# *Escherichia coli dinG* gene encodes a putative DNA helicase related to a group of eukaryotic helicases including Rad3 protein

Eugene V.Koonin
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bldg. 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA

Received January 15, 1993; Accepted January 28, 1993

*DinG* is an *Escherichia coli* DNA damage-inducible gene (1, 2). Sequence database searches (3, 4) revealed marginally significant similarities between the 637 amino acid DinG protein, yeast DNA helicases ChlI, Rad3 and Rad 15, and the human helicase ERCC2. Remarkably, the similarities included some of the motifs conserved in the 'DExD/H' helicase superfamily (5). However, DinG sequence did not contain the A motif of the purine NTP-binding pattern, which is conserved in all helicases (5). This motif was found in the sequence of an unidentified (ORF) upstream of *dinG*. Thus the two ORFs may encode a single protein consisting of 752 instead of 637 amino acid residues. Extended variants of the seven motifs typical of the 'DExD/H' helicase superfamily and two other conserved motifs were identified in DinG and the related proteins (Fig. 1). Statistical significance of the alignment was confirmed by two independent methods and four sequence fingerprints that are unique to this set of (putative) helicases were derived (see legend to Fig. 1), indicating that they comprise a distinct group within the 'DExD/H' superfamily including both eukaryotic and prokaryotic members.

Rad3 is a DNA helicase involved in the repair of damaged DNA and in chromosome DNA replication (6); similar functions have been reported for ChlI, Rad15, and ERCC2 (7–9). This is compatible with a function of the putative DinG DNA helicase in the repair and perhaps also in the replication of the *E.coli* chromosome.

Very recently the sequence of the region including the *dinG* gene has been reported from an independent study (H.Ohmori; GenBank L02123). This sequence confirmed that a frameshift existed in the upstream portion of the original sequence (2) but was itself frameshifted between the regions coding for motifs II and III. Combining the two sequences allowed a reconstruction of the amino acid sequence of the putative DinG helicase (submitted to SWISSPROT).

## ACKNOWLEDGEMENT

I am grateful to Kenn Rudd for valuable comments.

## REFERENCES

1. Lewis,L.K., *et al.* (1991) *J. Bacteriol.* **173**, 3377–3385.
2. Lewis,L.K. and Mount,D.W. (1992) *J. Bacteriol.* **174**, 5110–5116.
3. Altschul,S.F. *et al.* (1990) *J. Mol. Biol.* **215**, 403–410.
4. Henikoff,S. and Henikoff,J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
5. Gorbalenya,A.E *et al.* (1989) *Nucl. Acids Res.* **17**, 4713–4730.
6. Sung,P. *et al.* (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8951–8955.
7. Gerring,S.L., Spencer,F. and Hieter,P. (1990) *EMBO J.* **9**, 4347–4358.
8. Murray,J.M. *et al.* (1992) *Nucl. Acids Res.* **20**, 2673–2678.
9. Weber,C.A. *et al.* (1990) *EMBO J.* **9**, 1437–1447.
10. Schuler,G.D., *et al.* (1991) *Proteins Struct. Funct. Genet.* **9**, 180–190.
11. Gorbalenya,A.E. *et al.* (1989) *J. Molec. Evol.* **28**, 256–268.

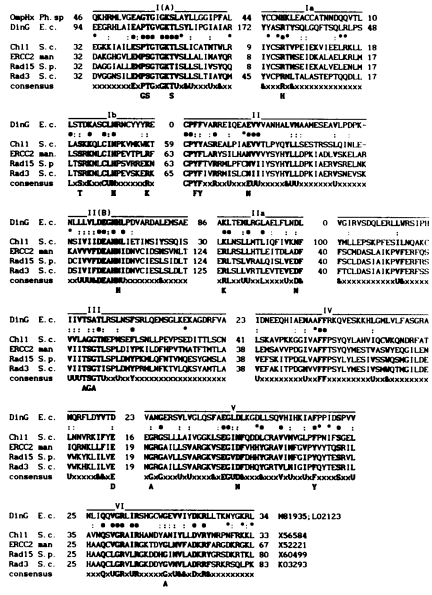

**Figure 1.** Conserved sequence blocks in DinG and related eukaryotic helicases. The alignment was generated using the MACAW program (10). The conserved motifs are designated as in ref. 5. Motifs Ib and IIa were specific for this group of proteins. For each of the blocks, the probability of all five sequences matching by chance was below $10^{-7}$. This may be an overestimate due to the high similarity between Rad3, Rad15, and ERCC2 (10). A more conservative estimate obtained by evaluating the triple alignment of DinG with ChlI and ERCC2 yielded values of $1.7\times10^{-9}$ for block II, $3.5\times10^{-5}$ for block I, and $1.3\times10^{-3}$ for block VI. In the other blocks the alignment of these three sequences was not significant at the 99 percent level. In an independent test the pairwise alignment of DinG and Chl1 in the region spanning motifs II through VI scored 7.1 standard deviations above the random expectation, suggesting that the similarity is unlikely to be due to chance (11). The following sequence signatures were found to be unique for the Rad3-related family of (putative) helicases (U—a bulky aliphatic residue, x—any residue; alternate residues are bracketed): UEx[PG][TS]GxGK[TS]U[ST]xU (motif I), Ux[TS]x$Kx_2$CU[HN]$x_6$[KR] (motif 1a), CP[FY][FY]$x_2$R$x_2$U (conserved region upstream of motif II), and QxUGRxUR$x_7$U[UFY][UF]xDxR (motif VI). Asterisks show identical amino acid residues and colons show similar residues in DinG, and Chl1. The 'consensus' shows the amino acid residues conserved in all of the aligned sequences; @ — an aromatic residue (F, Y, W), & —a bulky hydrophobic residue (aliphatic or aromatic). OmpX is a product of a partially sequenced unidentified ORF upstream of the *ompH* gene of *Photobacterium sp* (GenBank X67094), which was retrieved by a database search for sequences similar to DinG, and is likely to be another bacterial member of this family of helicases. The sequences were from GenBank; the accession number is indicated for each sequence. E.c. — *Escherichia coli*, S.c. — *Saccharomyces cerevisiae*, S.p — *Schizosaccharomyces pombae*, P.sp. — *Photobacterium sp.*