Nucleic Acids Research, 1993, Vol. 21, No. 1 1–3

Homology in accessory proteins of replicative polymerases—*E.coli* to humans

Mike O'Donnell[1,2], Rene Onrust[1], Frank B.Dean[3], Mei Chen[3] and Jerard Hurwitz[3]
[1]Microbiology Department and [2]Howard Hughes Medical Institute, Cornell University Medical College, 1300 York Avenue, New York, NY 10021 and [3]Program in Molecular Biology, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Research Center, 1275 York Avenue/Box 97, New York, NY 10021, USA

Received November 4, 1992; Accepted November 19, 1992 GenBank accession nos*

The basis for the remarkably high processivity of DNA polymerases that duplicate long chromosomes appears quite similar in prokaryotes and eukaryotes. In each of these cell types, the replicative polymerase has several accessory proteins which endow the polymerase subunit with its speed and processivity. The replicative polymerases of the well studied systems of bacteriophage T4, *E.coli* (DNA polymerase III holoenzyme) and humans (polymerase δ), contain accessory proteins which form a 'sliding clamp' on DNA that acts to tether the polymerase to the DNA for rapid and highly processive synthesis (1–4). In the *E.coli* system this sliding clamp has been shown to be a dimer of the β subunit, which is in the shape of a ring encircling the DNA (5). The functional homologue of β in the T4 system is the product of gene 45 (g45 protein) and in humans it is the proliferating cell nuclear antigen (PCNA). These proteins are homologous in function, and although they show no homology at the amino acid sequence level, a case has recently been made for a structural similarity of β to PCNA and to the T4 g45 protein on the basis of sequence using the crystal structure of β as a guide (5).

These replicative polymerases each require several other accessory proteins to assemble the sliding clamp around the DNA. A list of these proteins is presented in Table I. The requirement for more than one protein in this reaction may reflect a need for several functions, including recognition of a primed template, the opening and closing of the ring shaped clamp protein around the DNA, and the coupling of this process to ATP hydrolysis. In the *E.coli* system this 'clamp loader' is the 5-protein γ complex ($\gamma\delta\delta'\chi\psi$), in humans it is the 5-protein Activator 1 (A1), also referred to as RF-C, and in phage T4 it is the 5 subunit g44 protein/g62 protein complex (the g44/62 complex).

The mechanism of the 'clamp loader' in these three systems may be similar. The sites of binding to the primer-template junction are similar for the *E.coli* γ complex and β subunits, the human A1 and PCNA, and the g44/62 and g45 proteins (8–11). In the case of *E.coli*, the loading of the β subunit forms a 'preinitiation complex', dependent upon ATP hydrolysis, that can be isolated by gel filtration (12,13). The action of the γ complex in this reaction can be substituted by pairs of the subunits, γδ, or τδ′ (14). The preinitiation complex assembled by the yeast or human A1 and PCNA and dependent upon ATP hydrolysis can also be isolated by gel filtration (7,15). The complex assembled by the T4 proteins on primed DNA differs, in that while it can be cross-linked to the DNA, it is not stable enough to be isolated after gel filtration; a complex containing the g45 can only be isolated by gel filtration if the DNA polymerase (g43) is present with the g44/62 and ATP (9,16).

The ATPase activity of the multisubunit complex in each of the three systems is required to assemble the 'sliding clamp' and the polymerase onto the template, but is not required for the subsequent replication of long stretches of the template in vitro. This indicates that ATP hydrolysis by the accessory protein complex is not required for translocation by the polymerase, at least along a template unencumbered by other bound proteins.

Recently, the genes encoding most of these subunits have been identified. Comparison of the amino acid sequences in Fig. 1 shows several of these subunits are truly homologous between *E.coli*, humans and phage T4, especially in one region toward the middle (110–160), implying for the first time that the functional similarity among these systems has its basis in an evolutionarily conserved structure. Presumably all these proteins evolved from the same ancestral gene.

The homology between these protein subunits actually starts near the amino terminus, continues through the ATP-binding domain which starts approximately at the site of the 'GKT' box, and extends for about two hundred amino acid residues toward their carboxyl termini. All these proteins have a strictly conserved 'SRC' sequence of unknown function at residues 157–159. The presence of the four amino-acid 'DEAD' motif in three of the human protein subunits represent an unusual occurrence of this

Table I.

	E.coli	Eukaryotic	phage T4
DNA polymerase	Pol III core ($\alpha\epsilon\theta$)†	pol δ (pol ϵ)*	g43
Accessory complex	γ complex ($\gamma\delta\delta'\chi$)	Activator 1 (A1) (aslo called RF-C)	g44/62
Accessory protein clamp	β	PCNA	g45

†The 10th subunit of *E.coli* DNA polymerase holoenzyme, τ, binds together two molecules of pol III core (24, 25) presumably for coordinated synthesis of both strands of duplex DNA.
*Polϵ is a different molecule from polδ and it is also activated by the eukaryotic accessory proteins (6,7).

*L04577, L07540, L07541

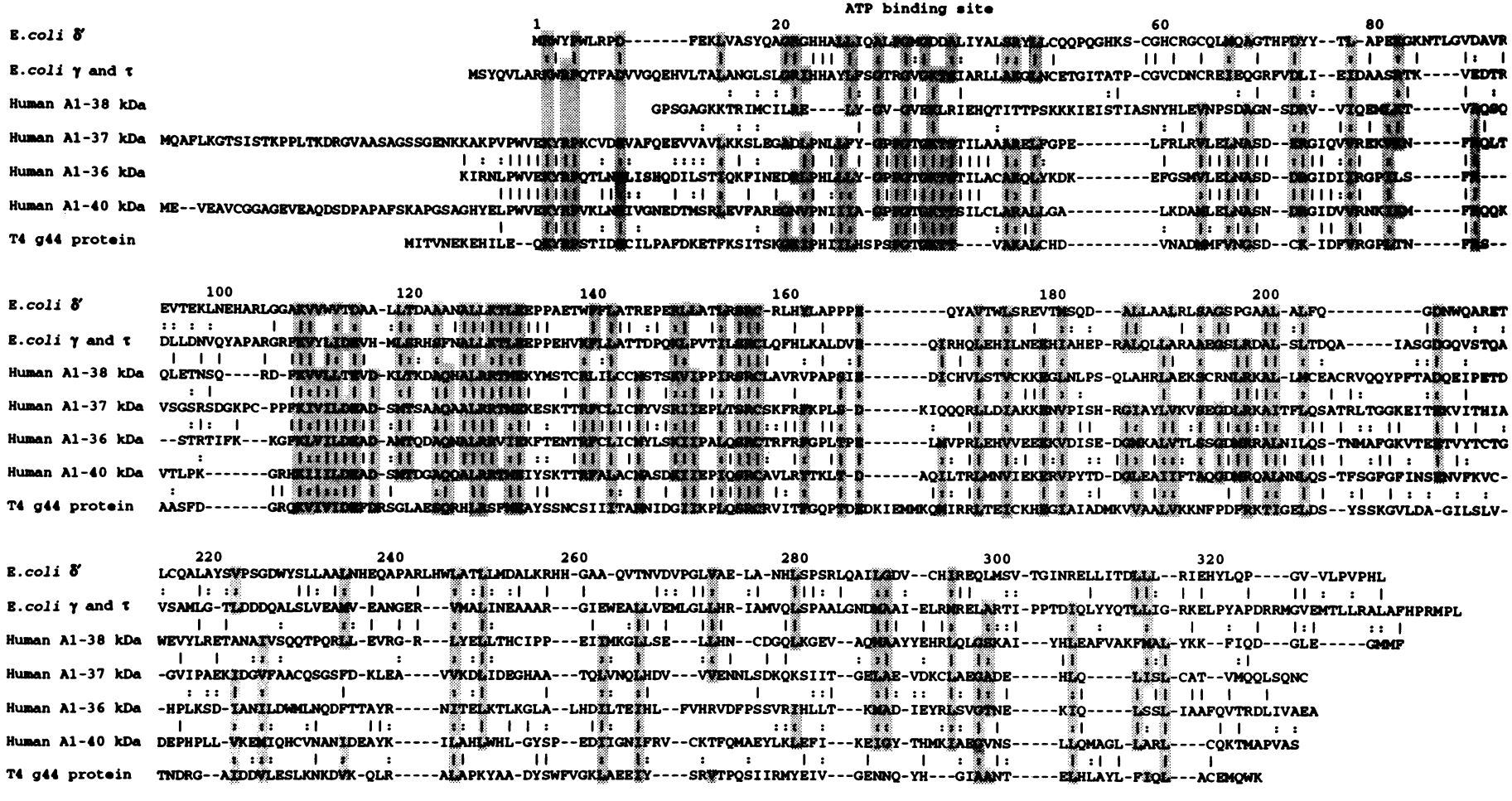

Figure 1. Homology between accessory proteins of replicative polymerases of *E.coli*, Humans, and phage T4. The seven proteins are aligned top to bottom to show maximum homology with proteins directly above and below. Identical residues between neighboring proteins have a vertical dash, and homologous residues have two dots. Gaps were introduced to maximize homology. Numbering of δ' is presented above the sequences. Shaded areas represent identical or conservative amino acids that are aligned in at least five of the seven sequences (most are in six or all seven). The ATP binding site sequence in γ, τ, A1-36 kDa protein, A1-37 kDa protein, A1-38 kDa protein, A1-40 kDa protein, and T4 g44 protein is indicated above the sequence. Only the first 351 amino acids of γ and τ are shown (431 and 643 amino acids total, respectively). Presented for the first time in Fig. 1 are the sequence of *E.coli* δ' (R. Onrust and M. O'Donnell, submitted) and the A1 38 and 36 kDa proteins (some of the N-terminal amino acids of both proteins are yet unknown, F. Dean, M. Chen and J. Hurwitz, unpublished). Other sequences are from the following references: γ and τ (18, 19), A1 40 kDa (20) and 37 kDa proteins (21), and T4 g44 protein (22).

sequence in proteins not having RNA-dependent ATPase or putative RNA helicase activities. However, as mentioned above, ATP hydrolysis by these protein complexes is not required for translocation along the DNA template by the polymerase.

An interesting feature to emerge from the sequence comparisons is the similarity among subunits present within one polymerase accessory factor. In *E.coli*, the δ' subunit shares 27% identity to the γ and τ subunits; γ and τ are derived from the same gene and therefore have identical sequences (a translational frameshift produces γ (47 kDa) which is 24 kDa shorter than τ, ref. 23). The gene encoding δ' (*holB*) also produces two subunits, δ'_{small} and δ'_{large}, which differ by 520 da. δ'_{small} is the product of the full gene and therefore δ'_{large} is produced by a modification, possibly a translational frameshift. In humans, four subunits of the A1 are highly homologous to one another. Hence, the 'clamp loader' complexes of *E.coli* and humans appear to contain a family of subunits of similar structure within them. Although the phage T4 g44 protein is not homologous to g62 protein, the subunit composition of the g44/g62 complex is 4:1 (g44-to-g62) (17). Thus, the g44/62 complex also has structural redundancy and may be compared to the case of the human proteins, in which four different proteins having homology, the 40, 38, 37, and 36 kDa subunits, are present in the complex with one additional subunit, of 145 kDa, that has little homology to the other four.

Why there is such structural redundancy within these complexes is unclear at present. However, these sequence alignments underscore the basic similarity in structure and function of the cellular replicase accessory proteins across the evolutionary spectrum.

REFERENCES

1. Stukenberg, P.T., Studwell, P.S. and O'Donnell, M. (1991) *J. Biol. Chem.* **266**, 11328–11334.
2. Hurwitz, J., Dean, F.B., Kwong, A.D. and Lee, S.-H (1990) *J. Biol. Chem.* **265**, 18043–18046.
3. Nossal, N.G., and Alberts, B.M. (1983) In Bacteriophage T4 (Mathews, C.K., Kutter, E.M., Mosig, G., and Berget, P.B., eds) pp. 71–81, American Society for Microbiology, Washington, D.C.
4. Kornberg, A. and Baker, T. (1991) DNA Replication. New York, W.H. Freeman.
5. Kong, X.-P., Onrust, R. O'Donnell, M. and Kuriyan, J. (1992) *Cell* 69, 425–437.
6. Lee, S.-H, Pan, Z.-Q., Kwong, A.D., Burgers, P.M.J. and Hurwitz, J. (1991) *J. Biol. Chem.* **266**, 22707–22717.
7. Burgers, P.M.J. (1991) *J. Biol. Chem.* **266**, 22698–22706.
8. Munn, M.M. and Alberts, B.M. (1991) *J. Biol. Chem.* **266**, 20024–20033.
9. Capson, T.L., Benkovic, S.J. and Nossal, N.G. (1991) *Cell* 65, 249–258.
10. Tsurimoto, T. and Stillman, B. (1991) *J. Biol. Chem.* **266**, 1950–1960.
11. Griep, M.A. and McHenry, C.S. (1992) *J. Biol. Chem.* **267**, 3052–3059.
12. Wickner, S.H. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3511–3515.
13. O'Donnell, M. (1987) *J. Biol. Chem.* **262**, 16558–16565.
14. O'Donnell, M. and Studwell, P.S. (1990) *J. Biol. Chem.* **265**, 1179–1187.
15. Lee, S.-H. and Hurwitz, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5672–5676.
16. Richardson, R.W., Ellis, R.L. and Nossal. N.G. (1990) *UCLA Symp. Mol. Cell. Biol.* 127, 247–259.
17. Jarvis, T.C., Paul, L.S. and von Hippel, P.H. (1989) *J. Biol. Chem.* **264**, 12709–12716.
18. Flower, A.M. and McHenry, C.S. (1986) *Nuc. Acids Res.* **14**, 6541–6549.
19. Blinkowa, A. and Walker, J. (1986) *Nuc. Acids Res.* 14, 6541–6549.
20. Chen, M, Pan, Z.-Q and Hurwitz, J. (1992) *Proc. Natl. Acad. Sci USA* **89**, 2516–2520.
21. Chen, M., Pan, Z.-Q and Hurwitz, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5211–5215.
22. Spicer, E.K., Nossal, N.G. and Williams, K.R. (1984) *J. Biol. Chem.*, **259**, 15425–15432.
23. Tsuchihashi, Z. and Kornberg, A. (1990) *Proc. Natl. Acad. Sci. USA* 87, 2516–2520.
24. McHenry, C.S. (1982) *J. Biol. Chem.* 257, 2657–2663.
25. Studwell-Vaughan, P.S. and O'Donnell, M (1991) *J. Biol. Chem.* **266**, 19833–19841.