Recognition of deaminated bases by archaeal family-B DNA polymerases

Bernard A. Connolly

Institute for Cell and Molecular Biosciences (ICaMB), University of Newcastle, Newcastle upon Tyne NE2 4HH, U.K.

Abstract

Archaeal family-B DNA polymerases interact specifically with uracil and hypoxanthine, stalling replication on encountering these deaminated bases in DNA template strands. The present review describes X-ray structural data which elucidate the mechanism of read-ahead recognition of uracil and suggests how this is coupled to cessation of polymerization. The possible role of read-ahead recognition of uracil/hypoxanthine in DNA repair is discussed, as is the observation that the feature appears to be limited to replicative polymerases of the archaeal domain.

PCR problems with archaeal DNA polymerases

Biochemists are most familiar with archaeal family-B DNA polymerases, such as the enzyme from *Pyrococcus furiosus* (Pfu-Pol), as reagents for PCR. In several PCR applications, especially the minimization of carry-over contamination, dUTP is used in place of dTTP, but archaeal polymerases show exceedingly poor PCR performance with dUTP [1]. Later, it was observed that archaeal polymerases were strongly inhibited by uracil-containing DNA, binding tightly to such sequences [2]. Other enzymes were unaffected by uracil and thermostable bacterial family-A polymerases, e.g. the enzyme from *Thermus aquaticus*, showed unchanged PCR properties with dUTP. These early studies clearly demonstrated that archaeal polymerases are subject to profound inhibition by uracil, but the mechanism by which this took place remained obscure.

Read-ahead recognition of deaminated bases

In 1999, it was demonstrated that archaeal polymerases are unable to replicate beyond template-strand uracil (the deamination product of cytosine), rather polymerization is halted four bases before its encounter [3]. Such read-ahead recognition (Figure 1) accounts for the inability to use dUTP in the PCR; although this base is incorporated, the resulting amplicons cannot be copied in subsequent cycles. Later, it was proposed that the N-terminal region, a domain found only in archaeal polymerases, was responsible for uracil binding. A pocket in this domain was suggested, using site-directed mutagenesis of Pfu-Pol, to interact with uracil and a plausible structural model was proposed to account for specific recognition of the base [4]. The binding pocket confers high affinity for uracil in single-stranded DNA, most pronounced for primer–templates with uracil at +4 in the template, the exact stalling position [5]. Read-ahead recognition is also seen with hypoxanthine (the deamination product of adenine), first demonstrated with the polymerase from *Sulfolobus solfataricus* [6]. An investigation using Pfu-Pol showed that interaction with hypoxanthine was a little weaker than with uracil and that the same pocket was responsible for specifically recognizing both bases [7]. Other atypical DNA bases, such as xanthine (the deamination product of guanine) and damaged pyrimidines, do not interact with the pocket [5,7].

Structure of an archaeal DNA polymerase in complex with a uracil-containing DNA

A recent X-ray crystallographic investigation, using the polymerase from *Thermococcus gorgonarius* (Tgo-Pol), has elucidated full details of uracil recognition [8] (PDB code 2VWJ). A structure of the apo-enzyme had been published previously [9] (PDB code 1TGO), but the present structure includes a primer–template containing uracil at the +4 position in the template. Figure 2(A) shows the five polymerase domains and the bound DNA, revealing that the N-terminal domain is indeed used for binding of the single-strand located uracil. The double-stranded DNA interacts, as expected, mainly with the thumb domain, which is rotated by 15° and makes less contact with the exonuclease domain compared with the apo-enzyme. Uracil is ‘flipped’ into its binding pocket, which is located just off the single-stranded DNA-binding channel, and the two adjacent bases stack loosely, replacing the weak stacking interaction typical of single-stranded DNA. Strong interactions are seen with the two phosphates immediately flanking uracil: Tyr7 and Arg97 contact the 5′ and 3′-phosphates respectively (Figure 2B). The pocket is shaped to accommodate uracil and reject the four standard DNA bases. The backbone amide linkages of Tyr77 and Ile114 recognize uracil by hydrogen-bonding to the exocyclic O-4 and O-2 groups respectively. Pro36, Pro40 and Pro43 are near C-5 of uracil and prevent......
stable binding of thymine, which contains an additional CH$_3$ function at this position, by steric exclusion (Figure 2C). Val$^{93}$ forms an unusual stacking interaction with the heterocyclic ring of uracil, the isopropyl side chain of Val$^{93}$ and the ring lying in the same plane (Figure 2D). Key uracil-recognizing amino acids are highly conserved in archaeal DNA polymerases, and their mutagenesis compromises uracil recognition [8]. The N-terminal domain is extremely rigid, barely changing conformation following uracil binding. Thus the polymerase appears to be set up to ensnare uracil as the base passes the pocket’s mouth during translocation. The mechanism by which uracil capture leads to cessation of replication is not fully elucidated. However, the enzyme–DNA structure is nearer an editing complex than one suited for polymerization, with the 3′-OH of the last base in the duplex region (the attachment site for the incoming dNTP) far from the active site. The first base in the template, which forms Watson–Crick base pairs with the incoming dNTP, is relatively disordered and not appropriately positioned to fulfil its templating role. Thus binding of uracil at the +4 position in the template may prevent the translocation step required for correct assembly of the polymerase active site.

**Distribution of read-ahead recognition**

Interaction with deaminated bases has been observed for all archaeal family-B polymerases investigated, including the single polymerase found in euryarchaea [3,6,13] and the two polymerases (B1 and B3) found in crenarchaea ([6,13] and S. Gill and B.A. Connolly, unpublished work). Many archaea live at elevated temperatures, conditions that greatly accelerate cytosine and adenine deamination to uracil and hypoxanthine [14,15]. Originally, it was proposed that read-ahead recognition may be a characteristic of hyperthermophiles and serve to protect against increased DNA deamination [3]. However, the family-B DNA polymerase from the mesophilic archaean *Methanosarcina acetivorans* recognizes uracil and hypoxanthine as efficiently as the polymerases from hyperthermophilic archaea [13]. The replicative polymerases from bacteria (Pol III from *Escherichia coli*) and eukaryotes (Pols ε and γ from *Saccharomyces cerevisiae*, responsible for nuclear DNA replication; Pol γ from *Homo sapiens*, responsible for mitochondrial DNA replication) are unable to recognize deaminated bases and do not stall replication in response to these bases [13]. Thus read-ahead recognition seems to be generic to the replicative family-B DNA polymerases of the archaean domain of life, rather than a property of hyperthermophiles. If read-ahead recognition does serve to repair the few deamination events that occur during replication, which cannot be dealt with easily by base excision repair, it remains unclear how bacteria and eukaryocytes cope with the problem.

**Improving PCR performance**

The discovery of read-ahead recognition by archaean DNA polymerases had its origin in difficulties observed during the PCR. One mutation to Pfu-Pol, used to establish the uracil-binding pocket, shows superior properties in PCR. Val$^{93}$ stacks with the aromatic ring of uracil (Figure 2D) and mutation to glutamine completely abolishes the ability to recognize uracil, while fully preserving polymerization activity [4,8]. As a consequence Pfu-Pol V93Q is active when dTTP is completely replaced by dUTP and can be applied in polymerases tends to give rise to irreversible mutation. Fortunately, most damage, e.g. DNA strand breaks which a polymerase cannot cross and bulky bases/abasic sites to which a polymerase cannot easily match an incoming dNTP, automatically stops the replication apparatus. Uracil, and, to a lesser extent, hypoxanthine, are exceptional, being reasonable mimics of thymine and guanine respectively. Thus these two damaged bases are capable of fooling the active sites of polymerases, leading to inappropriate replication and mutation. The provision of a specific uracil/hypoxanthine-recognizing pocket enables the two bases to be recognized as aberrant. The events following uracil/hypoxanthine-induced stalling in archaea await elucidation. However, most stalled replication forks are repaired by damage-tolerant recombination pathways, which preserve the mispair for repair post-replication [10–12]. Similar pathways seem likely in the archaea.
all PCR applications where the use of dUTP is advantageous [1]. Furthermore, V93Q shows superior performance in normal PCR, where dUTP is not added deliberately. Uracil can occur adventitiously during the PCR, as damage to the DNA to be amplified or as a dUTP contamination (resulting in uracil incorporation into amplicons) of the dNTP mixture. Alternatively, the heat–cool cycles of the PCR might deaminate DNA cytosine to uracil or dCTP to dUTP. Such processes give rise to ‘uracil poisoning’ and lead to reduced performance by wild-type archaeal polymerases [16]. A crystal structure of Tgo-Pol V93Q [8] (PDB code 2VWK) shows no conformational changes compared with the apo-enzyme: the slightly longer glutamine side chain simply extends further towards uracil and sterically excludes the base from its binding pocket.

**Unanswered questions**

The physiological role of read-ahead recognition is assumed to be DNA repair, but this hypothesis needs testing. With the increasing ability to manipulate archaea [17], genetic approaches, particularly the creation of a V93Q phenotype, should be applicable. Further investigation is also required to fully elucidate events that follow stalling of replication in response to uracil/hypoxanthine. Additional structural and kinetic data are needed to define exactly how the capture of a deaminated base switches off the polymerase active site. Finally, polymerases never act alone in vivo, but are a component of a multiprotein replication machine, the replisome [18,19]. The influence of other replisome components on the handling of deaminated bases remains to be established.

---

**Figure 2** | X-ray structural details of the interaction between Tgo-Pol and a primer-template containing uracil at the +4 position in the template

The DNA used has the sequence AAUGGAGACGGCTTTGGCCGTGTC, which forms a snap-back primer-template containing a (T)₄ loop. The single-stranded template region is underlined. (A) Overall structure with the polymerase domains colour-coded and the DNA shown in red. Uracil is located in the N-terminal domain (yellow). (B) Uracil flipped into its binding pocket. The hydrogen bonds between the 5′ and 3′-phosphates flanking the uracil and Tyr⁷ [2.7 Å (1 Å = 0.1 nm)] and Arg⁹⁷ (3.2 Å) are illustrated (broken lines). (C) Amino acids lining the uracil-binding pocket of Tgo-Pol. The amide nitrogens of Ile¹¹⁴ and Tyr¹⁷ form hydrogen bonds (broken lines) with uracil O-2 and O-4 respectively. Val⁹³ stacks over the heterocyclic ring of uracil and Pro¹⁵⁶, Pro⁹⁰ and Phe¹¹⁶ are adjacent to the uracil C-5. (D) Stacking of Val⁹³ over the uracil ring with surfaces shown as dots. (A–C) are reprinted from [8] with permission, © 2008 Elsevier.
Figure 3 Consequences of the conversion of cytosine (C) into uracil (U) and adenine (A) into hypoxanthine (H) by deamination

In double-stranded DNA, a U:G or H:T mispair are the immediate products. On replication, 50% of the progeny inherit a transition mutation.

Funding

Work was funded by the Biotechnology and Biological Sciences Research Council [grant numbers BBS/B/05060 and E19804, Cancer Research UK [grant number C5098] and the European Union [grant number QLK3-CT-2001-00448].

References


Received 22 July 2008
doi:10.1042/BST0370065