The ϕ 29 DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition

Satwik Kamtekar^{1,5}, Andrea J Berman^{1,5},
Jimin Wang¹, José M Lázaro², Miguel de
Vega², Luis Blanco², Margarita Salas²
and Thomas A Steitz^{1,3,4,*}

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA, ²Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), Universidad Autónoma, Canto Blanco, Madrid, Spain, ³Department of Chemistry, Yale University, New Haven, CT, USA and ⁴Howard Hughes Medical Institute, Yale University, New Haven, CT, USA

The absolute requirement for primers in the initiation of DNA synthesis poses a problem for replicating the ends of linear chromosomes. The DNA polymerase of bacteriophage ϕ 29 solves this problem by using a serine hydroxyl of terminal protein to prime replication. The 3.0 Å resolution structure shows one domain of terminal protein making no interactions, a second binding the polymerase and a third domain containing the priming serine occupying the same binding cleft in the polymerase as duplex DNA does during elongation. Thus, the progressively elongating DNA duplex product must displace this priming domain. Further, this heterodimer of polymerase and terminal protein cannot accommodate upstream template DNA, thereby explaining its specificity for initiating DNA synthesis only at the ends of the bacteriophage genome. We propose a model for the transition from the initiation to the elongation phases in which the priming domain of terminal protein moves out of the active site as polymerase elongates the primer strand. The model indicates that terminal protein should dissociate from polymerase after the incorporation of approximately six nucleotides.

The EMBO Journal (2006) 25, 1335–1343. doi:10.1038/sj.emboj.7601027; Published online 2 March 2006

Subject Categories: genome stability & dynamics; structural biology

Keywords: DNA polymerase; initiation; protein-primed replication; structure; terminal protein

Introduction

Polymerases use a variety of mechanisms to synthesize sequences at the 5' terminus of a genome. While some RNA-dependent RNA polymerases begin synthesis through

de novo initiation, this option is not available to DNA polymerases because DNA synthesis invariably requires a primer. Eukaryotes employ telomerase to overcome this problem (Greider and Blackburn, 1985; Cech, 2004). Bacteria and DNA viruses either avoid the problem with circular genomes, or follow strategies involving recombination, hairpin formation, or the use of a protein as a primer (Salas, 1991; Mosig, 1998; Chaconas and Chen, 2005).

Protein primers, also known as terminal proteins since they are covalently linked to the 5' ends of a genome, are used by a variety of polymerases during initiation of replication. These include B family DNA polymerases from adenovirus, and bacteriophages such as ϕ 29, Cp-1 and PRD1, as well as polymerases involved in plasmid replication and even perhaps chromosomal replication within the genus *Streptomyces* (Salas, 1991; Chaconas and Chen, 2005). They also include the reverse transcriptase of Hepatitis B virus, as well as the RNA-dependent RNA polymerases of poliovirus and rhinovirus (Salas, 1991; Paul, 2002). All of these systems appear to have converged on a similar solution to the 5' end replication problem for linear genomes: a protein side chain provides the necessary hydroxyl group for priming the initiation of DNA replication at the end of the genome.

Unlike other DNA polymerases, those polymerases that use proteins as primers have distinct initiation and elongation phases. During initiation, these protein-primed DNA polymerases begin replication at genomic termini and retain contact with their cognate protein primers. These contacts are broken in the transition to the elongation mode (King *et al*, 1997; Méndez *et al*, 1997). These features of protein-primed DNA polymerases suggest a functional similarity with multi-subunit RNA polymerases; initiation of transcription occurs at promoters and is mediated by transcription factors, which are released after promoter clearance.

Protein-primed DNA replication has been best studied in bacteriophage ϕ 29 (reviewed in Salas, 1991, 1999; Salas *et al*, 1996). This bacteriophage possesses a 19.3 kb linear double-stranded DNA genome with an origin of replication at each end. Phage-encoded DNA polymerase and terminal protein form a heterodimer in the absence of DNA (Blanco *et al*, 1987), and replication initiates with the polymerase-catalyzed addition of dAMP to serine-232 of terminal protein (Blanco and Salas, 1984; Hermoso *et al*, 1985). The template for this addition is the second base from the 3' end of the genome. Prior to further nucleotide addition, a 'sliding back' event occurs, resulting in the base pairing of the first dAMP with the 3' terminal base of the genome (Méndez *et al*, 1992). Subsequent synthesis leads to the dissociation of terminal protein from polymerase after 6–10 nucleotides have been incorporated (Méndez *et al*, 1997).

The crystal structure of ϕ 29 DNA polymerase provided the first view of a protein-primed DNA polymerase structure, and insight into how it is able to replicate the entire genome

*Corresponding author. Department of Molecular Biophysics and Biochemistry, Yale University, Room 418, Bass Center, 266 Whitney Avenue, New Haven, CT 06520-8114, USA.

Tel.: +1 203 432 5617/5619; Fax: +1 203 432 3282;

E-mail: eatherton@csb.yale.edu

⁵These authors contributed equally to this work

without accessory helicases or processivity factors (Kamtekar *et al*, 2004). Two sequence insertions, TPR1 and TPR2 (Blasco *et al*, 1990; Dufour *et al*, 2000), found only in the polymerase domains of protein-primed DNA polymerases, formed discrete subdomains in the structure (Kamtekar *et al*, 2004). While TPR1 abutted the polymerase domain in a position consistent with interaction with terminal protein and DNA, TPR2 appeared to contribute to both efficient strand displacement and processivity. Homology modeling of the substrate complex showed the template strand passing through a narrow tunnel formed in part by TPR2, prior to entering the polymerase active site, providing a plausible mechanism for the efficient separation of template and non-template strands. The model also showed encirclement of the duplex DNA product by polymerase in a manner reminiscent of clamp proteins, accounting for the extraordinary processivity of the polymerase on DNA (Blanco *et al*, 1989). Consistent with this modeling, deletion of the TPR2 subdomain has recently shown that it is required for efficient strand displacement and processivity (Rodríguez *et al*, 2005).

Here we describe the structure of φ29 DNA polymerase bound to terminal protein and its implications for the mechanism of initiation of replication. The conformation of the polymerase in the complex is largely unchanged from that of the *apo* polymerase. Terminal protein forms an extended structure containing an N-terminal domain, a domain that interacts with TPR1, and a priming domain. The priming domain appears to mimic duplex product DNA in its electrostatic profile and binding site in the polymerase. This spatial overlap of binding sites explains why the polymerase:terminal protein complex initiates exclusively at the ends of linear DNA. DNA synthesis by the heterodimer cannot begin at internal sites because the upstream 3' template would sterically clash with terminal protein. Additionally, the priming domain must back away from the active site as duplex product is synthesized. The modeling described below suggests that it can do so for approximately 6 bases before terminal protein must dissociate from polymerase, a result that is consistent with biochemical studies (Méndez *et al*, 1997).

Results

Structure determination

The structure of φ29 DNA polymerase bound to terminal protein was initially determined using multiple isomorphous replacement in a crystal form with a space group of I23. This crystal form contained a single copy of the polymerase:terminal protein heterodimer per asymmetric unit, and yielded data to 3.5 Å resolution. Improvement of the heavy-atom phases by solvent-flipping and cross-crystal averaging, together with *B*-factor sharpening of the amplitudes, provided maps into which the main chain could be positioned. The sequence register of polymerase in these maps could be established based on previous high-resolution structures of polymerase (Kamtekar *et al*, 2004), but the sequence for terminal protein could not be assigned at this resolution.

Subsequently, similar crystallization conditions gave rise to a closely related crystal form in the space group C2 that diffracted to 3.0 Å resolution and contained six heterodimers per asymmetric unit. The unit cell dimensions of this C2 crystal form, $a = 304$ Å, $b = 220$ Å, $c = 217$ Å, $\beta = 45^\circ$ are

Figure 1 Electron density for a helix of terminal protein near the active site of polymerase. On the left is a 3.5 Å resolution map contoured at 1σ using data from the I23 crystal form. It was calculated using amplitudes sharpened by a factor of 100 and experimentally phased with solvent-flattened heavy-atom phases. Side chains cannot be unambiguously assigned in this map. On the right is a composite omit map contoured at 1σ and calculated to 3 Å using data from the C2 crystal form. Side chain density is much better defined in this map. Figures 1, 2, 3A, 3B, and 4 were made using Pymol (<http://www.pymol.org>).

related to those of the I23 crystals, where $a_{I23} = 218$ Å. An edge of the C2 lattice, a_{C2} , corresponds to a diagonal on face of the I23 cell ($a_{C2} = \sqrt{2}a_{I23}$). Cross-crystal averaging with the I23 crystal form, along with noncrystallographic symmetry (NCS) averaging within the C2 crystal form was used to improve maps, and clear density could be observed for many side chains of terminal protein (Figure 1).

The structure in the C2 crystal form was refined to an R_{cryst} of 20.0% and an R_{free} of 22.9% (Table I). Residues 6–575 of polymerase and residues 71–119, 140–226, and 234–259 of terminal protein have been built for all six copies in the asymmetric unit. In addition, 34 residues of the N-terminal domain of terminal protein have been modeled as a poly-alanine sequence (Figure 2). All six polymerase:terminal protein heterodimers in the asymmetric unit are very similar (the RMSDs between them vary from 0.3 to 0.6 Å when calculated over all C_α atoms). The only significant differences in conformations between the crystallographically independent heterodimers in the C2 crystal form appear to be localized to residues 112–119 of terminal protein. Apart from this region of terminal protein and a few other loop regions, both positional and *B*-factor constraints were maintained between NCS-related regions of the polymerase:terminal protein heterodimer during refinement.

The structure of the polymerase:terminal protein heterodimer

The structure of φ29 DNA polymerase, as previously described, has the architecture of a canonical B-family DNA polymerase with two additional subdomains unique to protein-primed polymerases (Kamtekar *et al*, 2004). It contains

Table 1 Crystallographic statistics

	C2	I23 ^a
Space group	C2	I23 ^a
Unit cell dimensions (Å)	304 × 220 × 217, β = 134.6°	218 × 218 × 218
Resolution (Å)	50.0–3.0 (3.1–3.0)	50–3.45 (3.57–3.45)
R _{merge} ^b (%)	8.5	7.2
I/σ ^c	12.9 (1.2)	21.5 (1.2)
Completeness (%)	99.9 (100.0)	99.6 (100.0)
Unique reflections	202, 295	22, 628
Redundancy	3.8	6.2
Copies in AU	6	1
Phasing power (<i>acentric</i>) ^d		
Mercury acetate	—	1.1 (20.0–4.0 Å)
Ethyl mercury phosphate	—	1.2 (20.0–4.0 Å)
Trimethyl lead acetate	—	1.4 (20.0–4.0 Å)
Platinum tetrachloride	—	2.0 (20.0–4.3 Å)
Gold cyanide	—	1.6 (20.0–4.0 Å)
Figure of merit		
Heavy atoms combined	—	0.63 (20.0–4.0 Å)
Solvent flattened	—	0.64 (20.0–3.45 Å)
RMSD bond length (Å)	0.007	—
RMSD bond angle	1.00°	—
R _{cryst} ^e (%)	20.0	—
R _{free} ^e (%)	22.9	—
PDB ID	2EX3	2EX3

^aElectron density maps using data from this space group were used to construct preliminary models as described in the text.

^bR_{merge} is $\sum_i |I_i - \langle I \rangle|$, where I_i is the intensity of an individual reflection and $\langle I \rangle$ is the mean intensity for multiply recorded reflections.

^cI/σ is the mean of intensity divided by the standard deviation.

^dThe phasing power is the RMS isomorphous difference divided by the lack of closure.

^eR_{cryst} is $\sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o is the observed amplitude and F_c the calculated amplitude; R_{free} is the same statistic calculated over a subset of the data (10%) that has not been used for refinement.

an N-terminal exonuclease domain, which is responsible for proofreading (residues 1–189), and a C-terminal polymerase domain (residues 190–575). By analogy to a right hand, polymerase domains are conventionally divided into palm, finger, and thumb subdomains, which are involved in binding the catalytic metal ions, the incoming dNTP, and the duplex product DNA, respectively (Steitz *et al*, 1994). The two subdomains that are unique to protein-primed polymerases, TPR1 and TPR2, are insertions in the palm subdomain involved in interaction with terminal protein (Dufour *et al*, 2000, 2003), and in processivity and strand displacement (Kamtekar *et al*, 2004; Rodríguez *et al*, 2005).

Our current structure shows that the conformation of polymerase during initiation is very similar to that of the *apo* polymerase. Thus, apart from residues 304–314 in TPR1, only minor changes are apparent when the structures of the *apo* polymerase and the polymerase in the polymerase:terminal protein complex are superimposed (Figure 2C). In the *apo* polymerase structures, these residues of TPR1 form loops with varied conformations and high B-factors, indicating a substantial degree of flexibility. In the current structure, this loop has moved to allow terminal protein access to the active site of polymerase. The RMSD between all C_α atoms excluding residues 304–314 in the two structures is 1.0 Å.

In this complex with polymerase, the terminal protein forms an elongated three-domain structure (Figure 2D). The N-terminal domain (residues 1–73) contains disordered sequence as well as regions with high temperature factors, perhaps because it has few stabilizing contacts within the crystal lattice and is not interacting with the polymerase. Two helices within this domain have been built as poly-alanine. The intermediate domain (residues 74–172) contains two long α-helices and a short β-turn-β structure. It is connected

through a hinge region to the C-terminal priming domain (residues 173–266), which is comprised of a four-helix bundle. This four-helix bundle has a left turning up-down-up-down topology (Kamtekar and Hecht, 1995), with the third helix (residues 214–223) forming an unusually large angle with respect to the axis of the four-helix bundle. Serine-232, which provides the priming hydroxyl group for DNA synthesis, lies in a loop at the end of the priming domain closest to the active site of the DNA polymerase. The loop containing the priming serine (residues 227–233) is disordered in our structure, presumably because of the absence of a template and incoming deoxynucleoside triphosphate.

The disorder of the loop was confirmed by diffusing mercury chloride into crystals containing terminal protein with a S232C mutation. While mercury bound to the exposed cysteines in the polymerase yielded peaks with high density in difference maps, none were near cysteine-232. Since residue 232 of terminal protein appears to be solvent accessible, the absence of a peak near to this residue can be explained by its being disordered.

A comparison of the structure of terminal protein with those in the structural database using the program DALI (Holm and Sander, 1993) yielded only one match with a Z-score above 4. This alignment superimposes a four-helix bundle domain from Cbl, a ubiquitin ligase involved in cell signalling, onto the priming domain of terminal protein with a Z-score of 4.7 (PDB code 2CBL; Meng *et al*, 1999). Since the four-helix bundle is an abundant motif and Cbl is functionally unrelated to terminal protein, this structural similarity appears purely coincidental.

A number of transcription factors also appear to have more limited similarity to the priming domain of terminal protein. For example, the priming domain of terminal protein can be

Figure 2 The structure of the polymerase:terminal protein heterodimer. (A) A ribbon representation, with polymerase colored according to Kamtekar *et al*, 2004, and terminal protein shown with cylindrical helices. (B) A view of the complex rotated 90° from that shown in (A), with terminal protein shown as cylinders underneath a transparent surface. (C) A C α trace of polymerase from the polymerase:terminal protein complex (in color) superimposed on the *apo* polymerase structure. Significant differences in conformation occur only in a loop between residues 304 and 314 (shown in magenta in the complex and in black in the *apo* polymerase structure). The polymerase active site is marked by the space-filling representations of the carboxylates that coordinate the catalytic metal ions. (D) Terminal protein in the same orientation as in (B).

superimposed on transcription factor IIS (PDB code 1eo0; Booth *et al*, 2000). However, there is no evidence suggesting that the helical bundle domains in these transcription factors associate with RNA polymerase (Kettenberger *et al*, 2004) in ways that are analogous to the interaction between the priming domain of terminal protein and ϕ29 DNA polymerase.

Both the intermediate and priming domains of terminal protein make extensive contacts with polymerase accounting for its high affinity (80 nM) for the polymerase (Lázaro *et al*, 1995). The intermediate domain buries 575 Å² of surface area against the TPR1 subdomain of polymerase. Here, and elsewhere, buried surfaces are calculated per monomer. This interface has many charged residues and includes two salt bridges between arginine residues in terminal protein and glutamic acid residues in TPR1 (R158:E291; R169:E322). In contrast to the intermediate domain, the priming domain is highly electronegative and contains 15 acidic residues (Figure 3C). The structure shows interactions between

many of these acidic residues and positively charged residues of polymerase (e.g., between E191:K575 or D198:K557). Consistent with studies showing an impaired binding of R96A mutant polymerase and terminal protein (Rodríguez *et al*, 2004), this arginine residue hydrogen bonds with Q253 and forms a stacking interaction with Y250 of terminal protein. Part of the C-terminal helix of the priming domain packs against the TPR2 subdomain of polymerase, forming hydrogen bonds between residues E252, Q253 and R256 of terminal protein and L416, G417 and E419 of polymerase. Altogether, the exonuclease domain and the palm, thumb, TPR1 and TPR2 subdomains of polymerase encircle the priming domain of terminal protein, burying over 1500 Å² of its surface area in the process (Figure 3), which produces a total surface area of terminal protein buried upon complex formation of 2075 Å².

The location of the priming domain of terminal protein overlaps that of the primer:template DNA bound to the enzyme in a homology-modeled elongation complex. The

Figure 3 The priming domain of terminal protein binds polymerase in a fashion analogous to how primer:template DNA binds polymerase. (A) A space-filling representation of polymerase with a cylinder representation of terminal protein, both colored as in Figure 2. (B) Primer:template DNA from the ternary structure of RB69 DNA polymerase (Franklin *et al*, 2001) homology modeled onto the structure of ϕ 29 DNA polymerase. The binding site of the priming domain overlaps that of the DNA. (C) A space-filling representation of the priming domain of terminal protein, and a ribbon representation of polymerase. The electrostatic surface of the priming domain was calculated using GRASP, with a range of -10 to $+10$ kT (Nicholls *et al*, 1991). The portion of this domain in contact with polymerase is highly negatively charged. This domain of terminal protein also effectively seals a path, occupied by primer:template in (B), leading out of the polymerase active site. (C) was generated using MOLSCRIPT (Kraulis, 1991) and rendered using POV-RAY (<http://www.povray.org>).

catalytic palm subdomain of ϕ 29 DNA polymerase can be aligned with the palm of the B-family DNA polymerase from bacteriophage RB69 (Franklin *et al*, 2001; Kamtekar *et al*, 2004). As described previously, this alignment allows placement of the primer, template, and incoming nucleotide from a ternary complex of RB69 polymerase onto ϕ 29 DNA polymerase (Kamtekar *et al*, 2004). The spatial overlap of the priming domain of terminal protein with modeled primer:template as well as its negative charge and overall dimensions suggest that it mimics DNA in its interactions with polymerase (Figure 3).

Since the priming domain plays a central role in ϕ 29 protein-primed replication, we searched for evidence of a similar structure in the more complex protein-primed system of the eukaryotic virus, adenovirus. Secondary structure predictions indicate that the priming serine residue in adenoviral terminal protein is also in a loop surrounded on either side by helices (de Jong and van der Vliet, 1999; McGuffin *et al*, 2000). However, the low level of sequence identity between the terminal proteins of ϕ 29 and adenovirus make sequence alignments and structural modeling unreliable.

Discussion

The 3.0 Å resolution structure of the ϕ 29 DNA polymerase:terminal protein heterodimer provides insights into the specificity for initiation at the template terminus, the initiation phase of protein-primed replication, DNA packaging, and the varied, but analogous, strategies used by polymerases during the initiation of oligonucleotide synthesis. The division of the terminal protein structure into three domains and its extended overall fold have implications for the transition from the initiation of replication to the elongation of the complete genome, and for the packaging and ejection of the terminal protein–DNA covalent complex (TP–DNA). Finally, the overall processes of priming site recognition and the transition from initiation to elongation by ϕ 29 DNA polymerase manifests mechanistic themes previously described only in RNA polymerases.

Specificity for 3' template ends

The ϕ 29 DNA polymerase:terminal protein heterodimer initiates synthesis only at the 3' end of the phage genome, which functions as an origin of replication. Two sources of this specificity have been examined previously. First, the presence of parental terminal protein, which is covalently linked to the 5' end of the nontemplate strand by a previous cycle of replication, has been shown to increase the templating activity of a terminus, but only by 6–10-fold (Gutiérrez *et al*, 1986; González-Huici *et al*, 2000). Second, the nucleotide bases at the template 3' terminus have been altered to test whether recognition is sequence specific. These experiments show that the only necessary sequence requirement is that the two 3' terminal nucleotides be identical to allow for the sliding back by one nucleotide at initiation (Méndez *et al*, 1992). Thus, the heterodimer appears to be principally recognizing the 3' terminus of the template strand.

The structure of this complex is consistent with the observed lack of template sequence specificity. Since the polymerase must be sequence independent, only the terminal protein could be involved in sequence recognition. Template DNA modeled into the structure indicates the possibility of only a very limited contact between terminal protein and DNA. As the template in the tunnel formed by the palm, fingers, TPR2, and exonuclease regions is closely surrounded by polymerase, the nucleotides in the tunnel appear unable to contact terminal protein. Together with the biochemical data, which show that the second base of the bacteriophage genome usually acts as a template for addition of the first dAMP to terminal protein (Méndez *et al*, 1992), the structure thus suggests that, at most, these two terminal bases could directly interact with terminal protein in the polymerase active site under standard initiation conditions.

We conclude from our structure that the polymerase:terminal protein heterodimer selectively recognizes the initiation site of replication at the end of the genome through steric exclusion of alternative sites. As terminal protein almost completely blocks the upstream tunnel leading out of the polymerase active site, the polymerase:terminal protein heterodimer cannot accommodate upstream template and can only bind near the ends of DNA (Figures 2 and 3), thus

providing the absolute requirement for initiation at the ends of DNA. Similar steric mechanisms to ensure specificity for the ends of a genome have been employed by RNA viruses such as hepatitis C (Lesburg *et al*, 1999; Hong *et al*, 2001) and φ6 (Butcher *et al*, 2001).

A model for the transition from initiation to elongation

The transition from the initiation to the elongation phase of DNA synthesis must be accompanied by conformational changes in terminal protein, since, as the primer is elongated, the duplex product will increasingly occupy the upstream duplex tunnel; thus, it must progressively displace the priming domain of terminal protein as synthesis proceeds. We have modeled this displacement to generate a scheme for this transition.

The anticipated path of the primer strand backbone product DNA through the polymerase provides a major constraint on a model for the displacement of the priming domain (Figure 4). The priming serine attached to the first nucleotide traces a spiral path away from the active site after each nucleotide is incorporated into the primer strand as replication proceeds. For example, after incorporation of five nucleotides, the priming serine will have rotated by approxi-

mately 180° and translated by 17 Å, placing the attached nucleotide in contact with the thumb subdomain.

While the priming domain must move in response to the translocation of the DNA after nucleotide incorporation, the intermediate domain of terminal protein need not for the initial steps. This domain can be modeled in a fixed orientation on the polymerase, thereby preserving significant contact with the polymerase. As the intermediate domain is connected to the priming domain through a turn, or hinge, some displacement of the priming domain appears possible without dissociation.

The hinge between the priming domain and a fixed intermediate domain of terminal protein appears to allow the addition of up to approximately six to seven nucleotides (Figure 4). This hinging can occur for only about six nucleotides before further translocation would place the serine-linked α-phosphate closer to the hinge than the approximately 40 Å length of the four-helix bundle. The priming loop (residues 227–233) of terminal protein is disordered in our current structure, and this flexibility introduces additional uncertainty in our modeling and could allow incorporation of one or two additional nucleotides. This modeling implies that incorporation and translocation of the seventh or eighth nucleotide into the primer strand would require

Figure 4 A model for the transition from initiation of replication to elongation. Cut-away views of the polymerase (white/wheat) are shown. (A) Modeled primer (green) and template (orange) in an elongating complex are represented as sticks. The positions of a phosphate being incorporated, as well as after two, five, and eight cycles of incorporation, are indicated as red spheres, illustrating the spiral translocation of synthesized DNA. (B) A model for the covalent addition of the first nucleotide onto S232 of terminal protein. Terminal protein is colored by domain, with a plausible path for the loop containing S232 shown in red (this loop is disordered in our current structure). (C) Following the hinge motion indicated in (B), the interactions between the intermediate domain and polymerase can be maintained. The priming domain has backed away sufficiently from the polymerase active site to allow the incorporation of two more nucleotides. (D) A further hinge motion between the priming domain and the intermediate domain is consistent with the incorporation of five nucleotides. (E) A hinge motion alone is insufficient after the incorporation of seven bases. The example shown here is for nucleotide eight: the distance between the α-phosphate and the hinge is only 26 Å, which is shorter than the length of the priming domain (approximately 40 Å). Instead of a hinging motion that leaves interactions between the intermediate domain of terminal protein and polymerase intact, displacement of the intermediate domain leads to the dissociation of polymerase:terminal protein complex.

additional structural changes involving the release of the interaction between TPRI of polymerase and the intermediate domain of terminal protein, resulting in the dissociation of polymerase from terminal protein. Thus, this model is consistent with biochemical data indicating that dissociation occurs after 6–10 bases have been incorporated (Méndez *et al*, 1997).

Implications for packaging and ejection of the phage genome

The elongated structure of terminal protein may both facilitate the entry of TP-DNA into a preformed phage particle and the ejection of DNA through the tail of the phage. Terminal protein has a van der Waals diameter of 30–32 Å at its widest point in our structure. This is substantially smaller than would be expected for a globular protein of similar size. Carbonic anhydrase, for example, has 260 amino acids and is egg-shaped with a diameter of 45–55 Å. The TP-DNA genome is packaged into the tail-less capsid through the dodecameric head–tail connector, which at its narrowest has an inner diameter of 36 Å (Simpson *et al*, 2000). Assuming that the individual domains of terminal protein in TP-DNA resemble those in our structure, they could pass through the connector sequentially, changing their observed relative orientations at the connecting loops, without requiring either terminal protein to unfold or major conformational rearrangements in the connector ring. However, according to 3D reconstructions of the empty phage made from electron micrographs, the inner diameter of the narrowest part of the tail is 26 Å (Tao *et al*, 1998). This suggests that either terminal protein unfolds to fit through the tail during release of the genome, or that there is a transient conformational change in the tail that allows the TP-DNA to pass through without conformational rearrangements.

Common features of the initiation to elongation transition by polymerases

Polymerases that initiate synthesis at distinct sites face a common problem: they must possess a high affinity for the initiation sites during the initiation phase and this affinity must decrease for a successful transition to elongation mode. Furthermore, they must accommodate an elongating product duplex while continuing to remain attached to the initiation site. The structural basis of the transition from the initiation phase to the elongation phase is better understood in RNA polymerases than DNA polymerases.

The structures of the initiation and elongation complexes of the single-subunit RNA polymerase from bacteriophage T7 have been determined and show dramatic conformational change in the N-terminal domain of the polymerase (Cheetham *et al*, 1999; Cheetham and Steitz, 1999; Tahirov *et al*, 2002; Yin and Steitz, 2002). This change disrupts the promoter-binding site in the polymerase, elongates the heteroduplex product binding site from three to eight base pairs, and creates a tunnel through which the emerging RNA transcript can pass in the elongation complex. Models of the transition between the initiation and elongation phases have implied that the trigger for the conformational change is the force exerted by the elongating heteroduplex product on a helical subdomain of the N-terminal domain of T7 RNA polymerase (Tahirov *et al*, 2002; Yin and Steitz, 2002; Theis *et al*, 2004).

In contrast to T7 RNA polymerase, bacterial multi-subunit RNA polymerases recognize promoters using transcription factors encoded as discrete polypeptides. Sigma factor bound to bacterial RNA polymerases recognizes promoters and dissociates from the polymerases either during or following the transition to elongation phase. The synthesis of approximately 12 nucleotides of RNA has been proposed to displace the 3.2 loop of the sigma factor, which is hypothesized to weaken interactions between the polymerase and sigma, leading to promoter escape (Murakami and Darst, 2003).

Eukaryotic RNA polymerase II behaves in an analogous fashion. It is localized to promoters by general transcription factors that are discarded by the polymerase in elongation mode. The binding site of one of these factors, TFIIB, competes with the binding of RNA longer than 10 bases (Bushnell *et al*, 2004), providing a structural basis for the dissociation of the polymerase from promoter DNA.

φ29 and related protein-primed DNA polymerases appear to use a mechanism to recognize and then clear origins of replication that is, in principle, similar to that used by RNA polymerases to bind promoters and then enter an elongation phase. Both recognize initiation sites with high affinity. Synthesis of the duplex product progressively produces conformational changes, either in the polymerases themselves or in associated factors, moving the protein out of the path of the newly synthesized or translocated nucleic acid. Ultimately, this reduces the affinities of the polymerases for their initiation sites. These polymerases subsequently enter elongation mode, in which they possess an intrinsic strand displacement activity and are highly processive.

Materials and methods

Protein purification and crystallization

Exonuclease-deficient (D12A/D66A) φ29 DNA polymerase, wild-type terminal protein, and S232C mutant terminal protein were expressed and purified as described elsewhere (Zaballos *et al*, 1989; Lázaro *et al*, 1995) and stored as ammonium sulfate precipitated pellets at –80°C. The pellets of polymerase and terminal protein were resuspended in 50 mM NaCl, 50 mM Tris–HCl, pH 7.5, 20 mM ammonium sulfate, and 1 mM dithiothreitol at concentrations of 6.7 and 3.3 mg/ml, respectively, prior to crystallization. Crystals were grown by vapor diffusion at 4°C. Typically, 2 µl of a protein stock solution was mixed with an equal volume of well solution containing 1.8–2.0 M NaKPO₄, pH 7 and 100 mM Na-acetate. Small crystals appeared within a week, and increased in size over the course of a month. After several months, some of these crystals appeared to change space groups from I23 to C2. Crystals were stabilized by transferring to 2.2 M NaKPO₄ pH 7, 100 mM Na-acetate. The concentration of ethylene glycol in this stabilization solution was then increased stepwise to a final value of 15%, prior to freezing the crystals in liquid propane. Both the stabilization and cryoprotection solutions used to treat the crystal that yielded the highest resolution data also contained 20 mM trimethyl lead acetate. Diffraction data were integrated and scaled using the HKL software suite (Otwinowski and Minor, 1997).

Structure determination and refinement

Several heavy-atom-derivatized crystals were used to determine the phases of the cubic crystal structure. These included crystals incubated for 1–4 h with (1) 0.1 mM mercury acetate, (2) 0.1 mM ethyl mercury phosphate, (3) 1 mM platinum tetrachloride, (4) 20 mM trimethyl lead acetate, or (5) 2 mM gold cyanide. The heavy-atom phases were refined using SOLVE (Terwilliger and Berendzen, 1999), and improved through the use of cross-crystal averaging and solvent flattening using DMMLTI (CCP4, 1994). Sharpening of the data by *B*-factors of between 50 and 100 substantially improved the quality of the maps.

Using the preliminary structures determined in the cubic crystal forms as search models, molecular replacement with the program PHASER (McCoy *et al*, 2005) led to the solution of the C2 crystal form (Table I). NCS and cross-crystal averaging subsequently yielded better maps. The models were built using the program O (Jones *et al*, 1991). Refinement was performed in CNS (Brunger *et al*, 1998) and REFMAC (Murshudov *et al*, 1997). Tight NCS restraints were maintained throughout the refinement of the C2 crystal form. In structures containing high degrees of NCS, the intensities of reflections are correlated, making the choice of an unbiased test set more complicated. To help minimize this correlation, we used a test set generated in I23 and then symmetry expanded to cover the C2 unit cell.

References

- Blanco L, Bernad A, Lázaro JM, Martín G, Garmendia C, Salas M (1989) Highly efficient DNA synthesis by the phage φ29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* **264**: 8935–8940
- Blanco L, Prieto I, Gutiérrez J, Bernad A, Lázaro JM, Hermoso JM, Salas M (1987) Effect of NH₄⁺ ions on φ29 DNA-protein p3 replication: formation of a complex between the terminal protein and the DNA polymerase. *J Virol* **12**: 3983–3991
- Blanco L, Salas M (1984) Characterization and purification of a phage φ29 coded DNA polymerase required for the initiation of replication. *Proc Natl Acad Sci USA* **81**: 5325–5329
- Blasco MA, Blanco L, Parés E, Salas M, Bernad A (1990) Structural and functional analysis of temperature-sensitive mutants of the phage φ29 DNA polymerase. *Nucl Acids Res* **18**: 4763–4770
- Booth V, Koth CM, Edwards AM, Arrowsmith CH (2000) Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J Biol Chem* **275**: 31266–31268
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**: 905–921
- Bushnell DA, Westover KD, Davis RE, Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II–TFIIB cocrystal at 4.5 Å. *Science* **303**: 983–988
- Butcher SJ, Grimes JM, Makeyev EV, Bamford DH, Stuart DI (2001) A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**: 235–240
- Cech TR (2004) Beginning to understand the end of the chromosome. *Cell* **116**: 273–279
- Chaconas G, Chen CW (2005) Replication of linear bacterial chromosomes: no longer going around in circles. In *The Bacterial Chromosome*, Higgins NP (ed), pp 525–539. Washington, DC: American Society for Microbiology Press
- Cheetham GMT, Jeruzalmi D, Steitz TA (1999) Structural basis for initiation of transcription from an RNA-polymerase promoter complex. *Nature* **399**: 81–83
- Cheetham GMT, Steitz TA (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**: 2305–2309
- Collaborative Computational Project 4 (1994) The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr D* **50**: 760–763
- de Jong RN, van der Vliet PC (1999) Mechanism of DNA replication in eukaryotic cells: cellular host factors stimulating adenovirus DNA replication. *Gene* **236**: 1–12
- Dufour E, Méndez J, Lázaro JM, de Vega M, Blanco L, Salas M (2000) An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. *J Mol Biol* **304**: 289–300
- Dufour E, Rodríguez I, Lázaro JM, Blanco L, de Vega M, Salas M (2003) A conserved insertion in protein-primed DNA polymerases is involved in primer-terminus stabilization. *J Mol Biol* **331**: 781–794
- Franklin MC, Wang J, Steitz TA (2001) Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* **105**: 657–667
- González-Huici V, Salas M, Hermoso JM (2000) Sequence requirements for protein-primed initiation and elongation of phage φ29 DNA replication. *J Biol Chem* **275**: 40547–40553

Acknowledgements

We thank the staff at NSLS beamlines X25, X26C, and X29; at APS beamlines 19-ID and 8-BM; at CHESS beamlines A1 and F1; and at ALS beamlines 8.2.1, 8.2.2 and 8.3.1. We also thank Laurentino Villar for the purification of the terminal protein, members of the Steitz laboratory for help with data collection and useful discussions, Cathy Joyce for comments on this article, and the staff of the CSB core facility at Yale. We were funded by grants R01 GM 57510 from the National Institutes of Health (to TAS) and BMC 2002-03818 from the Spanish Ministry of Science and Technology (to MS), and an institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular ‘Severo Ochoa’.

- Greider CW, Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**: 405–413
- Gutiérrez J, García JA, Blanco L, Salas M (1986) Cloning and template activity of the origins of replication of phage φ29 DNA. *Gene* **43**: 1–11
- Hermoso JM, Méndez E, Soriano F, Salas M (1985) Location of the serine residue involved in the linkage between the terminal protein and DNA of phage φ29. *Nucl Acids Res* **13**: 7715–7728
- Holm L, Sander C (1993) Protein structure comparisons by alignment of distance matrices. *J Mol Biol* **233**: 123–138
- Hong Z, Cameron CE, Walker MP, Castro C, Yao N, Lau JYN, Zhong W (2001) A novel mechanism to ensure terminal initiation by hepatitis C virus NS5B polymerase. *Virology* **285**: 6–11
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47**: 110–119
- Kamtekar S, Berman AJ, Wang J, Lázaro JM, de Vega M, Blanco L, Salas M, Steitz TA (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage φ29. *Mol Cell* **16**: 609–618
- Kamtekar S, Hecht MH (1995) Protein motifs. 7. The four-helix bundle: what determines a fold? *FASEB J* **11**: 1013–1022
- Kettenberger H, Armache K, Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* **16**: 955–965
- King AJ, Teerstra WR, van der Vliet PC (1997) Dissociation of the protein primer and DNA polymerase after initiation of adenovirus replication. *J Biol Chem* **272**: 24617–24623
- Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J Appl Crystallogr* **24**: 946–950
- Lázaro JM, Blanco L, Salas M (1995) Purification of bacteriophage φ29 DNA polymerase. *Methods Enzymol* **262**: 42–49
- Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol* **6**: 937–943
- McCoy AJ, Grosse-Kunstleve RJ, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D* **61**: 458–464
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404–405
- Méndez J, Blanco L, Esteban JA, Bernad A, Salas M (1992) Initiation of φ29 DNA replication occurs at the second 3′ nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. *Proc Natl Acad Sci USA* **89**: 9579–9583
- Méndez J, Blanco L, Salas M (1997) Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. *EMBO J* **16**: 2519–2527
- Meng W, Sawasdikosol S, Burakoff SJ, Eck MJ (1999) Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. *Nature* **398**: 84–90
- Mosig G (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu Rev Genet* **32**: 379–413
- Murakami KS, Darst SA (2003) Bacterial RNA polymerase: the whole story. *Curr Opin Struct Biol* **13**: 31–39

- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr* **53**: 240–255
- Nicholls A, Sharp K, Honig B (1991) Protein folding and association: insights from interfacial and thermodynamic properties of hydrocarbons. *Prot Struct Funct Genet* **11**: 281–296
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326
- Paul AV (2002) Possible unifying mechanisms of picornavirus genome replication. In *Molecular Biology of Picornaviruses*, Semler BL, Wimmer E (eds), pp 227–246. Washington, DC: American Society for Microbiology Press
- Rodríguez I, Lázaro JM, Blanco L, Kamtekar S, Berman AJ, Wang J, Steitz TA, Salas M, de Vega M (2005) A specific subdomain in φ29 DNA polymerase confers both processivity and strand displacement capacity. *Proc Natl Acad Sci USA* **102**: 6407–6412
- Rodríguez I, Lázaro JM, Salas M, de Vega M (2004) φ29 DNA polymerase-terminal protein interaction. Involvement of residues specifically conserved among protein-primed DNA polymerase. *J Mol Biol* **337**: 829–841
- Salas M (1991) Protein-priming of DNA replication. *Annu Rev Biochem* **60**: 39–71
- Salas M (1999) Mechanisms of initiation of linear DNA replication in prokaryotes. In *Genetic Engineering*, Setlow KK (ed), Vol. 2, pp 159–171. New York: Kluwer Academic, Plenum Press
- Salas M, Miller JT, Leis J, DePamphilis ML (1996) Mechanisms for priming DNA synthesis. In *DNA Replication in Eukaryotic Cells*, DePamphilis ML (ed), pp 131–176. Cold Spring Harbor: Cold Spring Harbor Laboratory Press
- Simpson AA, Tao Y, Leiman PG, Badasso MO, He Y, Jardine PJ, Olson NH, Morais MC, Grimes S, Anderson DL, Baker TS, Rossmann MG (2000) Structure of the bacteriophage φ29 DNA packaging motor. *Nature* **408**: 745–750
- Steitz TA, Smerdon SJ, Jager J, Joyce CM (1994) A unified mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**: 2022–2025
- Tahirov TH, Temiakov D, Anikin M, Patlan V, McAllister WT, Vassilyev DG, Yokoyama S (2002) Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* **420**: 43–50
- Tao Y, Olson NH, Xu W, Anderson DL, Rossmann MG, Baker TS (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95**: 431–437
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D* **55**: 849–861
- Theis K, Gong P, Martin CT (2004) Topological and conformational analysis of the initiation and elongation complex of T7 RNA polymerase suggest a new twist. *Biochemistry* **43**: 12709–12715
- Yin WY, Steitz TA (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **298**: 1387–1397
- Zaballos A, Lázaro JM, Méndez E, Mellado RP, Salas M (1989) Effects of internal deletions on the priming activity of the phage φ29 terminal protein. *Gene* **83**: 187–195