Supplemental Figure Legends

Fig. S1. Schematic representation of multi-subunit RNAP large subunit sequence characteristics.

A. Schematic representation of the *Thermus* RNAP β subunit sequence. Regions shared among all RNAPs are color-coded and labeled. Additional regions shared among bRNAP are colored teal. Above each region shared among all RNAPs is a histogram showing the Blosum62 information score (scale on the left) for each residue, as determined by the program PFAAT [1]. The secondary structure is shown directly above the sequence bar (helices, black rectangles; β-strands, grey rectangles). Important structural features are denoted above that. Below the sequence bar, small grey numbers (vertically oriented 100, 200, etc.) denote the residue numbering of the *Thermus* β-subunit. The approximate insertion points of the lineage-specific insertions are denoted by cyan circles and labeled according to Lane & Darst [2]. The position and roles (when known) of highly conserved residues are denoted below. At the bottom, the horizontal lines denote segments of β shared regions that participate in conserved interactions with other shared regions in β or β'.

B. Schematic representation of the *Thermus* RNAP β' subunit sequence. As in (A) except regions shared among bRNAP are colored dark pink.

Fig. S2. Role of highly conserved βLeu997 (βa15) in helping to position highly conserved βGln567 (βa10), and βLys838 and βLys846 (βa14), which interact with the RNA transcript.

A. Schematic representation of the sequence context of βa10 (green), βa14 (red), and βa15 (blue) in *Thermus* RNAP. As in Fig. S1.

B. (Inset, lower-right) β-side view of the *Tth* bRNAP TEC structure (PDB 2O5J) [3]. The RNAP is shown as a backbone worm and color-coded as follows: αI, αII, ω, grey; β, cyan (except βa10, βa14, and βa15 are colored green, red, and blue, respectively); β', pink. The nucleic acids of the TEC are shown in CPK format: template DNA, dark green; nontemplate DNA, light green; RNA transcript, orange. The boxed region is magnified above. (Magnified view) βa10, βa14, and βa15 are shown as transparent backbone worms and color-coded as in (A). Other parts of RNAP are not shown for clarity. The active-site MgI and MgII are shown as yellow spheres. The incoming nucleotide substrate is shown in stick format with carbon atoms colored blue. Highly conserved residues βGln567, βLys838, βLys846, and βLeu997 are shown in CPK format. βGln567 interacts with O3' of the -3 RNA position. βLys838 and βLys846 interact with the RNA phosphate backbone at the -1 position. βLeu997 makes van der Waals contacts that appear to position βGln567, βLys838, and βLys846 for these interactions with the RNA transcript. The thick black arrow points in the downstream direction (the direction of RNAP transcription).

Fig. S3. Interaction between highly conserved residues β'Leu709 (β'a12), β'Thr1088 (β'a15; Bridge helix), and β'Thr1234 (β'a16; TL helix1).

A. Schematic representation of the sequence context of β'a12 (magenta), β'a15 (yellow), and β'a16 (cyan) in *Thermus* RNAP. As in Fig. S2.

B. (Inset, upper-right) β'-side view of the *Tth* bRNAP TEC structure (PDB 2O5J) [3]. The RNAP is shown as a backbone worm and color-coded as follows: αI, αII, ω, grey; β, cyan; β', pink (except β'a12, β'a15, and β'a16 are colored magenta, yellow, and cyan, respectively). The nucleic acids of the TEC are shown as backbone worms: template DNA, dark green; nontemplate DNA, light green; RNA transcript, orange. The boxed region is magnified below. (Magnified view) β'a12, β'a15, and β'a16 are shown as transparent backbone worms and color-coded as in (A). Other parts of RNAP are not shown for clarity. The active-site MgI and MgII are shown as yellow spheres. The incoming nucleotide substrate is shown in stick format with carbon atoms colored blue. Highly conserved residues β'Leu709, β'Thr1088, and β'Thr1234 are shown in CPK format. The thick black arrow points in the downstream direction (the direction of RNAP transcription).

Fig. S4. Dual DPBB architecture of the RNAP active site.

A. Schematic representation of the sequence context of the βDPBB (top) and the β'DPBB (bottom) in *Thermus* RNAP. Regions shared among all RNAPs are color-coded dark blue or magenta (regions within βDPBB or β'DPBB, respectively) or a lighter shade of blue or magenta (additional shared regions, β and β' respectively). The secondary structure is shown directly above the sequence bar (helices, black rectangles; β-strands, grey rectangles). The 6 core β-strands of the DPBB domain [4] are colored blue (βDPBB) or magenta (β'DPBB) and labelled (S1-S6). Important structural features are denoted above that. Below the sequence bar, small grey numbers (vertically oriented 100, 200, etc.) denote the residue numbering of the *Thermus* subunit. The approximate insertion points of the lineage-specific insertions are denoted by cyan circles (β) and labeled according to Lane & Darst [2]. The position and roles (when known) of highly conserved residues are denoted below. At the bottom, the horizontal lines denote segments of β shared regions that participate in conserved interactions with other shared regions in β or β'.

B. (Inset, lower-right) Front view of the *Tth* bRNAP structure (PDB 2O5J) [3] with the nucleic acids removed. The RNAP is shown as a backbone worm and color-coded as follows: αI, αII, ω, grey; β, cyan; β', pink (except the shared regions are colored according to Fig. S4A) The boxed region is magnified above. (Magnified view) Obscuring parts of the RNAP have been removed for clarity. The core β-strands of the β and β' DPBB domains are shown as backbone ribbon arrows (color-coded as in

Fig. S4A). The rest of the RNAP is shown as a backbone worm. The active-site MgI is shown as a yellow sphere.

Fig. S5. The βDPBB.

A. Schematic representation of the sequence context of the βDPBB in *Thermus* RNAP. Regions shared among all RNAPs are color-coded dark colors (regions within the βDPBB) or lighter colors (additional shared regions). The secondary structure is shown directly above the sequence bar (helices, black rectangles; β-strands, grey rectangles). The 6 core β-strands of the DPBB domain [4] are colored according to their shared region and labelled (S1-S6). Important structural features are denoted above that. Below the sequence bar, small grey numbers (vertically oriented 100, 200, etc.) denote the residue numbering of the *Thermus* subunit. The approximate insertion points of the lineage-specific insertions are denoted by cyan circles and labeled according to Lane & Darst [2]. The position and roles (when known) of highly conserved residues are denoted below. At the bottom, the horizontal lines denote segments of the shared regions that participate in conserved interactions with other shared regions.

B. (Inset, lower-left) Front view of the *Tth* bRNAP structure (PDB 2O5J) [3] with the nucleic acids removed. The RNAP is shown as a backbone worm and color-coded as follows: αI, αII, ω, grey; β, cyan; β', pink (except the shared regions are colored according to Fig. S5A) The boxed region is magnified above. (Magnified view) Obscuring parts of the RNAP have been removed for clarity. The core β-strands of the βDPBB domain are shown as backbone ribbon arrows (color-coded as in Fig. S5A). The

rest of the RNAP is shown as a backbone worm. The active-site MgI is shown as a yellow sphere.

Fig. S6. The β'DPBB.

A. Schematic representation of the sequence context of the β'DPBB in *Thermus* RNAP. Regions shared among all RNAPs are color-coded dark colors (regions within the β'DPBB) or lighter colors (additional shared regions). The secondary structure is shown directly above the sequence bar (helices, black rectangles; β-strands, grey rectangles). The 6 core β-strands of the DPBB domain [4] are colored according to their shared region and labelled (S1-S6). Important structural features are denoted above that. Below the sequence bar, small grey numbers (vertically oriented 100, 200, etc.) denote the residue numbering of the *Thermus* subunit. The position and roles (when known) of highly conserved residues are denoted below. At the bottom, the horizontal lines denote segments of the shared regions that participate in conserved interactions with other shared regions.

B. (Inset, lower-left) Front view of the *Tth* bRNAP structure (PDB 2O5J) [3] with the nucleic acids removed. The RNAP is shown as a backbone worm and color-coded as follows: αI, αII, ω, grey; β, cyan; β', pink (except the shared regions are colored according to Fig. S6A) The boxed region is magnified above. (Magnified view) Obscuring parts of the RNAP have been removed for clarity. The core β-strands of the β'DPBB domain are shown as backbone ribbon arrows (color-coded as in Fig. S6A). The rest of the RNAP is shown as a backbone worm. The active-site MgI is shown as a yellow sphere.

Fig. S7. Structural conservation between DDRPs and RDRPs.

A. Schematic representation of the sequence context of the *Thermus* DDRP β and β' subunits (top and bottom, respectively) and the catalytic domain of the RDRP QDE-1 (middle). For DDRP β and β', regions shared among all DDRPs are color-coded grey, except regions showing structural conservation with the RDRP are colored dark blue (βDPBB domain) or cyan (other regions of β), or magenta (β'DPBB) or pink (other regions of β'). Structurally conserved regions of QDE-1 are colored green (DPBB1), red (DPBB2) or orange (other conserved regions). The secondary structure of each protein is shown directly above the sequence bar (helices, black rectangles; β-strands, grey rectangles, or colored according to the structural conservation). The 6 core β-strands of the DPBB domains [4] are labelled (S1-S6). Important structural features of the DDRP β/β' subunits are denoted above the secondary structure. Below the sequence bars, small grey numbers (vertically oriented 100, 200, etc.) denote the residue numbering. The approximate insertion points of the lineage-specific insertions in the DDRP subunits are denoted by cyan (β) or magenta (β') circles and labeled according to Lane & Darst [2]. The position and roles (when known) of highly conserved residues shared between the DDRP and RDRP are denoted below the sequence bars.

B. Stereo view showing only the conserved structural features of the DDRPs (PDB 205J) [3] and the RDRP QDE-1 (PDB 2J7N) [5]. The two structures were superimposed on the two DPBB domains only. The proteins are shown as backbone ribbons and colored according to Fig. S7A. Side chains of shared, conserved residues are shown and labeled.

References

1. Caffrey, D. R., Dana, P. H., Mathur, V., Ocano, M., Hong, E.-J., Wang, Y. E., Somaroo, S., Caffrey, B. E., Potluri, S. & Huang, E. S. (2007). PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. *BMC Bioinformat.* 8, 381.

2. Lane, W. J. & Darst, S. A. (2009). Molecular evolution of multi-subunit RNA polymerases: sequence analysis.

3. Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I. & Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448, 163-168.

4. Iyer, L. M., Koonin, E. V. & Aravind, L. (2003). Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* 3, 1-23.

5. Salgado, P. S., Koivunen, M. R. L., Makeyev, E. V., Bamford, D. H., Stuart, D. I. & Grimes, J. M. (2006). The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol.* 4, 2274-2281.