, varicella-zoster virus, TVMV, tobacco vein mottling virus, YFV, yellow fever virus, BVDV, bovine viral diarrhea virus, BMV, brome mosaic virus, TMV, tobacco mosaic virus, SNBV, Sindbis virus, RUBV, rubella virus, BNYVV, beet yellow vein virus, PVX, potato virus X, IBV, infectious bronchitis virus, IFIVPI HsdR = *Eco*124/3 R subunit, *E coli* HsdR = *Eco*K R subunit, P1BP RES = *Eco*P1 R subunit

Correspondence address A E. Gorbalenya, Inst of Poliomyelitis and Viral Encephalitides, USSR Academy of Medical Sciences, Post-office Inst of Poliomyelitis, 142782 Moscow Region, USSR

2 MATERIALS AND METHODS

2 1 *Amino acid sequences*

For sources of the sequences see [7,8] except for *E coli* HsdR [3], P1BP RES [4], IFIVP1 HsdR [5], yeast putative DNA helicase RadH [9], spliceosome component PRP16 [10], and the putative helicase of RUBV [11]

2 2 *Comparative sequence analysis*

The SWISSPROT data bank version 12 was searched for sequences similar to a query sequence using the program QUICK, which is a module of the GENEBEE program package for biopolymer sequence analysis [12] Amino acid sequences were compared by the program OPTAL as previously described [13] using the amino acid residue comparison matrix MDM78 Program OPTAL, implementing the Sankoff algorithm, generates multiple sequence alignments in a stepwise manner and calculates adjusted alignment scores as the number of standard deviations (SD) over the mean of 25 random simulations

3 RESULTS AND DISCUSSION

3 1 *R subunit of* EcoK *is related to the RNA and DNA helicases of the 'DEAD' family*

The vast class of the purine NTP-binding pattern-

```
                  +          ++                                                  +
1:EIF-4AI  68 GYDVIAQAQS GTGKTATFAI SILQQI-ELD -------LKA ----TQA--- ---LVLAPTR
2:P68     110 GLDMVGVAQT GSGKTLSYLL PAIVHI-NHQ P----F-LER ----GDG--- PICLVLAPTR
3:VASA    281 GRDLMACAQT GSGKTAAFLL PILSKL-LED P----HELEL ----GRP--- QV-VIVSPTR
4:SrmB     40 GPDVLGSAPT QAGKTAAYLL PALQHL-LDF P----R-KKS ----GPP--- RI-LILTPTR
5:PL10    215 KRDLMACAQT GSGKTAAFLL PILSQIYTDG PGEALRAMKE NGKYGRRKQY PISLVLAPTR
6:MSS     142 DHDVIAPAKT GTGKTFAFLI PIFQHI-INT K--:----FDS Q---YMV--- KV-IVAAPTR
7:RecQ     41 GRDCLVVMPT GGGKSLCYQI PAL--L-LNG -------LT- ---------- ---VVVSPLI
                        * *****    *   *  *            *                  ***   *
8:HsdR    481 QQEILLAMAT GTGKTRT-AI AMMFRLIQSQ R------FK- ---------- RI-LFLVdRR

   +                                                          +
1:ELAQQIQKVV MAL-GDYMGA S---CHACIG GTNVRAEVQK LQMEAPHIIV GTPGRVF--D MLNR-RYLSP
2:ELAQQVQQVA AEY-CRACRL K---STCIYG GAPKGPQIRD LER-GVEICI ATPGRLI--D FLEC-GKTNL
3:ELAIQIFNEA RKF-AFESYL K---IGIVYG GTSFRHQNEC ITR-GCHVVI ATPGRLL--D FVDR-TFITF
4:ELAMQVSDHA REL-AKHTHL D----ATITG GVAYMNHAEV FSE-NQDIVV ATTGRLL--Q YIKE-ENFDC
5:ELAVQIYEEA RKF-SYRSRV R---PCVVYG GADIGQQIRD LER-GCHLLV ATPGRLV--D MMER-GHIGL
6:DLALQIEAEV KKIHDMNYGL KKYACVSLVG GTDFRAAMNK MNKLRPNIVI ATPGRLI--D VLEKYSNKFF
7:SL---MKDQV DQL-QANGSL N----STQIR EQQLEVMTGC RTG-QIRLLY IAPERLMLDN FLE---HLAH
   *  *                 *                      *       * * **   **     **    * *
8:SLGEQALGAF ED-----TRI N---GDTFNS IFDIKGLTDK FPEDSTKIHV ATVQsLV-KR TLQSDEPMPV

          ++ +      +  +                    +
1:KYIKMFVLDE ADEMLSRG-- FKDQ---IYD IFQK--LNSN --TQVVLLSA    -28-    LEGIRQFYIN
2:RRTTYLVLDE ADRMLDMG-- FEPQ---IRK IVDQ--IRPD --RQTLMWSA    -32-    ILQIVDVCHD
3:EDTRFVVLDE ADRMLDMG-- FSED---MRR IMTHVTMRPE --HQTLMFSA    -28-    CSDVKQTIYE
4:RAVETLILDE ADRMLDMG-- FAQD---IEH IAGE--TRWR --KQTLLFSA    -29-    RKKIHQWYYR
5:DFCKYLVLDE ADRMLDMG-- FEPQ---IRR IVEQDTMPPK GVRHTMMFSA    -28-    SENITQKVVW
6:RFVDYKVLDE ADRLLEIG-- FRDDLETISG ILNEKNSKSA DNIKTLLFSA    -32-    HERIDQSVVI
7:WNPVLLAVDE AHCISQWGHD FRPE----YA ALGQLRQRFP -TLPFMALTA    -25-    FDRRNIRYML
       ***** * *    *   *  *          *            ***                 *
8:ARYDCIVVDE AHR----G-- YILD---KEQ TEGE--LQFR --SQLDYVSA    -84-    LEDDQDFEVA

1:VEREEWKLDT LCDLY---ET L----TITQ- AVIFINTRRK VDWL--T-EK MHARDFT--- ---VSAMHGD
2:VEKDE-KLIR LMEEI---MS E----KENK- TIVFVETKRR CDEL--T-RK MRRDGWP--- ---AMGIHGD
3:VNKYA-KRSK LIE-I---LS E----QADG- TIVFVETKRG ADFL--A-SF LSEKEFP--- ---TTSIHGD
4:ADDLEHKTAL LVHLL---KQ P----EATR- SIVFVRNRLE AVCM--SWQT GCANGIN--- ---NCYLEGE
5:VEEAD-KRSF LLDLL---NA T----KGDSL ILVFVETKKG ADSL--E-DF LYHEGYA--- ---CTSIHGD
6:SEKFANSIFA AVEHI---KK QIKERDSNYK AIIFAPTVKF TSFLCSILKN EFKKDLP--- ---ILEFHGK
7:MEKFK-PLDQ LMRYV---QE Q----REKS- GIIYCNSRAK VEDT--A-AA LQSKGIS--- ---AAAYHAG
             *   *                  **** *     * *          *              * **
8:DFNRGLVIPA FNRAVCNELT NYLDPTGSQK TLVFCVTNAH ADMVVEELRA AFKKKYPQLE HDAIIKITGD

      +                          +      +         +     +      +    ++
1:MDQKERD-VI M-REFRSGSS R-VLITTDLL ARGIDVQQVS LVINYDLPTN RENYIHRIGR GGRF
2:KSQQERD-WV L-NEFKHGKA P-ILIATDVA SRGLDVEDVK FVINYDYPNS SEDYIHRIGR TARS
3:RLQSQRE-QA L-RDFKNGSM K-VLIATSVA SRGLDIKNIK HVINYDMPSK IDDYVHRIGR TGCV
4:MVQGKRN-EA I-KRLTEGRV N-VLVATDVA ARGIDIPDVS HVFNFDMPRS GDTYLHRIGR TARA
5:RSQRDRE-EA L-HQFRSGKS P-ILVATAVA ARGLDISNVK HVINFDLPSD IEEYVHRIGR TGRV
6:ITQNKRT-SL V-KRFKKDES G-ILVCTDVG ARGMDFPNVH EVLQIGVPSE LANYIHRIGR TARS
7:LENNVRA-DV QEQKFRDD-L Q-IVVATVAF GMGINKPNVR FVVHFDIPRN IESYYQETGR AGRD
       *      *  **         ***    *    ****  *   ** #          *     **    *
8:ADKDARKVQT MITRFNKERL PNIVVTvDLL TTGVDIPSIC NIVFLRKvRS RILYEQMKGR ATRL
```

containing proteins could be divided, on basis of the observed sequence similarities, into numerous groups of different ranks Each of these groups could be characterized by conservation of some additional sequence motifs, besides specific forms of the two motifs constituting the NTP-binding pattern [7] Upon screening the SWISSPROT data bank for similarity to the amino acid sequence of the *E coli HSDR* gene product (*Eco*K R subunit), considerable similarity was revealed with the mouse eIF-4A sequence, which is a member of the so- called 'DEAD' family of RNA and DNA helicases [8,14] Detailed sequence comparison between the R subunit and the 'DEAD' family helicases revealed two segments of over 140 amino acid residues each, displaying statistically significant similarity (Fig 1) Of the 21 invariant residues of the presented set of cellular proteins (one of these residues cannot be seen in Fig 1 as it lies within the spacer separating the two stretches shown), 10 were fully conserved and 6 were replaced by similar residues in the endonuclease

3.2 *All sequenced R subunits of type-I and type-III enzymes contain the motifs specific for the helicase superfamily II*

'DEAD' family is a subdivision of the so-called helicase superfamily II whose members share 7 highly conserved sequence motifs [8] Thus we searched the three known sequences of type-I and type-III R subunits for the presence of these motifs Five motifs (I, Ia, II, V and VI) could be readily identified in all the three sequences The counterparts to motif III could be confidently identified only in the *Eco*R124/3 and *Eco*P1 R subunit sequences, and counterpart to motif IV in that of *Eco*K (Fig 2) These results suggested that R subunits of type-I and type-III R-M complex might be peripheral members of the helicase superfamily II. However, as we have demonstrated previously that the helicase superfamilies I and II are distantly related [8], it seemed important to compare the sequences of the R subunits to those of helicases belonging to each superfamily Superposition of the conserved motifs typical of the two superfamilies with those found in endonucleases is shown in Fig. 2. It could be concluded that the R subunit sequences fitted better the consensus patterns of superfamily II. This was particularly obvious when motifs II, III and VI were compared Interestingly, however, in the latter motif, the sequence of the *Eco*124/3 R subunit encompassed the remarkable replacement of the otherwise fully conserved throughout superfamily II Gly residue by Ser The signature 'S/TR' is typical of the respective segment in superfamily I (Fig 2; [16,17]) Thus the structure of this segment in *Eco*124/3 might be considered intermediate between the two helicase superfamilies

In accord with the previous observations [3–5], and quite unexpectedly in view of their functional similarity, the R subunit sequences had little in common with each other, besides the conserved helicase motifs Moreover, despite the more or less uniform spacing of the motifs themselves, the relative locations of the putative helicase domains encompassing them were very different in the three R subunits (Fig 3). Apparently, these dissimilarities could account for the previously published erratic alignment of two R subunit sequences, in which neither of the helicase motifs (unnoticed at the time) matched in the compared sequences [5]

3.3 *The putative helicase activity of the R subunits may be required for local unwinding of double-stranded DNA*

These observations allow us to hypothesize that the R subunits of type-I and type-III R-M enzymes may possess helicase activity The most obvious role for the putative helicase in restriction is local unwinding of DNA in the cleavage sites In type-I enzymes the putative helicase activity might also be involved in DNA translocation It is to be noted that, although type-III enzymes appear to cleave DNA in the absence of ATP hydrolysis, this cleavage is never complete [1], suggesting that, for the optimal activity, the helicase might be required Type-II restriction enzymes, which function separately from the respective M systems, encompass no putative helicase motifs (unpublished observations) This can be tentatively connected with the fact that the recognition/cleavage sites for type-II enzymes are palindromic [18], whereas the cleavage sites for type-I and type-III enzymes reveal no symmetry [2] It can be speculated that cleavage by type-II endonucleases occurs at transient cross-structures in DNA, requiring no local unwinding, and consequently no helicase. Experimental search for the helicase activity in type-I and type-III R enzymes will probably shed new light on the mechanisms of their action

←
Fig 1 Alignment of *E coli* HsdR protein sequence with those of the 'DEAD' family helicases and RecQ protein The alignment is composed of two large stretches, each aligned by program OPTAL, and separated by a spacer where the sequences could not be aligned because of an approximately 50 amino acid residue insert in the HsdR protein (the numbers of residues in the spacer segments are indicated) The scores for each of the aligned segments were over 6 9 SD, which is indicative of a genuine relationship ([13], and references therein) It has been shown previously that *E coli* RecQ protein is closely related to the 'DEAD' helicase family [8] and exerts helicase activity [15] Asterisks = amino acid residues identical or similar in HsdR and the proteins of 'DEAD/RecQ' family (one exception allowed), plus signs = conserved residues in the proteins of the latter family, all residues identical or similar in HsdR and any of the other sequences are highlighted by boldface The grouping of amino acid residues was as follows D,E,N,Q, S,T, K,R; I,L,V,M,F,C; F,Y,W

			I		IA		II	
1	MOUSE EIF4AI	70	DVIAQAQSGTGKTATFAIS	11	TQALVLAPTRELAQQI	61	MFVLDEAd	23
2	YEAST MSS	144	DVIAPAKTGTGKTFAFLIP	15	VKVIVAAPTRDLALQI	66	ykVLDEAd	30
3	E.COLI SrmB	42	DVLgSAPTqAGKTAAYLLP	15	PRILILTPTRELAMQV	60	tLILDEAd	23
4	E.COLI RecQ	43	DCLVVMPTGGGKSLCYQIP	5	GLtVVVsPLiSLMKDQ	61	LLAVDEAH	27
5	M.LUT UvrB	42	DVVLMGATGTGKSAT--tA	7	RPtLVMVqNKTLAAQL	261	LLVVDEsH	46
6	YEAST PRP16	370	VVIIIGETGSGKTTQLAqD	11	KSIVVTqPRRVAAISV	55	CVIIDEAH	24
7	YEAST Rad3	36	NsILEMPSGTGKTVSL-LS	11	*RKIIyCsrTmSeIEKA*	148	IVIFDEAH	217
8	K2P1 P4	55	SLIVCYDvGLGKTYAAACL	8	*FKVLyLsnSlNsIDNF*	43	LIILDEvH	23
9	T5BP D10	102	TCIINGKpGFGKTILALAL	7	*QKtLVICtNtSIREMW*	50	tVIVDEvH	15
10	VV NTPase1	49	SLLLFHETGVGKTMT-TVY	10	*NWAIILLvKKALIEDP*	43	CVIIDEcH	32
11	VV NTPase2	39	SVLLFHImGSGKTIIALLF	7	KKVyILVPNiNILKIF	48	IFIVDEAH	21
12	VZV Gp51	61	VtVVRAPmGSGKTTAL-LE	9	ISVLVVscRRSFTQTL	47	VLILDEvm	27
13	TVMV CI	79	DIILMGAvGSGKSTG--LP	6	GGVLLLePTRPLAENV	51	FIIFDEfH	21
14	YFV NS3	192	TtVLDFHpGAGKTRRF-LP	10	LRtLVLAPTRVVLSEM	48	VIIMDEAH	23
15	BVDV P125	?	FkqITLATGAGKTTE--LP	10	KRVLVLIPLRAAAESV	48	yIFLDEyH	23
Superfamily II consensus			+++ tg GKt + s s		+++++p r + k		++++DEah	
16	P1BP RES	78	VIdVSMETGTGKTYTYTkT	10	NKFIIIVPTISIKAGT	98	FIIIDEpH	19
17	E.COLI HsdR	483	EILLAMATGTGKTRTAIAM	10	KRILFLVdRRSLGEQA	59	CIVVDEAH	19
18	IFIVP1 HsdR	301	GgYIWHTTGSGKTLTSFkA	10	DKVFFVVdRKDLDYQT	57	VFIFDEcH	19
Superfamily I consensus			a g s +++ g aG GKt		++++ ++ +		++++DE+	
19	E.COLI Rep	16	PCLVLAGAGSGKTRVITNK	12	RHIAAVTFTNKAAREM	146	yLLVDEyQ	19
20	E.COLI RecB	17	ErLIEASAGTGKTFTIAAL	20	EELLVVTFTEAAtAEL	307	VAMIDEFQ	21
21	E.COLI RecD	165	IsVISGGpGTGKTTTVAKL	12	CRIrLAAPTGKAAARL	52	VLVVDEAS	19
22	YEAST PIF	252	NIFyTGSAGTGKSILLREM	10	ENVAVtASTGLAACNI	39	ALVVDEIS	29
23	YEAST RadH	29	GLqVTAGpGTGKTKVLTSR	12	RDIIVtTFTNKAANEM	164	hVLVDEFQ	23
24	VZV Gp55	84	VyLISGNAGSGKSTCIQTL	3	IDCIItGSTRVAAQNV	116	VIVIDEAG	33
25	BMV 1a	680	ISMVDGVAGCGKTTAIKDA	3	GEdLVITANRKsAEDV	32	rLLVDEAG	18
26	TMV P126	828	VVLVDGVpGCGKTKEILSR	3	DEdLILVPGKQAAEMI	35	rLFIDEgL	18
27	SNBV nsP2	181	TIgVIGTpGSGKSAIIKST	2	ARdLVtSGKKEnCREI	29	VLyVDEAF	19
28	RUBV ?	?	IrVwNmAAGAGKTTRILAA	2	REdLYVCPTNALLHEI	29	rIyIDEAF	17
29	BNYVV1 P237	937	LeyVKGGpGTGKSFLIRSL	3	IRdLVVAPSIKLrSDy	27	IIFVDEFT	18
30	PVX1 P166	699	ACVIHGAGGSGKSHAIQKA	7	SDItVVLPTNELrLDw	25	IVIFDDyS	19
31	BNYVV2 P42	128	VgIVLGApGVGKSTSIKNL	7	HKMVLCLPFSQLLEGV	24	tMLVDEVT	18
32	PVX2 P25	24	PLVVHAVAGAGKSTALRKL	7	TNhtLgVPDKVsIRTr	12	FAILDEyT	9
33	IBV 'HEL'	?	RttVQGPpGSGKSHFAIGL	6	ARVVFtACSHAAVDAL	52	ILLVDEVS	18

Fig 2 (for legend, see next page)

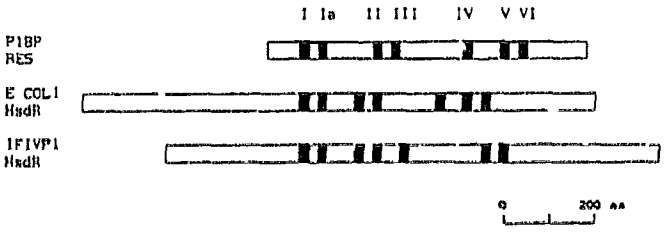

Fig 3 Location of the putative helicase domains in the three R subunits of type-I and type-III restriction-modification enzymes The polypeptide chains are designated by rectangles drawn to scale The seven conserved motifs (see text) are shown

REFERENCES

[1] Yuan, R (1981) Annu Rev Biochem 50, 285–315
[2] Bickle T A (1982) in The Nucleases (Linn S M and Roberts, R J eds) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1982, pp 85–108
[3] Loenen, W A M, Daniel, A S, Braymer, H D and Murray, N E (1987) J Mol Biol 198, 159–170
[4] Humberlin, M, Suri, B, Rao, D N Hornby, D P, Eberle, H, Pripfl, T, Kenel, S and Bickle, T A (1988) J Mol Biol 200, 23–29,
[5] Price, C, Lingner, J, Bickle, T A, Firman, K and Glover, S.W (1989) J Mol Biol 205, 115–125,
[6] Walker, J E, Saraste, M, Runswick, M J and Gay, N J (1982) EMBO J 1, 945–951

	III		IV		V		VI	
1:	VVLLSATMP	56	AVIFINTRRK	40	VLItTDLLARGIDVQQVSLVI	6	NRENYIhRIGRgGR	41
2:	tLLFSATLD	65	AIIFAPTVKF	43	ILVCTDVgARGMDFPNVHEVL	6	ELANYIhRIGRtAR	195
3:	tLLFSATLE	56	sIVFVRnRLE	41	VLVATDVAARGIDIPDVSHVF	6	SGDTYLhRIGRtAR	103
4:	FMALTATAD	52	gIIYCNSRAK	40	IVVATVAFGMGINkPNVRFVV	6	NIESYYQETGRAGR	279
5:	tVyLSATPG	45	VLVtTLTKRM	39	VLVgiNLLREGLDLPEVSLVA	11	STTSLIQTIGRAAR	129
6:	LIITSATMN	56	ILIFMTgQED	58	IIIATNIAETsLTIKGIRYVI	24	SKANADQRsGRAGR	380
7:	VIITSGTIS	72	MVVFFPSYLY	50	AILLSVArGEGIDFQYgRTVL	36	AMRHAAQCLGRVlR	107
8:	ILVITATPM	94	*InAFINSIKE*	71	ILLgSSVLSEsITLYRVKHLh	6	NYGQIKQSIGRAiR	201
9:	kIgLSGTLK	31	*VLIvSDrTEL*	41	LAAAqSIFSEGISLNELSCLI	6	NESLIEQLAGRVqR	59
10:	MICLSATPI	129	*yndFKNSLRD*	109	tCVFSSSgGEGISFFSINDIF	6	NEASLRQIVGRAiR	151
11:	FLLLSGSPI	159	*FkYFINrIQT*	79	FLFsSNIMSEsyTLKEVRHIw	6	TFSQYNQILGRsiR	180
12·	IIAMdATVN	80	ICIFSSTLSF	34	VLVyTTVVTVGLSF-DMAHFh	13	DMVSVYQSLGRVrl	466
13:	IIkVSATPP	43	ILVYVASYNE	40	FIVATNIIENGVTL-DVDVVV	26	SLGERIQRFGRVGR	276
14.	tILMTATPP	38	tAWFLPSIRA	36	FILATDIAEMGANL-CVERVL	27	SASSAAQRrGRIGR	156
15:	VVAMTATPA	48	MLVFVPTRNM	36	VIVATNAIESGVTLPDLDTVI	30	TVGEQAQRrGRVGR	?
	++++sat		+++f t		++++t + g+ + + ++		q +GR+gr	
	tgs		y s		s		a	
16.	IIrygATFS	142	*tLFFIDdIEG*	77	FIFskWTLREGwDnPNVFQIC	6	STTSKLQEVGRglR	433
17.	*LdyVSAyRR*	121	tLVFCVTNAH	52	IVVtvDLLTTGVDIPSICNIV	6	SRILYEQMkGRAtR	243
18:	qFgFTGTPI	32	*VLkFKVdYND*	170	LLIVvGMFLTGFDAPTLNTLF	5	RYHGLMQAFsRtnR	342
	a				s		f g t	
	++++gd Q		+ R		+ ++t+ qG+ + v +++		++yva+sRa	
19:	FTVVGDDDQ	24	VIKLEQNYRS	272	VQLMTLHASkGLEFPYVYMVg	19	ERRLAYVGITRAQK	32
20:	LLLIGDPKQ	22	HYTLDTNWRS	286	VQIVTIHKSkGLEYPLVWLpF	41	DLRLLYVALTRsVW	369
21:	VIFLGDRDQ	40	TGTeAASLRD	183	tWAMTVHKSQGsEFDHaALIL	8	TRELVYtAVTRARR	33
22·	IVCVGDFFQ	27	TIMLQKVFRQ	280	AWsLSIHKSQGqTLPKVKVdL	4	EKGqAYVALSRAVS	117
23:	MTIVGDPDQ	25	TIILVENYRS	329	VTIsTIHGAkGLEWPVVFIpg	49	ERRMFFVAqTRAKY	447
24.	IVCVGsPTQ	42	WAIFINNKRC	463	kLAMTIARSQGLSLEKVAICF	5	RLNsVYVAMSRtVS	36
25:	VLAFGDTEQ	21	RDVVHKTYRC	80	gHIkTVHEAQGISVDNVTLVr	10	HEEyClVALTRhKK	15
26:	AyVyGDTQQ	23	VETrRTTLRC	64	sDVhTVHEVQGeTYSDVSLVr	11	DSPhVlVALSRhTC	30
27:	VVLCGDPMQ	23	TFYkYISRRC	60	hEVMTAAASQGLTrKGVYAVr	11	TSEhVnVlLTRtED	378
28:	VICVGDRDQ	18	TERsRHTWRF	52	IRAyTVREAQGMSVGTaCIhV	11	TRDLAiVsLTRASD	?
29:	IyLVGDEQQ	22	THVpIMNFRN	63	VSktTVRANQGsTYDNVVLpV	9	SAELnlVALSRhRN	926
30·	VILTGDSRQ	26	RYYLNATHRN	48	nDtFTYAGCQGLTkPKVQIVL	7	SANVMYtALSRATD	502
31:	VICFGDPAQ	16	IAECYASRRF	58	IEsIlYSDAhGqTYDVVTIIL	10	DPNVraVILTRARK	30
32:	qALFADPYQ	9	FY-LETSFRV	54	VEFVkPCQVtGLEFKVVTVVs	7	QSTAFYnAITRsK-	9
33:	VVyVGDPAQ	29	DIFLAKCYRC	86	LNVqTVDSSQGsEYDYVIFCV	8	NINrFnVALTRAKR	28

Fig 2 Conserved motifs in the helicase superfamilies I and II and in the R subunits of type-I and type-III R-M enzymes For the sake of clarity, only representative samplings of the sequences of superfamilies I and II were included The conserved segments are designated I to VI according to [8] The numbers of amino acid residues between the segments, and those between the aligned regions and the protein termini are indicated The consensus residues defined as in [8] are indicated, the respective positions in all sequences are highlighted by boldface, and the deviations from the consensus are shown in lower case, plus signs = conserved hydrophobic residues The segments that have been only tentatively identified with the respective motifs are highlighted by italics

[7] Gorbalenya A E and Koonin, E V (1989) Nucleic Acids Res 17, 8413-8440

[8] Gorbalenya, A E , Koonin, E V., Donchenko, A P and Blinov, V M (1989) Nucleic Acids Res 17, 4713-4730

[9] Abousssekhra, A , Chanet, R , Zgaga, Z , Cassier-Chauvat, C , Heude, M and Fabre, F (1989) Nucleic Acids Res 17, 7211-7219.

[10] Burgess, S , Couto, J R and Guthrie, C (1990) Cell 60, 705-717

[11] Dominguez, G , Wang, C -Y and Frey, T K (1990) Virology 177, 225 238

[12] Brodsky, L I , Drachev, A L , Tatuzov, R L and Chumakov, K M (1991) Biopolimery i Kletka 7, 10 14.

[13] Gorbalenya, A E Blinov, V M , Donchenko A P and Koonin, E V (1989) J Mol Evol 28, 256-268

[14] Linder, P , Lasko, P F , Ashburner, M , Leory, P , Nielsen, P J , Nishi, K , Schnier, J and Slonimski, P (1989) Nature 337, 121-122

[15] Umezu, K , Nakayama, K and Nakayama, H (1990) Proc Natl Acad Sci USA 87, 5363-5367

[16] Gorbalenya, A E , Koonin, E V., Donchenko, A P and Blinov, V M (1988) FEBS Lett 239, 16-24.

[17] Hodgman, T C (1988) Nature 333, 22 23; 579

[18] Roberts, R J (1988) Nucleic Acids Res. 16, Suppl , r271-r313