FEBS Letters 588 (2014) 219–229

journal homepage: www.FEBSLetters.org

Review

Directed polymerase evolution

Tingjian Chen, Floyd E. Romesberg *

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, United States

ARTICLE INFO

Article history:
Received 19 October 2013
Revised 28 October 2013
Accepted 29 October 2013
Available online 5 November 2013

Edited by Wilhelm Just

Keywords:
Directed evolution
Polymerase
Modified nucleotide
Protein engineering

ABSTRACT

Polymerases evolved in nature to synthesize DNA and RNA, and they underlie the storage and flow of genetic information in all cells. The availability of these enzymes for use at the bench has driven a revolution in biotechnology and medicinal research; however, polymerases did not evolve to function efficiently under the conditions required for some applications and their high substrate fidelity precludes their use for most applications that involve modified substrates. To circumvent these limitations, researchers have turned to directed evolution to tailor the properties and/or substrate repertoire of polymerases for different applications, and several systems have been developed for this purpose. These systems draw on different methods of creating a pool of randomly mutated polymerases and are differentiated by the process used to isolate the most fit members. A variety of polymerases have been evolved, providing new or improved functionality, as well as interesting new insight into the factors governing activity.

1. Introduction

In addition to mediating the storage, retrieval, and transfer of information in all cells, polymerases are the cornerstone of a variety of technologies, ranging from the now indispensible polymerase chain reaction (PCR) [1] to DNA sequencing, and their availability has revolutionized virtually all areas of the biological and medical sciences. Other applications include cloning [2], next generation sequencing by synthesis [3], the diagnosis of genetic diseases [4], the detection of single-nucleotide polymorphisms [5,6], personal identification [7], and systematic evolution of ligands by exponential enrichment (*SELEX*) [8]. In all of these applications, the polymerase is used to copy a deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) template into its complement strand of DNA or RNA using monomeric deoxynucleotide triphosphate (dNTP) or ribonucleotide triphosphate (rNTP) precursors. Under physiological conditions, polymerases are remarkably efficient at utilizing their natural substrates, with polymerization rates as high as 10^3 nucleotides per second and typical fidelities of less than one error for every 10^7 nucleotides replicated (polymerases with proofreading ability can reach a fidelity of less than one error for every 10^9 nucleotides replicated) [9]. However, polymerases are poorly active under some conditions of interest due to instability or the presence of inhibitors, and their optimization for better performance under these conditions would enable or facilitate a variety of additional applications. Also, the ability to control transcription from multiple promoters not used by natural RNA polymerases would increase our ability to manipulate cellular pathways for synthetic biology applications [10].

The four nucleotide monomers of DNA and RNA are composed of a nucleobase (guanine, cytosine, adenine, and thymine or uracil), a ribotyl or deoxyribotyl sugar, and a phosphate that links the nucleotide to the preceding nucleotide. Possibly because nucleic acids evolved to be stable, their functionality is actually quite limited from a physiochemical perspective, and this precludes or limits many of the possible applications of polymerases as well as the biopolymers they synthesize. The ability to recognize modified triphosphates would enable synthesis of the corresponding polymers, and the ability to recognize these polymers as templates would enable their amplification or conversion into DNA or RNA.

A related and particularly interesting class of nucleotide modification is replacement of the entire natural nucleobase with an unnatural one, because two such unnatural nucleobases that are efficiently and selectively paired during enzymatic synthesis could form the foundation of an expanded genetic alphabet [11,12]. In addition to providing unnatural codons that might be used to direct the incorporation of unnatural amino acids into proteins, the modification of unnatural base pairs with linkers could enable the attachment of different functionalities of interest for materials, diagnostics, or SELEX applications [13–16]. Poor recognition of the modified nucleotides by natural polymerases currently limits the development of such expanded genetic systems and thus is a major obstacle toward achieving these ambitious goals.

It is clear that despite the remarkable ability of polymerases to efficiently synthesize natural DNA or RNA, and the revolutionary

* Corresponding author. Address: Department of Chemistry, CB262R, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, United States. Fax: +1 858 784 7472.
E-mail address: floyd@scripps.edu (F.E. Romesberg).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2013.10.040

applications that this has enabled, even more applications would be possible if the properties and activities of polymerases could be appropriately tailored. In nature, this tailoring works by adaptive evolution, an iterative process of selective amplification of the "most fit" members of a population of mutants and their further optimization by mutation and recombination. This process may be recapitulated and accelerated in the laboratory setting, a process commonly known as directed evolution (directed in the sense that a selection pressure is specified by the experimentalist). Over the last several decades, directed evolution has produced a myriad of proteins with optimized properties and reactivities. In addition, unlike with natural evolution, it is straightforward to identify the mutations responsible for the selected phenotype, making it possible to gain unique insights into mechanism.

Herein, we review the directed evolution of polymerases. After a brief discussion of the structures, properties, and substrate repertoires of the natural polymerases, which serve as starting points for directed evolution efforts, we describe the approaches that have been used to generate diversity. We then review in more detail the strategies that have been employed to selectively amplify the most fit mutants. Lastly, we review the polymerase variants that have been evolved from a perspective of the applications that they may eventually enable and the insight that they provide into polymerase function.

2. Polymerases in nature

In general, evolution in nature is thought to proceed from progenitor proteins where some level of the desired activity is already present [17,18]. While it is obvious that directed polymerase evolution should begin with a natural polymerase, the natural polymerases constitute an ancient class of proteins that diverged long ago into different families with subtly different properties and activities, which might make one more suitable than another as starting points. Most generally, polymerases can be classified into three categories based on their substrates and products: DNA polymerases (DNAPs), RNA polymerases (RNAPs), and reverse transcriptases (RTs). DNAPs synthesize strands of DNA complementary to DNA templates (following Watson–Crick base pairing), and are responsible for all DNA replication during cell division as well as the DNA replication involved in DNA recombination and repair, and lateral gene transfer in prokaryotes. Generally, polymerases fold into a common overall structure that resembles a right hand with a palm, fingers, and thumb subdomain (Fig. 1). DNAP-catalyzed synthesis of DNA generally includes five steps: (1) binary complex formation between the DNAP and a DNA template; (2) ternary "open" complex formation upon binding a dNTP; (3) a conformational change from the open to a closed state when the bound dNTP forms a correct Watson–Crick base pair with the template nucleotide in the active site; (4) catalysis of phosphodiester bond formation between the dNTP and the 3′OH of the primer terminus; and (5) a conformational change from the closed to the open state and release of an inorganic pyrophosphate [19–22]. During processive synthesis, the DNAP then translocates the newly elongated primer/template to position the next templating nucleotide in the active site.

Based on sequence homology, DNAPs are classified into six families: A, B, C, D, X, and Y [23]. Classification within families A, B, and C is based on homology with *Escherichia coli* polymerase genes, *polA*, *polB*, and *polC*, respectively, which correspondingly encode Pol I, Pol II, and the α subunit of Pol III [24]. Extensive effort has been focused on rationally optimizing family A DNAPs, especially *E. coli* Pol I, or its truncated variant Klenow fragment (Kf); *Taq* DNAP from *Thermus aquaticus* (*Taq*), or one of its truncated variants Stoffel fragment (Sf) or *KlenTaq*; and the DNAP from T7 bacteriophage, due to their already widespread use and because they are functional in the absence of accessory proteins (with the exception of T7 DNAP, which requires *E. coli* thioredoxin as a cofactor). Family A DNAPs have six common structural motifs: A, B, C, 1, 2, and 6 that surround and form the active site and interact with substrates (Figs. 1A and 2) [25]. Motifs A, C, and 2 are located in the palm domain; Motifs B and 6 are located in the fingers domain; and motif 1 is located in thumb domain. Motifs A, B, and C are more conserved than motifs 1, 2, and 6, especially at the amino acid level, consistent with their functional importance.

RNAPs synthesize RNA molecules complementary to DNA templates, and are responsible in all cells for the initiation of DNA synthesis and all RNA synthesis, and in RNA viruses for information storage. RNAPs recognize specific sequences of DNA, referred to as promoters and terminators, to initiate and halt transcription, respectively. Relatively simple, single subunit RNAPs include those from many viruses, such as T3, T7, SP6, and K11 phages, and from the mitochondria of eukaryotic cells [26], which like their DNAP counterparts, assume a right-hand-like overall structure (Fig. 1B). However, core bacterial RNAPs, such as that from *E. coli*, include five subunits (β', β, α^{I}, α^{II}, ω), as well as another small subunit, σ, which is responsible for promoter recognition [27,28]. Eukaryotic genomes encode five multiple subunit RNAPs, RNAP I–V, which are responsible for the synthesis of different RNAs [29–33]. The RNAPs of bacteria and archaea are most structurally and mechanically similar to eukaryotic RNAP II [34–36]. Regardless of their subunit architecture, RNAPs function similarly to DNAPs, except that full length transcription is preceded by an abortive phase, during which short RNAs are synthesized [37].

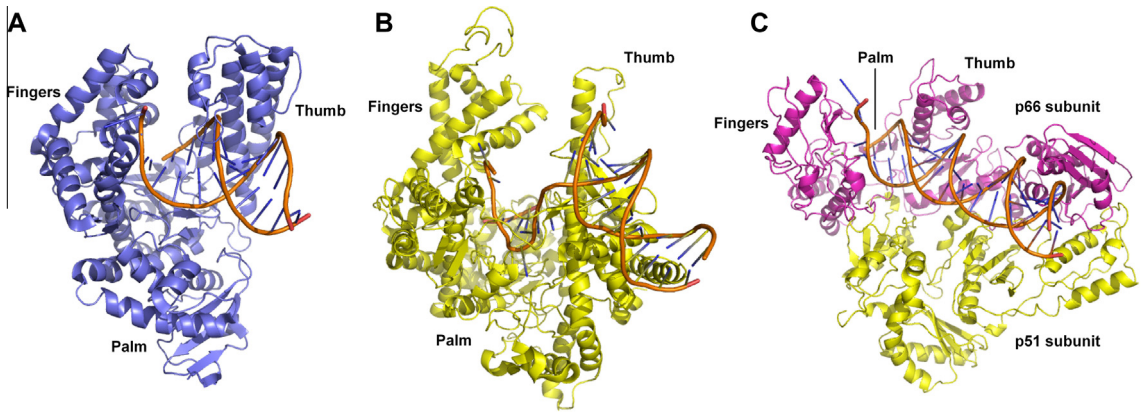

Fig. 1. Common polymerase fold, with palm, thumb, and fingers domains, illustrated with SF DNAP (PDB ID: 1QSY) (A), T7 RNAP (PDB ID: 1QLN) (B), and RT from HIV-1 (PDB ID: 1RTD) (C). Primer/template is shown in orange.