SUPPORTING INFORMATION

Homology Modeling of the OB-fold Domain of *E. coli* Polymerase III α

**Subunit.** Known structures related to the putative OB-fold domain of *E. coli* Pol III α subunit were identified by searching with PSI-BLAST (1) in the non-redundant protein sequence database taken from NCBI (ftp:// ftp.ncbi.nih.gov/blast/db/). A number of matching OB-fold domains could be reliably detected including anticodon binding domains of tRNA synthetases, ssDNA binding proteins and the RecG “wedge” domain. However, these relatives share at most only ~20% identical residues with the OB-fold domain of *E. coli* Pol III α subunit. At this level of sequence similarity (often referred to as the “twilight zone”) the accuracy of a homology model to a large degree is determined by the correctness of underlying sequence-structure alignment (2). Therefore, modeling of the OB-fold domain was performed using a procedure in which the alignment is refined iteratively by building and assessing 3D models corresponding to different alignment variants. Briefly, reliable alignment regions are first delineated with the PSI-BLAST-ISS tool (3) and only the alignment in unreliable regions is adjusted. Models corresponding to different alignment variants are then built automatically with Modeller (4) and assessed both with Prosa2003 (5) and visually. Iterations are repeated until Prosa2003 energy scores cannot be improved any further and the visual assessment reveals no significant flaws in the modeled structure. Previously, this procedure proved to be successful both in a blind-mode test (6) and in detecting sequence-structure mapping errors in OB-fold domain crystal structures during large scale analysis (7).
Since the use of multiple templates may also help to improve the accuracy of the model (8), three structural templates belonging to different OB-fold families were used: anticodon binding domain of *E. coli* Aspartyl-tRNA synthetase (PDB code: 1C0A) (9), the RecG “wedge” domain from *Thermotoga maritima* (1GM5) (10) and the DNA binding domain A (DBD-A) of human Replication Protein A 70 KDa subunit (1FGU) (11) (see Figure S1). More recently, as the X-ray structure of Pol III α subunit from *Thermus aquaticus* became available (2HPI) (12) its OB-fold domain was added as the evolutionarily closest structural template. In addition to corroborating the accuracy of the conserved β-barrel core of the model, it enabled us to better represent the conformation of the flanking regions. Side chains in the final 3D model were positioned with SCWRL (13).

A favorable Prosa2003 energy Z-score (-8.0) suggests that the model (Figure 1b) is an accurate representation of the *E. coli* Pol III α subunit OB-fold domain structure. For comparison, three out of four OB-fold domains used as structural templates display worse Z-scores (1C0A: -8.2; 1GM5: -7.7; 1FGU: -7.5 and 2HPI: -7.3). Note that the poor Z-score for the OB-fold domain of *T. aquaticus* Pol III α subunit (2HPI) is due to a large number of missing residues in loops.

Bound ssDNA fragment was modeled by similarity to the structure of the related RecG “wedge” domain/DNA complex (1GM5). The OB-domain model was first superimposed with the RecG “wedge” domain. Next, the RecG structure and most of the DNA were deleted, leaving only the nucleotide making a base
stacking interaction with the conserved phenylalanine and two adjacent nucleotides.

Bound dsDNA to the (HhH)$_2$ motif was modeled by superimposing this structural motif of α with the corresponding motif of rat pol β complexed with DNA (PDB code: 2BPF) (14). Then the structure of pol β was deleted and the DNA position was optimized. The orientation of this model (Figure 1c) shows that each of the individual HhH motifs may potentially interact with the minor groove, providing non-sequence specific binding to dsDNA.

**Plasmid Construction.** Plasmids encoding *E. coli* full-length Pol III α subunit and Pol III α1-917 with an N-terminal histidine tag (pET28a-α and pET28α-α917, respectively) were generous gifts from Prof. J. Kuriyan (UC-Berkeley) (15). Plasmids encoding *E. coli* α917-1160 (pAlpha917-1160), α978-1160 (pAlpha978-1160), and α1076-1160 (pAlpha1076-1160) were constructed from plasmid pET28α-α, as follows: NdeI restriction sites were introduced at residues 917, 978, or 1076 using the QuikChange site-directed mutagenesis kit (Stratagene) and the following primers with the respective reverse complements, respectively:

5′-GAAAGCGGAACATATGGGTCAGGCCGATATG-3′, 5′-GAGATTGAGCGTTATCATATGGGCGTAAGGCTGAAAGAC-3′, and 5′-GGATATTGACGAACATATGGAAAAATATGCTCGCGGGC-3′. After digesting with NdeI, the plasmids were then ligated to yield α917-1160, *E. coli* α978-1160, and *E. coli* α1076-1160 with amino acid residues 917-1160, 978-1160, and 1076-
1160 of wild-type *E. coli* Pol III, respectively. The plasmid encoding α1-835 was constructed by introducing stop codons in pET28a-α in all three reading frames at position 836 using the QuikChange site-directed mutagenesis kit and the following primer with its reverse complement: 5′-CGAAATCGTGTATGGTATTGTAGCTAACTAGGTAATCGGTGAAGGTCC-3′. The integrity of the constructs was confirmed by using automated DNA sequencing analysis (MGH Core Facility or MIT Biopolymers Laboratory, Cambridge, MA).

**Protein Preparation.** Wild-type *E. coli* Pol III, Pol III α1-835, and Pol III α1-917 were expressed in Tuner (Novagen), Tuner pGro7 (Takara), or BL21(DE3) pLysS strains, while Pol III α917-1160, Pol III α978-1160, and Pol III α1076-1160 were expressed in Tuner pGro7. Competent cells were prepared by using the CaCl$_2$ method (16). All proteins were expressed by inducing cultures at OD$_{600}$= 0.8 with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 30 °C for 3-4 h. The cells were harvested, resuspended by sonication, and lysed by treatment with lysozyme and DNase in lysis buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 10% Glycerol, and 2 mM β-mercaptoethanol). Clarified cell extract was loaded onto 2 mL Ni-NTA-Agarose (Qiagen) and bound proteins were eluted with a step gradient at 0.01, 0.025, 0.1, 0.25, 0.5 M imidazole in lysis buffer. Pure fractions identified via SDS-PAGE were concentrated in Vivaspin centrifugal concentrators (Vivascience) and the buffer exchanged to 50 mM HEPES pH 7.5, 250 mM NaCl,
and 10% Glycerol to remove imidazole. We were unable to purify a construct consisting of only the predicted OB-fold region (residues 978-1078) or an active construct that lacked only the C-terminal 81 residues (residues 1-1079), so a C-terminal construct excluding the OB-fold (residues 1076-1160) was purified as a negative control. For experiments conducted in low salt, the buffer was exchanged to 50 mM HEPES, pH 7.5, 50 mM NaCl, 10% Glycerol by using Vivaspin centrifugal concentrators. Purified protein concentrations were determined by Bradford protein assay (Bio-Rad), and concentrated proteins were stored at -80 °C. Protein activity was confirmed over a range Na⁺ concentration (Figure S2).

**Dual Beam Optical Tweezers.** Two continuous wave infrared lasers (JDSU, Newport) are counter propagated using polarization optics (17, 18). A confocal pair of high numerical aperture water immersion microscope objectives (Nikon) focuses the beams to a diffraction limited diameter of ~1 μm. Polystyrene beads 5 μm in diameter (Bangs Labs, Spherotech) are held at the focus and any force upon the sphere is determined by the deflection of the trapping lasers, which is measured upon a pair of lateral effect photodiode detectors (Duma, Melles Griot). The deflection is proportional to the applied force once a correction is made for the finite stiffness of the trap.

Phage λ-DNA (Roche) is biotinylated (Invitrogen) on the complimentary 3′ cos overhang (19, 20). Streptavidin coated beads (Bangs Labs) are held upon a
fixed micropipette (WPI) and within the optical trap. The micropipette is glued into a custom made flow cell. Movement of this cell is controlled by a feedback controlled piezoelectric translation stage (Thor Labs, Melles Griot). DNA is ‘caught’ between the two beads and stretched as the stage is extended. Thus tension in the double helix is measured as a function of position, to precisions of 0.1 pN and 2.0 nm. A buffer of 10 mM HEPES, pH 7.5 and 100 mM Na$^+$ was used unless otherwise noted.

Data is collected by moving the stage (typically 100 nm lengths, with a resolution of 5 nm), pausing briefly (usually 1 second) to average the measured position on the lateral effect detectors. This measurement is corrected for finite trap stiffness then calibrated by noting the midpoint of the melting transition (0.45 nm per base pair, as seen in Figure 3 of the text) and setting it to a force of 62.6 ± 0.2 pN (the force is linear about the midpoint, and may be averaged over the range 0.44 to 0.46 nm per base pair) (21). The low force data (< 45 pN) may also be checked against the Worm-Like Chain model (see below) to verify the response of the instrument and the tethered DNA molecule. Finally, data were collected at varying pulling/relaxation rates, and the melting force was not observed to change within uncertainty until rates of 1000 nm/s (10× the standard rate of 100 nm/s) were exceeded.

The reproducibility of the binding to single-stranded DNA is illustrated in Figure S3. In these experiments, it is important to note that though binding and unbinding of the protein are slow, both events may be observed. To facilitate binding to ssDNA, the DNA molecule is partially melted and held (at the
extension marked by the arrow). Stabilization of single-stranded DNA by bound protein is evident upon relaxation. However, protein unbinding takes place as the tension is reduced below the melting force. If the DNA molecule is held at low tension for a fixed time, the protein may fully unbind. Extending the DNA molecule may then allow the protein to re-bind and the process repeated. Thus binding is reversible and the overlap of the extension/relaxation cycles shows the reproducibility of the experiment.

**Modeling the Elasticity of DNA.** DNA is often characterized as a flexible polymer with an additional elasticity. The continuum Worm-Like Chain model characterizes the measured, force-dependent length of dsDNA ($b_{ds}$) as a function of a (force-independent) contour length ($B_{ds}$), elastic modulus ($S_{ds}$) and persistence length ($P_{ds}$). A solution for this model in the high force limit appears as

$$b_{ds}(F) = B_{ds} \left[1 - \frac{1}{2} \left(\frac{k_B T}{P_{ds} F}\right)^{1/2} + \frac{F}{S_{ds}}\right]. \quad (S1)$$

Fits to this function for dsDNA are shown in Figure 3 and Figure 4b, yielding typical values of $B_{ds} = 0.340 \pm 0.001$ nm/bp, $P_{ds} = 48 \pm 2$ nm and $S_{ds} = 1200 \pm 100$ pN (as shown in Figure 3) (17, 21). The Freely-Jointed Chain model describes ssDNA as a sequence of flexible segments;

$$b_{ss}(F) = B_{ss} \left[\coth\left(\frac{2P_{ss} F}{k_B T}\right) - \frac{1}{2} \frac{k_B T}{P_{ss} F}\right] \left[1 + \frac{F}{S_{ss}}\right]. \quad (S2)$$
Values for ssDNA have been found previously \((B_{ss} = 0.575 \text{ nm/bp}, P_{ss} = 0.75 \text{ nm} \text{ and } S_{ss} = 800 \text{ pN})\), and are shown graphically in Figure 3 (22). These values for \(b_{ds}\) and \(b_{ss}\) are held constant for the fits of Equation 1 (which utilize equations S1 and S2), though the value for \(P_{ss}\) has been altered to 0.90 nm to reflect the persistence length of ssDNA plus bound protein. This value was used to improve the overall quality of the fits (to lower \(\chi^2\)), and this value was fixed for all of the fits to Eq. S2 in the presence of bound protein.

REFERENCES

1. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25, 3389-3402.

2. Venclovas, Č., Zemla, A., Fidelis, K., and Moult, J. (2001) Comparison of performance in successive CASP experiments, *Protein. Struct. Funct. Genet.* 45, 163-170.

3. Margelevičius, M., and Venclovas, Č. (2005) PSI-BLAST-ISS: an intermediate sequence search tool for estimation of the position-specific alignment reliability, *BMC Bioinformatics* 6, 185.

4. Šali, A., and Blundell, T. L. (1993) Comparative protein modeling by satisfaction of spatial restraints, *J. Mol. Biol.* 234, 779-815.
5. Sippl, M. J. (1993) Recognition of errors in three-dimensional structures of proteins, *Protein. Struct. Funct. Genet.* 17, 355-362.

6. Venclovas, Č., and Margelevičius, M. (2005) Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment, *Protein. Struct. Funct. Genet.* 61, 99-105.

7. Venclovas, Č., Ginalska, K., and Kang, C. (2004) Sequence-structure mapping errors in the PDB: OB-fold domains, *Protein Sci.* 13, 1594-1602.

8. Venclovas, Č. (2001) Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment, *Protein. Struct. Funct. Genet.* 45, 47-54.

9. Eiler, S., Dock-Bregeon, A., Moulinier, L., Thierry, J. C., and Moras, D. (1999) Synthesis of aspartyl-tRNA(Asp) in Escherichia coli--a snapshot of the second step, *EMBO J.* 18, 6532-6541.

10. Singleton, M. R., Scaife, S., and Wigley, D. B. (2001) Structural analysis of DNA replication fork reversal by RecG, *Cell* 107, 79-89.

11. Bochkareva, E., Belegu, V., Korolev, S., and Bochkarev, A. (2001) Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding, *EMBO J.* 20, 612-618.

12. Bailey, S., Wing, R. A., and Steitz, T. A. (2006) The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases, *Cell* 126, 893-904.
13. Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction, *Protein Sci.* 12, 2001-2014.

14. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP, *Science* 264, 1891-1903.

15. Lamers, M. H., Georgescu, R. E., Lee, S. G., O'Donnell, M., and Kuriyan, J. (2006) Crystal structure of the catalytic alpha subunit of *E. coli* replicative DNA polymerase III, *Cell* 126, 881-892.

16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

17. McCauley, M. J., and Williams, M. C. (2007) Mechanisms of DNA binding determined in optical tweezers experiments, *Biopolymers* 85, 154-168.

18. Smith, S. B., Cui, Y., and Bustamante, C. (2003) Optical-trap force transducer that operates by direct measurement of light momentum, *Methods Enzymol.* 361, 134-162.

19. Williams, M. C., Wenner, J. R., Rouzina, I., and Bloomfield, V. A. (2001) The effect of pH on the overstretching transition of dsDNA: Evidence of force-induced DNA melting, *Biophys. J.* 80, 874-881.

20. Williams, M. C., Wenner, J. R., Rouzina, I., and Bloomfield, V. A. (2001) Entropy and heat capacity of DNA melting from temperature dependence of single molecule stretching, *Biophys. J.* 80, 1932-1939.
21. Wenner, J. R., Williams, M. C., Rouzina, I., and Bloomfield, V. A. (2002) Salt dependence of the elasticity and overstretching transition of single DNA molecules, *Biophys. J.* 82, 3160-3169.

22. Smith, S. B., Cui, Y. J., and Bustamante, C. (1996) Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules, *Science* 271, 795-799.

23. Beuning, P. J., Simon, S. M., Godoy, V. G., Jarosz, D. F., and Walker, G. C. (2006) Characterization of *Escherichia coli* translesion synthesis polymerases and their accessory factors *Methods Enzymol.* 408, 318-340.

**FIGURE LEGENDS**

**Figure S1.** Multiple sequence alignment of OB-fold domains, including residues 978-1078 of the α subunit. Aligned sequences include other bacterial homologs of *E. coli* pol III α subunit (DPO3A_ECOLI) and structural templates used in modeling. Conserved residues shared by more than half of aligned sequences are highlighted in red (identical) and pink (similar). The predicted secondary structure corresponding to the OB-fold model is indicated above the alignment and displayed in Figure 1b of the main text. ECOLI, *Escherichia coli*; HAEIN, *Haemophilus influenzae*; NEIMA, *Neisseria meningitides*; MYCTU, *Mycobacterium tuberculosis*; DEIRA, *Deinococcus radiodurans*; AQUAE, *Aquifex aeolicus*; STAAS, *Staphylococcus aureus*; BACHD, *Bacillus halodurans*. 
Figure S2. Primer extension products of *E. coli* pol III alpha subunit determined as a function of [Na⁺]. Assays were carried out as described in (23) with 600 nM alpha, time points of 0, 1, 10, and 20 min, without KCl, and with the NaCl concentrations indicated. Products were analyzed by 10% polyacrylamide gel electrophoresis.

Figure S3. The reproducibility of force extension and relaxation data in the presence of the full-length α subunit is shown as solid and dotted lines, respectively. After data collection on bare DNA as an instrument calibration, a buffer containing 100 nM full-length α was added. The first extension/relaxation cycle included a 30 minute pause at the maximum extension shown at the blue arrow, and described in the text. The relaxation clearly shows evidence of ssDNA binding by the protein. The DNA was held at low tension (≤ 1 pN) for ~ 20 minutes, then another cycle was collected, with no pause at longer extensions (green). No binding to ssDNA is observed in this cycle, as the protein has become unbound in the interval between cycles. A subsequent cycle includes another 30 minute pause at maximum extension (yellow and arrow), followed by a 20 minute rest at low tension, and a final cycle (red) show this effect to be completely reproducible and that the protein may be completely unbound if the DNA is held at low tension.
Supplemental Figure 1
Supplemental Figure 2
Supplemental Figure 3