Getting it Right: How DNA Polymerases Select the Right Nucleotide

Samra Ludmann and Andreas Marx*

Abstract: All living organisms are defined by their genetic code encrypted in their DNA. DNA polymerases are the enzymes that are responsible for all DNA syntheses occurring in nature. For DNA replication, repair and recombination these enzymes have to read the parental DNA and recognize the complementary nucleotide out of a pool of four structurally similar deoxynucleotide triphosphates (dNTPs) for a given template. The selection of the nucleotide is in accordance with the Watson-Crick rule. In this process the accuracy of DNA synthesis is crucial for the maintenance of the genome stability. However, to spur evolution a certain degree of freedom must be allowed. This brief review highlights the mechanistic basis for selecting the right nucleotide by DNA polymerases.

Keywords: DNA polymerases · DNA replication · Hydrogen bonding · Modified nucleotides · Selectivity of DNA polymerases

The DNA polymerases’ ability to achieve a balance between selective DNA synthesis whilst allowing at the same time some degree of error, was investigated intensively in the past by employing biochemical, biophysical, and chemical means. Scientists of various disciplines have undertaken enormous efforts to elucidate the complex mechanism of these molecular machines. With simple organisms having already more than one polymerase, these projects proved to be rather complex and complicated. Currently, more than a dozen human DNA polymerases and five DNA polymerases in E. coli are known, varying significantly in their selectivity.[1–6] Replicative DNA polymerases are high fidelity DNA polymerases with low error rates, making only one mistake within one million synthesized nucleotides. This high accuracy derives from the polymerase’s intrinsic fidelity during the polymerization reaction, combined with the error-correcting 3'-5' exonuclease activity (‘proofreading’) and mismatch repair mechanisms. The replication in E. coli and bacteriophages, for example, displays in vivo a base substitution error rate of $10^{-2}$–$10^{-8}$ per nucleotide in the absence of mismatch repair.[7] This error rate can be improved up to $10^{-8}$–$10^{-10}$ in E. coli by proofreading, mismatch repair and numerous other factors.[8,9] In contrast to replicative DNA polymerases, DNA polymerases involved in DNA repair and particularly in translesion synthesis exhibit relatively low fidelity, i.e. an error rate of approximately 1/10, which is a remarkable one in ten catalyzed phosphodiester bond formations error rate for human DNA polymerase η (eta).[10] Even more remarkable is human DNA polymerase ι (iota), inserting G opposite T rather than A opposite T in vitro.[11]

This wide range of variation in fidelity among the numerous DNA polymerases raises the question: Which determinants affect fidelity of these enzymes? During the past years, valuable new insights have been gained into the selectivity process of DNA polymerases by employing carefully designed synthetic nucleotide and oligonucleotide analogs in functional enzyme studies.[11] The following covers some recent approaches and highlights the mechanistic model drawn from conclusions from experimental data.

Hydrogen Bonding

Fidelity is the DNA polymerases’ ability to discriminate between a complementary base pair and a mismatch. Thus, DNA polymerases have to choose the correct (i.e. Watson-Crick-paired) nucleotide out of a pool of four structurally similar dNTPs. It was a common prediction that nucleotide selectivity of DNA polymerases is manifested in the hydrogen bonding capacity of the nucleobases and the formation of the correct base pair according to Watson and Crick (Fig. 1a).[12] Nevertheless, the small free energy barriers between matched and mismatched base pairs in solution could account for error rates of about 1:100, revealing that the DNA polymerases’ selectivity does not primarily depend on the hydrogen bonding capability.[13]

In order to investigate the impact of hydrogen bonding on DNA replication selectivity mechanisms, isosteric nucleotide analogs, lacking hydrogen bonding...
Steric Effects

Based on the studies employing the isosteric probes, Kool postulated the model of ‘active site tightness and substrate fit in DNA replication’[14,15], emphasizing that especially steric effects explain the highly accurate performance of DNA polymerases. The analysis of crystal structures suggests that the active sites of selective DNA polymerases form tight binding pockets with a geometry which is complementary to the respective incoming canonical dNTP[16] and the respective templating nucleobase.[14,17] In general, Fig. 1 depicts that the four combinations of Watson-Crick base pairs have only little variation in their overall shape. Their geometric constraints show minor alterations in the minor and major groove, but no variation in the overall length (Fig. 1b).[14,18] In the postulated geometric model, the incoming nucleotide, placed opposite the templating nucleobase, has to fit into the geometric constraints of the consensus base pair shape (Fig. 1c). In case of a mismatched nucleobase pair steric clashes may occur. Thus, either the incoming nucleotide cannot enter the binding pocket or if it partially does, the triphosphate moiety cannot align correctly for efficient phosphodiester bond formation. However, in this scenario it must be considered that the tightness of the active site binding pocket is only defined by the closed conformation of the enzyme. The closure of the active site, a prerequisite for efficient catalysis, might also depend on whether a canonical or non-canonical nucleobase pair is formed and thereby has impact on the outcome.

Further evidence for the steric model was derived from insertion studies opposite an abasic site has been termed the ‘A-rule’. Since the Watson-Crick recognition cannot take place in the presence of an abasic site, it is obvious that other properties account for the selectivity. Therefore, superior stacking as well as solvation properties of adenine have been discussed to be the driving force behind the adenine selection[15,19,20] Considering the geometric model as selection criteria opposite abasic sites, a non-natural nucleotide analog, being almost as large as a natural base pair, should show the highest incorporation efficiency. Indeed, Matray and Kool identified the pyrene nucleoside triphosphate (dPTP) as a perfect match in the absence of a templating base (Fig. 2). They could show that the pyrene-modified nucleotide is incorporated by DNA polymerase I from E. coli with higher efficiency than any other natural nucleotide, corroborating the steric model for selection of the right nucleotide.

Varied Selectivity among DNA Polymerases: Probing Active Site Tightness

DNA polymerase selectivity often varies significantly depending on the DNA polymerase. It is assumed that these different error propensities result from different properties like shape and tightness among the nucleotide binding pockets.[14] Thus, high-fidelity DNA polymerases are believed to form more rigid binding pockets, tolerating less geometric deviation, whereas low-fidelity enzymes exhibit more flexibility, resulting in decreased fidelity. To investigate this hypothesis, we developed nucleotide analogs of increasing steric demand[21] by replacing the standard (C)4-hydrogen with alkyl groups of increasing size (Fig. 3). Structural and functional studies on DNA polymerases showed that the sugar residues of the nucleotides are part of the substrate recognition process. Experiments on the selectivity, employing the Klenow fragment, which is a high-fidelity DNA polymerase, and the probes T4-dTP and T6-dTP showed that with increasing size of the sugar moiety, the tolerance for non-canonical conformations for nascent nucleotide pairs decreases. In consequence, the selectivity of the DNA polymerase reaction increases, which supports the geometric model for achieving DNA polymerase selectivity. Furthermore, the action of the steric probes on the Klenow fragment was compared with their action on archaeal Y-family DinB homolog (Dbh) of Sulfolobus solfataricus as a representative of a low-fidelity DNA polymerase.[22] Indeed, the substitution of a hydrogen atom by a methyl group at the 4-position in the nucleotide analog resulted in a reduction of the maximum nucleotide incorporation rate by about 40-fold and 12-fold for Klenow fragment and Dbh, respectively. In the presence of dT4-dTP, the incorporation rate of both enzymes are further twofold reduced. A drastic decline in incorporation rates for both enzymes is observed in case of misincorporation of either TTP or T6-dTP opposite a templating G. Interestingly, the binding affinity for these nucleotides drops about 16- and fourfold for the Klenow fragment and Dbh, respectively. The varied proficiencies of the two enzymes in processing the steric probes indicate varied flexibility of the enzymes’ active sites. These findings again support the model of active site tightness as criteria for varied DNA polymerase selectivity.

Further evidence along these lines came from Kool and colleagues by usage of gradually expanding thymine nucleobase analogs.[17] They increased the size of the base analog incrementally over a 1.0 Å range by the substitution of the oxygen atoms of thymine with H, F, Cl, Br, or I, respectively (Fig. 4). The kinetics with the Klenow fragment illustrated that incorporation efficiency opposite adenine increased throughout the series, reaching a peak at the chlorinated compound. Thus, high-fidelity DNA polymerases are believed to form more rigid binding pockets, tolerating less geometric deviation, whereas low-fidelity enzymes exhibit more flexibility, resulting in decreased fidelity. To investigate this hypothesis, we developed nucleotide analogs of increasing steric demand[21] by replacing the standard (C)4-hydrogen with alkyl groups of increasing size (Fig. 3). Structural and functional studies on DNA polymerases showed that the sugar residues of the nucleotides are part of the substrate recognition process. Experiments on the selectivity, employing the Klenow fragment, which is a high-fidelity DNA polymerase, and the probes T4-dTP and T6-dTP showed that with increasing size of the sugar moiety, the tolerance for non-canonical conformations for nascent nucleotide pairs decreases. In consequence, the selectivity of the DNA polymerase reaction increases, which supports the geometric model for achieving DNA polymerase selectivity. Furthermore, the action of the steric probes on the Klenow fragment was compared with their action on archaeal Y-family DinB homolog (Dbh) of Sulfolobus solfataricus as a representative of a low-fidelity DNA polymerase.[22] Indeed, the substitution of a hydrogen atom by a methyl group at the 4-position in the nucleotide analog resulted in a reduction of the maximum nucleotide incorporation rate by about 40-fold and 12-fold for Klenow fragment and Dbh, respectively. In the presence of dT4-dTP, the incorporation rate of both enzymes are further twofold reduced. A drastic decline in incorporation rates for both enzymes is observed in case of misincorporation of either TTP or T6-dTP opposite a templating G. Interestingly, the binding affinity for these nucleotides drops about 16- and fourfold for the Klenow fragment and Dbh, respectively. The varied proficiencies of the two enzymes in processing the steric probes indicate varied flexibility of the enzymes’ active sites. These findings again support the model of active site tightness as criteria for varied DNA polymerase selectivity.
and co-workers investigated the Y-family DNA polymerase Dpo4 towards the acceptance of the five nonpolar thymidine analogs (Fig. 4). At first, Dpo4 followed the same trend as the high-fidelity Klenow fragment, preferring the same size optimum. However, the distinctive difference between these enzymes is manifested by the range of the incorporation efficiencies for the most to the least efficiently processed thymidine analog. While for Dpo4 the variation differs only by 30–35-fold, for the Klenow fragment a 1100-fold difference was observed. These results indicate that Dpo4 has, in comparison to the Klenow fragment, decreased steric constraints in the active site, which indicates a loose active site that can accommodate non-canonical conformations with smaller energetic costs.

In summary, the presented studies highlight that marginal changes of the overall shape and size of the substrate have significant impact on nucleotide selection. Furthermore, these data provide strong experimental evidence that variations of steric constraints within the nucleotide binding pockets of DNA polymerases account for the differences in nucleotide incorporation selectivity.

**Minor Groove Scanning**

In addition to Watson-Crick hydrogen bonds another type of hydrogen bond contributes to DNA polymerase selectivity. The recognition of the minor groove, also known as minor groove scanning, is mediated by hydrogen bonds between the minor groove side of the nucleobases and the enzyme. These H-bonds are not only formed with the newly established base pair in the insertion site, they are also present in the post-insertion site and several positions downstream of the DNA duplex probing correct base pairing. Ternary complexes of DNA polymerases with DNA and dNTP substrate showed that functional groups at the minor groove site of the nucleobases (O(2) atoms of pyrimidines and N(3) atoms of purines) specifically interact with the side chains of amino acids in proteins either directly or via water molecules. As all four nucleotides carry these groups donating lone-pair electrons at similar relative positions from the nucleic acid backbone (Fig. 1c) DNA polymerases can recognize all four bases with the same interaction mechanism. Disruption of the interaction patterns in the minor groove can lead to misalignment of the components involved in catalysis in the active site and therefore hinder insertion of non-canonical nucleotides or elongation of non-canonical primer/template complexes.

The impact of minor groove interactions was investigated by Kool and colleagues using two non-polar nucleoside isosteres of dATP: dZTP is devoid of Watson-Crick hydrogen bonding capability and dQTP, which also lacks the hydrogen bonding capability at the Watson-Crick side but retains the nitrogen N(3) of adenine as a minor groove acceptor (Fig. 5a). In vitro DNA polymerase reactions employing dATP, dZTP, and dQTP or the residues dA, dZ and dQ in the template position, revealed that minor groove interaction show minor impact on the incorporation efficiency by Klenow fragment, T7 DNA polymerase, Taq DNA polymerase and HIV1-RT (viral DNA polymerase).

However, other polymerases showed improved incorporation efficiencies in presence of a single minor groove interaction at either the template position or the incoming triphosphate or at both positions. Further elongation studies indicated that a hydrogen bond between the primer-terminal nucleobase (position n-1) and the polymerase is essential for the elongation reaction employing Klenow fragment, KlenTaq and HIV1-RT. In contrast, a H-bond acceptor at the n-1 template nucleotide is not markedly promoting the reaction. In the last years, the suitability of dQ and dZ as purine analogs was questioned as their C(1)-H groups can clash with the N(3)-H of a pairing dT causing a twist of the dT nucleobase by around 25°, as was shown in a ternary structure of RB69 DNA polymerase with a dT-dQTP pair in the insertion site. Therefore, effects observed with the dZ analog might also result from the different geometry of the base pairs formed and not only from its missing minor groove H-bonding capability.

Next, 3-deazaadenine (3DA) or 3-deazaguanine (3DG) were used structurally and functionally to investigate the role of minor groove H-bonds in DNA synthesis by DNA polymerases (Fig. 5b). These studies revealed that minor groove hydrogen bonding interactions can be crucial for efficient nucleotide incorporation. However, the interaction patterns differ among DNA polymerases, so that their impact on selectivity has to be regarded individually for each enzyme. Although several minor groove interactions were observed in crystal structures of various DNA polymerases only a few of them have been investigated regarding their relevance on their selectivity profile by mutation studies.

**Nucleobase Stacking**

Besides hydrogen bonding between base pairs, base stacking interactions between the extended pi-surface of the nucleobases play an important role in dNTP binding and especially in stabilizing the whole DNA duplex. Stacking interactions could, in principle, enforce the formation of non-canonical nucleobase pairs due to stabilizing complexes between incoming dNTP and primer/template that do not obey the Watson-Crick rule. Therefore, non-ca-
nonol nucleotides are incorporated with a higher frequency resulting in a higher error rate. Recent structural and functional data employing hydrophobic nucleobase surrogates with increased stacking ability reveal that DNA polymerase active sites are not only capable of selecting for a correct structure among the pairing nucleotides but also, at least in some cases, capable of enforcing it.[33] This suggests that DNA polymerases might have evolved to favor a coplanar geometry to prevent natural mispairing via cross-strand intercalation and instead allow only the more specific, edge-to-edge hydrogen-bonding interactions of the Watson-Crick type.[33]

Conclusion

In summary the depicted examples demonstrate that the determinants of fidelity of DNA polymerases are highly diverse. For elucidation of the mechanistic basis of the DNA polymerases selectivity chemically modified nucleotide analogs proved to be valuable tools to gain more insights into these complex molecular machineries. However, we are far from a complete picture of this process, since often an organism expresses several DNA polymerases and crucial auxiliary factors. Moreover, studies indicated that auxiliary proteins show a high impact on the selectivity process involving DNA polymerases[34] which further adds to the complexity of the system.

Received: November 3, 2015