DnaC protein contains a modified ATP-binding motif and belongs to a novel family of ATPases including also DnaA

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

Submitted February 21, 1992

DnaC protein is required for initiation of DNA replication at oriC and for primosome assembly in *E.coli* and is thought to ensure the delivery of DnaB helicase to the replication fork (1, 2). This process is ATP-dependent (1). Inspection of the amino acid sequence of DnaC revealed the presence of a sequence resembling the A motif of the purine NTP-binding pattern (3, 4), with the exception that DnaC contained Asn instead of Ser or Thr in the GKT/S signature. Screening the Non-redundant amino acid sequence data base (NRDB; National Center for Biotechnology Information) using the program BLASTP (5) and the DnaC sequence as the probe showed marginally significant similarity with several NTP-binding pattern-containing proteins, with the highest score observed with the product of an unassigned reading frame (ORF301) adjacent to the DnaB gene of *Bacillus subtilis*. Recently I have reported that ORF301 protein sequence is distantly related to those of DnaA proteins (6). In addition, a BLASTP search with the ORF301 sequence showed a significant similarity to the ISTB proteins encoded by *E.coli* insertion sequence IS21 and its homologue from *Bacillus thuringiensis*. I performed a more detailed comparative analysis of the whole set of these (putative) NTPases.

When aligned using the program OPTAL (7), the DnaC sequence scored 7.7 standard deviations (SD) above the random expectation with ORF301 protein, 8.7 SD with two aligned ISTB sequences, and 8.9 SD with six aligned DnaA sequences, indicative of comparable, and convincing, level of similarity to each of these proteins. Inspection of the resulting alignment confirmed this, showing identical spacing of the A and B motifs of the NTP-binding pattern and revealing additional patches of conservation (Figure). Search of the NRDB for the signature UUUxGx_2GxGKT/NHL (U — bulky hydrophobic residue, x — any residue) that is conserved in the A motif of all this set of sequences except ORF301 protein failed to reveal it in any other protein. I conjecture that these proteins comprise a novel family of related ATPases involved in different steps of DNA replication initiation.

These findings show that DnaC contains a modified purine NTP-binding pattern and suggest that it may possess an ATPase activity that might be required for binding to the replication fork and/or for the release of DnaC from the DNA-protein complex upon the delivery of DnaB (2). Furthermore, it appears that DnaA and DnaC proteins that function at two consecutive steps of replication initiation (1) may share a common ancestry.

REFERENCES

1. McMacken,R. *et al.* (1987) In Neidhardt,F.C., Ingraham,J.L., Low,K.B., Magasanik,B., Schaechter,M. and Umbarger,H.E., (eds), *Escherichia coli and Salmonella typhimurium*. ASM, Washington, D.C., pp. 564–612.
2. Allen,G.C. and Kornberg,A. (1991) *J. Biol. Chem.* **266**, 22096–22101.
3. Walker,J.E. *et al.* (1982) *EMBO J.* **1**, 945–951.
4. Gorbalenya,A.E. and Koonin,E.V. (1989) *Nucleic Acids Res.* **17**, 8413–8440.
5. Altschul,S.F. *et al.* (1990) *J. Mol. Biol.* **215**, 403–410.
6. Koonin,E.V. (1992) *Nucleic Acids Res.* in press.
7. Gorbalenya,A.E. *et al.* (1989) *J. Mol. Evol.* **28**, 256–268.

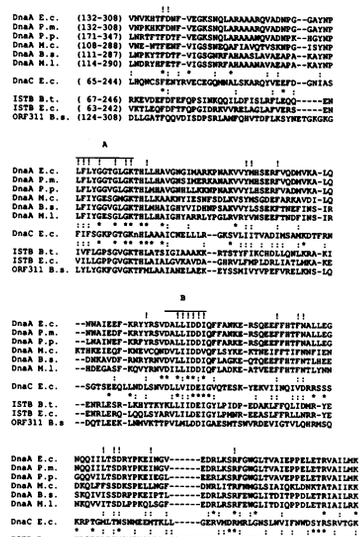

Figure 1. Alignment of the amino acid sequences of DnaC and other proteins of the new family of (putative) DnaA-related ATPases. The boundaries of the aligned segments in each protein are indicated in parentheses. Asterisks denote identical, and colons similar amino acid residues in DnaC and at least five out of six DnaA proteins, or in DnaC and at least two sequences of the group of three proteins comprised by ORF311 product and ISTB. Exclamation marks — identical or similar residues in all aligned sequences. The following groups of similar residues were considered: 1) G,A; 2) S,T; 3) D,E,N,Q; 4) K,R; 5) I,L,V,M,F,Y,W. The two sequence motifs constituting the NTP-binding pattern are delineated. The N residue in the DnaC sequence that replaces the otherwise conserved T/S in the A motif is shown in lower case. E.c., — *E.coli*, P.m. — *Proteus mirabilis*, P.p. — *Pseudomonas putida*, M.c. — *Mycoplasma capricoli*, B.s. — *Bacillus subtilis*, M.l. — *Micrococcus luteus*. The sequences were from PIR protein sequence bank (Release 30.0).

*On leave from Institute of Microbiology, Academy of Sciences, 117811 Moscow, Russia