The Extended "Two-Barrel" Polymerases Superfamily: Structure, Function and Evolution

Ludovic Sauguet

The extended 'two-barrel' polymerases superfamily: structure, function and evolution

Ludovic Sauguet[1]

1 Institut Pasteur, Unité de Dynamique Structurale des Macromolécules, 75015 Paris, France

Correspondence to Ludovic Sauguet: lsauguet@pasteur.fr

Abstract

DNA and RNA polymerases play central roles in genome replication, maintenance and repair, as well as in the expression of genes through their transcription. Multisubunit RNA polymerases (msRNAPs) carry out transcription and are represented, without exception, in all cellular life forms as well as in nucleo-cytoplasmic DNA viruses. Since their discovery, msRNAPs have been the focus of intense structural and functional studies revealing that they all share a well-conserved active site region called the two-barrel catalytic core. The two-barrel core hosts the polymerase active site, which is located at the interface between two double-psi β-barrel (DPBB) domains that contribute distinct amino acid residues to the active site in an asymmetrical fashion. Recently, sequencing and structural studies have added a surprising variety of DNA and RNA polymerases (DNAPs and RNAPs) to the two-barrel superfamily, including the archaeal replicative DNAP (PolD), which extends the family to DNA-dependent DNAPs involved in replication. While all these polymerases share a minimal core that must have been present in their common ancestor, the two-barrel polymerase superfamily now encompasses a remarkable diversity of enzymes, including DNA-dependent RNAPs, RNA-dependent RNAPs and DNA-dependent DNAPs, which participate in critical biological processes such as DNA transcription, DNA replication, and gene silencing. The present review will discuss both common features and differences among the extended two-barrel polymerase superfamily, focusing on the newly discovered members. Comparing their structures provides insights on the molecular mechanisms evolved by the contemporary two-barrel polymerases to accomplish their different biological functions.

Introduction

Polymerases are essential for life being responsible for preserving genetic information by replicating and repairing nucleic acid molecules as well as for the expression of genes through transcription (1). While all nucleotide polymerases likely use a two-metal ion catalytic general mechanism (2–4), nucleotide polymerization has been invented several times during evolution. Most nucleotide polymerases can be divided into three major groups (5,6). The first group contains nucleotide polymerases that structurally resemble the *E. coli* Pol-I Klenow-fold (7). Their overall fold is characterized by thumb, finger and palm subdomains and is shared by most eukaryotic, archaeal, and viral replicative DNA polymerases (DNAPs), the bacteriophage single-subunit RNA polymerase (RNAP), the mitochondrial RNAP, and reverse-transcriptases (8,9). The second group, often referred to as Polβ-like polymerases, includes the bacterial replicative DNAPs and several eukaryotic DNAPs involved in DNA repair (10). While they share a similar three-dimensional arrangement of catalytic aspartates in the active site with the first group, they show a completely different topology of the palm subdomain. The third group includes multisubunit RNA polymerases (msRNAPs), which carry out transcription and is represented, without exception, in all cellular life forms as well as in nucleo-cytoplasmic DNA viruses. Since their discovery about 60 years ago, cellular msRNAPs have been the focus of intense structural and functional studies revealing that they all share a well-conserved active site region, consisting of two double-psi β-barrel (DPBB) subdomains (11,12). However, sequencing and structural studies have added a surprising variety of nucleotide polymerases to the two-barrel superfamily (13–15) (**Table 1**). Recently, the two-barrel polymerase superfamily was substantially diversified with the structure elucidation of PolD, an archaeal replicative DNAP, which extends the family to DNA-dependent DNAPs involved in replication (16,17).

All two-barrel polymerases share a common catalytic core, which is located at the interface between two DPBB subdomains that contribute distinct amino acid residues to the active site in an asymmetrical fashion (**Table 1**). The first DPBB (DPBB-1, respectively DPBB-B in msRNAPs) harbors two canonical lysine residues involved in substrate binding, while the second DPBB (DPBB-2, respectively DPBB-A in msRNAPs) contributes invariant aspartic residues that coordinate a catalytic Mg^{2+}. The present review will discuss both common features and differences among members of the extended two-barrel polymerase superfamily, focusing on the newly discovered members. While all these polymerases share a minimal core that must have been present in their common ancestor, the two-barrel polymerase superfamily now encompasses a remarkable diversity of enzymes, including DNA-dependent RNAPs, RNA-dependent RNAPs (18,19), and DNA-dependent DNAPs, which participate to critical biological processes such as DNA transcription, DNA replication, and gene silencing. Comparing their structures provides insights on the remarkable variety of molecular mechanisms, which were evolved by the contemporary two-barrel polymerases in order to achieve their different biological functions.

Multi-subunit RNA polymerases: the remarkably conserved cellular transcriptases

In all forms of cellular life, the genomes are transcribed by a highly conserved family of msRNAPs, which was discovered about 60 years ago and has since then been the focus of intense structural and functional studies (11). Bacteria and archaea have one single type of msRNAP to transcribe their gene repertoire, whereas eukaryotes have three RNAPs, responsible mainly for the synthesis of ribosomal RNA (RNAPI), pre-messenger RNA (RNAPII), and small RNAs including transfer RNAs (RNAPIII) respectively (**Table 1**). Plastids encode two types of RNAPs: two different single-subunit, nuclear-encoded phage-type RNAPs, and one plastid-encoded multi-subunit RNAPs, whose subunits are homologous to the bacterial RNAP (20). Over the past two decades, X-ray crystallographic and cryo-EM structures have been determined for msRNAPs from the three domains of life, revealing that these enzymes operate by closely related molecular mechanisms (12).

All cellular msRNAPs share an exceptionally conserved large active site region, composed of five core subunits, which resemble a crab claw, the jaws of which interact with the downstream duplex DNA template (21–25)(26,27) (**Fig. 1**). The two large subunits form the core of the enzyme and host the polymerase catalytic center (Rpb1 and Rpb2 in Eukaryotic RNAPII, Rpo1 and Rpo2 in archaeal RNAP, β and β' in bacterial RNAP), which have been pinpointed by extensive biochemical and genetic analyses (28–30) (**Table 1**). These two catalytic subunits are anchored and stabilized by three smaller "assembly" subunits (the Rpb3/Rpb11 heterodimer and Rpb6 in Eukaryotic RNAPII; the Rpo3/Rpo11 heterodimer and Rpo6 in archaeal RNAP; the α/α homodimer and ω subunits in bacterial RNAP) (11). In addition, the activity of msRNAPs is modulated by several transcription factors during the three distinct phases of the transcription cycle: initiation, elongation and termination. The active site of msRNAPs is located at the interface between the two core subunits and can be dissected within several functionally critical regions (21,31) (**Fig. 1**)

In the transcribing RNAP, the downstream DNA traverses the major DNA-binding channel, lining the "floor", until it encounters the active center of the RNAP "wall". The DNA-binding channel forms a positively charged cleft, lined from one side by a mobile clamp, and from the other side by the lobe and protrusion domains. The DNA-RNA hybrid rises up from the active site perpendicular to the downstream duplex DNA. The hybrid and clamp domains secure about one helical turn of the DNA-RNA hybrid, whose strands are separated by the RNAP lid. The catalytic core harbors the active site, including the Mg^{2+}-bound two-barrel catalytic core, the bridge, the DNA-RNA hybrid-binding helix and trigger helix, which are essential for the nucleotide translocation cycle (32–36). The two-barrel catalytic center is organized around two catalytic Mg^{2+} ions, each Mg^{2+} being harbored by one of the two barrels. The DPBB-A of the largest RNAP catalytic subunit carries three invariant aspartic residues in the highly conserved NADFDGD motif, which coordinates a catalytic Mg^{2+} ion. A second Mg^{2+} is required for catalysis but comes with the substrate. The DPBB-B of the second catalytic subunit possesses two invariant lysine residues, whose side chains face the last two nucleotides of the RNA primer in the elongating bacterial and eukaryotic RNAPs (37). The bridge helix and the two-barrel catalytic center line a perforation in the floor of the cleft, which widens towards the exterior, creating an inverted funnel, named the pore, or secondary channel. The

pore allows substrates and cleavage factors to access the active site, and allows extrusion of the transcript during backtracking. The outer rim of the funnel is lined by the funnel region.

Non-canonical multi-subunit RNA polymerases from DNA viruses

In addition to being present in all forms of cellular life, msRNAPs are encoded by several lineages of nucleo-cytoplasmic DNA viruses (NCLDVs), which are one of the largest viral division that includes viruses with large DNA genomes that infect diverse eukaryotes (38). NCLDVs encode their own msRNAP, which presumably derives from the eukaryotic RNAPI form (39) and operates during the cytoplasmic phase of their infectious cycle, when the viral DNA is not accessible to the host nuclear transcription system. Their multi-subunit structure systematically includes the two largest subunits, Rpb1 and Rpb2, as well as several 'assembly' subunits, which play a critical role in the organization of the active site region (**Table 1**) (40). Interestingly, several DNA viruses encode non-canonical msRNAPs that do not include any classical 'assembly' subunits.

For example, the well-characterized insect Baculovirus encodes a non-canonical msRNAP, which is responsible for transcribing a relatively small number of viral late genes, including the polyhedrin gene, which is used to overexpress foreign proteins in the baculovirus expression systems (41–43). These msRNAPs are made of four subunits: p47, LEF4, LEF8 and LEF9 (LEF stands for late expression factor), all with unknown structures (**Table 1**) (44). These subunits are unrelated to canonical RNAPs, except for a limited similarity of its LEF-9 and LEF-8 subunits to Rpb1 and Rpb2, respectively (45–47). An updated sequence analysis revealed that this region of homology not only includes the two barrels, but encompasses also the switch 2 domain and the funnel domain (13). However, the two subunits (LEF4 and p47) have no sequence similarity with any classical 'assembly' subunit, nor with any known msRNAP subunits (45,48,49).

Recent sequencing and biochemical studies have revealed that giant phages encode non-canonical msRNAPs (50). While the transcription of DNA phages usually depends on a phage-encoded Klenow-like monomeric RNAP or on the host transcription system, the sequencing of the giant phages from *Pseudomonas aeruginosa* (ΦEL, ΦKZ180) and *P. chlororaphis* (201Φ2) have revealed two sets of proteins distantly related to the β' and β subunits of bacterial msRNAPs (50,51). Biochemical characterization of the non-canonical msRNAP encoded by the ΦKZ giant phage has revealed that it specifically initiates transcription of late phage promoters in a rifamycin-resistant manner (52). The ΦKZ phage-encoded msRNAP includes five subunits: four subunits are homologous to the cellular msRNAPs; the fifth subunit is a protein of unknown function (**Table 1**). Like the msRNAP from Baculovirus, the ΦKZ msRNAP lacks identifiable 'assembly' and promoter specificity subunits characteristic for cellular msRNAPs.

Non-canonical msRNAPs from insect viruses (baculoviral and nudiviral lineages) and giant phages may respectively originate from eukaryotic (insect viruses) and bacterial (phages) RNAP ancestor genes, which underwent a fast and highly divergent evolution obliterating anything that was not strictly needed for DNA transcription (13). Strikingly, none of these enzymes harbors a subunit that may play a role in RNAP assembly. The reduced repertoire of the different viral msRNAPs supports the idea that primordial msRNAPs were mainly

composed of the two catalytic core subunits. This hypothesis is further supported by the fact that a variety of two-barrel polymerases carry the entire active site on a single polypeptide chain, suggesting that they are closer to the ancestral protein than their msRNAPs counterparts.

Single-chain RNA- and DNA-dependent RNA polymerases

Eukaryotic cells use small RNAs (sRNAs) guides to limit proliferation of viruses and transposable elements, maintain proper chromosomal structure, and control gene expression (53–55). In Fungi, plants, and several animal systems, sRNA production often depends on template-dependent RNA synthesis catalyzed by single-chain RNA-dependent RNAPs (56). RNA-dependent RNAPs have been initially described in plants as a cellular RNA polymerase activity induced upon viral infection that synthesized antisense RNA in a primer-dependent or independent manner (57–59). Subsequent genetic studies showed that mutations in RNA-dependent RNAP genes impaired gene silencing in a variety of systems (57,60). Within the same organism, they often coexist under structurally and functionally diverse forms such as the three paralogous genes present in *Neurospora crassa*, one of the best studied models for gene silencing (60)(61). *N. crassa* thus encodes three distinct RNA-dependent RNAPs: SAD-1, that is involved in meiotic silencing (62); RRP-3 and QDE-1, which participate in post-transcriptional gene silencing (60,61). Isolation and biochemical characterization of recombinant QDE-1 orthologs from related fungi shows that their mode of action differs considerably. While *N. crassa* QDE-1 efficiently produces full-length copies of short 9-21 nucleotides copies scattered throughout the input ssRNA templates and is capable to extend complementary primers (63), *Thielavia terrestris* QDE-1 produces predominantly short RNA copies via primer independent initiation (64).

Crystal structures of large C-terminal fragments of QDE-1 from both *N. crassa* and *T. terrestris* have been determined by X-ray crystallography (19,64). QDE-1 forms a functional homodimer, which was confirmed by gel filtration, sedimentation assays (19), and electron microscopy (64). Consistently, crystal structures of the QDE-1 dimers show a pyramidal shape with the two-barrel catalytic core located at the base of each subunit (**Fig. 2A**). Each subunit includes a catalytic domain that houses the two-barrel catalytic core and the catalytic loop. The catalytic loop hosts three canonical aspartate residues that coordinate a Mg^{2+} ion, which has been shown to be required for enzymatic activity (**Fig. 2B**) (63). While they are arranged sequentially on a single polypeptide chain, the DPBBs subdomains in QDE-1 and msRNAPs are structurally very similar and almost identically disposed. The two-barrel catalytic core forms a cleft that is surrounded by the head and slab domains, which form an extensive groove that is well adapted for accommodating the template and the nascent RNA product. In addition, QDE-1 also contains the flap subdomain peripheral to the active site cleft and the neck domain that connects the catalytic domain to the head domain. The QDE-1 structure shows a highly-positively charged channel formed between the slab and head of each subunit, which leads to the active site and has been proposed to accommodate the nascent dsRNA product (19). A second charged tunnel may be a route for the entry of NTPs. QDE-1 shares with its DNA-dependent msRNAPs counterparts the two-barrel catalytic core (18), the bridge helix and the proposed NTP tunnel that matches with the one proposed in yeast RNAPII (37). The mobile head domain may be equivalent to the clamp domain in multi-subunit RNAPs by closing down on the slab domain to stabilize the RNA product during polymerization. Interestingly, the

proposed dsRNA product-binding channels are not identical in the two subunits and adopt either an open or a closed conformation. This is mainly due to the disposition of the head domains that is different in the two subunits. In the subunit showing a closed conformation, the head domain is clamped down on the active site cleft, while in the subunit showing an open conformation, rotation of the head domain provides space for an RNA duplex (19) (**Fig. 2B**). The structure suggests that only one catalytic site is active, represented by the closed conformation, with the inactive subunit being held open. The dimeric QDE-1 polymerase has been proposed to work as a "two-stroke" engine (19). In this mechanism, substrate binding to one active site would prime the other, thereby enabling an efficient re-initiation, leading to the effective production of appropriate-length dsRNA triggers.

Iyer and colleagues (18) identified a similarity between the single-chain RNA-dependent RNAPs and the YonO protein of *B. subtilis*, originating from an inserted prophage SPβ. Yono-like open-reading frames have since then been found in several other firmicute phages (13,65). Recently, YonO was been shown to be a highly processive DNA-dependent RNAP that specifically transcribes the late genes of the SPβ prophage (66). Strikingly, the homology of YonO to msRNAPs is restricted to only few amino acids, and most conserved domains of msRNAPs are absent in YonO. Instead, as noted previously (18), YonO shows a much stronger similarity to its single-chain two-barrel RNA-dependent RNAPs counterparts. Interestingly, a putative 3D model covering ~70% of the YonO sequence could be predicted using Phyre 2 (67), based on the crystal structure of *N. crassa* QDE-1 (**Fig. 2C**). This similarity encompasses most of the catalytic and clamp domains of QDE-1. Interestingly, QDE1 regions contributing to the interactions between the two monomers corresponds to insertions that appear to be absent in YonO, thereby suggesting that homodimerization is a specificity of RNA-dependent RNAPs. However, biochemical and structural data are required to confirm this model and decipher the molecular specificities of single-chain YonO-like DNA-dependent RNAPs over single-chain RNA-dependent RNAPs.

PolD: the archaeal DNA-dependent two-barrel replicative DNAP

In all forms of cellular life, DNAPs play central role in genome replication, maintenance and repair. Over the years, all DNAPs have been grouped in different families, using sequence alignments: PolA, PolB, PolC, PolD, PolX, PolY and reverse transcriptases (8,9,68). The main replicative DNAPs from Eukarya are found in family B, Bacteria in family C, and Archaea in families B and D (6). PolD is a heterodimeric replicative DNAP composed of a large catalytic subunit (DP2) and a smaller subunit with 3′-5′ proofreading exonuclease activity (DP1) (69–71). PolD has been shown to be essential for cell viability (72–74) and is widely distributed among Archaea, being present in all four major superphyla: Euryarchaeota (including the methanogenic human symbionts); DPANN (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota); the emerging Asgard superphylum; and TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota), only absent from Crenarchaeota (75). Based on biochemical evidence, it has been proposed that PolD may act soon after initiation by the primase (76,77). Resolving the individual crystal structures of the DP1 and DP2 catalytic cores from the archaeon *Pyrococcus abyssi* revealed that PolD is an atypical DNAP that has all functional properties of a replicative DNAP but with the catalytic

core of a two-barrel RNAP (16). Recently, the cryo–electron microscopy (cryo-EM) structure of the DNA-bound heterodimeric DP1–DP2 PolD complex was reported (17). This structure sheds light on DNA-binding domains evolved by PolD to perform DNA replication and extends the repertoire of protein domains known to be involved in DNA replication (**Fig. 3A**).

All replicative DNAPs have evolved protein domains named palm, fingers, and thumb domains arranged to form the DNA-binding cleft. The palm domain carries the catalytic residues, the fingers domain drapes over the nascent base pair, and the thumb domain holds the DNA duplex during replication and contributes to processivity (78,79). PolD has evolved a specific claw-shaped active site, at the center of which is the two-barrel catalytic core and at the edge of which are three zinc-binding modules named the Zn-I, Zn-II, and Zn-III domains. The DNA substrate is cradled between a bipartite clamp domain, named 1 and 2, emanating from DPBB-1 and DPBB-2, respectively (**Fig. 3B**). Clamp-1 and clamp-2 domains contribute a central cleft with a diameter of 30 Å, which is located upstream of the DP2 polymerase catalytic center. The Zn-III domain plays a critical role in replication by interacting with the minor groove of the nascent duplex (80). The Zn-I and Zn-II domains contribute to stabilizing clamp-2, which is otherwise composed mainly of loops. The clamp domain is barricaded from one side by a KH-like domain located in the N-terminal region of DP2, which is ideally located to orient the DNA template in the active site. This domain shares structural homology with the archaeal/bacterial type-II KH domains, which are ancestral single-stranded nucleic acid–binding folds (81). KH domains had been predicted in the N-terminal region of bacterial PolC (82), and the cryo-EM structure of PolD shows in a structural context, a KH domain associated with a replicative DNAP. It is noteworthy that KH domains are found in transcription factors, including in the NusA elongation factor, which harbors a type-I KH domain (83,84). Two accessory domains, named -1 and -2, emanating from insertions in DPBB-1 and DPBB-2 subdomains, respectively, play a structural role by scaffolding together the essential two-barrel catalytic cores and the clamp-1 and clamp-2 domains. The KH-like domain is connected to the anchor domain, which firmly attaches it to the clamp-2 domain and orients it in the active site (**Fig 3A**).

The exonuclease and polymerase active sites of PolD are hosted by two distinct subunits (71,85) (**Fig. 3**). DP1 shows an oligonucleotide binding (OB) domain that is inserted within the N-terminal region of a large Mre11-like nuclease phosphodiesterase domain (PDE) (86) whose active site entry faces the 3′ end of the nascent DNA strand. The DP1 nuclease and DP2 polymerase active sites sandwich the 3′ end of the nascent DNA strand and are about 40 Å distant from each other, a feature shared with other replicative DNAPs (79). Although the cryo-EM PolD structure shows no contacts between DP1 and the bound DNA, the exonuclease active site is suitably located to catch the 3′ end of the nascent DNA strand. The phosphate moiety of the 3′-terminal nucleotide of the primer lies at 25 Å away from the exonuclease active site of DP1, which could be accounted for by a 4-nt-long single-stranded DNA, a value shared with other DNAPs with proofreading activity (79). It is noteworthy that msRNAPs have evolved a proofreading mechanism called backtracking, which involves a nucleolytic reaction that occurs in the two-barrel catalytic core, at the same site as polymerization (87). While PolD also hosts a polymerase two-barrel core, it evolved another proof-reading mechanisms, which recruited a

PDE domain. In that sense, PolD differs from its two-barrel counterparts, but resembles other replicative DNAPs, which contain a distinct nuclease-specific active site.

The minimal catalytic core shared by all two-barrel polymerases

The two-barrel polymerase superfamily encompasses DNA-dependent RNAPs, RNA-dependent RNAPs (13,18,19), and DNA-dependent DNAPs (16). Comparing their structures enables us to delineate the minimal core shared by all these distantly related DNAPs or RNAPs and discuss the molecular basis for their respective substrate specificities (**Fig. 4**). All these polymerases share a two-barrel catalytic core with several invariant amino acids located in the immediate vicinity of their catalytic Mg^{2+} cations. Superimposing these two-barrel polymerases reveales that the relative disposition of their DPBB subdomains is highly conserved. Interestingly, perturbing the disposition of the two barrels in the individual DP2 crystal structure of PolD severely impairs the activity of PolD (16,17). All two-barrel polymerases share an α-helix that is connected to the N-terminal end of DPBB-1 (DPBB-B, in msRNAPs) and shows a conserved orientation with respect to the two-barrel catalytic core and the nascent duplex in both PolD and RNAPII DNA-bound structures (**Fig. 4A**). This helix, arbitrarily named duplex-binding helix, contains two canonical basic residues and shares a similar orientation with respect to the two-barrel catalytic core and the nascent duplex in both families of enzymes (88). Finally, PolD and two-barrel RNA polymerases share a loop, arbitrarily named DPBBs-connector, that emanates from the C-terminal end of DPBB-1 (DPBB-B, in msRNAPs) and is connected to the adjacent DPBB through secondary structure interactions. The DPBBs-connector likely participates in maintaining the correct relative disposition of the two DPBBs: in the individual crystal structure of DP2, the DPBBs-connector is found to be disordered and the canonical disposition of the two DPBBs is lost. It is noteworthy that the active sites of both PolD and msRNAPs orient the DNA template entry and the nascent duplex exit in a same relative axis with respect to the two-barrel catalytic core.

A structure-based alignment of the two-barrel polymerases sequences over the structurally homologous catalytic core reveals that their common architecture is underpinned by the conservation of hydrophobic, polar and charged residues (**Fig. 4B**). In particular, several positively charged residues occupy conserved positions in this shared minimal core. These residues are located in several conserved motifs of the two DPBBs subdomains, the duplex-binding helix and the DPBBs-connector. They contribute to form a circular clamp that circumvents the nascent duplex in both RNAPII and PolD DNA-bound structures. In particular, DPBB-B in msRNAPs and DPBB-1 in PolD, carry two extremely conserved positively charged residues (RPB2-K979 and RPB2-K987 in *S. cerevisiae* RNAPII; DP2-K386 and DP2-K392 in *P. abyssi* PolD). In both DNA-bound structures, the positively charged side chains of these residues is found to contact the phosphate moieties of the 3' end of the nascent RNA or DNA. Genetic studies in *S. cerevisiae* (89) and in *E. coli* (90) revealed that even precise inter-atomic distance between the two residues is highly-critical for RNAP activity, given the lethality of lysine to arginine mutations.

Comparison of two-barrel polymerases also reveals structural determinants that are specific to DNAPs versus RNAPs. The catalytic site of all RNA- or DNA-dependent two-barrel

RNAPs is characterized by two catalytic Mg^{2+} that are coordinated by 4 invariant aspartic residues (479DFDGD483 in RPB1 and D837 in RPB2 in *S. cerevisiae* RNAPII; 1007DYDGD1011 and D709 in *N. crassa* QDE-1) (37)(91). Only two aspartic residues are strictly conserved in PolD: 956DGD958, respectively corresponding to 481DGD483 in *S. cerevisiae* RPB1 and 1009DGD1011 in *N. crassa* QDE-1, possibly explaining why no density accounting for the presence of Mg^{2+} has been observed so far in the active site of PolD, neither in the DNA-bound PolD cryo-EM structure (17) nor in the DP2 individual crystal structure (16). Phosphates of the incoming nucleotide may thus be required in order to bind the two catalytic Mg^{2+} ions in the active site, as observed for other DNAPs (92).

A conserved arrangement of the genes encoding the structural elements of the two-barrel minimal core

Interestingly, the structural elements composing the two-barrel minimal core are clustered and share a similar organization in the primary sequence of all single-chain two-barrel polymerases: DNA-dependent DNAPs (PolD), RNA-dependent RNAPs (QDE-1, SAD-1 and RPR-3) and DNA-dependent RNAPs (YonO-like polymerases) (**Fig. 5**). In contrast to single-chain two-barrel polymerases, the gene organization of msRNAPs reveals a much higher level of complexity. In most cellular msRNAPs, the two-barrel minimal core is split within two distinct subunits: RPB1 and RPB2 in eukaryotic RNAPII, RPO1 and RPO2 in archaea, β and β' in bacteria (**Table 1**). However, it is striking that the regions coding for the two halves of the two-barrel minimal core, are respectively clustered at the C-terminal end of one core subunit and at the N-terminal end of the second core subunit, thereby suggesting that although msRNAPs share a conserved ancestor with their single-chain counterparts, diversification and complexification may have resulted in a scission of their minimal catalytic core into two separate genes (**Fig. 5**). Splitting the two-barrel catalytic core onto separate subunits mays thus have facilitated the radical evolution of the common single-chain two-barrel polymerase ancestor into the complex msRNAPs (14). This hypothesis is supported by the fact that epsilon proteo-bacteria harbor a fused single catalytic subunit encompassing both β and β' subunits (93). Furthermore, the yeast *K. lactis* killer DNA plasmids encodes a non-canonical two-barrel RNAP that resembles a fusion of portions of the two RPB1 and RPB2 core subunits (94–96). It is puzzling that the splits occurred in the middle of the two-barrel minimal core, leaving two distinct subunits, each carrying one half of the active site. Splitting the active site in two distinct subunits ensures that only correctly-assembled msRNAPs are active and may be important for regulating the activity of these enzymes.

Evolution of two-barrel polymerases

All two-barrel polymerases evolved from a common ancestor, probably at a very early stage of evolution. One might hypothesize that their common ancestor has been responsible for transmission of RNA genomic information in a RNA-based world. The capacity of PolD to use RNA-primed DNA and occasionally incorporate ribonucleotides may be a property inherited from its common ancestor to RNAPs (76,97). Similarly, the potentially ancient RNA-dependent activity of msRNAPs is illustrated by the fact that they can use RNA templates in special cases

(98,99). Two-barrel polymerases most likely evolved from a single DPBB ancestor (**Fig. 6**). Indeed, DPBB is an ancient fold that is found as single-copy in the active site of several metabolic enzymes in all domains of life (18). Gene duplication led to a molecule with two DPBBs on a single polypeptide chain, which differentiated and recruited both DPBBs-connector and duplex-binding structural elements to produce a molecule capable of efficient RNA polymerization. This primordial two-barrel minimal core may have served as a RNA-binding protein cofactor to a self-replicating ribozyme in an ancient RNA world, before taking over the polymerization catalytic activity.

While the structures of two-barrel polymerases have undergone significant diversifications from their common ancestor, these distantly-related polymerases show intriguing similarities. All two-barrel polymerases thus evolved a nucleic acid binding clamp domain, which occupies the same relative location with respect to the two-barrel active site. In PolD and msRNAPs, clamp domains host several loops that participate in binding the nascent DNA–DNA or RNA–DNA duplex and several zinc-binding domains (Zn-I and Zn-II domains in PolD; Zn-7, Zn-8, and Zn-6 in *S. cerevisiae* RNAP-II), which stabilize the loop-rich clamp domains. Whereas these clamps do not share sequence motifs, Zn-I of PolD and Zn-7 of RNAP-II share the same topology, which is conserved in all eukaryotic/archaeal msRNAPs. Consistently, when the structures of PolD and the elongation complex of *S. cerevisiae* RNAPII are superimposed on their two-barrel catalytic cores, the nascent RNA/DNA hybrid in RNAPII and the nascent DNA duplex in PolD show a very similar orientation with respect to their two-barrel catalytic core (17).

The two-barrel catalytic core is found to be associated to a wide range of structurally distinct DNA-binding domains, which contribute to the wide range of substrate specificities (DNA-dependent or RNA-dependent), activities (RNAPs and DNAPs), and functions (DNA transcription, DNA replication and gene silencing) exhibited by contemporaneous two-barrel polymerases. The two-barrel minimal core is remarkably versatile and evolved toward a wide range of biological functions in modern DNA-world by accretion of distinct sets of additional domains. Comparing their structures illustrates how this minimal catalytic core has recruited different domains in order to perform distinct biological functions. As an example, msRNAPs that are involved in DNA transcription have evolved the lid domain (absent in PolD and QDE-1), which is inserted in DPBB-B, to prevent the formation of an extended upstream RNA–DNA hybrid. In PolD, two ancestral single-stranded DNA binding domains, OB and KH-like, have been recruited for DNA replication to guide the DNA template into the active sites, and a phosphodiesterase domain to perform proof-reading. Although conserved in all two-barrel RNAPs (33), the bridge helix is notably absent in PolD. Instead, the DNA template seems to be guided in the active site of PolD by a KH-like domain (17). Both the bridge helix in RNAPs and the KH-like domain of PolD display a canonic motif of residues that are colocalized with respect to the DNA substrate, suggesting that they may play a similar role in guiding the DNA substrate into the active site.

Since their discovery, cellular msRNAPs have been the focus of intense structural and functional studies revealing, with an unprecedented level of details, how these molecular

machines perform DNA transcription. However, many questions remain unanswered regarding the molecular specificities of their non-canonical two-barrel RNA and DNA polymerases counterparts. Further structural and functional studies are thus required to decipher the molecular mechanisms evolved by these polymerases for polymerization, template binding, nucleotide-selection, and proof-reading. In particular, no structure of the non-canonical viral msRNAPs has been reported yet and their elucidation may provide valuable models for comparative analysis of the mechanism and evolution of the extended two-barrel polymerases superfamily.

Ackonwledgements
The author wishes to thank Pierre Béguin, Pierre Raia, Clément Madru and Marc Delarue for helpful discussions. The work is funded by an ANR JCJC grant ANR-17-CE11-0005-01 and Institut Pasteur.

Keywords
RNA polymerase, DNA polymerase, double-psi β-barrel, two-barrel minimal core, evolution

Abbreviations used
cryo-EM, cryo–electron microscopy; msRNAPs, multi-subunit RNA polymerases; NTP, nucleic acid triphosphate; NCLDVs, nucleo-cytoplasmic DNA viruses; LEF, late expression factor; DNAP, DNA polymerase; DPANN, Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota; DPBB, double-psi β-barrel; dsDNA, double-stranded DNA; KH, K-homology; OB, oligonucleotide binding; PDB, Protein Data Bank; PDE, phosphodiesterase domain; QDE-1, quelling defective phenotype; RNAP, RNA polymerase; ssDNA, single-stranded DNA; TACK, Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota.

Table 1: Subunit composition and activity of two-barrel polymerases. NCLDVs, nucleo-cytoplasmic DNA viruses. Alternative names that are common in the literature are shown in brackets.

Figure 1: Structure of the msRNAP catalytic subunits. (A) The central panel is a simplified diagram that illustrates the overall architecture of msRNAPs. Important structural and functional features discussed in the main text are highlighted. Top and bottom boxes show enlarged cartoon representations of functionally-important regions of the RNAP: the hybrid duplex exit channel, the DNA-binding channel, the two-barrel catalytic center and the pore. Structural information is derived from *S. cerevisiae* RNAPII (PDBid:2e2i) (88). Individual domains are named according to ref. (21) (B) Distribution of the conserved domains and motifs on the RPB1 and RPB2 subunits. Domains are colored as in (A).

Figure 2: Structures of the single-chain two-barrel RNAPs. (A) Ribbon diagrams of the *N. crassa* QDE-1 homodimer (PDBid:2J7N) (19) showing the subunit A colored according to domains and subunit B colored grey. Mg^{2+} ions are marked as spheres. (B) Cartoon representation of the QDE-1 subunit in its "closed" conformation colored according to domains. For comparison, the head domain in its "open' conformation is colored in grey. The direction

of the movement of the head domain is indicted as a grey arrow. The RNA duplex is modeled based on a superposition with an elongation complex of *S. cerevisiae* RNAPII (88). (C) Comparison of the structure of a *N. crassa* QDE-1 monomer (left panel) with a homology model of the YonO protein (right panel), originating from the prophage SPβ. The homology model that covers about 70% of the YonO sequence was predicted using Phyre 2 (67), based on the crystal structure of *N. crassa* QDE-1. Distribution of the domains on the QDE-1 and YonO polymerases is shown in the bottom panels. The RNA/DNA hybrid duplex is modeled based on a superposition with an elongation complex of *S. cerevisiae* RNAPII (88).

Figure 3: Overall structure and surface architecture of PolD: the two-barrel DNAP. (A) Ribbon diagrams of the DNA-bound PolD structure highlighting the domains and domain-like regions that compose the DP1 and DP2 subunits. Zn^{2+} and Fe^{3+} ions are shown as spheres (PDBid: 6hms). The bottom panels show the distribution of the conserved domains and motifs on the DP1 and DP2 subunits of PolD. Domains are colored as above. (B) Surface charge distribution of the PolD-DNA-binding site. The surface of PolD is colored according to the electrostatic surface potential, with negative, neutral, and positive charges shown in red, white, and blue, respectively. The electrostatic potential was calculated as discussed in ref. (17) using a PolD model containing all side-chain atoms. The clamp-1 and clamp-2 domains are highlighted by dotted lines. The DNA backbone is colored red. The right-panel shows a cutaway view of PolD showing a putative path for the DNA being digested by the proof-reading exonuclease domain.

Figure 4: The minimal catalytic core shared by all two-barrel polymerases: structure and sequence conservation. (A) Cartoon representations showing the shared structural elements of the two-barrel catalytic core between *P. abyssi* PolD (PDBid:6hms) (17), *S. cerevisiae* RNAPII (PDBid:2e2i) (88), and *N. crassa* QDE1 (PDBid:2j7n) (19). The two-barrel core is colored according to the conserved structural elements: the two DPBBs (purple), the duplex-binding helix (yellow), and the DPBBs-connector (blue). The invariant aspartic and lysine residues harbored by the two DPBBs are colored in cyan. Mg^{2+} ions are marked as spheres (grey). (B) Structure-based alignment of the shared structural elements within the two-barrel catalytic core of PolD, msRNAPs and single-chain RNA-dependent RNAPs. The shared secondary structure elements are shown above the alignment. Yellow boxes highlight hydrophobic residues, green boxes highlight polar residues, grey boxes highlight conserved glycine or proline positions, and red boxes highlight highly-conserved catalytic motifs. The following structures were used for the structure-based alignment: *S. cerevisiae* RNAPI (100) (PDBid:5m5x), *S. cerevisiae* RNAPII (88) (PDBid:2e2i), *S. cerevisiae* RNAPIII (101) (PDBid:6f41), *T. kodakarensis* RNAP (102) (PDBid:4qiw), *E. coli* RNAP (103) (PDBid:6asx), *T. thermophylus* RNAP (104) (PDBid:2o5i), *N. crassa* QDE-1 (19) (PDBid: 2j7o), and *P. abyssi* PolD (17) (PDBid:6hmf).

Figure 5: Distribution of the conserved structural elements forming the catalytic core in the extended two-barrel polymerase superfamily. The conserved structural elements are colored as in **Fig. 4**: the two DPBBs (purple), the duplex-binding helix (yellow), and the DPBBs-connector (blue). The structural elements composing the two-barrel minimal core are clustered and share a similar organization in the primary sequence of all single-chain two-barrel polymerases. In msRNAPs, regions coding for the two halves of the two-barrel minimal core, are respectively clustered at the C-terminal end of one core subunit and at the N-terminal end of the second core subunit.

Figure 6: A hypothetical scheme of evolution of the two-barrel catalytic core.

References

1. Kornberg A, Baker TA. DNA Replication. University Science Books; 2005. 964 p.
2. Steitz TA. A mechanism for all polymerases. Nature. 1998 Jan 15;391(6664):231–2.
3. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A. 1993 Jul 15;90(14):6498–502.
4. Yang W, Lee JY, Nowotny M. Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. Mol Cell. 2006 Apr 7;22(1):5–13.
5. Wu S, Beard WA, Pedersen LG, Wilson SH. Structural Comparison of DNA Polymerase Architecture Suggest a Nucleotide Gateway to the Polymerase Active Site. Chem Rev. 2014 Mar 12;114(5):2759–74.
6. Raia P, Delarue M, Sauguet L. An updated structural classification of replicative DNA polymerases. Biochem Soc Trans. 2019 Jan 15;
7. Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA. Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature. 1985 Mar 28;313(6005):762–6.
8. Patel PH, Loeb LA. Getting a grip on how DNA polymerases function. Nat Struct Biol. 2001 Aug;8(8):656–9.
9. Delarue M, Poch O, Tordo N, Moras D, Argos P. An attempt to unify the structure of polymerases. Protein Eng. 1990 May;3(6):461–7.
10. Bailey S, Wing RA, Steitz TA. The Structure of T. aquaticus DNA Polymerase III Is Distinct from Eukaryotic Replicative DNA Polymerases. Cell. 2006 Sep 8;126(5):893–904.
11. Cramer P. Multisubunit RNA polymerases. Curr Opin Struct Biol. 2002 Feb 1;12(1):89–97.
12. Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol. 2011 février;9(2):85–98.
13. Ruprich-Robert G, Thuriaux P. Non-canonical DNA transcription enzymes and the conservation of two-barrel RNA polymerases. Nucleic Acids Res. 2010 Aug;38(14):4559–69.
14. Fouqueau T, Blombach F, Werner F. Evolutionary Origins of Two-Barrel RNA Polymerases and Site-Specific Transcription Initiation. Annu Rev Microbiol. 2017 Sep 8;71:331–48.
15. Forrest D. Unusual relatives of the multisubunit RNA polymerase. Biochem Soc Trans. 2018 Dec 21;
16. Sauguet L, Raia P, Henneke G, Delarue M. Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. Nat Commun. 2016 Aug 22;7:12227.
17. Raia P, Carroni M, Henry E, Pehau-Arnaudet G, Brûlé S, Béguin P, et al. Structure of the DP1-DP2 PolD complex bound with DNA and its implications for the evolutionary history of DNA and RNA polymerases. PLoS Biol. 2019 Jan 18;17(1):e3000122.
18. Iyer LM, Koonin EV, Aravind L. 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. 2003 Jan 28;3:1.
19. Salgado PS, Koivunen MRL, Makeyev EV, Bamford DH, Stuart DI, Grimes JM. The structure of an RNAi polymerase links RNA silencing and transcription. PLoS Biol. 2006 Dec;4(12):e434.
20. Pfannschmidt T, Blanvillain R, Merendino L, Courtois F, Chevalier F, Liebers M, et al. Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. J Exp Bot. 2015 Dec;66(22):6957–73.

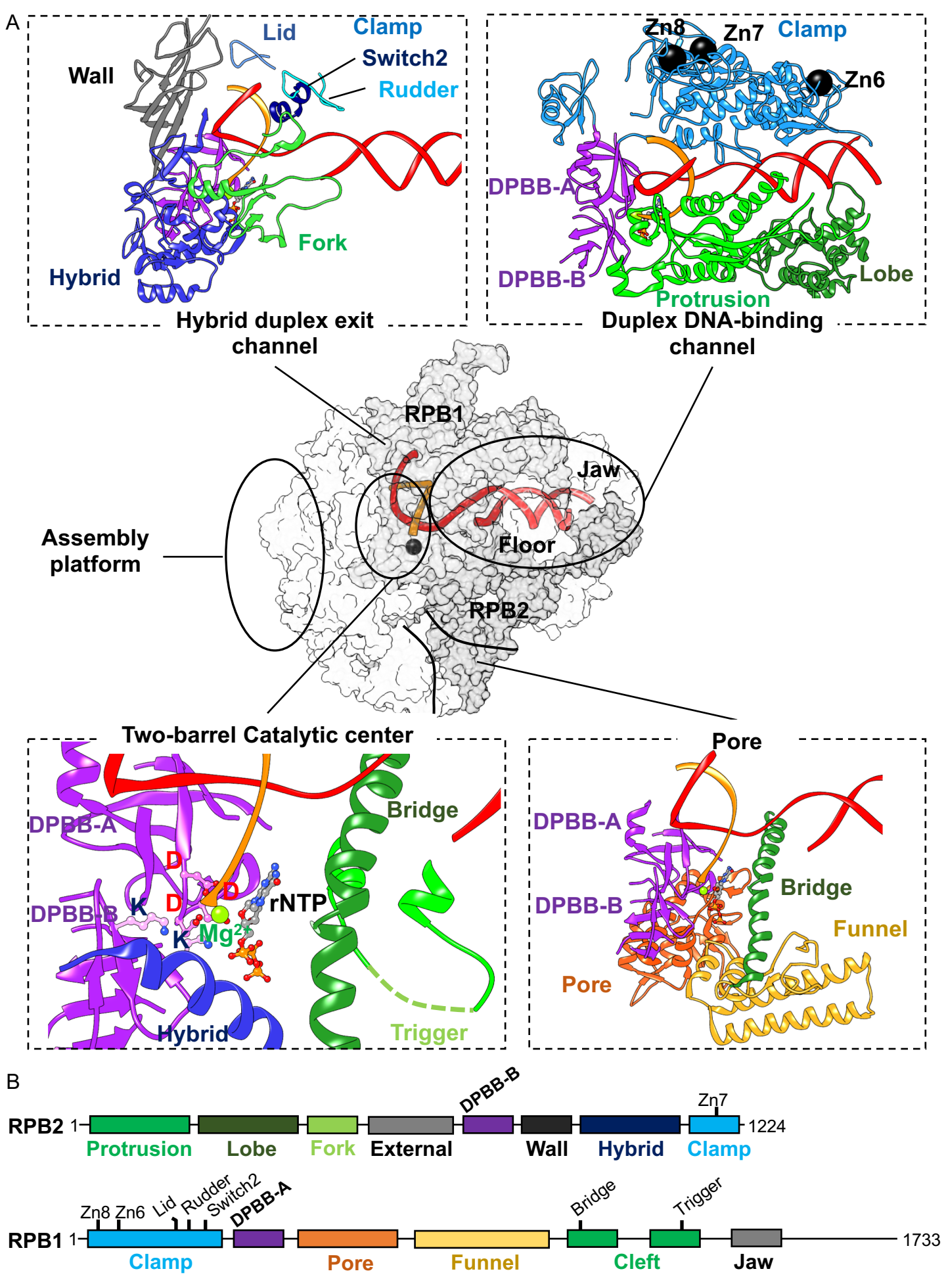
21. Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science. 2001 Jun 8;292(5523):1863–76.
22. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. Science. 2001 Jun 8;292(5523):1876–82.
23. Hirata A, Klein BJ, Murakami KS. The X-ray crystal structure of RNA polymerase from Archaea. Nature. 2008 Feb 14;451(7180):851–4.
24. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell. 1999 Sep 17;98(6):811–24.
25. Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, et al. Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. PLoS Biol. 2009 May;7(5):e1000102.
26. Lane WJ, Darst SA. Molecular evolution of multisubunit RNA polymerases: sequence analysis. J Mol Biol. 2010 Jan 29;395(4):671–85.
27. Lane WJ, Darst SA. Molecular evolution of multisubunit RNA polymerases: structural analysis. J Mol Biol. 2010 Jan 29;395(4):686–704.
28. Markovtsov V, Mustaev A, Goldfarb A. Protein-RNA interactions in the active center of transcription elongation complex. Proc Natl Acad Sci U S A. 1996 Apr 16;93(8):3221–6.
29. Zaychikov E, Martin E, Denissova L, Kozlov M, Markovtsov V, Kashlev M, et al. Mapping of catalytic residues in the RNA polymerase active center. Science. 1996 Jul 5;273(5271):107–9.
30. Nudler E. Transcription elongation: structural basis and mechanisms. J Mol Biol. 1999 Apr 23;288(1):1–12.
31. Nudler E. RNA Polymerase Active Center: The Molecular Engine of Transcription. Annu Rev Biochem. 2009;78:335–61.
32. Nedialkov YA, Opron K, Assaf F, Artsimovitch I, Kireeva ML, Kashlev M, et al. The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. Biochim Biophys Acta. 2013 Feb;1829(2):187–98.
33. Weinzierl ROJ. The Bridge Helix of RNA Polymerase Acts as a Central Nanomechanical Switchboard for Coordinating Catalysis and Substrate Movement [Internet]. Archaea. 2011 [cited 2018 Sep 7]. Available from: https://www.hindawi.com/journals/archaea/2011/608385/
34. Fouqueau T, Zeller ME, Cheung AC, Cramer P, Thomm M. The RNA polymerase trigger loop functions in all three phases of the transcription cycle. Nucleic Acids Res. 2013 Aug;41(14):7048–59.
35. Kaplan CD, Larsson K-M, Kornberg RD. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol Cell. 2008 Jun 6;30(5):547–56.
36. Zhang J, Palangat M, Landick R. Role of the RNA polymerase trigger loop in catalysis and pausing. Nat Struct Mol Biol. 2010 Jan;17(1):99–104.
37. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell. 2004 Nov 12;119(4):481–9.
38. Koonin EV, Yutin N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology. 2010;53(5):284–92.
39. Yutin N, Koonin EV. Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. Virol J. 2012 Aug 14;9:161.
40. Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 2006 Apr;117(1):156–84.

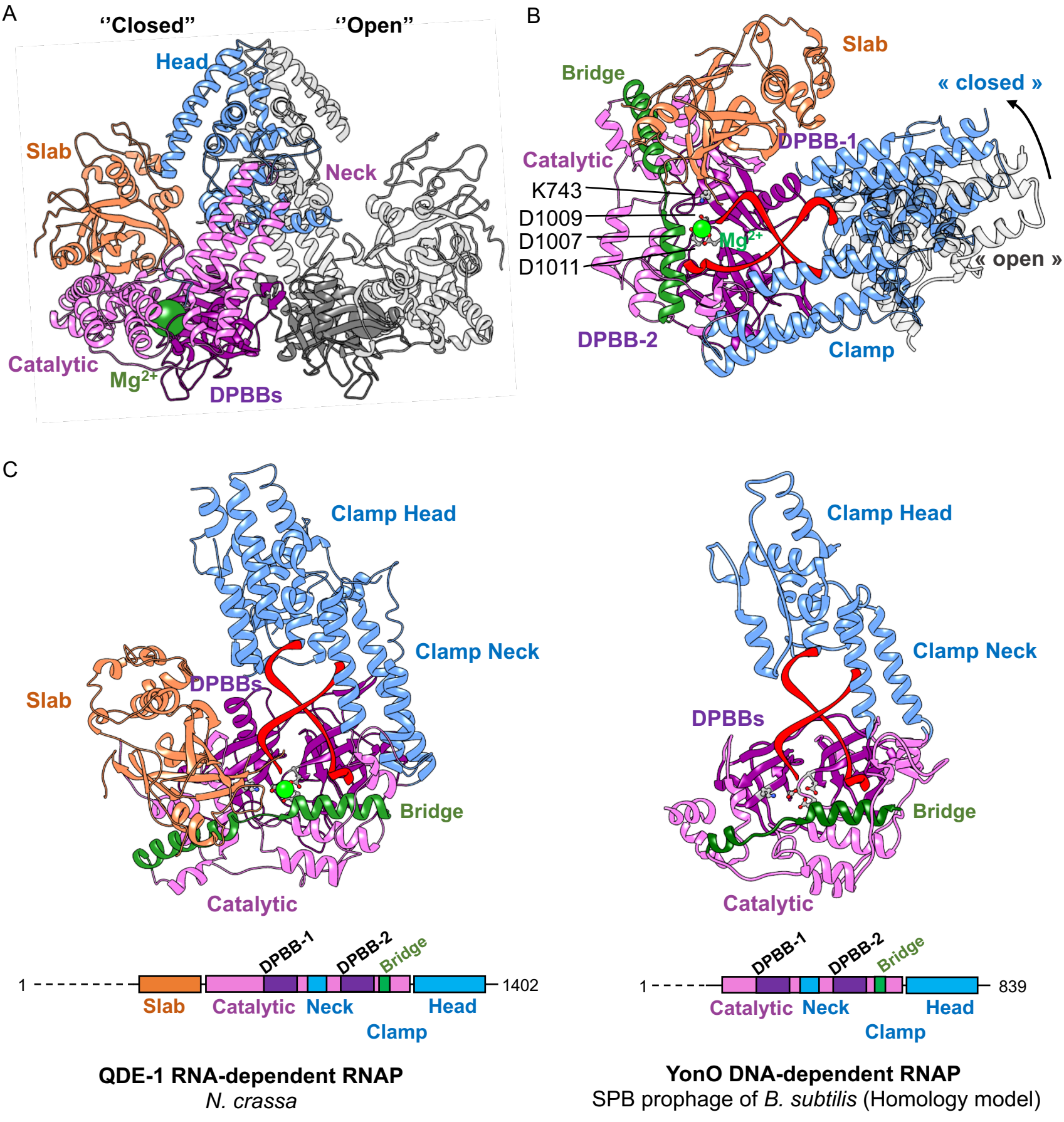
41. Grula MA, Buller PL, Weaver RF. alpha-Amanitin-Resistant Viral RNA Synthesis in Nuclei Isolated from Nuclear Polyhedrosis Virus-Infected Heliothis zea Larvae and Spodoptera frugiperda Cells. J Virol. 1981 Jun;38(3):916–21.
42. Fuchs LY, Woods MS, Weaver RF. Viral Transcription During Autographa californica Nuclear Polyhedrosis Virus Infection: a Novel RNA Polymerase Induced in Infected Spodoptera frugiperda Cells. J Virol. 1983 Dec;48(3):641–6.
43. Huh NE, Weaver RF. Identifying the RNA polymerases that synthesize specific transcripts of the Autographa californica nuclear polyhedrosis virus. J Gen Virol. 1990 Jan;71 (Pt 1):195–201.
44. Guarino LA, Xu B, Jin J, Dong W. A virus-encoded RNA polymerase purified from baculovirus-infected cells. J Virol. 1998 Oct;72(10):7985–91.
45. Passarelli AL. Baculovirus RNA polymerase: Activities, composition, and evolution. Virol Sin. 2007 Apr 1;22(2):94–107.
46. Titterington JS, Nun TK, Passarelli AL. Functional dissection of the baculovirus late expression factor-8 gene: sequence requirements for late gene promoter activation. J Gen Virol. 2003 Jul;84(Pt 7):1817–26.
47. Lu A, Miller LK. Identification of three late expression factor genes within the 33.8- to 43.4-map-unit region of Autographa californica nuclear polyhedrosis virus. J Virol. 1994 Oct;68(10):6710–8.
48. Gross CH, Shuman S. Characterization of a baculovirus-encoded RNA 5'-triphosphatase. J Virol. 1998 Sep;72(9):7057–63.
49. Jin J, Dong W, Guarino LA. The LEF-4 subunit of baculovirus RNA polymerase has RNA 5'-triphosphatase and ATPase activities. J Virol. 1998 Dec;72(12):10011–9.
50. Thomas JA, Rolando MR, Carroll CA, Shen PS, Belnap DM, Weintraub ST, et al. Characterization of Pseudomonas chlororaphis myovirus 201φ2-1 via genomic sequencing, mass spectrometry, and electron microscopy. Virology. 2008 Jul 5;376(2):330–8.
51. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, et al. Genome comparison of Pseudomonas aeruginosa large phages. J Mol Biol. 2005 Dec 2;354(3):536–45.
52. Yakunina M, Artamonova T, Borukhov S, Makarova KS, Severinov K, Minakhin L. A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. Nucleic Acids Res. 2015 Dec 2;43(21):10411–20.
53. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 2015 Feb;16(2):71–84.
54. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009 Feb;10(2):94–108.
55. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009 Feb;10(2):126–39.
56. Wassenegger M, Krczal G. Nomenclature and functions of RNA-directed RNA polymerases. Trends Plant Sci. 2006 Mar;11(3):142–51.
57. Schiebel W, Pélissier T, Riedel L, Thalmeir S, Schiebel R, Kempe D, et al. Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. Plant Cell. 1998 Dec;10(12):2087–101.
58. Astier-Manifacier S, Cornuet P. RNA-dependent RNA polymerase in Chinese cabbage. Biochim Biophys Acta. 1971 Mar 25;232(3):484–93.
59. Schiebel W, Haas B, Marinković S, Klanner A, Sänger HL. RNA-directed RNA polymerase from tomato leaves. I. Purification and physical properties. J Biol Chem. 1993 Jun 5;268(16):11851–7.
60. Cogoni C, Macino G. Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature. 1999 May 13;399(6732):166–9.

61. Pickford AS, Catalanotto C, Cogoni C, Macino G. Quelling in Neurospora crassa. Adv Genet. 2002;46:277–303.
62. Shiu PK, Raju NB, Zickler D, Metzenberg RL. Meiotic silencing by unpaired DNA. Cell. 2001 Dec 28;107(7):905–16.
63. Makeyev EV, Bamford DH. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. Mol Cell. 2002 Dec;10(6):1417–27.
64. Qian X, Hamid FM, El Sahili A, Darwis DA, Wong YH, Bhushan S, et al. Functional Evolution in Orthologous Cell-encoded RNA-dependent RNA Polymerases. J Biol Chem. 2016 Apr 22;291(17):9295–309.
65. Thomas JA, Hardies SC, Rolando M, Hayes SJ, Lieman K, Carroll CA, et al. Complete Genomic Sequence and Mass Spectrometric Analysis of Highly Diverse, Atypical Bacillus thuringiensis phage 0305φ8-36. Virology. 2007 Nov 25;368(2):405–21.
66. Forrest D, James K, Yuzenkova Y, Zenkin N. Single-peptide DNA-dependent RNA polymerase homologous to multi-subunit RNA polymerase. Nat Commun. 2017 06;8:15774.
67. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845–58.
68. Braithwaite DK, Ito J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. 1993 Feb 25;21(4):787–802.
69. Imamura M, Uemori T, Kato I, Ishino Y. A non-alpha-like DNA polymerase from the hyperthermophilic archaeon Pyrococcus furiosus. Biol Pharm Bull. 1995 Dec;18(12):1647–52.
70. Ishino Y, Komori K, Cann IK, Koga Y. A novel DNA polymerase family found in Archaea. J Bacteriol. 1998 Apr;180(8):2232–6.
71. Takashima N, Ishino S, Oki K, Takafuji M, Yamagami T, Matsuo R, et al. Elucidating functions of DP1 and DP2 subunits from the Thermococcus kodakarensis family D DNA polymerase. Extrem Life Extreme Cond. 2019 Jan;23(1):161–72.
72. Berquist BR, DasSarma P, DasSarma S. Essential and non-essential DNA replication genes in the model halophilic Archaeon, Halobacterium sp. NRC-1. BMC Genet. 2007 Jun 8;8:31.
73. Cubonová L, Richardson T, Burkhart BW, Kelman Z, Connolly BA, Reeve JN, et al. Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in Thermococcus kodakarensis. J Bacteriol. 2013 May;195(10):2322–8.
74. Birien T, Thiel A, Henneke G, Flament D, Moalic Y, Jebbar M. Development of an Effective 6-Methylpurine Counterselection Marker for Genetic Manipulation in Thermococcus barophilus. Genes [Internet]. 2018 Feb 7;9(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5852573/
75. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell. 2018 Mar 8;172(6):1181–97.
76. Henneke G, Flament D, Hübscher U, Querellou J, Raffin J-P. The hyperthermophilic euryarchaeota Pyrococcus abyssi likely requires the two DNA polymerases D and B for DNA replication. J Mol Biol. 2005 Jul 1;350(1):53–64.
77. Greenough L, Kelman Z, Gardner AF. The roles of family B and D DNA polymerases in Thermococcus species 9°N Okazaki fragment maturation. J Biol Chem. 2015 May 15;290(20):12514–22.
78. Rothwell PJ, Waksman G. Structure and mechanism of DNA polymerases. Adv Protein Chem. 2005;71:401–40.
79. Doublié S, Zahn KE. Structural insights into eukaryotic DNA replication. Front Microbiol [Internet]. 2014 Aug 25 [cited 2016 Feb 1];5. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142720/

80. Abellón-Ruiz J, Waldron KJ, Connolly BA. Archaeoglobus Fulgidus DNA Polymerase D: A Zinc-Binding Protein Inhibited by Hypoxanthine and Uracil. J Mol Biol. 2016 Jul 17;428(14):2805–13.
81. Grishin NV. KH domain: one motif, two folds. Nucleic Acids Res. 2001 Feb 1;29(3):638–43.
82. Timinskas K, Venclovas Č. The N-terminal region of the bacterial DNA polymerase PolC features a pair of domains, both distantly related to domain V of the DNA polymerase III τ subunit. FEBS J. 2011 Sep;278(17):3109–18.
83. Worbs M, Bourenkov GP, Bartunik HD, Huber R, Wahl MC. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol Cell. 2001 Jun;7(6):1177–89.
84. Zhou Y, Mah T-F, Greenblatt J, Friedman DI. Evidence that the KH RNA-binding domains influence the action of the E. coli NusA protein. J Mol Biol. 2002 May 17;318(5):1175–88.
85. Cann IK, Komori K, Toh H, Kanai S, Ishino Y. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. Proc Natl Acad Sci U S A. 1998 Nov 24;95(24):14250–5.
86. Aravind L, Koonin EV. Phosphoesterase domains associated with DNA polymerases of diverse origins. Nucleic Acids Res. 1998 Aug 15;26(16):3746–52.
87. Sydow JF, Cramer P. RNA polymerase fidelity and transcriptional proofreading. Curr Opin Struct Biol. 2009 Dec;19(6):732–9.
88. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell. 2006 Dec 1;127(5):941–54.
89. Treich I, Carles C, Sentenac A, Riva M. Determination of lysine residues affinity labeled in the active site of yeast RNA polymerase II(B) by mutagenesis. Nucleic Acids Res. 1992 Sep 25;20(18):4721–5.
90. Kashlev M, Lee J, Zalenskaya K, Nikiforov V, Goldfarb A. Blocking of the initiation-to-elongation transition by a transdominant RNA polymerase mutation. Science. 1990 May 25;248(4958):1006–9.
91. Sosunov V, Zorov S, Sosunova E, Nikolaev A, Zakeyeva I, Bass I, et al. The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. Nucleic Acids Res. 2005;33(13):4202–11.
92. Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature. 1998 Jan 15;391(6664):251–8.
93. Zakharova N, Hoffman PS, Berg DE, Severinov K. The largest subunits of RNA polymerase from gastric helicobacters are tethered. J Biol Chem. 1998 Jul 31;273(31):19371–4.
94. Tommasino M, Ricci S, Galeotti CL. Genome organization of the killer plasmid pGK12 from Kluyveromyces lactis. Nucleic Acids Res. 1988 Jul 11;16(13):5863–78.
95. Wilson DW, Meacock PA. Extranuclear gene expression in yeast: evidence for a plasmid-encoded RNA polymerase of unique structure. Nucleic Acids Res. 1988 Aug 25;16(16):8097–112.
96. Sýkora M, Pospíšek M, Novák J, Mrvová S, Krásný L, Vopálenský V. Transcription apparatus of the yeast virus-like elements: Architecture, function, and evolutionary origin. PLoS Pathog [Internet]. 2018 Oct 22;14(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6211774/
97. Lemor M, Kong Z, Henry E, Brizard R, Laurent S, Bossé A, et al. Differential Activities of DNA Polymerases in Processing Ribonucleotides during DNA Synthesis in

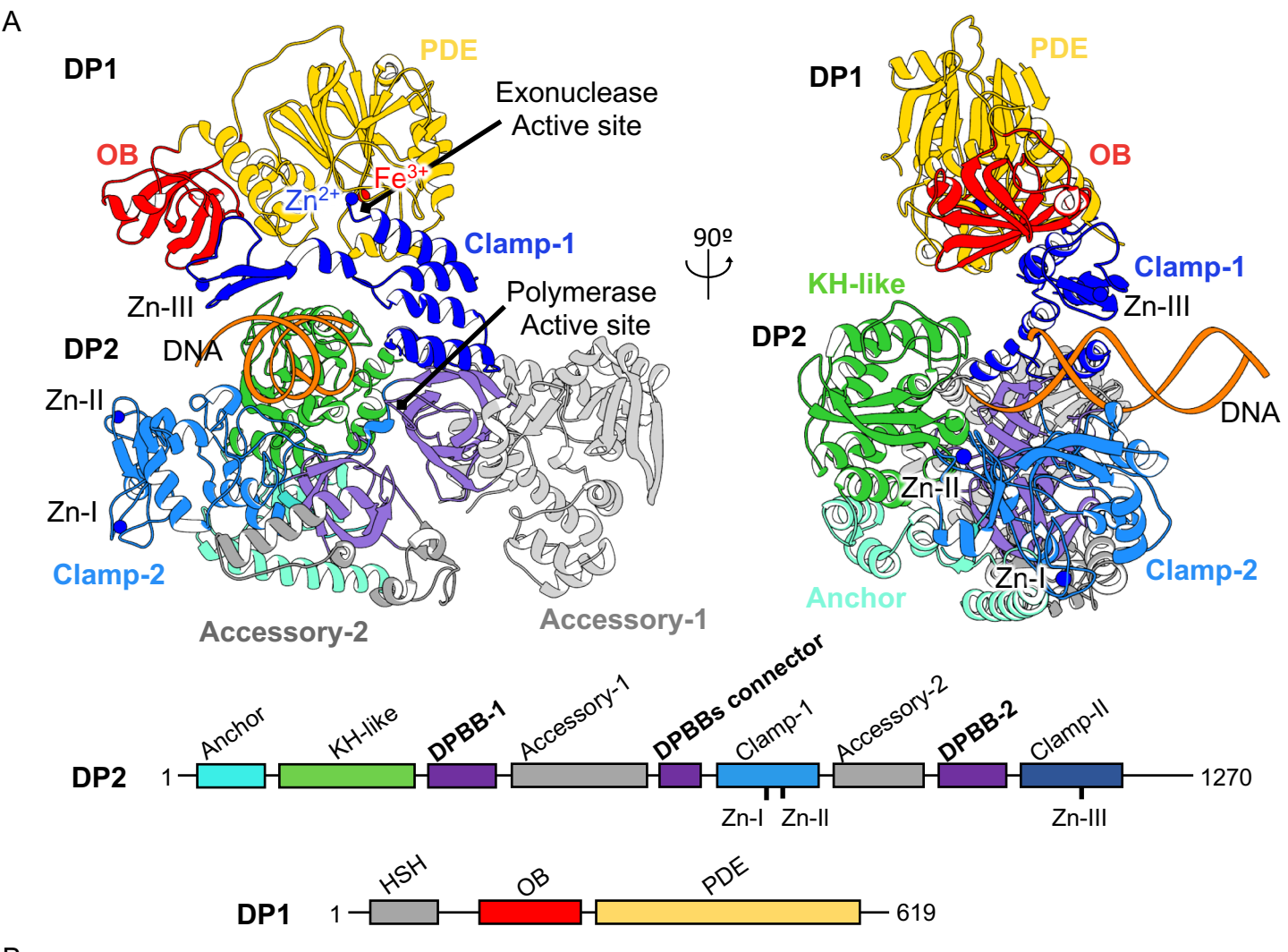
Archaea. J Mol Biol [Internet]. 2018 Oct 31; Available from: http://www.sciencedirect.com/science/article/pii/S0022283618307277
98. Wagner SD, Yakovchuk P, Gilman B, Ponicsan SL, Drullinger LF, Kugel JF, et al. RNA polymerase II acts as an RNA-dependent RNA polymerase to extend and destabilize a non-coding RNA. EMBO J. 2013 Mar 20;32(6):781–90.
99. Rackwitz HR, Rohde W, Sänger HL. DNA-dependent RNA polymerase II of plant origin transcribes viroid RNA into full-length copies. Nature. 1981 May 28;291(5813):297–301.
100. Tafur L, Sadian Y, Hoffmann NA, Jakobi AJ, Wetzel R, Hagen WJH, et al. Molecular Structures of Transcribing RNA Polymerase I. Mol Cell. 2016 Dec 15;64(6):1135–43.
101. Vorländer MK, Khatter H, Wetzel R, Hagen WJH, Müller CW. Molecular mechanism of promoter opening by RNA polymerase III. Nature. 2018 17;553(7688):295–300.
102. Jun S-H, Hirata A, Kanai T, Santangelo TJ, Imanaka T, Murakami KS. The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. Nat Commun. 2014;5:5132.
103. Kang JY, Mishanina TV, Bellecourt MJ, Mooney RA, Darst SA, Landick R. RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. Mol Cell. 2018 01;69(5):802–15.e1.
104. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. Nature. 2007 Jul;448(7150):157–62.

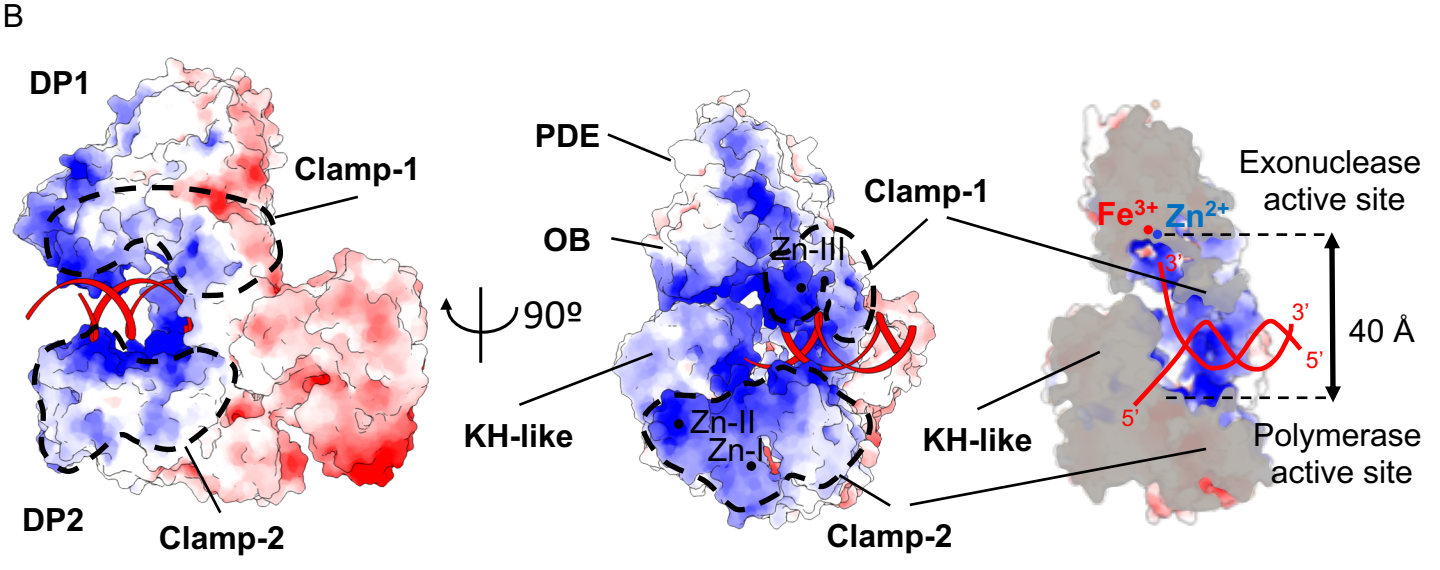

A
Lid
Clamp
Switch2
Wall
Rudder
Fork
Hybrid
Zn8
Zn7
Clamp
Zn6
DPBB-A
DPBB-B
Lobe
Protrusion
Hybrid duplex exit channel
Duplex DNA-binding channel
RPB1
Jaw
Assembly platform
Floor
RPB2
Two-barrel Catalytic center
Pore
DPBB-A
Bridge
D
DPBB-B
K
D
D
rNTP
K
Mg2+
Hybrid
Trigger
DPBB-A
Bridge
DPBB-B
Funnel
Pore
B
RPB2
1
DPBB-B
Zn7
1224
Protrusion
Lobe
Fork
External
Wall
Hybrid
Clamp
Zn8 Zn6
Lid
Rudder
Switch2
DPBB-A
Bridge
Trigger
RPB1
1
1733
Clamp
Pore
Funnel
Cleft
Jaw



QDE-1 RNA-dependent RNAP
N. crassa

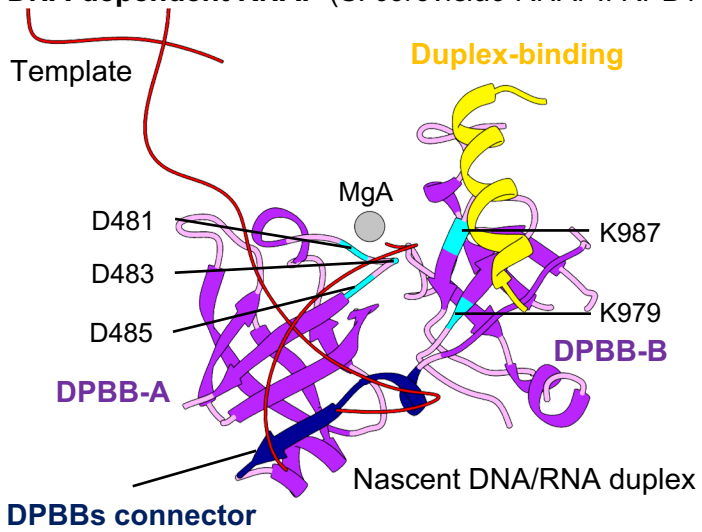
YonO DNA-dependent RNAP
SPB prophage of *B. subtilis* (Homology model)


A
DP1
PDE
Exonuclease
Active site
OB
Zn2+
Fe3+
Clamp-1
Zn-III
Polymerase
Active site
DP2
DNA
Zn-II
Zn-I
Clamp-2
Accessory-2
Accessory-1
90º
KH-like
Anchor
DP2
1
1270
DPBB-1
DPBBs connector
DPBB-2
Clamp-II
Zn-I
Zn-II
Zn-III
HSH
619

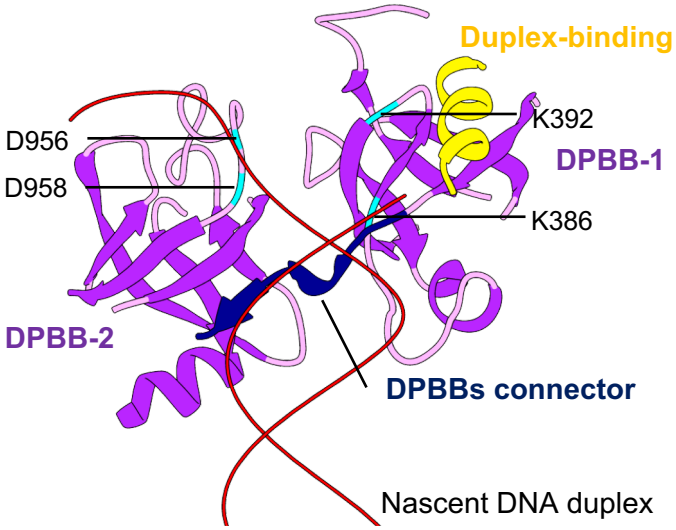

B
DP1
Clamp-1
DP2
Clamp-2
90º
PDE
OB
Zn-III
KH-like
Zn-II
Zn-I
Exonuclease
active site
Fe3+
Zn2+
3'
5'
40 Å
Polymerase
active site

A

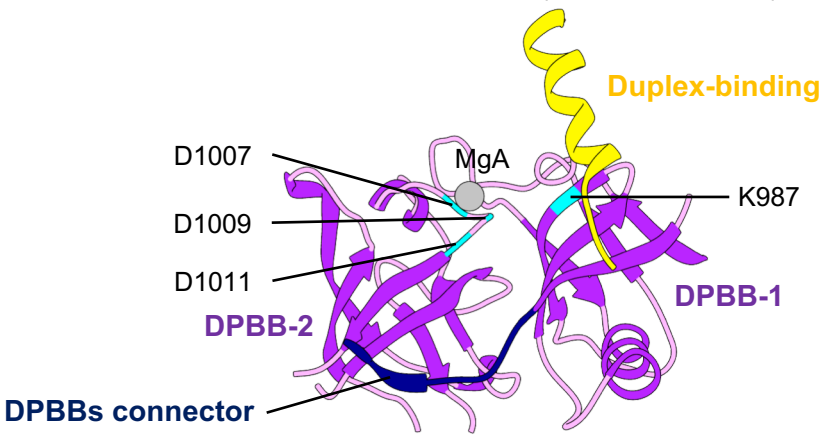
DNA-dependent RNAP (*S. cerevisiae* RNAPII RPB1-2)



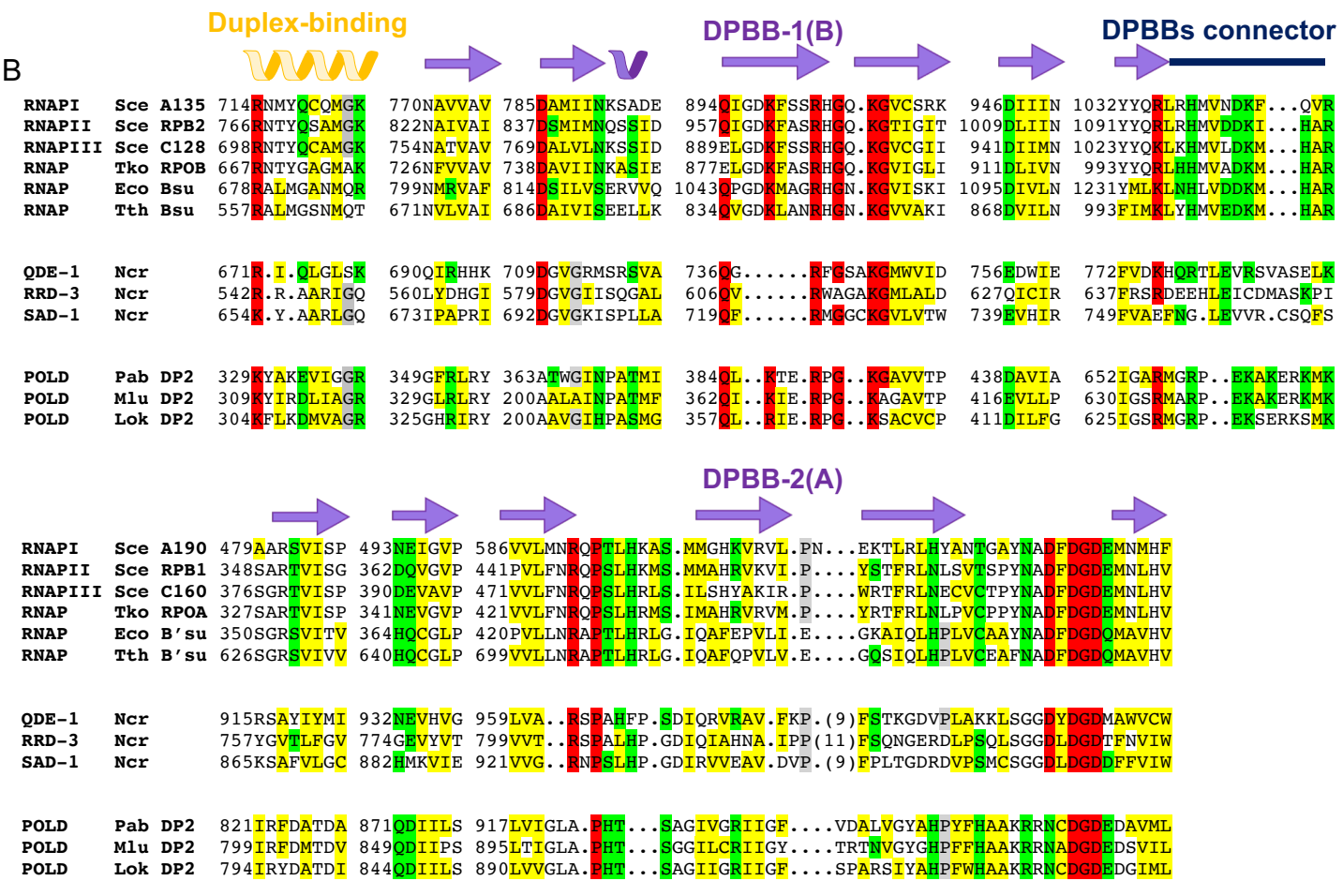
DNA-dependent DNAP (*P. abyssi* PolD DP2)

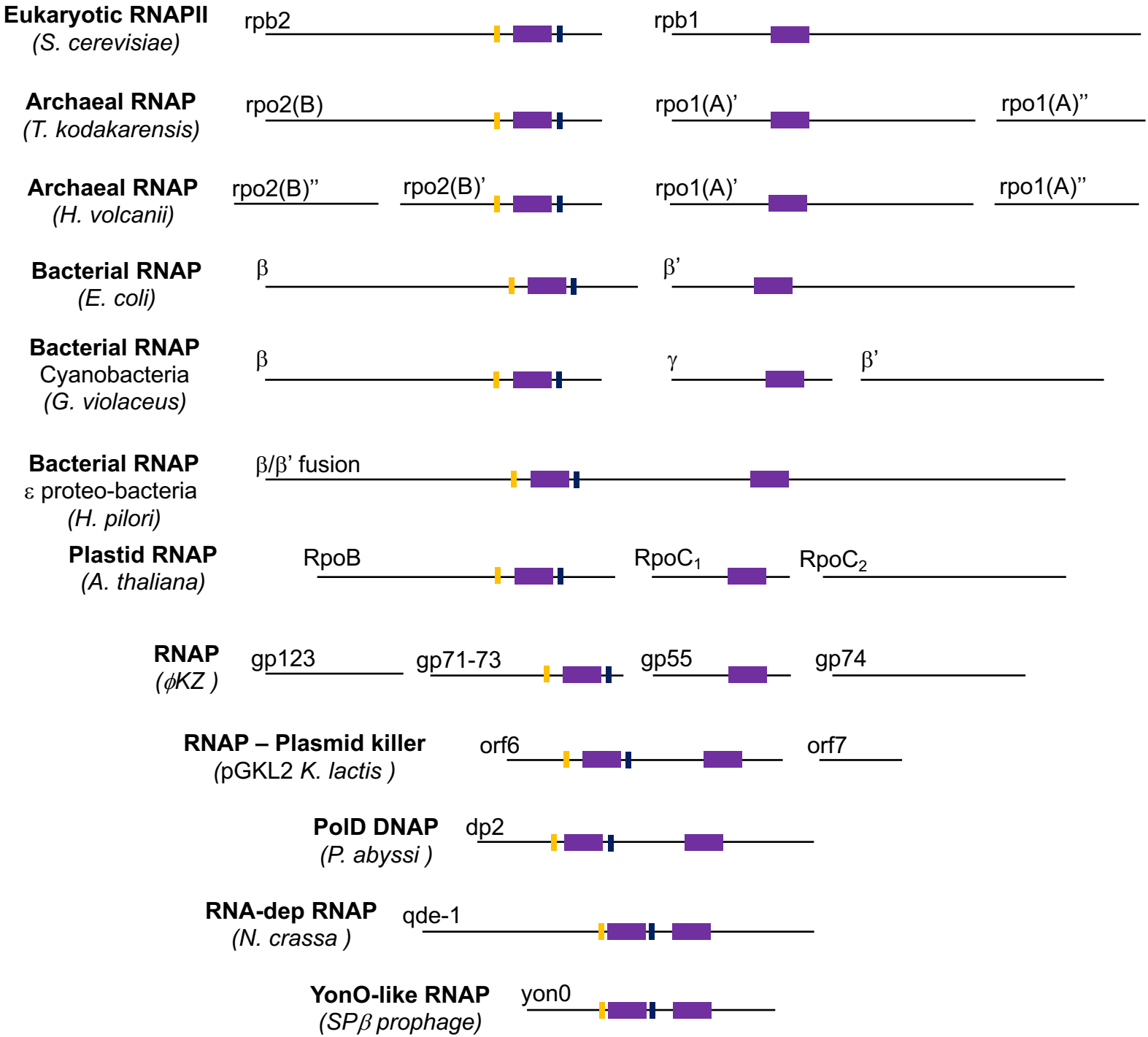


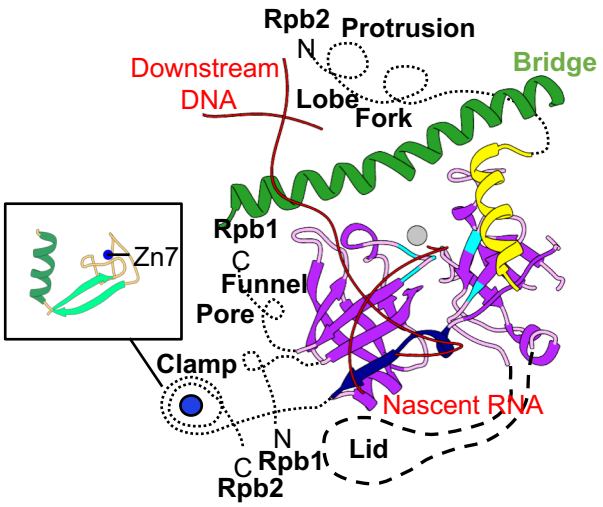
RNA-dependent RNAP (*N. crassa* QDE-1)



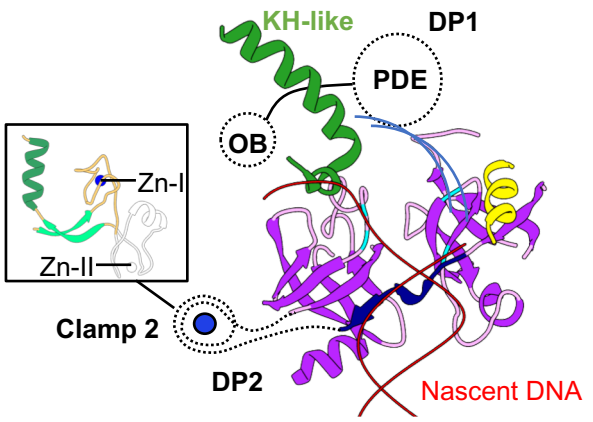
B



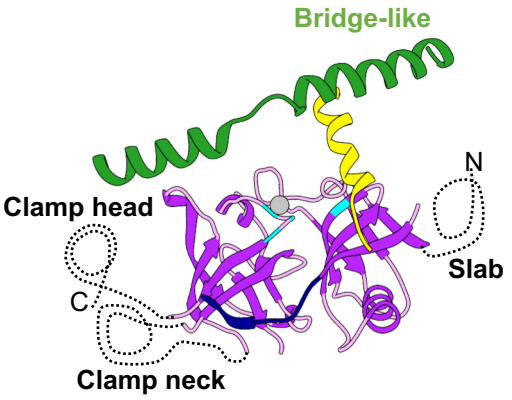

Eukaryotic RNAPII
(S. cerevisiae)
rpb2
rpb1
Archaeal RNAP
(T. kodakarensis)
rpo2(B)
rpo1(A)'
rpo1(A)"
Archaeal RNAP
(H. volcanii)
rpo2(B)"
rpo2(B)'
rpo1(A)'
rpo1(A)"
Bacterial RNAP
(E. coli)
β
β'
Bacterial RNAP
Cyanobacteria
(G. violaceus)
β
γ
β'
Bacterial RNAP
ε proteo-bacteria
(H. pilori)
β/β' fusion
Plastid RNAP
(A. thaliana)
RpoB
RpoC1
RpoC2
RNAP
(ϕKZ)
gp123
gp71-73
gp55
gp74
RNAP – Plasmid killer
(pGKL2 K. lactis)
orf6
orf7
PolD DNAP
(P. abyssi)
dp2
RNA-dep RNAP
(N. crassa)
qde-1
YonO-like RNAP
(SPβ prophage)
yon0



msRNAPs
(DNA-dependent RNAPs)
DNA transcription
(Euk., Arch., Bact. DNA viruses)

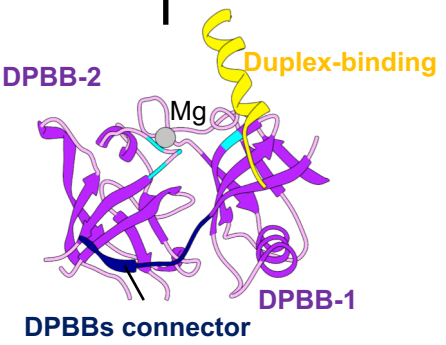


PolD
(DNA-dependent DNAP)
DNA replication
(Arch.)



Single-chain RNAPs
(RNA-dependent RNAPs)
Gene silencing
(Euk.)

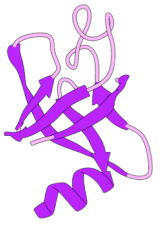
Divergent evolution
Acquisition of specialized
DNA- and RNA-binding domains



DPBBs connector

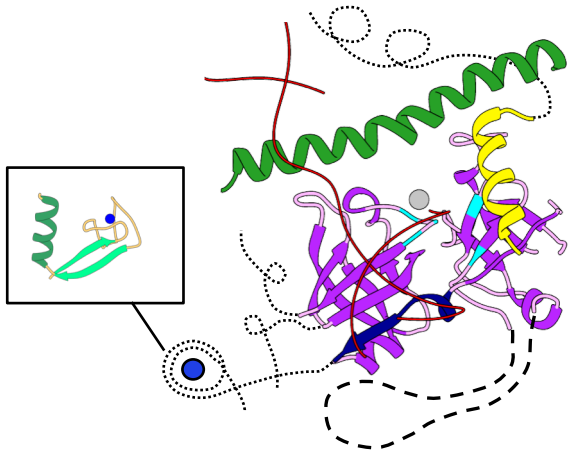
Two-barrel polymerase ancestor
(RNA-dependent RNAP ?)
RNA replicase in a RNA world ?

- Duplication & differentiation of the DPBBs
- Acquisition of the DPBBs-connector
- Acquisition of the duplex-binding helix

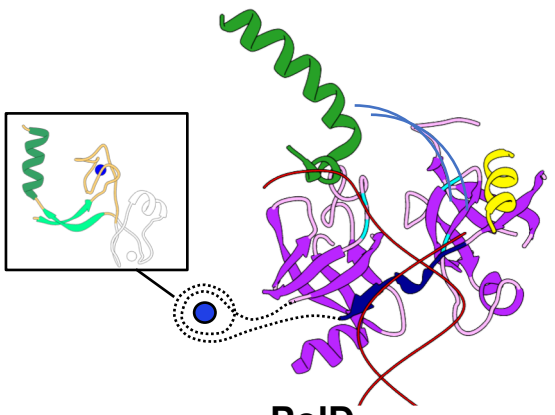


DPBB ancestor

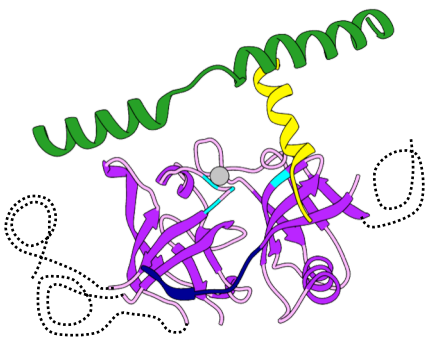
	Organisms		Two-barrel core subunits		Other core subunits	Biological function
			DPBB-1 (DPBB-B)	**DPBB-2 (DPBB-A)**		
DNA-dependent RNAP	Eukaryotes	RNAPI	A135	A190	AC19, AC40, RPB4-10, RPB12	Transcription of ribosomal RNA
		RNAPII	RPB2	RPB1	RPB3-12	Transcription of pre-messenger RNA
		RNAPIII	C128	C160	AC19, AC40, RPB5-6, RPB8, RPB10, RPB12, C11, C17, C25	Transcription of small RNAs
		RNAPIV (Plant)	NRPD/E2	NRPD1	RPB3, RPB5, RPB8, RPB10-12, NRPD/E4, NRPD7, NRPD9b	Production of small interfering RNAs
		RNAPV (Plant)	NRPD/E2	NRPE1	RPB3,NRPE5,RPB8, RPB9-12, NRPD/E4, NRPE7	
	Archaea	*M. jannaschii*	RPO2(A)	RPO1(A)'	RPO1(A)'', RPO3(D), RPO4(F), RPO5(H), RPO6(K), RPO7(E), RPO8(G), RPO10(N), RPO12(P)	Transcription of the whole gene repertoire
	Bacteria	*(e.g. E. coli)*	β-subunit	β'-subunit	α-subunit, ω-subunit	
		Epsilonproteobacteria	fusion β/β'-subunit		α-subunit, ω-subunit	
	Chloroplasts	Plastid-encoded prokaryotic-type RNAP	RPOB	$RPOC_1$	RPOA, $RPOC_2$, PAP1-10	Major RNAP activity in mature chloroplasts
	Viruses	NCLDVs *(e.g poxviruses)*	RPO2	RPO1	RPO7, RPO22, RPO30	Transcription (cytoplasmic phase of the virus cycle)
		ϕKZ giant phage	GP71-73	GP55	GP68, GP74, GP123	
		Baculoviruses	LEF-8	LEF-9	LEF-4, P47	Late gene expression
		Firmicutes phages	YonO			Late gene expression
	Yeast Plasmids	pGKL2 (e.g. *K. lactis*)	ORF6		ORF7	pGKL1 (and 2 ?) gene expression
RNA-dependent RNAP	Eukaryotes	*N. crassa*	QDE-1			PTG silencing
		N. crassa	SAD-1			Meiotic silencing
		N. crassa	RRP-3			PTG silencing ?
DNA-dependent DNAP	Archaea	PolD	DP2		DP1	DNA replication



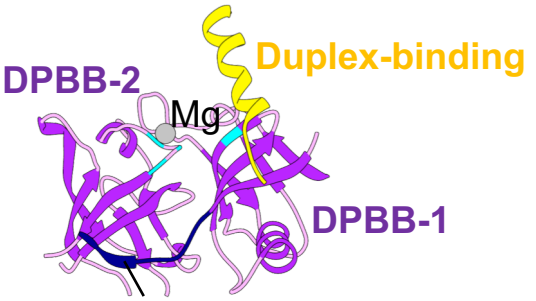
msRNAPs
(DNA-dependent RNAPs)
DNA transcription



PolD
(DNA-dependent DNAP)
DNA replication

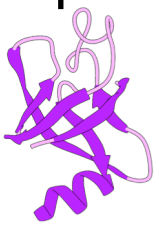


Single-chain RNAPs
(RNA-dependent RNAPs)
Gene silencing



DPBBs connector

Two-barrel polymerase ancestor
RNA replicase in a RNA world ?



DPBB ancestor