



CHALMERS
UNIVERSITY OF TECHNOLOGY

Recognition Mechanism of Complementary Nucleobases and Sequences in DNA and RNA: Interplay of Watson-Crick Hydrogen Bond Formation and

Downloaded from: <https://research.chalmers.se>, 2026-05-25 13:48 UTC


Citation for the original published paper (version of record):

Takahashi, M., Nordén, B. (2026). Recognition Mechanism of Complementary Nucleobases and Sequences in DNA and RNA: Interplay of Watson-Crick Hydrogen Bond Formation and Base Stacking Interactions. *DNA*, 6(1).
<http://dx.doi.org/10.3390/dna6010013>

N.B. When citing this work, cite the original published paper.

Review

Recognition Mechanism of Complementary Nucleobases and Sequences in DNA and RNA: Interplay of Watson–Crick Hydrogen Bond Formation and Base Stacking Interactions

Masayuki Takahashi ^{1,2,*}  and Bengt Nordén ³

¹ Cell Biology Center, Institute of Science Tokyo, Yokohama 226-8503, Japan

² Faculty of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto 606-8522, Japan

³ Department of Chemical and Biological Engineering, Chemistry, Chalmers University of Technology, 412 96 Gothenburg, Sweden; norden@chalmers.se

* Correspondence: takahashi.m.ay@m.titech.ac.jp or takahashi.m@kpu.ac.jp

Abstract

A/T(U) and G/C nucleobase pair formation in DNA and RNA is crucial to numerous fundamental biological processes, including replication, transcription, and translation. The specificity of A/T(U) and G/C base pairing is used for the recognition of complementary sequences in medical and biotechnological applications, such as PCR, nucleic acid drugs, and CRISPR–Cas9-based gene editing. It is essential to understand and predict fidelity of biological reactions, avoiding off-target binding, in order to improve the accuracy and efficacy of applications. In particular, recognition mechanisms of complementary bases or whole sequences must be understood in detail. Despite the prevailing view that Watson–Crick hydrogen bonding is a primary mechanism for complementary base recognition, several experiments have shown that DNA polymerase does not require hydrogen bonding to select complementary bases. Other factors, such as the shape and geometric fitting of the bases and the base stacking, also appear to be crucially involved in the selection. E.g., artificial bases lacking the ability to form hydrogen bonds can still be recognized by DNA polymerase solely based on base-pair geometry. However, hydrogen bonding also contributes importantly to recognition. The accuracy of selecting a complementary nucleobase or sequence varies depending on reactions, suggesting the co-existence of multiple selection mechanisms. This review provides an overview of biological processes and applications involving base pairing and discusses the molecular mechanism underlying complementary base recognition.

Keywords: nucleobase pair; DNA; RNA; Watson–Crick hydrogen bond; base stacking; complementary nucleotide sequence; DNA polymerase; gene targeting



Academic Editor: Mauricio Alcolea Palafox

Received: 19 December 2025

Revised: 2 February 2026

Accepted: 12 February 2026

Published: 4 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The discovery of the double-helical structure of DNA marks the dawn of modern biology. DNA comprises two strands going in opposite directions, which are connected by A/T and G/C base pairs [1]. This structure, especially the exclusive A/T and G/C base-pair formation, allowed Professor Crick to propose the famous Central Dogma, which explains how DNA is duplicated and how genetic information dictates the amino acid sequence of proteins and, thus, their phenotype [1,2]. Later experiments further supported this theory [3–8]. DNA polymerase binds to one of the DNA strands and uses it as a template to synthesize the complementary strand. The enzyme selects and adds a base

that fits the corresponding base on the template DNA by base-pairing to it. For protein synthesis, messenger RNA (mRNA) is first produced similarly by RNA polymerase using one of the DNA strands as a map. Subsequently, the ribosome binds the mRNA, selects a transfer RNA (tRNA) that forms base pairs with three consecutive nucleotides on the mRNA (codon/anticodon interactions), and ligates the amino acids carried by the tRNA to form a polymer. Base pair formation is thus essential for several fundamental biological reactions comprising replication, transcription, and translation.

Base pair formation has been later found to be involved in many other critical biological processes, such as regulation of translation by small-interfering RNA (siRNA), genomic maintenance and meiosis via homologous recombination, and disruption of viral DNA in bacteria by the clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) system [9–14].

Base pairing has found crucial applications in medicine and biotechnology, including use of the polymerase chain reaction (PCR), siRNA-like drugs, CRISPR–Cas9-based gene editing, and DNA microdevices [15–19]. PCR is based on the binding of primers complementary to parts of the target gene. Moreover, nucleic acid drugs that bind to a target mRNA via sequence complementarity and inhibit its translation into the target protein are currently in development. Homologous recombination and CRISPR–Cas9-based gene editing involve base-pairing to target the editing site. Similarly, DNA nano-machines have been constructed exploiting organization of DNA fragments through specific base pairs. All these applications are based on efficient and correct interactions in pairs of complementary nucleotides. Therefore, understanding the mechanistic details of how the complementary bases (or nucleotide sequences) are crucial.

The selection of complementary bases is frequently considered to be based entirely on hydrogen bond formation between complementary bases (Watson–Crick hydrogen bonds). However, there is no rigorous proof of this claim. On the contrary, several studies have demonstrated that DNA polymerase can select complementary bases in the absence of hydrogen bond formation, thereby suggesting the presence of additional factors [20–22]. This concept is supported by the observation of correct base-pair recognition also in the absence of protein, as thermal stability of double-stranded oligonucleotides, depending on the type of mismatch [23].

The goal of this paper is to review the various biological reactions that involve specific base-pair formation and discuss the mechanisms by which recognition of complementary base (or sequence) is achieved. It is our hope that a better understanding and generalization of mechanisms will help development of new applications.

2. Biological Roles of Base Pair Formation

Base pairing is essential for replication, transcription, and translation. In addition to these fundamental biological reactions, several other biological processes involve base pairing between two complementary sequences, as described below.

2.1. Homologous Recombination for DNA Repair and Meiosis

Homologous recombination consists of strand exchange between two DNA molecules with identical nucleotide sequences. The reaction is essential for amending stalled replication forks and repairing double-strand breaks [11,12]. It also plays a role in the formation of chromosome pairs during meiosis [13]. The recognition of sequence homology between two DNA molecules can be achieved through base-pairing between their complementary strands (Figure 1a).

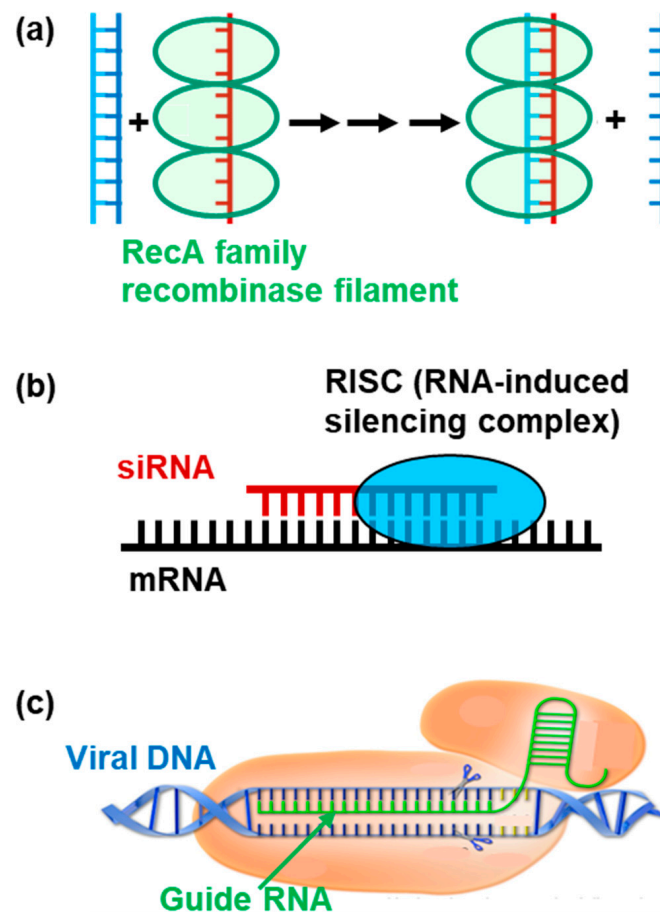


Figure 1. Schematic presentation of biological reactions involving base pair formation: (a) Recombinase-promoted DNA strand exchange for homologous recombination. The reaction steps are based on Ito et al. [24]. (b) CRISPR Cas9 nuclease bound to the target DNA. The figure is produced from the photo licensed under CC BY-SA-NC. (c) siRNA-DISC complex bound to the target mRNA. The figure is produced from the photo licensed under CC BY-SA-NC.

2.2. Regulation of Translation by siRNA and microRNA

siRNA is a non-coding RNA of 20–24 bases that silences the expression of specific genes [9,10]. It is produced as a self-complementary double-stranded RNA. After entering a cell, the siRNA molecule is incorporated into the RNA-induced silencing complex (RISC) and processed into a single-stranded RNA. This complex subsequently binds to the target mRNA by forming base pairs via sequence complementarity and cleaves it (Figure 1b). This results in the repression of translation of the target genes and the obviated production of the target protein. Consequently, this system is widely used in biological research and biomedical applications [16,17].

MicroRNAs (miRNAs) are small, highly conserved non-coding RNA molecules that—like siRNAs—regulate the translation of specific genes [25,26]. miRNAs are produced endogenously, unlike siRNAs, and are processed by the Microprocessor complex (Drosha/DGCR8) and Dicer. Thereafter, one strand (‘guide strand’) of a 20 to 24-nucleotide-long miRNA duplex becomes part of the miRNA-induced silencing complex (miRISC) and binds to the target mRNAs via sequence complementarity. The complex represses protein production through mRNA degradation or translational inhibition. Furthermore, the level of complementarity between the miRNA and mRNA target determines which silencing mechanism will be employed.

2.3. Disruption of Virus DNA by the CRISPR–Cas9 System

CRISPR–Cas9 system is a bacterial defense system that functions against bacteriophage (virus) infection. Typically, bacteria and archaea conserve DNA segments of bacteriophages that have previously infected them [14]. The guide RNA in CRISPR–Cas9 is produced from these DNA segments and incorporated into the Cas9 nuclease. With the help of the guide RNA, Cas9 detects and destroys DNA from related bacteriophages during subsequent infections (Figure 1c). This system is used explicitly in gene editing [18,27,28].

2.4. Structure of Non-Coding RNA

Some RNAs—such as tRNA, ribosomal RNA (rRNA), and spliceosomal RNA (snRNA)—function in various biological processes without being translated into proteins. These RNAs form specific three-dimensional structures to perform their functions [29,30]. The RNA structure is supported by the formation of partial double-helical structures (stems) by base pairing within the RNA molecule (Figure 2).

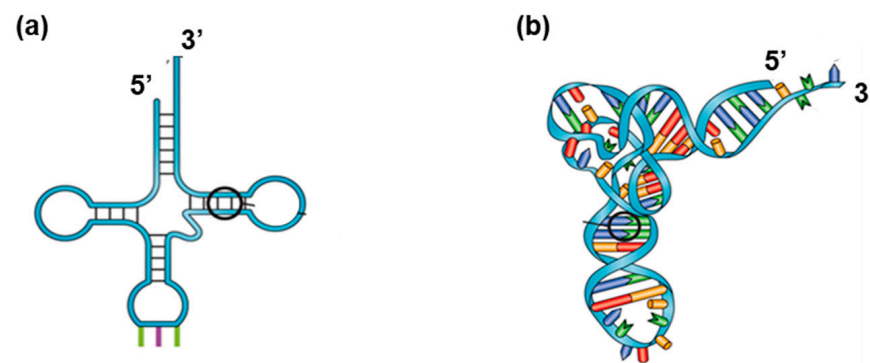


Figure 2. Structure of tRNA stabilized by internal base pair formation: (a) Interactions between bases. (b) 3-D structure. One of the base-paired parts is circled. The figures are produced from the photo licensed by CC-BY.

2.5. Size and Recognition Accuracy of Sequence Complementarity in the Reactions

A specific enzyme (or protein complex) catalyzes each of these biological reactions involving base pairing. Clarifying the mechanisms by which these enzymes recognize complementary single bases or sequences is essential to understanding the high fidelity of these reactions and the maintenance of life. DNA polymerase typically makes only one mistake over 10^7 events. Interestingly, we note that DNA polymerase does not incorporate incorrect bases evenly [31]; some mismatches occur more frequently than others. Moreover, CRISPR–Cas9 and siRNA recognize 20-base-pair complementarity, indicating that these reactions discard DNA or RNA fragments with one mismatch over 20 bases. Consequently, these features cannot account for the recognition of target sequences solely through hydrogen bonding.

The difference in the size of the complementary sequences suggests that reaction mechanisms can vary from reaction to reaction. DNA and RNA polymerases recognize only one complementary base at each step, whereas the ribosome recognizes three base pairs. Similarly, siRNA reactions and CRISPR–Cas9 require more than 20 nucleotide complementarities, whereas miRNA reactions accept partial sequence complementarity. Homologous recombination recombinases potentially recognize approximately six nucleotide pairs to initiate the reaction [32,33]. The recognition accuracy also varies with the reaction. Wherein replication appears to be more accurate than transcription [34,35]. Moreover, the replication fidelity varies among DNA polymerases [36].

3. Biotechnological Applications Using Base Pair Formation

3.1. Antisense- and siRNA-Based Gene Regulation and Drugs

Antisense and siRNA technologies use oligonucleotides complementary to the target mRNA, thereby regulating the production of the corresponding protein. This strategy is used in biological research to examine the cellular function of a gene by observing biological effects resulting from reduced expression of the target gene [37]. It has led to the development of oligonucleotide-based drugs for treating infectious diseases, genetic disorders, and cancer by reducing the production of disease-associated proteins [16,17,38]. In addition to binding to the mRNA target in the cytoplasm, antisense RNA can enter the nucleus and bind to DNA or pre-mRNA, thereby affecting transcription and splicing of the target mRNA. siRNA functions only in the cytoplasm and regulates translation. These systems allow rapid drug development without requiring target characterization and complex chemical synthesis of drug candidates. The technique can also easily adapt to the drug resistance caused by mutations of the target.

Although the therapy demonstrates high selectivity, off-target effects persist. Accordingly, a partial sequence of the complementary off-target mRNA can be repressed by siRNA and antisense therapies, consequently causing secondary effects. Due to the extensive size of the human genome, several similar sequences exist [39]. The presence of many off-targets reduces the drug's availability to the target and compromises the treatment's efficacy. Therefore, resolving this issue is essential to developing better treatment methods [40].

3.2. Homologous Recombination- and CRISPR–Cas9-Assisted Gene Editing

Homologous recombination is used for gene editing [41]. The system can insert a desired DNA segment into a defined position or replace a target segment with another one using sequence homology between the target segment and the corrector oligonucleotide. Gene editing is used to cure genetic diseases or to create genetically modified animals and crops. To increase homologous recombination efficacy, a double-strand break is induced at a desired position using CRISPR–Cas9 or other systems. The Cas9 enzyme cuts genomic DNA at a specific DNA sequence guided by the guide RNA, resulting in a double-stranded break.

However, homologous recombination occurs infrequently and is ineffective for treating human somatic cells. Consequently, a novel approach using only the CRISPR–Cas9 system was developed. In this system, a nucleotide-modifying enzyme—such as a deaminase—is fused to CRISPR–Cas9 and selectively “corrects” the point mutation [26,27].

3.3. DNA Hybridization

DNA hybridization is a simple molecular biology technique that involves the association of two complementary single strands of DNA (or RNA) via the specific base-pairing rules governing A/T (U) and G/C. The ability to detect specific DNA sequences has made hybridization a powerful technique with a wide range of applications, especially prior to the development of sequencing techniques.

This technique uses a labeled oligonucleotide to facilitate the detection of a DNA fragment containing the target segment among many pieces or the localization of the target segment within a long chromosomal DNA molecule [42–44]. It is used to identify homologous genes across various organisms and to trace their evolution. Additionally, it can be used to detect the genetic material of pathogens—such as viruses and bacteria—in patient samples. This technique is beneficial for identifying genetic mutations, diagnosing inherited diseases, and determining a person's carrier status for specific medical conditions. It is also used for DNA fingerprinting to identify individuals based on their unique DNA profile.

3.4. PCR for Amplification of a Specific DNA Segment (Or Fragment)

PCR is a technique used to rapidly obtain numerous copies of a specific segment of DNA from a small sample. This process, called DNA amplification, uses DNA polymerase enzymes to create new DNA strands, thereby yielding sufficient material for detailed study, analysis, or diagnosis [15,45]. Because it requires a primer that binds to a template DNA strand for DNA polymerase to initiate replication, this technique can amplify a specific segment by selecting appropriate primers. PCR tests are used for medical diagnostics, forensic analysis and genetic testing, along with the detection of infectious agents such as viruses.

3.5. DNA Nanodevices

DNA nanodevices are artificial nanoscale machines constructed by the designed association of several DNA fragments via sequence complementarity. The association creates a specific 3-D structure [19] that can be altered by changes in the environment (pH or ion concentration) and other factors. The device can be used as a drug delivery system, a sensor, and a motor in therapeutics [46–48]. The system is also exploited for computation (DNA computers) [49].

3.6. Aptamers (Chemical Antibodies)

Aptamers are single-stranded DNA or RNA that are 20–80 nucleotide long. They form a specific structure through internal interactions and bind a particular target. Double-helical stems formed within the molecules—as in tRNA—support their structures [48]. Some aptamers are developed as drugs or drug delivery systems to treat diseases such as age-related macular degeneration (AMD) and other forms of eye disease, cancer, and infectious diseases [50–52].

3.7. Problems with Applications

The primary concern with these applications is the off-target effect. Owing to the enormous size of the human genome, several sequences may share similarities with the target sequence. Off-target binding reduces the amount of siRNA available to the target, thereby reducing the overall efficacy. Additionally, binding to off-targets inhibits the translation of unexpected genes and provokes unintended adverse effects. Especially in CRISPR–Cas9 applications, irreversible genome modification can occur, resulting in permanent damage. Consequently, understanding the mechanism of complementary sequence recognition and improving the design of complementary sequences are essential for the further development of these technologies [53–55].

4. Molecular Mechanism of Complementary Base Recognition

Despite the importance of base pairing in several essential biological processes and in the development of biotechnological applications, the mechanism for recognizing complementary bases (or sequences) remains unknown.

4.1. Existence of Factors Other than the Hydrogen Bond

The formation of the Watson–Crick hydrogen bond is frequently considered as a primary element in the selection of the complementary nucleobase (or nucleotide sequence). Base pairs are usually held by hydrogen bonds, which suggests that hydrogen bond formation ensures the specificity of A/T and G/C base pairing. These hydrogen bonds exist in both DNA and RNA. However, there is no experimental evidence proving that complementary bases are selected solely through hydrogen bond formation.

The experimental examination of whether DNA polymerase selects complementary bases solely through hydrogen bonding was addressed 40 years after the discovery of the double-helical structure of DNA. Kool et al. used modified nucleobases that cannot form hydrogen bonds with any nucleobases during the replication [21,23]. They replaced the hydrophilic groups of T and A bases, which form hydrogen bonds, with nonpolar chemical groups (Figure 3a). These modifications disrupt hydrogen bond formation with any base. Subsequently, these modified bases were incorporated into a template DNA strand and tested for replication with a DNA polymerase *in vitro*. It was observed that DNA polymerase incorporated an A opposite the modified T, and a T opposite the modified A, despite the absence of hydrogen bonding (Figure 3b,c). In conclusion, the DNA polymerase selects the “correct” bases without the need to form hydrogen bonds. The results indicate the presence of another factor (or factors) that direct the complementary base section.

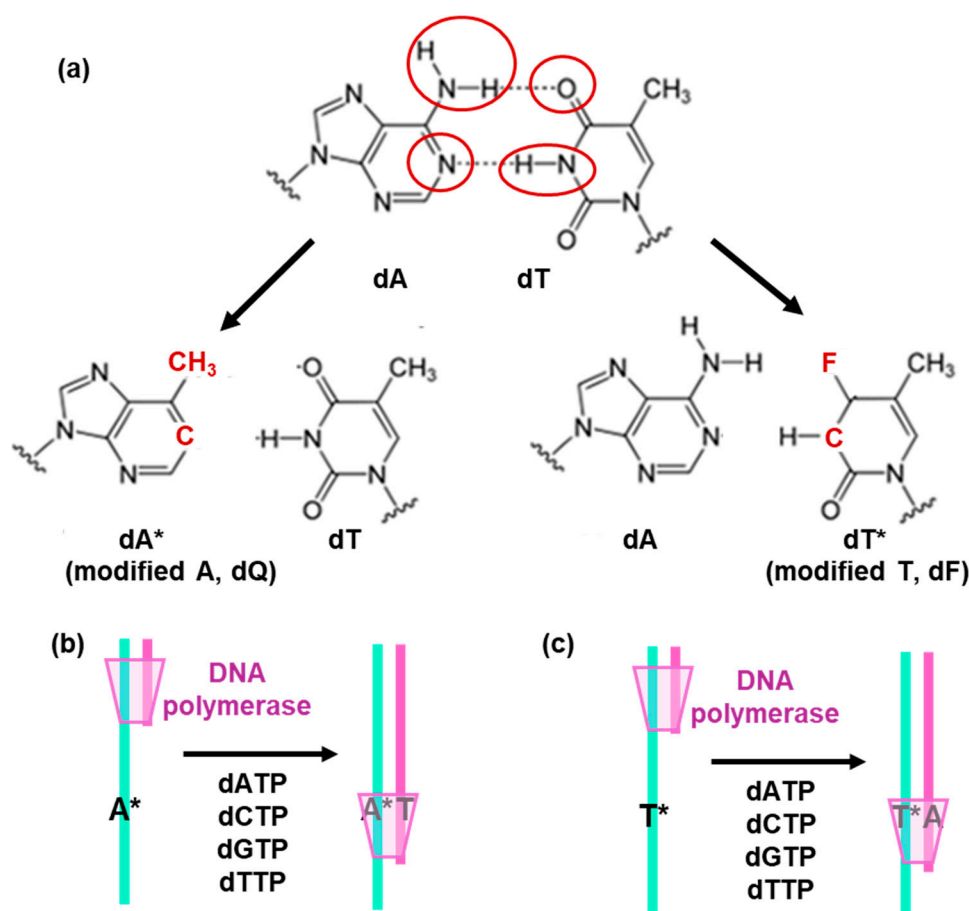


Figure 3. Replication of modified DNA bases that do not form any hydrogen bonds: (a) Replacement of hydrophilic groups engaged in hydrogen bonds (circled red) to disrupt them (b,c) “correct” DNA polymerase synthesizes a purple-colored strand using cyan-colored DNA containing a modified A base (A*) or modified T base (T*). DNA polymerase selects T opposite A* and A opposite T*. The figures are based on Moran et al. [20].

The same study also successfully performed replication using a non-modified DNA template, but with modified A and T trinucleotides. The modified T was revealed to be incorporated opposite A, and the modified A was incorporated opposite T [20,21]. Additionally, the DNA polymerase was found to select “correct” complementary bases also *in vivo* [56]. In conclusion, it was clearly demonstrated that DNA polymerase does not require hydrogen bonds to select complementary bases.

Notably, even in the absence of any stabilizing protein, the double-stranded oligonucleotide—in which the modified A pairs with T—was found to present a higher

thermal stability than other oligonucleotides in which the modified A pairs with a mismatching base [20].

4.2. Geometry of Base Pairs

A decisive factor for the selection is the combined shape of the bases and base pairs. The modifications of the bases that disrupt hydrogen bond formation did not affect their shape or size. Furthermore, the size of the base pair is essential for the regular arrangement of the two strands in the DNA (Figure 4). A/T and G/C base pairs each consist of one purine and one pyrimidine, and thus their size (the distance between the points attached to each DNA backbone) is similar: 12.0 nm for A/T and 11.9 nm for G/C pair [56]. This size similarity ensures a regular, smooth phosphodiester backbone in B-form DNA, regardless of its nucleotide sequence (Figure 4).

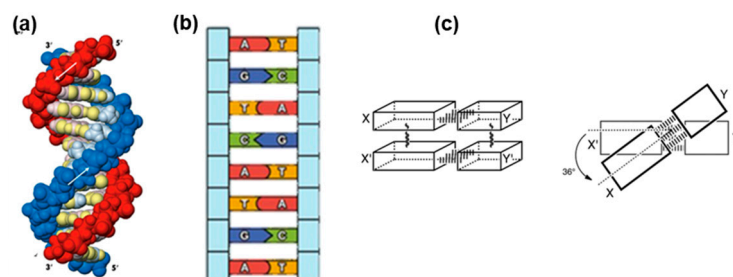


Figure 4. Smooth, regular helical backbone of DNA ensured by base pair geometry and base stacking: (a) Regular helical structure of DNA. The arrows present the strand orientation. (b) Constant strand-strand distance ensured by base pair geometry. (c) Regular stacking of base pairs. The figures are produced from the photos licensed under CC BY-NC.

The smooth and regular helical structure of the phosphodiester backbone is also related to ordered base-pair stacking (Figure 4). Base pairs are hydrophobic and firmly stack on top of each other, thereby stabilizing the double-helical dsDNA structure. Interestingly, base pairs regularly rotate around the DNA filament axis (Figure 4c). This regularity and the independence of nucleotide sequence in the base pair stacking are due to the similar orientation of bases relative to the phosphodiester backbone. The angle between a base and ribose is ≈ 55 degrees for any base [57]. Collectively, the geometry and organization of base pairs (base stacking) are essential for the formation of regular double-stranded DNA and potentially contribute to its stability.

4.3. The Geometry of Base Pairs Explains the Mutagenic Effects of O6-Methyl G and 8-Oxo G

The importance of base-pair geometry in replication is further supported by observations of nucleotide misincorporation by DNA polymerase opposite mutagenic bases. O6 methyl G and 8-oxo G are produced naturally by the modification of G and are both highly mutagenic. DNA polymerase selects A opposite O6 methyl G and T opposite 8-oxo-G instead of C. This results in mutations that can lead to cancer [57]. Structural analysis revealed that O6-methyl G can form a pair with C via two hydrogen bonds (Figure 5a). However, the geometry of the base pair differs from a normal G/C pair in size and orientation relative to the backbone. In contrast, the geometry of the O6 methyl G/A pair is similar to that of the regular base pairs. In vitro analysis indicated that DNA polymerase extends replication past the O6-methylated G/A pair, but disrupts replication past the O6-methylated G/C pair.

In the case of 8-oxo-G, it could form a pair with C, along with three hydrogen bonds (Figure 5b). However, a potential steric clash between the C8 carbonyl oxygen and the O4' of the sugar moiety may prevent this base pair [58]. 8-oxo-G can also create a base pair with T and maintain the regular base-pair geometry, although the base is in a syn configuration.

This observation explains the mutagenic effect of 8-oxo-G. Thus, the base-pair geometry is an essential factor in the selection of complementary bases by DNA polymerase and in maintaining the regular structure of DNA.

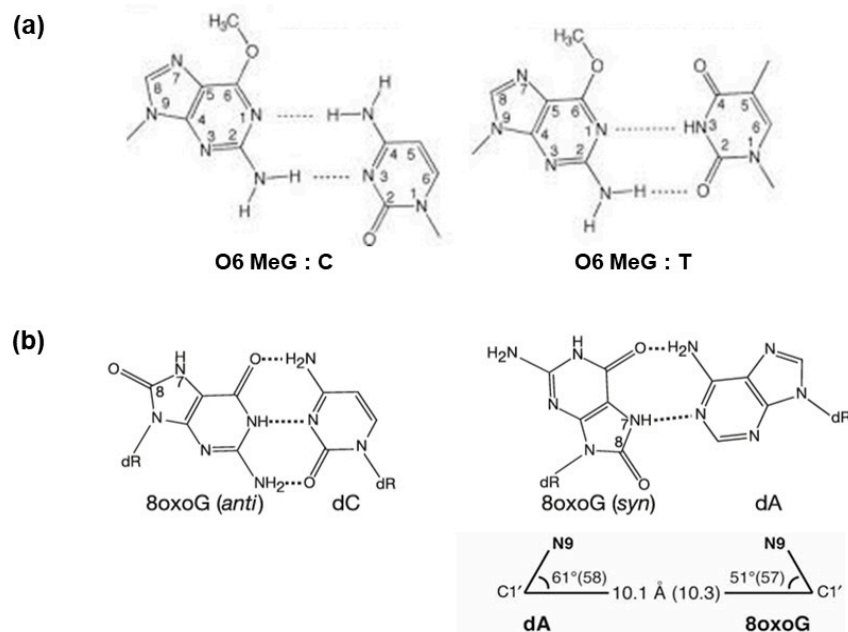


Figure 5. Possible base pair formation with mutagenic bases: (a) Base pair formation of O6-methyl-G with C and A. The figure is reproduced from Chavarria et al. [57]. (b) Base pair formation of 8-oxo-G with C and T. The angles between the bases and the deoxyribose for the 8-oxoG/A pair (61 and 51°) are close to those of standard A/T base pair (58 and 57°). The figures are reproduced from Hsu et al. [58].

4.4. Artificial Base Pairs Without Hydrogen Bond Formation

Hirao and colleagues developed an artificial base pair to enlarge the variation in aptamers [59]. Aptamers are single-stranded DNA or RNA molecules that bind to specific targets. However, the diversity of aptamers is limited, as there are only four nucleobases with similar shapes and chemical properties. More variation in bases, especially hydrophobic bases, can expand the potential of aptamers. One advantage of the aptamer strategy is the ability to select optimal aptamers via SELEX. In this approach, potential oligonucleotides are selected from a vast pool and amplified by replication for another round of selection. To use this selection method, newly designed nucleobases should be replicated by DNA polymerase with specifically pairing bases. A novel base pair design is required for this purpose. Hirao and colleagues designed hydrophobic base pairs in which two bases are complementary in shape and form a base pair that respects the regular base-pair geometry. They succeeded in obtaining such pairs that were correctly replicated in vitro and in vivo. Hydrogen bond formation does not appear to be required for base pairing during replication, underscoring the role of base geometry in the selection of “complementary” bases by DNA polymerase. NMR analysis showed that an oligonucleotide containing this artificial base pair forms a standard B-form structure [60].

4.5. Role of Base Stacking

The above observations that hydrogen bonding is not required for the selection of complementary bases do not imply that hydrogen bonding plays no role under “normal” situations. To test the role of hydrogen bond formation in selection and confirm the importance of base-pair geometry in the recognition of complementary bases, selection of complementary bases in the absence of geometry selection should be examined. Weakening

base stacking could affect the selection by geometry. In the absence of strong stacking interactions, base pairs can be accommodated more freely in DNA, and selection by geometry will be weakened.

Hydrophobic interactions mainly support base stacking. Since hydrophobic interactions occur between nonpolar (hydrophobic) molecules in a polar (hydrophilic) environment, base stacking can be weakened by either attaching hydrophilic groups to the molecules or using a less hydrophilic solvent. Nordén et al. successfully weakened base stacking using a semi-hydrophobic cosolvent, PEG [61]. PEG destabilizes double-stranded DNA and facilitates DNA strand exchange. They claimed that weakening base stacking does not eliminate the specificity of base pair formation [62]. Both hydrogen bonding and base stacking are involved in the selection of complementary bases.

5. Discussion

The studies indicate an interplay between hydrogen bond formation and base-pair geometry in the recognition of complementary bases or sequences. Further studies are required to clarify the exact mechanism underlying this phenomenon.

5.1. Kinetic Analyses

It is now crucial to find out whether the recognition of complementary bases begins with hydrogen-bond formation or with a selection search based on base-pair geometry. More precise kinetic studies will be required to elucidate this question. Both interactions are short-distance and directional. The action of enzymes is to bring two bases close enough and in an optimal mutual orientation. Recombinases that catalyze DNA strand exchange bind two DNA molecules in their filament and orient DNA bases in an appropriate direction [63]. Structural analyses of reaction intermediates may hopefully shed more light on the reaction mechanism [64].

5.2. Existence of Multiple Mechanisms

The recognition mechanism can vary from one reaction to another. The recognition accuracy and size of complementary sequences differ depending on the enzyme that catalyzes the reaction. Individual analyses of recognition mechanisms for each reaction may be required. However, some common aspects in the recognition mechanism should exist. The association of two complementary oligonucleotides occurs naturally without any protein. Furthermore, pairing of the modified A with T promotes more stable double-stranded DNA, and DNA polymerase selects T opposite the modified A. This coincidence suggests that DNA polymerase uses the characteristic nature of DNA stability to select complementary bases.

5.3. Limitations of Perturbing Approaches

The roles of hydrogen bond formation and base stacking in base pairing have primarily been analyzed by disrupting hydrogen bonds and weakening base stacking interactions [21,22,61,62], which poses certain limitations. The disruption of hydrogen-bond formation by replacing hydrophilic nucleobase groups with hydrophobic ones enhances hydrophobic interactions. This base modification strengthens base stacking and emphasizes selection by geometry. Weakening base stacking with PEG reduces the activity of water molecules, thereby enhancing hydrogen bonding by diminishing competition with them. Therefore, the selection by hydrogen bond formation may be exaggerated. It is necessary to quantitatively estimate the contribution of each force to the selection of complementary bases under natural conditions.

5.4. Examining the Strand Separation Step

Several biological reactions involve strand separation of double-stranded DNA prior to the formation of new base pairs. For instance, RNA polymerase promotes strand separation for transcription. Similarly, homologous recombination and CRISPR–Cas9 involve strand separation of double-stranded DNA to interact with the primary bound single-stranded DNA or guide RNA. Moreover, DNA strands can associate with wrong fragments or sites due to partial complementarity during hybridization or other reactions. Separation from such situations is crucial for the formation of the correct double-stranded structure.

In the absence of any proteins, strand separation occurs at high temperatures, thereby suggesting a requirement for significant energy. However, in biological systems, strand separation occurs under mild conditions with the help of enzymes. The mechanism of strand separation must be analyzed in more depth. Notably, semi-hydrophobic solvent PEG facilitates strand separation by weakening base stacking [61]. Hydrophobic residues are present in the DNA-binding sites of proteins. These residues may create hydrophobic environments and weaken base stacking, thereby facilitating strand separation.

5.5. Further Structure Investigations and Analytical Tools

Studying the biophysical underpinnings of gene regulation requires a deep understanding of how nucleic acids recognize and bind to one another. Recent research into CRISPR–Cas9 systems has revealed that base-pairing between the guide RNA (sgRNA) and target DNA is not a simple static match; it involves a “zipper-like” mechanism where the protein facilitates DNA melting by inducing base-flipping and rotational shifts to allow for sequential RNA–DNA heteroduplex formation. Similarly, in RNA silencing (RNAi), the Argonaut–RISC enforces a “seed region” recognition where the first few base pairs determine the kinetic stability of the entire silencing event. These intricate interactions are often studied using a combination of high-resolution structural biology and dynamic simulation [65–71].

The key analytical tools for studying base-pairing are x-ray crystallography, cryo-EM, Fluorescence Resonance Energy Transfer (FRET) and molecular dynamics (MD). X-ray is traditionally used to provide the first high-resolution “snapshots” of these complexes, such as the initial structures of Cas9 bound to its guide RNA, showing how the protein’s lobes physically separate DNA strands to allow for base-pairing. Cryo-EM has become the gold standard for capturing large, flexible complexes in near-native states. Cryo-EM is particularly useful for visualizing different “conformational checkpoints” during the CRISPR search process that X-ray crystallography might miss due to crystallization constraints. Single-Molecule FRET (smFRET) has revealed the timing and kinetics of base-pairing. By attaching fluorophores to specific sites, they can measure distance changes in real-time, revealing how the HNH nuclease domain of Cas9 shifts from an “inactive” to an “active” conformation only after successful base-pairing is completed. Artificial ‘recombination’ with PEG as a catalyst was studied using fluorescent oligonucleotides [62]. Because experimental methods have limits in temporal or spatial resolution, MD simulations are used to “fill in the gaps.” These simulations allow scientists to observe the sub-nanosecond “fraying” of terminal base pairs and the specific energy barriers that prevent off-target DNA from being cleaved [69–71].

6. Conclusions

The recognition of complementary nucleic bases and sequences is involved in the fundamental biological processes that almost all biotechnological and medical applications are based on. Therefore, an understanding of the recognition mechanisms in detail can be anticipated to be a prerequisite for the further development of these applications. Various

studies indicate that the selection of complementary bases and sequences is a complex process involving both Watson–Crick hydrogen bonding and geometric criteria.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: M.T. acknowledges Masayuki Oda (Kyoto Prefectural University) and Hiroshi Iwasaki (Institute of Science Tokyo) for discussion.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Watson, J.; Crick, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)] [[PubMed](#)]
2. Crick, F. On protein synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163.
3. Meselson, M.; Stahl, F. The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 671–682. [[CrossRef](#)]
4. Cairns, J. The bacterial chromosome and its manner of replication as seen by autoradiography. *J. Mol. Biol.* **1961**, *6*, 208–213. [[CrossRef](#)]
5. Lehman, I.R.; Bessman, M.J.; Simms, E.S.; Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J. Biol. Chem.* **1958**, *233*, 163–170. [[CrossRef](#)]
6. Weiss, S.B.; Gladstone, L.A. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *J. Am. Chem. Soc.* **1958**, *81*, 4118–4119. [[CrossRef](#)]
7. Chapeville, F.; Lipmann, F.; Ehrenstein, G.V.; Weisblum, B.; Ray, W.J., Jr.; Benzer, S. On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1086–1092. [[CrossRef](#)] [[PubMed](#)]
8. Matthaei, H.J.; Jones, O.W.; Martin, R.G.; Nirenberg, M.W. Characteristics and Composition of RNA Coding Units. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 666–677. [[CrossRef](#)] [[PubMed](#)]
9. Fire, A.; Xu, S.; Montgomery, M.K.; Kostas, S.A.; Driver, S.E.; Mello, C.C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **1998**, *391*, 806–811. [[CrossRef](#)]
10. Dana, H.; Chalbatani, G.M.; Mahmoodzadeh, H.; Karimloo, R.; Rezaiean, O.; Moradzadeh, A.; Mehmandoost, N.; Moazzen, F.; Mazraeh, A.; Marmari, V.; et al. Molecular Mechanisms and Biological Functions of siRNA. *Int. J. Biomed. Sci.* **2017**, *13*, 48–57. [[CrossRef](#)]
11. Wright, W.D.; Shah, S.S.; Heyer, W.D. Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* **2018**, *293*, 10524–10535. [[CrossRef](#)]
12. Chakraborty, S.; Schirmeisen, K.; Lambert, S.A. The multifaceted functions of homologous recombination in dealing with replication-associated DNA damages. *DNA Repair* **2023**, *129*, 103548. [[CrossRef](#)]
13. Székvölgyi, L.; Nicolas, A. From meiosis to postmeiotic events: Homologous recombination is obligatory but flexible. *FEBS J.* **2010**, *277*, 571–589. [[CrossRef](#)]
14. Mojica, F.J.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **2005**, *60*, 174–182. [[CrossRef](#)]
15. Kleppe, K.; Ohtsuka, E.; Kleppe, R.; Molineux, I.; Khorana, H.G. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* **1971**, *56*, 341–361. [[CrossRef](#)]
16. Kang, H.; Ga, Y.J.; Kim, S.H.; Cho, Y.H.; Kim, J.W.; Kim, C.; Yeh, J.-Y. Small interfering RNA (siRNA)-based therapeutic applications against viruses: Principles, potential, and challenges. *J. Biomed. Sci.* **2023**, *30*, 88. [[CrossRef](#)] [[PubMed](#)]
17. Chi, X.; Gatti, P.; Papoian, T. Safety of antisense oligonucleotide and siRNA-based therapeutics. *Drug Discov. Today* **2017**, *22*, 823–833. [[CrossRef](#)] [[PubMed](#)]
18. Cox, D.B.; Platt, R.J.; Zhang, F. Therapeutic genome editing: Prospects and challenges. *Nat. Med.* **2015**, *21*, 121–131. [[CrossRef](#)] [[PubMed](#)]
19. Fan, Q.; Yang, L.; Chao, J. Recent Advances in Dynamic DNA Nanodevice. *Chemistry* **2023**, *5*, 1781–1803. [[CrossRef](#)]
20. Moran, S.; Ren, R.X.; Kool, E.T. A thymidine triphosphate shape analog lacking Watson–Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506–10511. [[CrossRef](#)]

21. Morales, J.C.; Kool, E.T. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nat. Struct. Biol.* **1998**, *5*, 950–954. [[CrossRef](#)]
22. Kool, E.T.; Sintim, H.O. The Difluorotoluene Debate—A Decade Later. *Chem. Commun.* **2006**, *38*, 3665–3675. [[CrossRef](#)] [[PubMed](#)]
23. Tikhomirova, A.; Beletskaya, I.V.; Chalikian, T.V. Stability of DNA Duplexes Containing GG, CC, AA, and TT Mismatches. *Biochemistry* **2006**, *45*, 10563–10571. [[CrossRef](#)]
24. Ito, K.; Murayama, Y.; Takahashi, M.; Iwasaki, H. Two three-strand intermediates are processed during Rad51-driven DNA strand exchange. *Nat. Struct. Mol. Biol.* **2018**, *25*, 29–36. [[CrossRef](#)] [[PubMed](#)]
25. Bartel, D.P. Metazoan MicroRNAs. *Cell* **2018**, *173*, 20–51. [[CrossRef](#)]
26. O'Brien, J.; Hayder, H.; Zayed, Y.; Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **2018**, *9*, 402. [[CrossRef](#)]
27. Komor, A.C.; Kim, Y.B.; Packer, M.S.; Zuris, J.A.; Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016**, *533*, 420–424. [[CrossRef](#)]
28. Li, T.; Yang, Y.; Qi, H.; Cui, W.; Zhang, L.; Fu, X.; He, X.; Liu, M.; Li, P.-F.; Yu, T. CRISPR/Cas9 therapeutics: Progress and prospects. *Signal Transduct. Target. Ther.* **2023**, *8*, 36. [[CrossRef](#)] [[PubMed](#)]
29. Biela, A.; Hammermeister, A.; Kaczmarczyk, I.; Walczak, M.; Koziej, L.; Lin, T.; Glatt, S. The diverse structural modes of tRNA binding and recognition. *J. Biol. Chem.* **2023**, *299*, 104966. [[CrossRef](#)]
30. Leitão, A.L.; Enguita, F.J. The Unpaved Road of Non-Coding RNA Structure–Function Relationships: Current Knowledge, Available Methodologies, and Future Trends. *Non-Coding RNA* **2025**, *11*, 20. [[CrossRef](#)]
31. Sloane, D.L.; Goodman, M.F.; Echols, H. The fidelity of base selection by the polymerase subunit of DNA polymerase III holoenzyme. *Nucleic Acids Res.* **1988**, *16*, 6465–6475. [[CrossRef](#)]
32. Anand, R.; Beach, A.; Li, K.; Haber, J. Rad51-mediated double-strand break repair and mismatch correction of divergent substrates. *Nature* **2017**, *544*, 377–380. [[CrossRef](#)] [[PubMed](#)]
33. Takahashi, M. RecA kinetically selects homologous DNA by testing a five- or six-nucleotide matching sequence and deforming the second DNA. *Q. Rev. Biophys.* **2018**, *51*, e11. [[CrossRef](#)]
34. Kunkel, T.A.; Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **2000**, *69*, 497–529. [[CrossRef](#)] [[PubMed](#)]
35. Imashimizu, M.; Oshima, T.; Lubkowska, L.; Kashlev, M. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res.* **2013**, *41*, 9090–9104. [[CrossRef](#)]
36. Fijalkowska, I.J.; Schaaper, R.M.; Jonczyk, P. DNA replication fidelity in *Escherichia coli*: A multi-DNA polymerase affair. *FEMS Microbiol. Rev.* **2012**, *36*, 1105–1121. [[CrossRef](#)]
37. Han, H. RNA Interference to Knock Down Gene Expression. *Methods Mol. Biol.* **2018**, *1706*, 293–302.
38. Lauffer, M.C.; van Roon-Mom, W.; Aartsma-Rus, A. Possibilities and limitations of antisense oligonucleotide therapies for the treatment of monogenic disorders. *Commun. Med.* **2024**, *4*, 6. [[CrossRef](#)]
39. Ratilainen, T.; Lincoln, P.; Nordén, B. A simple model for gene targeting. *Biophys. J.* **2001**, *81*, 2876–2885. [[CrossRef](#)] [[PubMed](#)]
40. Neumeier, J.; Meister, G. siRNA Specificity: RNAi Mechanisms and Strategies to Reduce Off-Target Effects. *Front. Plant Sci.* **2021**, *11*, 526455. [[CrossRef](#)]
41. Hoshijima, K.; Juryneć, M.J.; Grunwald, D.J. Precise genome editing by homologous recombination. *Methods Cell Biol.* **2016**, *135*, 121–147. [[PubMed](#)]
42. Hoheisel, J.D. Application of hybridization techniques to genome mapping and sequencing. *Trends Genet.* **1994**, *10*, 79–83. [[CrossRef](#)]
43. Lao, A.I.K.; Su, X.; Aung, K.M.M. SPR study of DNA hybridization with DNA and PNA probes under stringent conditions. *Biosens. Bioelectron.* **2009**, *24*, 1717–1722. [[CrossRef](#)]
44. Furst, A.L.; Klass, S.H.; Francis, M.B. DNA Hybridization to Control Cellular Interactions. *Trends Biochem. Sci.* **2019**, *44*, 342–350. [[CrossRef](#)]
45. Bej, A.K.; Mahbubani, M.H.; Atlas, R.M. Amplification of Nucleic Acids by Polymerase Chain Reaction (PCR) and Other Methods and their Applications. *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26*, 301–334. [[CrossRef](#)] [[PubMed](#)]
46. Guan, C.; Zhu, X.; Feng, C. DNA Nanodevice-Based Drug Delivery Systems. *Biomolecules* **2021**, *11*, 1855. [[CrossRef](#)]
47. Kong, L.; Kim, S.; Wang, C.; Lee, S.Y.; Oh, S.; Lee, S.; Jo, S.; Kim, T. Understanding nucleic acid sensing and its therapeutic applications. *Exp. Mol. Med.* **2023**, *55*, 2320–2331. [[CrossRef](#)]
48. Sherman, W.B.; Seeman, N.C. A precisely controlled DNA biped walking device. *Nano Lett.* **2004**, *4*, 1203–1207. [[CrossRef](#)]
49. Parker, J. Computing with DNA. *EMBO Rep.* **2003**, *4*, 7–10. [[CrossRef](#)] [[PubMed](#)]
50. Shraim, A.S.; Abdel Majeed, B.A.; Al-Binni, M.A.; Hunaiti, A. Therapeutic Potential of Aptamer-Protein Interactions. *ACS Pharmacol. Transl. Sci.* **2022**, *5*, 1211–1227. [[CrossRef](#)]
51. Nimjee, S.M.; White, R.R.; Becker, R.C.; Sullenger, B.A. Aptamers as Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **2017**, *57*, 61–79. [[CrossRef](#)]

52. Yang, W.; Ran, C.; Lian, X.; Wang, Z.; Du, Z.; Bing, T.; Zhang, Y.; Tan, W. Aptamer-based targeted drug delivery and disease therapy in preclinical and clinical applications. *Adv. Drug Deliv. Rev.* **2025**, *226*, 115680. [[CrossRef](#)]
53. Ali Zaidi, S.S.; Fatima, F.; Ali Zaidi, S.A.; Zhou, D.; Deng, W.; Liu, S. Engineering siRNA therapeutics: Challenges and strategies. *J. Nanobiotechnol.* **2023**, *21*, 381. [[CrossRef](#)] [[PubMed](#)]
54. Guo, C.; Ma, X.; Gao, F.; Guo, Y. Off-target effects in CRISPR/Cas9 gene editing. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1143157. [[CrossRef](#)]
55. Cai, Y.; Yu, X.; Hu, S.; Yu, J. A Brief Review on the Mechanisms of miRNA Regulation. *Genom. Proteom. Bioinform.* **2009**, *7*, 147–154. [[CrossRef](#)]
56. Delaney, J.C.; Henderson, P.T.; Helquist, S.A.; Morales, J.C.; Essigmann, J.M.; Kool, E.T. High-fidelity in vivo replication of DNA base shape mimics without Watson–Crick hydrogen bonds. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4469–4473. [[CrossRef](#)]
57. Chavarria, D.; Ramos-Serrano, A.; Hirao, I.; Berdis, A.J. Exploring the roles of nucleobase desolvation and shape complementarity during the misreplication of O6-methylguanine. *J. Mol. Biol.* **2011**, *412*, 325–339. [[CrossRef](#)]
58. Hsu, G.W.; Ober, M.; Carell, T.; Beese, L.S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **2004**, *431*, 217–221. [[CrossRef](#)]
59. Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I. Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **2012**, *40*, 2793–2806. [[CrossRef](#)] [[PubMed](#)]
60. Guckian, K.M.; Krugh, T.R.; Kool, E.T. Solution structure of a nonpolar, non-hydrogen-bonded base pair surrogate in DNA. *J. Am. Chem. Soc.* **2000**, *122*, 6841–6847. [[CrossRef](#)] [[PubMed](#)]
61. Feng, B.; Sosa, R.P.; Mårtensson, A.K.F.; Jiang, K.; Tong, A.; Dorfman, K.D.; Takahashi, M.; Lincoln, P.; Bustamante, C.J.; Westerlund, F.; et al. Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17169–17174. [[CrossRef](#)]
62. Nordén, B.; Brown, T.; Feng, B. Mismatch detection in homologous strand exchange amplified by hydrophobic effects. *Biopolymers* **2021**, *112*, e23426. [[CrossRef](#)] [[PubMed](#)]
63. Takahashi, M.; Ito, K.; Iwasaki, H.; Nordén, B. Linear dichroism reveals the perpendicular orientation of DNA bases in the RecA and Rad51 recombinase filaments: A possible mechanism for the strand exchange reaction. *Chirality* **2024**, *36*, e23664. [[CrossRef](#)]
64. Freudenthal, B.D.; Beard, W.A.; Wilson, S.H. Structures of dNTP intermediate states during DNA polymerase active site assembly. *Structure* **2012**, *20*, 1829–1837. [[CrossRef](#)]
65. Bhattacharya, S.; Satpati, P. Insights into the mechanism of CRISPR/Cas9-based genome editing from molecular dynamics simulations. *ACS Omega* **2022**, *8*, 1817–1837. [[CrossRef](#)]
66. Jiang, F.; Doudna, J.A. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **2015**, *30*, 100–111. [[CrossRef](#)]
67. Lee, J.; Jeong, C. Single-molecule perspectives of CRISPR/Cas systems: Target search, recognition, and cleavage. *BMB Rep.* **2025**, *58*, 8–16. [[CrossRef](#)] [[PubMed](#)]
68. Müller, M.; Fazi, F.; Ciaudo, C. Argonaute proteins: From structure to function in development and pathological cell fate determination. *Front. Cell Dev. Biol.* **2020**, *7*, 360. [[CrossRef](#)] [[PubMed](#)]
69. Palermo, G. Establishing the allosteric mechanism in CRISPR-Cas9. *Biophys. J.* **2022**, *121*, 145a–146a. [[CrossRef](#)]
70. Singh, D.; Wang, Y.; Mallon, J.; Yang, O.; Fei, J.; Poddar, A.; Ceylan, D.; Bailey, S.; Ha, T. Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **2018**, *25*, 347–354. [[CrossRef](#)]
71. Wang, J.; Arantes, P.R.; Ahsan, M.; Sinha, S.; Kyro, G.W.; Maschietto, F.; Allen, B.; Skeens, E.; Lisi, G.P.; Batista, V.S.; et al. Twisting and swiveling domain motions in Cas9 to recognize target DNA duplexes, make double-strand breaks, and release cleaved duplexes. *Front. Mol. Biosci.* **2023**, *9*, 1072733. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.