

The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: Implications for a dual role as a primer recognition factor and processivity clamp

LI FAN, PAUL C. SANSCHAGRIN, LAURIE S. KAGUNI*, AND LESLIE A. KUHN*

Department of Biochemistry, Michigan State University, East Lansing, MI 48824-1319

Communicated by I. Robert Lehman, Stanford University School of Medicine, Stanford, CA, June 18, 1999 (received for review March 7, 1999)

ABSTRACT The accessory subunit of the heterodimeric mtDNA polymerase (pol γ) from *Drosophila* embryos is required to maintain the structural integrity or catalytic efficiency of the holoenzyme. cDNAs for the accessory subunit from *Drosophila*, man, mouse, and rat have been identified, and comparative sequence alignment reveals that the C-terminal region of about 120 aa is the most conserved. Furthermore, we demonstrate that the accessory subunit of animal pol γ has both sequence and structural similarity with class IIa aminoacyl-tRNA synthetases. Based on sequence similarity and fold recognition followed by homology modeling, we have developed a model of the three-dimensional structure of the C-terminal region of the accessory subunit of pol γ . The model reveals a rare five-stranded β -sheet surrounded by four α -helices with structural homology to the anticodon-binding domain of class IIa aminoacyl-tRNA synthetases. We postulate that the accessory subunit plays a role in the recognition of RNA primers in mtDNA replication, to recruit pol γ to the template–primer junction. A similar role is served by the γ -complex in *Escherichia coli* DNA polymerase III, and indeed our accessory subunit model shows structural similarity with the N-terminal domain of the δ' subunit of the γ -complex. Structural similarity is also found with *E. coli* thioredoxin, the accessory subunit and processivity factor in bacteriophage T7 DNA polymerase. Thus, we propose that the accessory subunit of pol γ is involved both in primer recognition and in processive DNA strand elongation.

mtDNA polymerase (pol γ) is the key enzyme involved in the replication of the mtDNA genome, and its catalytic features appear to be conserved from yeast to man. In studying *Drosophila* as an animal model of mitochondrial function, we have shown that *Drosophila* pol γ is a heterodimer of catalytic and accessory subunits of 125 and 35 kDa, respectively (1, 2). The large catalytic subunit contains both 5' \rightarrow 3' DNA polymerase and 3' \rightarrow 5' exonuclease activities (3). Typical of replicative DNA polymerases, the two-subunit *Drosophila* pol γ is both highly processive and highly accurate in nucleotide polymerization (4–6). With the high processivity determined in *in vitro* biochemical studies, *Drosophila* pol γ could replicate the entire mtDNA molecule on binding of the initiating primer on each DNA strand. Thus, the recruitment of pol γ to the initiating primer and maintenance of enzyme processivity may be critical for efficient and faithful mtDNA replication.

To elucidate structure–function relationships in *Drosophila* pol γ , we have cloned cDNAs for its subunits (3, 7). The catalytic subunit shares a high degree of sequence similarity with other DNA polymerase catalytic cores, both from the pol γ family (3, 8) and from other members of the family A class of DNA polymerases that includes the two-subunit bacterio-

phage T7 DNA polymerase (9). In contrast, we found that the cDNA for the accessory subunit encodes a novel protein (7). We have identified human, mouse, and rat homologs of the *Drosophila* cDNA, providing evidence for a common heterodimeric structure among animal pol γ s (7).

In parallel with biochemical studies of pol γ and its isolated subunits, we have pursued molecular modeling of the accessory subunit. Here we report a striking structural similarity between the C-terminal region of the accessory subunit of *Drosophila* pol γ and those of several class IIa aminoacyl (aa)-tRNA synthetases (RSs). That this domain has been shown to be involved in tRNA binding in the latter suggests the accessory subunit of pol γ plays a role in recognition of the unusual RNA primers generated at the mtDNA replication origin (10) to guide the catalytic subunit to the primer terminus. In addition, the accessory subunit model shows structural homology with thioredoxin, the accessory subunit of T7 DNA polymerase, despite their lack of amino acid sequence homology. Thus, we predict a dual role for the accessory subunit as a processivity clamp and as a primer recognition factor/clamp loader in mtDNA replication.

MATERIALS AND METHODS

Sequence Analysis of the Accessory Subunit of *Drosophila* pol γ . The SwissProt protein sequence database was searched by using the BLAST algorithm (11) through the National Center for Biotechnology Information web server (<http://www.ncbi.nlm.nih.gov/>) for homologs to the sequence of the accessory subunit of pol γ (7). Sequence alignments were obtained with the GAP and BESTFIT programs of the University of Wisconsin Genetics Computer Group software package (GCG) (12) by using the Needleman and Wunsch algorithm (with gap weight = 5 and length weight = 4). Residues 254–361 of the accessory subunit of pol γ (pol γ - β) were aligned by using GAP to residues 288–386 of the sequence from the crystal structure of the *Thermus thermophilus* prolyl tRNA synthetase (Pro-RS; generously provided by S. Cusack of European Molecular Biology Laboratory, Grenoble, France) (13), corresponding to an independent folding domain in *Escherichia coli* pro-RS identified by the BLAST search as being similar to the pol γ - β sequence. Minor modifications of this alignment were made so that insertions and deletions fell into regions between secondary structures.

Structural Modeling of the Accessory Subunit. Initial prediction of the three-dimensional (3D) fold of the accessory subunit was obtained from the Protein Fold Recognition server (<http://www.doe-mbi.ucla.edu/people/frsvr/preds.html>) based on the algorithm of Fischer and Eisenberg (14). This approach uses both the secondary structure assigned to the sequence by PHDSEC (15) (<http://www.embl-heidelberg.de/>

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: 3D, three dimensional; PDB, Protein Data Bank; RS, tRNA synthetase; aaRS, aminoacyl-RS; pol γ , mtDNA polymerase.

*To whom reprint requests should be addressed. E-mail: kuhn@sol.bch.msu.edu or lskaguni@pilot.msu.edu.

predictprotein/predictprotein.html) and the Gonnet amino acid substitution matrix to assess the similarity between the probe sequence (e.g., the accessory subunit) and sequences from the Protein Data Bank (PDB) (16, 17) of 3D protein structures. The use of both primary and predicted secondary structures in the search enhances its sensitivity in identifying the protein with the most similar 3D fold to that of the probe. When tested on a set of 68 proteins whose 3D structures were known but ignored, the fold recognition server can recognize successfully the 3D fold of a new sequence in 71% of the cases (14).

3D modeling was carried out by using Molecular Simulations' (Waltham, MA) INSIGHTII and BIOPOLYMER software on a Silicon Graphics (Mountain View, CA) Indigo2 graphics workstation. By using the sequence alignment between the accessory subunit of *Drosophila melanogaster* poly and the 2.4-Å crystal structure of residues 288–386 of *T. thermophilus* pro-RS, the backbone of pro-RS was used as a structural template, and side chains differing in poly- β were substituted individually. For two of the three short loops between regular secondary structures that differ between poly- β and pro-RS, the SEARCHLOOP function in BIOPOLYMER was used to identify a loop of known structure with appropriate length, geometry, and packing to use as a structural template; for the third loop, between β -strands 4 and 5, the corresponding loop in the structure of *T. thermophilus* histidyl-RS (PDB entry 1adj) (18) was used as a template, followed by side-chain substitution where necessary. When side-chain substitutions resulted in steric collisions, they were resolved by minimal torsional-angle adjustments in INSIGHTII, followed by energy minimization on the loop regions only, by 100 steps of steepest descents minimization with the CVFF consistent valence force field by using DISCOVER software (Molecular Simulations, Waltham, MA). The stereochemistry of the final structural model of the C-terminal domain of poly- β was validated by using PROCHECK (19, 20), and the favorability of amino acid environments within this structure was assessed by using the 3D Profile

method (21) via the VERIFY3D server (<http://www.doe-mbi.ucla.edu/Services/Verify3D.html>). 3D structural comparisons between the accessory subunit model and the available crystallographic structures of thioredoxin and the δ' subunit of DNA polymerase III were performed with the DALI server (22) (<http://www.embl-ebi.ac.uk/dali>).

RESULTS AND DISCUSSION

Identification of Sequence Homologs of the Accessory Subunit of Poly. We have identified human (7) (GenBank accession no. U94703), mouse (GenBank accession no. AF006072), and rat (GenBank accession no. AA892950) homologs of the *Drosophila* cDNA of the accessory subunit of poly. Comparative sequence alignment reveals that the C-terminal region of ≈ 120 aa is the most conserved (Fig. 1A).

BLAST search in the SwissProt database suggested moderate sequence homology between a region of the accessory, or β , subunit of *Drosophila* poly (poly- β) and aaRSs from both prokaryotes and eukaryotes. The most similar was a probable glycyl-tRNA synthetase from *Methanococcus jannaschii* (SwissProt accession code Q57681), with a score of 84 and likelihood of 0.0018 over two aligned segments, reflecting 31% identity to residues 211–254 of poly- β and 25% identity to residues 290–355. The top 10 matches, with up to 35% identity over ≈ 40 residues, were all members of the class II aaRS family (23). The BLAST sequence alignments (data not shown) between the accessory subunit of *Drosophila* poly and these aaRSs were mostly limited to the C terminus (residues 290–360) of poly- β (Fig. 1B).

Modeling the Accessory Subunit of Poly Based on Structural Homology with tRNA Synthetases. To obtain insights into the structure and function of the accessory subunit of poly, 3D fold recognition (14) was used to identify the best structural match for poly- β , based on the best alignment between its sequence and secondary structures with those of all the distinct structural folds in the PDB. The major benefit

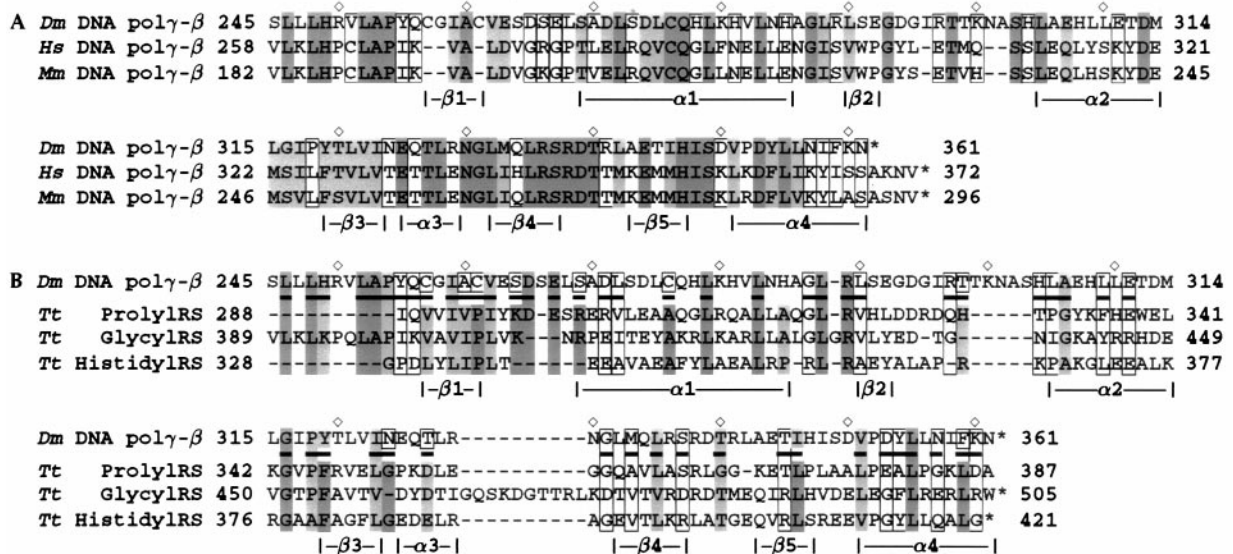


FIG. 1. Sequence comparisons of the accessory subunit of *Drosophila* poly with mammalian homologs and with class IIa aaRSs. (A) Sequence alignment between the C-terminal regions of the accessory subunit of *D. melanogaster* poly (*Dm*) (7) and its homologs from *Homo sapiens* (*Hs*) (7) and *Mus musculus* (*Mm*, GenBank accession no. AF006072). Residues are shaded based on degree of similarity, with dark shading indicating identical residues and light shading indicating conservative substitutions within the categories of the simplification matrix from the UW-GCC package (12): negatively-charged residues and derivatives (Asp, Asn, Glu, Gln); positively charged residues (His, Lys, Arg); small hydrophobic residues (Leu, Ile, Val, Met); large hydrophobic residues (Phe, Tyr, Trp); Cys; and other residues (Pro, Ala, Gly, Ser, Thr). Open boxes indicate loosely conserved residues, i.e., those conserved as hydrophobic or hydrophilic. \diamond marks every 10th residue in the *Drosophila* poly- β sequence. (B) Comparison of the *Drosophila* poly- β sequence with the *T. thermophilus* (*Tt*) prolyl-RS (13), glycyl-RS (PDB ID 1ati) (24), and histidyl-RS sequences (PDB ID 1adj) (18). Sequences are shaded as in A, with underlined residues in poly indicating conservation of hydrophobic/hydrophilic character between the accessory subunit and the RSs.

of the fold recognition server in the case of poly- γ - β was the identification of a structural homolog of the poly- γ - β protein that was not yet included in sequence databases, and therefore not identified by BLAST search. By using the sequence of the accessory subunit, the server indicated 3D structural similarity for poly- γ - β residues 24–359 with residues 73–504 of the glycyl-tRNA synthetase from *T. thermophilus* (PDB entry 1ati) (24). The Z score was 11.5, more than twice the threshold value (4.8 ± 1.0) for significant similarity.

aaRSs are enzymes that catalyze the attachment of each amino acid to its cognate tRNA (23, 25). The 20 aaRS enzymes can be divided into two different families of 10 members each based on their amino acid sequence and structural similarity, and crystallographic structures have been determined for members of both classes (refs. 13 and 23, and refs. therein). Besides Gly-RS, class II aaRSs include Ser-, Thr-, Pro-, His-, Asp-, Asn-, Lys-, Phe-, and Ala-RS. All are homodimers, with the catalytic domain built around a six-stranded antiparallel β -sheet. Most class II aaRSs have three conserved, class-defined motifs: motifs 2 and 3 are responsible for ATP binding and catalysis, whereas motif 1 is involved in dimerization essential for the enzymatic activity. Based on amino acid sequence comparison, Thr-, Pro-, His-, and most Gly-RS (including Gly-RS from *T. thermophilus*) have a homologous C-terminal domain attached to the catalytic core identifying them as members of subclass IIa. This domain interacts with the anticodon loop and stem of the tRNA substrate as elucidated by crystallography (13).

The sequence–structure alignments from fold recognition (data not shown) and sequence alignment (Fig. 1B) reveal that residues 254–360 of poly- γ - β map onto the anticodon-binding domain of class II aaRSs. The match of residues 24–253 of poly- γ - β to regions of the catalytic core of the synthetase proved more difficult to interpret because critical motifs 2 and 3 in the synthetase were not matched, and generally the alignment was substantially fragmented. Two β -strands in the catalytic core of Gly-RS are absent from poly- γ - β , which suggests that it would be difficult for the two to fold similarly in this region. In addition, the motif 1 structure in Gly-RS is missing in poly- γ - β . These results indicate that poly- γ - β lacks the central core structure for catalysis of tRNA aminoacylation.

Although these results suggest that the accessory subunit of poly- γ is not enzymatically homologous to the aaRS family, it is likely to share structural homology with the anticodon-binding domain of the synthetases, based on significant sequence identity in this region. Of the known structures of aaRSs, *T. thermophilus* Pro-RS has the highest sequence identity to poly- γ - β , with 28% identity (42.7% similarity) between the 81 C-terminal residues of *Drosophila* poly- γ - β and residues 302–381, which form the anticodon-binding domain in Pro-RS. The 3D structure of this C-terminal region of poly- γ - β is therefore expected to have a closely superimposable 3D structure with that of the anticodon-binding domain. This expectation is based on exceeding the degree of sequence identity required for structural similarity found from thousands of pairwise sequence and structural alignments between known protein structures (26) in the PDB. Once an appropriate structural template has been identified based on sequence identity, a homology model may be built by using this template for the main chain and by substituting side chains according to the sequence alignment.

We used the anticodon-binding domain of *T. thermophilus* Pro-RS (coordinates provided by S. Cusack, European Molecular Biology Laboratory, Grenoble, France) as the structural template to build a homology model of the 3D structure of the C-terminal domain of the accessory subunit of poly- γ . Amino acid sequence comparison (Fig. 1A) indicated that the C-terminal domain of the accessory subunit of *Drosophila* poly- γ is also the most conserved region relative to its human and murine homologs, supporting its functional importance. Mod-

eling was based on the BESTFIT sequence alignment between poly- γ - β and Pro-RS (Fig. 1B) and adjusted so that deletions or insertions fell in loop regions rather than in regular secondary structures. Residues 288–301 of Pro-RS were aligned manually to poly- γ - β based on the fold recognition alignment between Gly-RS and poly- γ - β . The aligned region shows 23–29% pairwise sequence identity between the aaRS sequences (Fig. 1B). Based on this alignment of Pro-RS to poly- γ - β , we used INSIGHTII to substitute poly- γ - β side chains on the Pro-RS main-chain structure, followed by side-chain and loop optimization (as described in *Materials and Methods*). The main-chain of the resulting model of the C-terminal region of poly- γ - β (residues 255–361; Fig. 2A) is the same as that of the anticodon binding domain of Pro-RS (0.21 Å rms deviation), except for three loop insertions: loop 1 between strand β 1 and helix α 1, loop 2 between β 2 and α 2, and loop 3 between β 4 and β 5 (Fig. 1B). This structural model was validated by using two techniques, PROCHECK, which evaluates stereochemistry (favorable bond lengths, angles, and atomic packing), and VERIFY3D, which assesses the favorability of residue environments within the folded structure. PROCHECK found all residues to have favored or highly favored main-chain dihedral (Φ , Ψ) angles, and the model to have significantly better overall stereochemistry (G factor of 0.07) than typical crystal structures solved at the resolution of the Pro-RS structure (2.4 Å). Evaluation of residue environments with VERIFY3D gave the accessory subunit model structure a score matching those of higher resolution (2.0 Å) crystal structures.

The structural model of the C-terminal domain of the accessory subunit of *Drosophila* poly- γ reveals a rare α/β -fold comprising a five-stranded mixed β -sheet surrounded by four α -helices (Fig. 2A) found only in the anticodon-binding domains of class II aaRSs (13). The anticodon-binding pocket of the class II RSs, which faces the viewer in the lower half of Fig. 2B, is formed predominantly by the central β -sheet and by α -helices 2 and 3 (13); this β -sheet appears at center, with helix 2 appearing as the long helix at right and helix 3 as the short helix at lower left. Whereas residue conservation is often associated with side chains that bind substrates or ligands, and thus the anticodon-binding pocket might be expected to be conserved among aaRSs and between aaRSs and the accessory subunit, this is explicitly not the case. Different tRNA synthetases must recognize different anticodons, and the accessory subunit of poly- γ must recognize at least two distinct primers; thus, their ligand-binding residues are expected to vary to confer this specificity. Of the residues identical in the aligned region between poly- γ - β and Pro-RS (Fig. 1B), half are surface-exposed and half are somewhat or completely buried in the protein and do not colocalize to a specific region of the structure; for the additional residues that are chemically similar but not identical, \approx two thirds are exposed and the remainder are buried. Conserved residues at the surface of this domain in Pro-RS are not associated with an interface with the region N-terminal to this domain because the two domains are connected only by a loose tether in class IIa synthetases. Conserved surface-exposed residues of poly- γ - β (yellow side chains in Fig. 2B) may be involved in interactions with the N-terminal domain of the accessory subunit or with the catalytic subunit or nucleic acid substrates. An RNA-binding structural fold in the C terminus of the accessory subunit corresponding to that in class IIa aaRSs was unanticipated but is consistent with a function of the accessory subunit in interacting with RNA molecules containing tRNA-like structures.

The modeled region of the accessory subunit also includes two previously noted sequence motifs, a putative zinc finger and a putative leucine zipper (7). The segment containing the leucine zipper sequence in the model structure (residues 319–347) consists of one α -helical segment and two short β -strands, and thus it cannot function as a leucine zipper. The

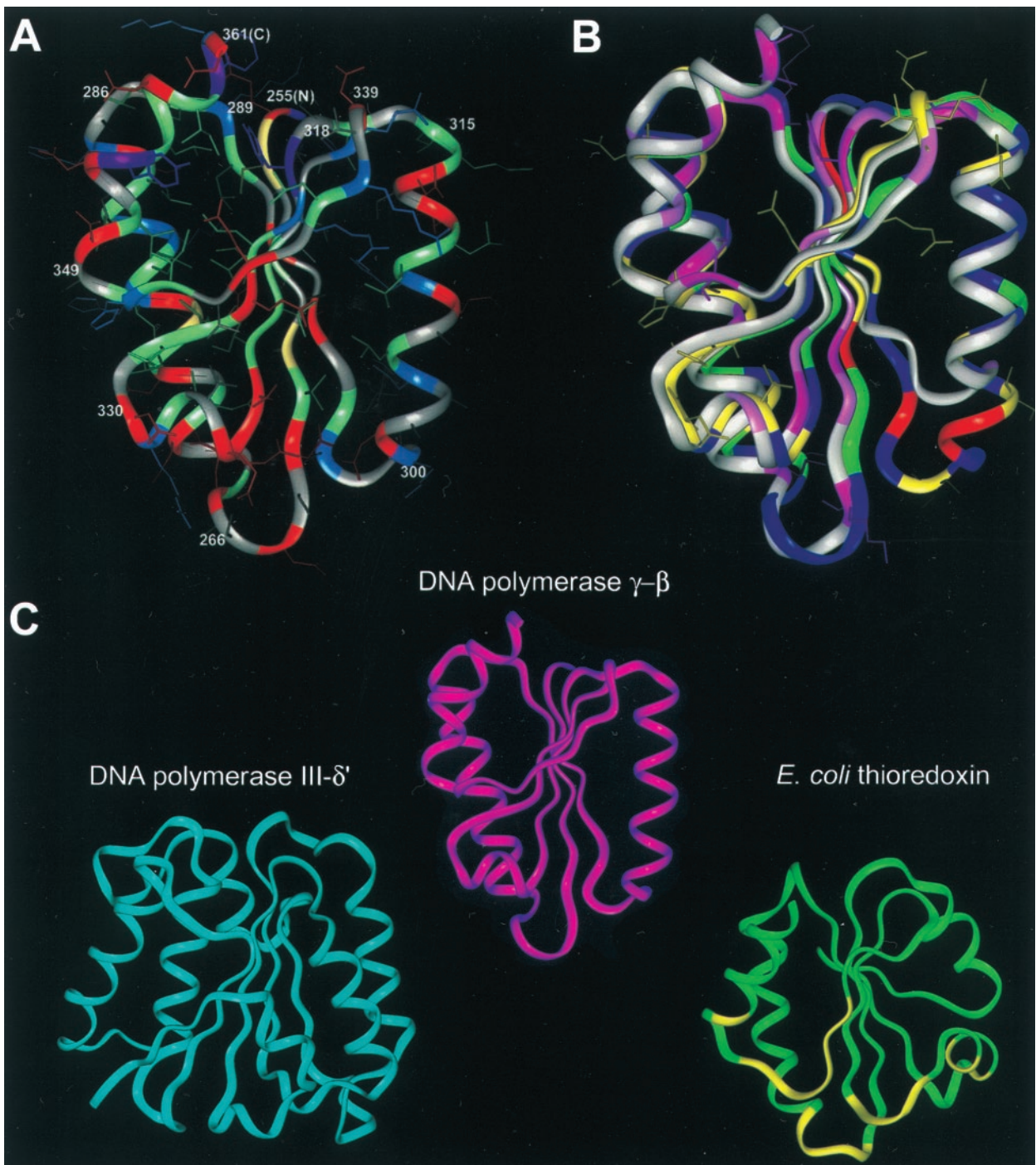


FIG. 2. Structural modeling and comparison of the C-terminal domain of the accessory subunit of *Drosophila* pol γ with other DNA polymerase accessory subunits. (A) Molecular model of the C-terminal domain of *Drosophila* pol γ - β . The main-chain ribbon is shown along with protein side-chain atoms colored according to chemistry: negatively charged residues and derivatives (Asp, Asn, Glu, Gln), red; positively charged residues (His, Lys, Arg), blue; small hydrophobic residues (Leu, Ile, Val, Met), green; large hydrophobic residues (Phe, Tyr, Trp), purple; cysteine, yellow; and other residues (Pro, Ala, Gly, Ser, Thr), gray. The termini of the α -helices and β -strands are numbered according to the *Drosophila* pol γ - β sequence in Fig. 1, with the N and C termini labeled as N and C. (B) Comparison of the structures of the modeled C-terminal domain of *Drosophila* pol γ - β and the C-terminal domain of prolyl-RS, used as a template for the pol γ - β model. The main-chain structure of the model, oriented as in A, is shown superimposed onto the anticodon-binding domain of *T. thermophilus* prolyl-RS. The accessory subunit is colored according to sequence conservation between the *Drosophila* and human sequences; residues conserved between the two sequences are shown in yellow along with their surface-accessible side chains, conservatively substituted residues are in green, nonconserved residues are in blue, and insertions in the fly sequence relative to the human sequence are in red. (There are no deletions in the fly sequence relative to human.) Conservation is defined as in Fig. 1. The prolyl-RS main-chain ribbon is colored by sequence conservation among the three related aaRSs shown in Fig. 1, with conserved residues in dark pink, similar residues in light pink, and nonconserved residues in grey. (C) Comparison of the structures of the accessory subunit of pol γ , the δ' -subunit of *E. coli* DNA polymerase III, and *E. coli* thioredoxin. The N-terminal domain of DNA polymerase III δ' (PDB ID 1a5t; ref. 36) is shown in blue, pol γ - β in magenta, and thioredoxin (PDB ID 1t7p, chain B) (40) in green, with thioredoxin residues interacting with the catalytic subunit of T7 DNA polymerase shown in yellow. Optimally similar orientations were produced by using DALI (22, 47).

segment for the putative zinc-finger motif in poly γ - β (residues 256–286) consists of a β -strand, a loop, and an α -helix, which is similar to the structure observed for some known zinc fingers (27). However, the side chains of Cys and His residues in the poly γ - β structural model are not positioned such that they could bind zinc.

Biochemical Implications of the Structural Model. *The accessory subunit of poly γ as a specialized primer recognition factor.* mtDNA replication is thought to involve different priming mechanisms at the two origins. The leading DNA strand origin is located within the major noncoding region of the mtDNA genome in all animals examined to date (28). In man and mouse, transcripts initiated from the L strand promoter are processed for use as primers for leading DNA strand synthesis; RNase mitochondrial RNA processing cleaves the transcripts at one of several discrete sites, \approx 100 nucleotides from the promoter (29, 30). The 3' termini of these processed transcripts are then presumably elongated by poly γ , but data regarding the mechanism of poly γ -binding at the template–primer junction are lacking. Our structural model suggests that the accessory subunit may play a role in the recognition of these unusual RNA–DNA hybrids to recruit poly γ to the template–primer junction. In particular, we propose that by virtue of its conserved C-terminal domain, homologous to the anticodon-binding domain in class IIa aaRSs, the accessory subunit recognizes secondary structural features in the processed RNA, thereby enhancing the template–primer binding affinity of poly γ .

A similar mode of RNA–DNA hybrid recognition by poly γ may occur at the lagging DNA strand origin. Priming of lagging DNA strand synthesis in mammals appears to be catalyzed by a primase (31) that recognizes a specific stem–loop structure formed in the displaced parental strand (32). Because the lagging DNA strand initiation site is nested within a cluster of five tRNA genes, the involvement of tRNA molecules cannot be excluded. Avian mtDNA genomes lack the noncoding sequence within the tRNA cluster that contains the stem–loop structure, suggesting a replication function for the tRNA genes (or a corresponding tRNA) (33). In *Drosophila*, the lagging DNA strand initiation site has been mapped within the noncoding A+T region, and in several species a stem–loop structure has been identified that may serve a replication function (34). Thus, although only in mammals has the stem–loop structure been documented to function in lagging DNA strand initiation, the situation may be similar in birds and in *Drosophila*. If so, the function of the accessory subunit of poly γ in specialized primer recognition is likely very similar in both leading and lagging strand initiation in mtDNA replication and is generally conserved in animals.

The postulated role of the accessory subunit of poly γ as a primer recognition factor parallels the function of the nonhomologous γ complex of *E. coli* DNA polymerase III and eukaryotic replication factor C. In contrast to the unidirectional and asymmetric mode of mtDNA replication, chromosomal replication in both prokaryotes and eukaryotes proceeds bidirectionally with symmetric synthesis of leading and lagging DNA strands. This involves multiple cycles of recognition of short oligonucleotide primers by the γ -complex and replication factor C, respectively. The γ -complex and replication factor C are multisubunit proteins that share both sequence and structural similarity (35, 36). Our structural model of the C terminus of the accessory subunit of poly γ shows some similarity with the crystal structure determined for the δ' subunit of the γ -complex, which forms a C shaped structure comprising three domains (36) (Fig. 2C Left). Although the N-terminal domain of δ' is not highly similar to the C-terminal domain of poly γ - β based on DALI comparison (Z score of 1.0, below the significance threshold of 2.0), they are qualitatively similar, sharing an α/β -fold with a central β -sheet. This is consistent with our hypothesis of the accessory subunit as a

primer-recognition factor and may also indicate a general structure–function theme in primer recognition factors in various replicative systems.

The accessory subunit of poly γ as a processivity clamp. The structural and functional features of *Drosophila* poly γ are similar to those of the bacteriophage T7 DNA polymerase. The T7 DNA polymerase and poly γ catalytic cores are members of the family A group of DNA polymerases (9). T7 DNA polymerase has a low intrinsic processivity, dissociating from the DNA after incorporation of only a few nucleotides (37). On infection, T7 DNA polymerase recruits a host-encoded protein, *E. coli* thioredoxin (12 kDa), to form a heterodimer that is capable of polymerizing thousands of nucleotides without dissociation (37). Likewise, we have found that the heterodimeric poly γ is a highly processive enzyme (6). In the *E. coli* and eukaryotic replicative systems, DNA polymerase III and DNA polymerase δ use ring-shaped accessory subunit complexes, β (38) and proliferating-cell nuclear antigen (39), respectively, to convert them into highly processive enzymes. These “processivity clamps” are directed to the template–primer by the γ -complex and replication factor C, discussed above, which actually serve dual roles in primer recognition and clamp loading.

The recently determined crystal structure of a T7 DNA polymerase–thioredoxin–DNA complex revealed that thioredoxin is not a ring-shaped protein (40). Instead, thioredoxin binds to the tip of the thumb in the T7 DNA polymerase structure and from this position presumably clamps the template–primer. Corresponding biochemical studies indicate that this association results in an \approx 80-fold increase in the affinity of T7 DNA polymerase for the primer terminus (37). Thus, the accessory subunit of T7 DNA polymerase apparently serves a dual function in primer binding and processivity enhancement. In fact, the mechanism of initiation at the T7 primary DNA replication origin shares similarity with that at the leading DNA strand origin in mitochondrial replication. In T7 replication, the T7 RNA polymerase initiates transcription at a site upstream of the origin and proceeds through it; T7 DNA polymerase then displaces the RNA polymerase and extends the 3' end of the transcript (41). We find that although the accessory subunit of poly γ and thioredoxin have low sequence similarity, our structural model of the C terminus of the accessory subunit is quite similar to the structure of thioredoxin (Fig. 2C Right), which consists of a central core of five β -strands surrounded by four helices. Although thioredoxin (PDB ID 1t7p) and the C-terminal domain of poly γ - β do not have identical fold topologies, DALI structural alignment indicates significant structural similarity between them (Z score of 2.7 over 71 residues).

Herpes simplex virus type 1 DNA polymerase is also a two-subunit enzyme (42, 43). The catalytic subunit is a member of the family B class of DNA polymerases (9). Its accessory subunit, the processivity factor UL42, binds double-stranded DNA (44). Comparative sequence analysis reveals that residues 115–174 of UL42 bear 35% identity with the anticodon-binding domain (residues 329–383) of *T. thermophilus* ProRS (data not shown). Furthermore, UL42 residues 80–289 aligned with poly γ - β residues 105–300, and of the 45 of these residues contained in the structural model, 31% are identical. Residues 129–163 and 202–337 in UL42 are required for stimulation of DNA polymerase activity (45), and residues within both regions are also required for DNA binding by UL42 (46). Taken together, these findings support our hypothesis of the role of the accessory subunit in poly γ function.

In summary, the C-terminal domain of the accessory subunit of *Drosophila* poly γ reveals a rare α/β -fold topology comprising a five-stranded mixed β -sheet surrounded by four α -helices, that is found only in the anticodon-binding domains of class IIa aaRSs (13). This structure may endow poly γ - β with the capacity to bind RNA molecules with tRNA-like structures, as in

aaRSs. This and the structural similarity with the δ' subunit of *E. coli* DNA polymerase III, and particularly with thioredoxin, suggest that the accessory subunit may play a dual role in loading poly γ onto the primer terminus and in enhancing its processivity. Considering that all of the proteins required for mtDNA replication are the products of nuclear genes, it is not surprising that the mitochondrion has evolved to require only a single polypeptide to perform the functions provided by both the multisubunit clamp loaders and processivity clamps in prokaryotic and nuclear replicative systems. Furthermore, this conservative mechanism for a cellular DNA polymerase can likely be generalized to animal polys, because our sequence-based structural comparisons indicate that the proposed C-terminal structure is conserved in the accessory subunits from fly to man. Baculovirus overexpression shows that the accessory subunit is largely insoluble when expressed independently, whereas a purified form of the catalytic subunit alone exhibits a very low specific activity (48). Thus, the apparently critical functions of the accessory subunit may be assessed in mutant holoenzyme derivatives.

We thank Dr. Stephen Cusack (European Molecular Biology Laboratory, Grenoble, France) for generously providing structural coordinates for the anticodon binding domain of prolyl-RS and for comments on the manuscript. We also thank Dr. Michael Kron (Michigan State University) for sharing his expertise on aaRSs and for reviewing the manuscript. L.F. thanks Ya He for her enthusiastic support. This research was supported by National Institutes of Health Grant GM45295 to L.S.K. and an American Heart Association Established Investigatorship to L.A.K.

- Wernette, C. M. & Kaguni, L. S. (1986) *J. Biol. Chem.* **261**, 14764–14770.
- Olson, M. W., Wang, Y., Elder, R. H. & Kaguni, L. S. (1995) *J. Biol. Chem.* **270**, 28932–28937.
- Lewis, D. L., Farr, C. L., Wang, Y., Lagina, A. T., III & Kaguni, L. S. (1996) *J. Biol. Chem.* **271**, 23389–23394.
- Wernette, C. M., Conway, M. C. & Kaguni, L. S. (1988) *Biochemistry* **27**, 6046–6054.
- Kaguni, L. S., Wernette, C. M., Conway, M. C. & Yang-Cashman, P. (1988) in *Cancer Cells: Eukaryotic DNA Replication*, eds. Kelly, T. & Stillman, B. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 6, pp. 425–432.
- Williams, A. J., Wernette, C. M. & Kaguni, L. S. (1993) *J. Biol. Chem.* **268**, 24855–24862.
- Wang, Y., Farr, C. L. & Kaguni, L. S. (1997) *J. Biol. Chem.* **272**, 13640–13646.
- Lecrenier, N., Van Der Bruggen, P. & Foury, F. (1997) *Gene* **185**, 147–152.
- Braithwaite, D. K. & Ito, J. (1993) *Nucleic Acids Res.* **21**, 787–802.
- Lee, D. Y. & Clayton, D. A. (1998) *J. Biol. Chem.* **273**, 30614–30621.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
- Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Cusack, S., Yaremchuk, A., Krikliiviy, I. & Tukalo, M. (1998) *Structure (London)* **6**, 101–108.
- Fischer, D. & Eisenberg, D. (1996) *Protein Sci.* **5**, 947–955.
- Rost, B. & Sander, C. (1994) *Proteins* **19**, 55–72.
- Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998) *Acta Crystallogr. D* **54**, 1078–1084.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *Eur. J. Biochem.* **80**, 319–324.
- Aberg, A., Yaremchuk, A., Tukalo, M., Rasmussen, B. & Cusack, S. (1997) *Biochemistry* **36**, 3084–3094.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291.
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992) *Proteins* **12**, 345–364.
- Luthy, R., Bowie, J. U. & Eisenberg, D. (1992) *Nature (London)* **356**, 83–85.
- Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138.
- Cusack, S. (1995) *Nat. Struct. Biol.* **2**, 824–831.
- Logan, D. T., Mazauric, M. H., Kern, D. & Moras, D. (1995) *EMBO J.* **14**, 4156–4167.
- Arnez, J. G. & Moras, D. (1997) *Trends Biochem. Sci.* **22**, 211–216.
- Sander, C. & Schneider, R. (1991) *Proteins* **9**, 56–68.
- Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
- Clayton, D. A. (1982) *Cell* **28**, 693–705.
- Chang, D. D. & Clayton, D. A. (1987) *Science* **235**, 1178–1184.
- Lee, D. Y. & Clayton, D. A. (1997) *Genes Dev.* **11**, 582–592.
- Wong, T. W. & Clayton, D. A. (1985) *J. Biol. Chem.* **260**, 11530–11535.
- Wong, T. W. & Clayton, D. A. (1985) *Cell* **42**, 951–958.
- Desjardins, P. & Morais, R. (1990) *J. Mol. Biol.* **212**, 599–634.
- Clary, D. O. & Wolstenholme, D. R. (1987) *J. Mol. Evol.* **25**, 116–125.
- O'Donnell, M., Onrust, R., Dean, F. B., Chen, M. & Hurwitz, J. (1993) *Nucleic Acids Res.* **21**, 1–3.
- Guenther, B., Onrust, R., Sali, A., O'Donnell, M. & Kuriyan, J. (1997) *Cell* **91**, 335–345.
- Huber, H. E., Tabor, S. & Richardson, C. C. (1987) *J. Biol. Chem.* **262**, 16224–16232.
- Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992) *Cell* **69**, 425–437.
- Krishna, T. S., Fenyo, D., Kong, X. P., Gary, S., Chait, B. T., Burgers, P. & Kuriyan, J. (1994) *J. Mol. Biol.* **241**, 265–268.
- Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) *Nature (London)* **391**, 251–258.
- Kornberg, A. & Baker, T. A. (1992) *DNA Replication* (Freeman, New York).
- Gottlieb, J., Marcy, A. I., Coen, D. M. & Challberg, M. D. (1990) *J. Virol.* **64**, 5976–5987.
- Hernandez, T. R. & Lehman, I. R. (1990) *J. Biol. Chem.* **265**, 11227–11232.
- Gallo, M. L., Jackwood, D. H., Murphy, M., Marsden, H. S. & Parris, D. S. (1988) *J. Virol.* **62**, 2874–2883.
- Monahan, S. J., Barlam, T. F., Crumpacker, C. S. & Parris, D. S. (1993) *J. Virol.* **67**, 5922–5931.
- Chow, C. S. & Coen, D. M. (1995) *J. Virol.* **69**, 6965–6971.
- Holm, L. & Sander, C. (1994) *Nucleic Acids Res.* **22**, 3600–3609.
- Wang, Y. & Kuhn, L. S. (1999) *J. Biol. Chem.* **274**, in press.