*Nucleic Acids Research, 1993, Vol. 21, No. 3 751–752*

# An expanding family of helicases within the 'DEAD/H' superfamily

Peer Bork[+], and Eugene V.Koonin[1,*,§]
Max-Delbrueck-Centre for Molecular Medicine, 1115 Berlin-Buch, Germany and [1]National Center for Biotechology Information, National Library of Medicine, National Institutes of Health, Bldg. 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA

Received November 16, 1992; Accepted December 16, 1992

'DEAD/H' DNA and RNA helicases constitute a vast superfamily, with both prokaryotes, eukaryotes and viruses encoding numerous members involved in various aspects of genome replication, repair and expression (1, 2). Recently a variety of eukaryotic proteins with similarity to yeast protein Snf2 have been described (3–8), some of them involved in transcription regulation (Snf2, Sth1, YAL001, and Mot1 from yeast, human hSnf2I, and BRM from *Drosophila*), and in DNA recombination and repair (yeast Rad54, Rad16, Rad5, and *Drosophila* LDR). Specific variants of the seven sequence motifs typical of the 'DEAD/H' helicase superfamily (1) are conserved in these proteins.

Using the sequence pattern of the variant of the ATP-binding motif A (helicase motif I) characteristic of this new family (uuxD[DE]uGuGKT; u–a bulky aliphatic residue, i.e. I, L, V, M) to screen the non-redundant amino acid sequence data base (National Center for Biotechnology Information), we found it in only one additional sequence, the putative *E.coli* helicase HepA (9). Alignment using the MACAW program (10) revealed four helicase motifs in the HepA sequence closely resembling the respective segments in the proteins of the new family (Figure). An additional motif shared by all these putative helicases was identified (1b in the Figure). Motifs IV to VI were absent in the published HepA sequence (9). Analysis of the sequence of the product of a long open reading frame (ORF) located downstream from the HepA ORF (11) using TFASTA (12) and TBLASTN (13) programs revealed significant similarity to the helicases of the new family, with obvious counterparts to motifs V and VI (Figure). Thus, we predict that these two ORFs actually comprise a single gene encoding a helicase belonging to the new family. Similarly, a probable frameshift was detected in the LDR gene sequence (14) allowing identification of motifs V and VI in the encoded helicase (Figure).

HepA is the first prokaryotic member of the new helicase family. Very recently, a *Bacillus cereus* gene has been identified encoding another related putative helicase (15). The prokaryotic members greatly increased the sequence diversity within the new helicase family and allowed a better assessment of the conserved blocks. The highest conservation could be assigned to the block including motifs V and VI (Figure). Interestingly, these motifs were shared with a Chilo iridescent virus protein that is likely to be yet another member of the new helicase family (Darai *et al.*, in preparation).

We have described here a rapidly growing family of (putative) helicases whhin the 'DEAD/H' superfamily that included eukaryotic, prokaryotic and viral members. The signatures conserved in motifs V and VI are unique identifiers allowing the convenient separation of the members of this family from other 'DEAD/H' helicases. The possibility of a common biological function for the helicases of the new family is suggested by the abundance of transcription regulators among them.

This paper is the result of two independent, nearly identical studies.

## ACKNOWLEDGEMENTS

We are grateful to Amos Bairoch and Kenn Rudd for helping us to combine these efforts, and to Gholamreza Darai and Anne-Brit Kolsto for communicating results prior to publication.

## NOTE ADDED IN PROOF

While this paper was being processed for publication, the sequence of human DNA repair gene ERCC6 encoding yet another putative helicase of the family described here, which is specifically involved in repair of actively transcribed gene, has been reported (Troelstra,C. *et al.* (1992) *Cell* **71**, 939–953). These authors also noticed the probable frameshift in the LDR gene.

## REFERENCES

1. Gorbalenya,A.E. *et al.* (1989) *Nucleic Acids Res.* **17**, 4713–4730.
2. Schmid,S.R. and Linder,P. (1992) *Mol. Microbiol.* **6**, 283–292.
3. Davis,J.L. *et al.* (1992) *Mol. Cell. Biol.* **12**, 1879–1892.
4. Laurent,B.C. *et al.* (1992) *Mol. Cell. Biol.* **12**, 1893–1902.
5. Johnson,R.E. et al. (1992) *Mol. Cell. Biol.* **12**, 3807–3818.
6. Okabe,I. *et al.* (1992) *Nucleic Acids Res.* **20**, 4649–4655.
7. Clark,M.W. *et al.* (1992) *Yeast* **8**, 133–145.
8. Mannhaupt,G. *et al.* (1992) *Yeast* **8**, 397–408.
9. Lewis,L.K. *et al.* (1992) *J. Bacteriol.* **174**, 3377–3385.
10. Schuler,G.D. *et al.* (1991) *Proteins Struct. Funct. Genet.* **9**, 180–190.
11. Yura,T. *et al.* (1992) *Nucleic Acids Res.* **20**, 3305–3308.

* To whom correspondence should be addressed
+ Present address: European Molecular Biology Laboratory, Heidelberg, Germany
§ On leave from Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia

12. Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
13. Altschul,S.F. *et al.* (1990) *J. Mol. Biol.* **215**, 403–410.
14. Girdham,C.H. and Glover,D.M. (1991) *Gen. Dev.* **5**, 1786–1799.
15. Kolstø,A.-B. *et al.* (1993) *J. Mol. Biol.* (in press).
16. Tsuchiya,E. *et al.* (1992) *EMBO J.* **11**, 4017–4026.

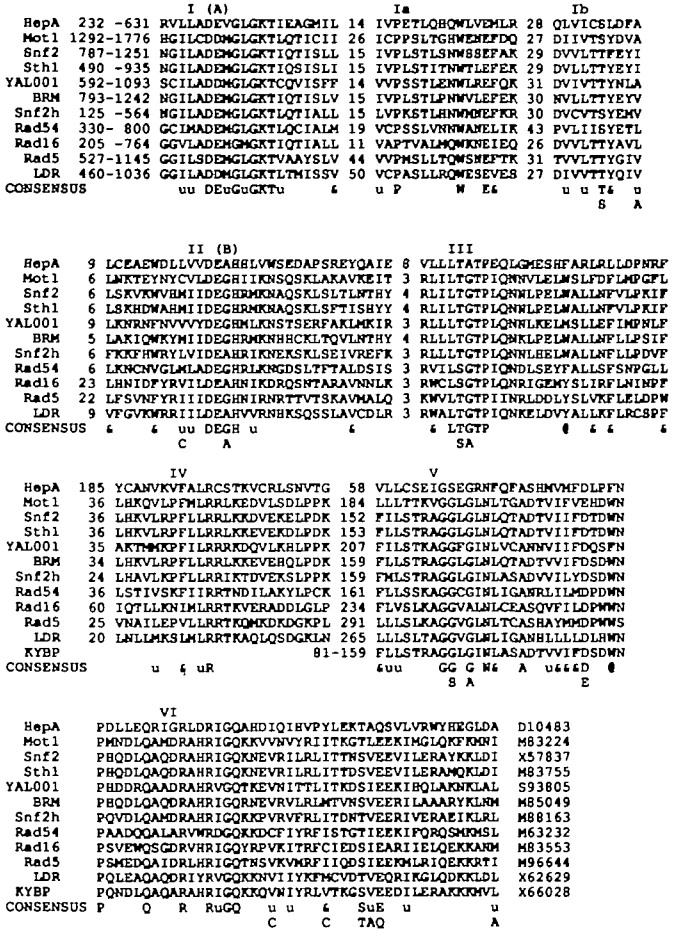

**Figure 1.** Conserved sequence blocks in the new family of helicases. Only the blocks that show the highest conservation are presented, with their boundaries adjusted using the MACAW program so as to achieve the maximal possible statistical significance. The probability that the observed similarity is due to chance was below $10^{-6}$ for each of the aligned blocks. The designation of the conserved motifs is according to ref. 1. The alignment differs in part from those in refs. 3–9 allowing identification of additional conserved amino acid residues, e.g. Pro and Trp in motif 1a. The number of amino acid residues between the conserved blocks and the positions of the aligned regions in the proteins are indicated. The 'consensus' shows the invariant amino acid residues and conserved residue properties: u — a bulky aliphatic residue (I, L, V, M), @ — an aromatic residue (F, Y, W), and & — a bulky hydrophobic residue (either aliphatic or aromatic). In the HepA sequence a frameshift is assumed between motifs III and IV, resulting in protein consisting of 969 amino acid residues instead of 529. In the LDR sequence the frameshift is between motifs IV and V yielding a protein of 1059 residues instead of a 974, the sequence of Sth1 appears to be identical to that of the newly reported yeast protein Nps1 (16). KYBP is the product of a partially sequenced murine gene. Snf2h is the purported human homologue of Snf2. The sequences were taken from GenBank; the accession number is indicated for each sequence. S.c. — *Saccharomyces cerevisiae*, D.m. — *Drosophila melanogaster*; E.c. — *Escherichia coli*.