

Emerging Roles of DNA Glycosylases and the Base Excision Repair Pathway

Elwood A. Mullins,¹ Alyssa A. Rodriguez,¹ Noah P. Bradley,¹ and Brandt F. Eichman^{1,*}

¹Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37232, USA

*Correspondence: brandt.eichman@vanderbilt.edu (B.F. Eichman)

Keywords: DNA Damage, DNA Repair, Base Excision Repair, DNA Glycosylase, Interstrand Crosslink, Secondary Metabolite

ABSTRACT

The base excision repair (BER) pathway historically has been associated with maintaining genome integrity by eliminating nucleobases with small chemical modifications. In the past several years, however, BER was found to play additional roles in genome maintenance and metabolism, including sequence-specific restriction modification and repair of bulky adducts and interstrand crosslinks. Central to this expanded biological utility are specialized DNA glycosylases, enzymes that selectively excise damaged, modified, or mismatched nucleobases. In this review, we discuss the newly identified roles of the BER pathway and examine the structural and mechanistic features of the DNA glycosylases that enable these functions.

GLOSSARY

C–H/ π interaction: a weak electrostatic interaction between the dipole of a C–H bond and the quadrupole of a π system.

Fanconi anemia (FA) pathway: a complex network for coordination of DNA repair pathways, often associated with repair of interstrand crosslinks and found only in vertebrates.

Genotoxin: a chemical agent that damages DNA.

Homologous recombination (HR): a type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA, commonly used by cells to repair double-strand breaks.

Intercalating residue: a protein residue that is inserted between stacked bases in the DNA duplex, often associated with base flipping.

Interstrand crosslink (ICL): a highly toxic lesion that covalently links opposing strands of DNA, preventing separation into the single-stranded templates that are necessary for transcription and replication.

Monoadduct: a lesion in which the modifying agent is covalently attached to a single nucleotide, often used in association with compounds capable of simultaneously and covalently attaching to two nucleotides, thereby generating an interstrand crosslink.

Natural product: a chemical compound produced by a living organism.

Nucleotide excision repair (NER): a versatile mechanism for removal of diverse bulky and helix-distorting/destabilizing lesions in which a short segment of the damaged strand is excised and replaced with newly synthesized DNA.

Replication fork: a Y-shaped DNA structure formed by separation of double-stranded DNA into two single-stranded templates during replication.

Replisome: a large multiprotein complex assembled at replication forks and comprised of the enzymes necessary to replicate the genome, including a DNA helicase to unwind the parental duplex, multiple DNA polymerases to synthesize the daughter strands, and numerous accessory proteins with various functions.

Secondary metabolite: an organic compound produced by a microorganism but not required for normal growth or reproduction, frequently associated with interspecies defense mechanisms.

Sister chromatid: one of two identical copies of a chromosome formed during replication.

Translesion synthesis (TLS): error-prone DNA synthesis by a low-fidelity polymerase to bypass DNA damage in the template strand, often introducing a mutation in the daughter strand.

Ubiquitylation: conjugation of one or more molecules of the small regulatory protein ubiquitin with a target protein to modulate its function or to mark it for processing.

MAIN TEXT

Recognition and Repair of DNA Damage

DNA is a reactive molecule that is continually challenged by both endogenous and exogenous insults [1, 2]. Cellular metabolites and their by-products, environmental toxins, and radiation alter the chemical structure of DNA, producing a wide spectrum of DNA damage. Single and double-strand breaks (DSBs) are generated by hydrolysis of the phosphodeoxyribose backbone, nucleotide mismatches are introduced by replication errors, and nucleobases are modified by alkylation, oxidation, and deamination (Figure 1A). Chemical adducts range in size from a single non-hydrogen atom (e.g., 8-oxoguanine, 3-methyladenine) to bulky lesions with helix-distorting properties, such as those produced by polyaromatic hydrocarbons and crosslinking agents. These chemically diverse lesions interfere with normal cellular processes through inhibition of replication, transcription, and chromosome maturation, leading to chromosome rearrangements and instability, cell death, ageing, and diseases including cancer [3].

Several DNA repair pathways exist to eliminate specific types of damage from the genome [3]. Pathway choice is dictated in part by the enzymes that recognize or initiate repair of a particular type of damage. By and large, the base excision repair (BER) pathway (Figure 1B) eliminates nucleobases with small modifications, abasic sites, and single-strand breaks, while **nucleotide excision repair (NER)** (see Glossary) removes bulky, helix-destabilizing lesions. BER is initiated by lesion-specific DNA glycosylases that excise the modified nucleobase from the DNA by catalyzing hydrolysis of the N-glycosidic bond (Figure 1B and Figure 2A). The resulting apurinic/apyrimidinic (AP) site is incised by an AP endonuclease (or a bifunctional DNA glycosylase), which generates a 3'-hydroxyl group needed for polymerase-dependent synthesis of new DNA [for a detailed overview of BER, see 4, 5, 6]. Almost every DNA glycosylase, regardless of its specificity or structural architecture, uses a similar overall strategy in which the

1
2
3
4 aberrant nucleotide is flipped out of the duplex and trapped in a nucleobase binding pocket on
5
6 the protein surface, while the resulting void left in the DNA is filled by one or more **intercalating**
7
8 **residues** that stabilize the extrahelical conformation (Figure 2B,C) [7-9]. Remodeling of the DNA
9
10 substrate through bending of the helical axis and widening of the minor groove promotes base
11
12 flipping by decreasing the energetic barrier to base pair opening, while also inducing strain that
13
14 allows the glycosylase to detect altered base stacking, base pairing, or solvation resulting from
15
16 chemical modification of the nucleobase [7, 10-13]. In addition to allowing for more stringent
17
18 substrate recognition, base flipping facilitates excision by enabling catalytic activation of the lesion
19
20 through chemical complementarity between the active site and the target nucleobase and
21
22 improving the reaction geometry between the glycosidic bond and an attacking water molecule
23
24 (Box 1) [9]. Nucleobase binding pockets are generally too small to accommodate more than a
25
26 small modification, and thus the discovery of DNA glycosylases capable of removing bulky or
27
28 crosslinked lesions associated with other types of repair, as well as non-toxic or unmodified bases,
29
30 has been unexpected. This review describes these new glycosylases with a focus on the
31
32 structural mechanisms that enable their activities.
33
34
35
36
37
38
39
40

41 **Self-Resistance to Genotoxic Secondary Metabolites**

42
43 Bacterial **secondary metabolites** are often used as defense mechanisms in microbial
44
45 warfare. To withstand the toxicity of their own natural products, antibiotic-producing bacteria
46
47 require self-resistance mechanisms, such as sequestration, efflux, and degradation of the toxin,
48
49 and protection and repair of the target (e.g., DNA) [14]. Two target-repair mechanisms for the
50
51 genotoxins yatakemycin (YTM) and azinomycin B (AZB) were recently found to involve the
52
53 evolutionarily unrelated DNA glycosylases YtkR2 and AlkZ [15, 16], which are encoded by genes
54
55 embedded within the *ytk* and *azi* biosynthesis clusters [15, 17]. Homologs of these enzymes are
56
57 present in diverse bacterial species, but only some of which are known to produce antibiotics [16,
58
59
60
61
62
63
64
65

1
2
3
4 18-20]. Homologs of YtkR2 are also present in archaea and lower eukaryotes. It is unclear if these
5
6 organisms encounter compounds similar to YTM and AZB in their environments, or if these
7
8 homologs have evolved to perform different functions.
9

10 11 12 13 14 *Excision of Bulky Adducts by YtkR2*

15
16
17 Streptomyces produce a staggering number and variety of secondary metabolites,
18
19 including many highly genotoxic antibiotics [21]. Among the most potent are those in the
20
21 spirocyclopropylcyclohexadienone family, which consists of CC-1065, duocarmycin A and SA,
22
23 and YTM [22, 23]. These bulky molecules undergo binding-induced conformational changes in
24
25 the minor groove of AT-rich regions of DNA that increase their reactivity, resulting in sequence-
26
27 specific alkylation of adenine at the N3 position (Figure 3A) [24]. While only a single covalent bond
28
29 is formed, a large number of non-covalent contacts, principally **C–H/ π interactions**, are created
30
31 between the antibiotic and deoxyribose groups in both DNA strands. Individually each of these
32
33 interactions is weak, but together they create a strong energetic barrier to duplex unwinding that
34
35 inhibits both transcription and replication, while also making these lesions poor substrates for the
36
37 NER pathway [25, 26].
38
39
40
41

42 The YTM-producer *Streptomyces* sp. TP-A0356 has evolved multiple resistance
43
44 mechanisms to minimize YTM toxicity [15, 27]. Within the gene cluster for YTM synthesis are
45
46 seven resistance genes that play roles in efflux and degradation of YTM and repair of YTM–DNA
47
48 adducts. YtkR2 is a DNA glycosylase that hydrolyzes the glycosidic bond of 3-yatakemycinyl-2'-
49
50 deoxyadenosine (YTMA) lesions (Figure 3A), enabling the BER pathway to provide self-
51
52 resistance to YTM toxicity [15]. Similarly, the CC-1065 biosynthesis cluster of *Streptomyces*
53
54 *zelensis* contains a homolog of YtkR2 (C10R5) that likely provides resistance to CC-1065 [19].
55
56
57
58
59
60
61
62
63
64
65

YtkR2 was identified as a YTM resistance protein by its similarity to AlkD, a previously discovered DNA glycosylase with the ability to excise bulky pyridyloxobutyl (POB) adducts of guanine and cytosine (Figure 3A) [18, 28]. Despite being present in *Bacillus cereus*, which is not a producer of YTM, AlkD efficiently removes YTMA lesions in vitro [26, 29]. As such, its structural and mechanistic characterization is a basis for understanding how YtkR2 provides self-resistance to YTM. AlkD has a structural architecture distinct from those of other DNA glycosylases. The HEAT-like repeat (HLR) fold lacks both a nucleobase binding pocket and apparent intercalating residues [9, 18, 30, 31]. Accordingly, studies using either methylated or yatakemycinylated nucleobases showed that AlkD recognizes and excises lesions without first sequestering the modified base from the DNA duplex [26, 29]. AlkD instead interacts with the phosphodeoxyribose backbone of the lesion, while leaving the nucleobase, which is not contacted by the enzyme, stacked in the duplex and paired with the opposing nucleotide (Figure 3B–E). In this conformation, the deoxyribose moiety of the lesion is in contact with three residues: Trp109, Asp113, and Trp187. Surprisingly, all three, including the two tryptophan residues, are catalytic [26, 32, 33]. As with catalytic residues in other DNA glycosylases (Box 1), Asp113 stabilizes the sugar as positive charge develops during cleavage of the glycosidic bond, while also pre-organizing the water nucleophile. Uniquely, Trp109 and Trp187 form C–H/ π interactions with the deoxyribose group. Electrostatic in nature, these interactions become stronger as the C–H bond becomes increasingly polarized [29], such as when positive charge develops on the sugar. As such, the C–H/ π interactions formed with Trp109 and Trp187 preferentially stabilize the transition state, reducing the activation barrier and accelerating excision.

AlkD was crystallized in two product complexes, containing either 3-methyladenine (3mAde) or 3-yatakemycinyladenine (YTMAde) nucleobase (Figure 3B–E) [26, 29]. In both complexes, the only contacts between the protein and the modified nucleobases are with the bulky YTM moiety, which is located in an extended cavity between the enzyme surface and the

1
2
3
4 minor groove (Figure 3C,E). In the complex with the much smaller 3mAde nucleobase, this cavity
5
6 is filled by solvent, as is the major groove in both complexes (Figure 3B–E). This lack of steric
7
8 restraint is likely what allows AlkD to remove nucleobases with bulky YTM or POB modifications
9
10 located in either groove (Figure 3A). However, in contrast to DNA glycosylases that flip the lesion
11
12 into a more restrictive nucleobase binding pocket, which contains catalytic residues that activate
13
14 the base for excision (Box 1), AlkD is limited to removing positively charged lesions with
15
16 nucleobases that are inherently good leaving groups. Moreover, this lack of steric restriction
17
18 seems to make DNA glycosylases in this family well-suited to remove diverse lesions, limited
19
20 primarily by the mechanistic requirement for a positively charged substrate.
21
22
23
24
25
26

27 *Unhooking of Interstrand Crosslinks by AlkZ*

30 Along with the spirocyclopropylcyclohexadienone compounds, bifunctional alkylating
31
32 agents are among the most potent antibiotics produced by Streptomyces [34]. AZB is a
33
34 genotoxic nonribosomal peptide/polyketide secondary metabolite with a densely functionalized
35
36 backbone that scaffolds electrophilic aziridine and epoxide moieties [35, 36]. This arrangement
37
38 creates a molecule that is ideally suited to bind in the major groove of GNC and GNT sequences,
39
40 and to undergo nucleophilic addition at the N7 positions of guanine and adenine (Figure 4A) [37-
41
42 39]. By covalently tethering opposing strands, bifunctional alkylating agents generate **interstrand**
43
44 **crosslinks (ICLs)**, which block transcription and replication, and generally necessitate repair by
45
46 a combination of repair pathways [40-42]. In bacteria, the primary mechanism of ICL repair
47
48 requires the combined actions of NER and either **translesion synthesis (TLS)** or **homologous**
49
50 **recombination (HR)** [1, 43, 44]. Unlike the primary mechanisms of ICL repair in vertebrates,
51
52 which involve DNA structures generated only during replication [45-48], ICL repair in bacteria
53
54 occurs in the context of duplex DNA [1, 43, 44]. The lesion is unhooked from one strand by the
55
56 NER machinery, the gap filled by either HR-dependent synthesis or TLS, and the monoadduct
57
58
59
60
61
62
63
64
65

1
2
3
4 repaired by a second round of NER. Recently, however, a DNA glycosylase was found to unhook
5
6 ICLs produced by AZB, implicating BER in bacterial ICL repair [16].
7
8

9 Much like *Streptomyces* sp. TP-A0356, the AZB-producer *Streptomyces sahachiroi* has
10 evolved multiple mechanisms for self-resistance, including sequestration and efflux of AZB, and
11 repair of AZB ICLs [16, 49]. Originally annotated as a putative winged helix-turn-helix (wHTH)
12 transcription factor, AlkZ is a novel DNA glycosylase that excises ICLs formed by AZB, cleaving
13 the glycosidic bonds of both modified nucleotides [16]. The resultant AP sites, despite their
14 proximity on opposing strands, were shown to be substrates for the bacterial AP endonuclease
15 Endo IV, suggesting AZB ICLs can be repaired by the BER pathway alone. Consistent with this
16 role, genetic knockouts of *alkZ* were highly susceptible to AZB toxicity, and resistance could be
17 restored with a plasmid expressing AlkZ [16]. In addition to its repair role, AlkZ has been proposed
18 to provide self-resistance to AZB by blocking target sites, thereby preventing crosslinks from
19 forming [16]. However, the experiments that served as the basis for this proposal were not
20 designed to distinguish between protection of target sites and unhooking of ICLs. Both activities
21 would allow for unwinding of double-stranded DNA after incubation with AZB, which was the
22 experimental observation. Given the proven excision activity of AlkZ, unhooking seems to be the
23 more probable explanation.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 Even more so than excision of bulky adducts, unhooking of ICLs, which are tethered to
44 both DNA strands, is incompatible with the base-flipping mechanisms used by most DNA
45 glycosylases. Correspondingly, the structural architecture of AlkZ most closely resembles that of
46 AlkD, although the two folds are distinct [26, 50]. Like AlkD, AlkZ is C-shaped with a positively
47 charged concave surface ideal for binding DNA, and lacks a nucleobase binding pocket and
48 apparent DNA intercalating residues. AlkZ is the defining member of the HTH_42 superfamily of
49 proteins, which are predicted to contain tandem wHTH motifs [16]. Interestingly, the structure of
50 AlkZ shows that the enzyme uses three such motifs to create the DNA binding scaffold, although
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 none seem to engage DNA in the same manner as WTH motifs found in transcription factors
5
6 [50]. While no structure of an AlkZ–DNA complex is available, mutational analysis indicated three
7
8 essential catalytic elements within the putative substrate binding cleft, including a β -hairpin
9
10 (β 11/ β 12) and two glutamine residues (Gln37 and Gln39). Docking of a computationally derived
11
12 model of an AZB ICL revealed two nearly equivalent DNA orientations, related by the
13
14 pseudosymmetric nature of the crosslink relative to the dyad axis, and suggested possible roles
15
16 for the three catalytic elements (Figure 4B) [38, 50]. In each of the binding orientations, Gln37
17
18 interacts with a phosphate group in the DNA backbone, while the β 11/ β 12 hairpin inserts into the
19
20 minor groove and interacts with the backbones of both strands and one of the modified
21
22 nucleobases. Gln39 is located just outside the minor groove and is ideally positioned to catalyze
23
24 hydrolysis of the glycosidic bond (Box 1).
25
26
27
28

29 As modeled, no contacts are present between AlkZ and AZB in the hypothetical protein–
30
31 DNA complexes, which could otherwise favor one binding orientation over the other. This is
32
33 consistent with biochemical data indicating excision of both modified nucleotides [16]. Strikingly,
34
35 the two binding orientations are not mutually exclusive, allowing for the possibility of dimer
36
37 formation at the crosslink (Figure 4C,D) [50]. Moreover, docking of two AlkZ molecules on a single
38
39 AZB ICL revealed remarkable self-complementarity at the protein interface, including a potential
40
41 network of salt bridges (Glu152 and Arg153), hydrogen bonds (Ser304 and Arg308), and
42
43 hydrophobic contacts (Ala309 and Pro340). However, there is currently no available experimental
44
45 data to support or refute dimerization. Further work will be necessary to determine whether dimer
46
47 formation plays a role in recognition or excision of AZB ICLs, as well as the mechanism by which
48
49 downstream enzymes in the BER pathway process the AP products generated by AlkZ (Figure
50
51 4E).
52
53
54
55
56
57
58

59 **Restart of Stalled Replication Forks**

60
61
62
63
64
65

1
2
3
4 By preventing the **replisome** from unwinding duplex DNA, ICLs stall progression of the
5
6 **replication fork**, which in turn results in genomic instability [41, 42]. As in prokaryotes, the
7
8 conventional mechanism of ICL repair in eukaryotes involves the combined activities of NER, TLS,
9
10 and HR. In higher eukaryotes, these activities are coordinated by the **Fanconi anemia (FA)**
11
12 **pathway** and are coupled with DNA replication [51, 52]. Following convergence of two replication
13
14 forks at a crosslink, eukaryotic ICL repair is initiated by unloading of replisome components from
15
16 the DNA template [46, 47, 53]. Fanconi proteins then recruit endonucleases associated with the
17
18 NER pathway to create incisions in one strand flanking the lesion [45, 48, 54]. These incisions
19
20 generate a DSB in one **sister chromatid** and leave a **monoadduct** on the other. The former is
21
22 repaired by HR, and the latter bypassed by TLS. Recently, however, the vertebrate DNA
23
24 glycosylases NEIL1 and NEIL3 were found to unhook ICLs, as well as to excise bulky adducts,
25
26 providing an alternative mechanism to restart stalled replication forks [54-57].
27
28
29
30

31
32 Identified by sequence similarity to bacterial DNA glycosylases in the Fpg/Nei superfamily,
33
34 the Nei-like enzymes NEIL1, NEIL2, and NEIL3 are the most recent DNA glycosylases to be
35
36 discovered in vertebrates [58-61]. Like the related bacterial enzymes, NEIL1–3 were originally
37
38 linked to repair of oxidized nucleobases [62]. However, several subsequent findings have called
39
40 into question whether this is the primary function of the vertebrate enzymes. First, the small
41
42 oxidative lesions excised by NEIL1–3 are also removed by other DNA glycosylases present in
43
44 higher eukaryotes, namely oxidized pyrimidines by NTHL1 and oxidized purines by OGG1 [60].
45
46 Second, NEIL1 and NEIL3 form specific interactions with key proteins required for replication, and
47
48 NEIL2 interacts with a number of proteins necessary for transcription, including RNA polymerase
49
50 II [63-65]. Third, expression of NEIL1 and NEIL3 is cell-cycle dependent, and induced during S
51
52 phase [66-68]. Fourth, NEIL2 and NEIL3 preferentially excise lesions from bubble, fork, and
53
54 single-stranded DNA structures [64, 69-72]. Together, these findings suggest roles for NEIL1 and
55
56 NEIL3 in replication-associated repair, and for NEIL2 in transcription-coupled repair [73].
57
58
59
60
61
62
63
64
65

The different functions of NEIL1 and NEIL3 during replication were first suggested when NEIL1 was shown to remove psoralen crosslinks from triplex DNA (Figure 5A,B) [74, 75]. During normal ICL repair, crosslinked triplex structures are generated following unhooking of the ICL by NER and bypass of the resulting monoadduct by TLS (Supplemental Figure S1). While NER is able to repair these structures [76, 77], the activity of NEIL1 suggests BER may be an alternative [57, 75, 78]. Importantly, NEIL1 is not able to perform the initial unhooking of the ICL in duplex DNA, or in the X-shaped structures formed when replication forks converge at a crosslink (Figure 5C) [55, 74, 75]. Conversely, NEIL3 was recently shown to unhook both psoralen and AP crosslinks at convergent replication forks (Figure 5A,C), providing an incision-independent alternative for repair of ICLs [54]. Both incision-dependent and -independent repair take place at replication forks that have converged at a crosslink (Figure 5C, Supplementary Figure S1) [47, 54]. Pathway choice is determined by **ubiquitinylation** of replisome components (Supplemental Figure S1) [46]. Short polyubiquitin chains recruit NEIL3 to the ICL, whereas longer chains, which form in the absence of unhooking by NEIL3, lead to unloading of the replisome and unhooking of the ICL by structure-specific endonucleases. NEIL3 recruitment and unhooking prior to initiation of the incision-dependent pathway suggests the glycosylase-mediated pathway is the preferred mechanism of ICL repair during S-phase in vertebrates [46, 54]. One key distinction between the two pathways is that the incision-independent mechanism avoids the DSBs that are integral to the incision-dependent mechanism (Supplemental Figure S1). However, as a bifunctional DNA glycosylase, NEIL3 possesses AP lyase activity (Figure 2A), which could potentially generate a DSB at convergent replication forks, negating the apparent benefit of an otherwise incision-independent repair pathway. The AP lyase activity of NEIL3, however, is weak relative to those of NEIL1 and NEIL2 [70], and appears to be further reduced in this context, as DSBs were not observed when ICL repair was initiated by NEIL3 in nuclear extracts [54].

1
2
3
4 The mechanisms by which NEIL1 and NEIL3 excise bulky adducts and unhook ICLs are
5
6 unclear. Structures of NEIL1 are available only with DNA containing small lesions [79-81], and
7
8 the single available structure of NEIL3 lacks DNA (Figure 5D,E) [73]. Nonetheless, comparison
9
10 of the two structures, as well as structures of the bacterial enzymes, provides a basis for
11
12 speculation. Both NEIL1 and NEIL3 share the same structural architecture and arrangement of
13
14 N-terminal catalytic residues (Pro2/Val2 and Glu3) as Fpg and Nei. Like the bacterial enzymes
15
16 [9], NEIL1 also utilizes three intercalating residues to stabilize the extrahelical conformation of the
17
18 DNA substrate (Figure 2B and Figure 5D). The capping loop in NEIL1, which contacts the flipped
19
20 lesion in the bacterial enzymes, is similar to that of Nei [9, 79-81]. In all structures of Nei and most
21
22 structures of NEIL1, this loop is largely disordered (Figure 5D), suggesting an inherent flexibility
23
24 that may allow for recognition and excision of diverse lesions. In contrast to the well-ordered loop
25
26 in Fpg [82], which contacts extrahelical nucleobases through amide groups in the protein
27
28 backbone (Figure 2B,C), the partially disordered loop in NEIL1 interacts with the lesion through
29
30 the sidechain of a conserved arginine residue [81]. The flexibility of this loop may enable
31
32 expansion of the nucleobase binding pocket to accommodate bulky adducts, such as nitrogen
33
34 mustard and aflatoxin B₁ derivatives of N⁶-(2'-deoxyribosyl)-2,6-diamino-4-oxo-5-
35
36 formamidopyrimidine (FapyG) and psoralen monoadducts (Figure 5A) [56, 74]. This flexibility may
37
38 also allow NEIL1 to bind triplex substrates, by pulling the short, third strand away from the duplex
39
40 and binding the duplex and the crosslink in the same manner as a bulky adduct in double-stranded
41
42 DNA.
43
44
45
46
47

48
49 In contrast, NEIL3 shares fewer structural features with the bacterial enzymes [9, 73].
50
51 Unlike Fpg and Nei, NEIL3 appears to lack intercalating residues, with the possible exception of
52
53 Met99 (Figure 5E) [73]. In NEIL1 and the bacterial enzymes, the two other intercalating residues
54
55 primarily interact with the complementary DNA strand in duplex substrates [9, 81]. The absence
56
57 of these residues is consistent with the established preference of NEIL3 for lesions in single-
58
59
60
61
62
63
64
65

1
2
3
4 stranded DNA [71, 73], and suggests NEIL3, like AlkZ, may unhook ICLs using a non-base-
5
6 flipping mechanism. However, the non-base-flipping mechanisms employed by AlkZ and
7
8 AlkD/YtkR2 limit these proteins to excising positively charged lesions, which do not require
9
10 activation of the nucleobase through contacts in the nucleobase binding pocket (Box 1). NEIL3
11
12 acts on neutral lesions and crosslinks [54, 56, 62], which would seemingly require activation by
13
14 the enzyme. Yet, the capping loop that performs this function in NEIL1, Nei, and Fpg is strikingly
15
16 short in NEIL3, and seems unlikely to contact even a fully flipped nucleotide (Figure 5E). While
17
18 such an open nucleobase binding pocket is consistent with removal of bulky adducts, including
19
20 nitrogen mustard derivatives of FapyG [56], how NEIL3 activates neutral lesions for excision is
21
22 unclear, and will require additional structural experiments with appropriate DNA substrates.
23
24 Similarly, mechanistic understanding of NEIL2 has been severely limited by a lack of available
25
26 structural information. However, the primary structure of NEIL2 shares all features of NEIL3 that
27
28 seem to enable unhooking of ICLs, and NEIL2, like NEIL3, also shares a preference for non-
29
30 duplex structures [62]. Based on these common features and the interaction of NEIL2 with RNA
31
32 polymerase II [65], we speculate NEIL2 may be involved in transcription-coupled repair of ICLs.
33
34
35
36
37
38
39
40

41 **Protection from Foreign DNA**

42
43 Restriction-modification systems protect prokaryotes from the potentially deleterious
44
45 effects of foreign DNA, regulating genetic exchange among bacteria and guarding against
46
47 infection by phage [83-85]. At a minimum, these systems are comprised of two components: a
48
49 restriction endonuclease that recognizes a specific DNA sequence and introduces one or more
50
51 strand breaks, and a DNA methyltransferase that modifies the same sequence to render it
52
53 resistant to the endonuclease. Type II restriction endonucleases use a Mg^{2+} -dependent
54
55 mechanism to cleave both strands of DNA, usually within a short palindromic sequence,
56
57 producing a DSB [86]. However, genomic analyses recently identified a novel type II enzyme in
58
59
60
61
62
63
64
65

1
2
3
4 hyperthermophilic archaea—with homologs in both thermophilic and mesophilic bacteria—that
5
6 generates strand breaks without requiring Mg^{2+} [87-89]. Unlike other type II enzymes, which
7
8 directly hydrolyze the phosphodiester backbone, R.PabI hydrolyzes the glycosidic bond of 2'-
9
10 deoxyadenosine, producing nearby AP sites on opposing strands. A DSB is then generated by
11
12 heat-promoted β -elimination or enzymatic AP endonuclease or AP lyase activities (Figure 6A)
13
14 [90-92]. Similar to restriction endonucleases, R.PabI cuts exclusively within palindromic GTAC
15
16 sequences, making it the only known sequence-specific DNA glycosylase. To avoid formation of
17
18 DSBs in the genome, R.PabI is coexpressed with the DNA methyltransferase M.PabI, which forms
19
20 N^6 -methyl-2'-deoxyadenosine (N^6 -mA) within the same GTAC sequences, preventing excision by
21
22 R.PabI [91, 93].
23
24
25
26

27 The unique ability of R.PabI to specifically excise adenine from a palindromic sequence is
28
29 based on a novel structural architecture, distinct from not only other DNA glycosylases but other
30
31 proteins [88]. In the absence of DNA, R.PabI forms a dimer with a central, highly twisted β -sheet,
32
33 comprised of six β -strands from each subunit. This structural element is the defining feature of
34
35 the HALFPIPE superfamily and plays a critical role in substrate recognition, providing much of the
36
37 positively charged DNA binding interface [88, 90]. In the presence of non-specific DNA, R.PabI
38
39 dimers associate to form a homotetramer that fully encircles the DNA duplex (Figure 6B) [94]. The
40
41 interface between dimers is small, consisting only of four salt bridges between Arg70 and Asp71.
42
43 Nonetheless, tetramerization appears to facilitate the search for GTAC sequences, as mutation
44
45 of Arg70 decreased the relative rate of adenine excision in the presence of increasing amounts
46
47 of non-specific DNA and reduced the proportion of higher oligomeric species bound to DNA
48
49 lacking the recognition sequence [94]. As in other complexes between DNA glycosylases and
50
51 DNA [9], nearly all DNA interactions in the tetrameric search complex are non-specific and
52
53 mediated by hydrogen-bonds between R.PabI and phosphoryl groups in the DNA backbone [94].
54
55 Perturbation of the duplex in this complex is modest, consisting of a 20° bend along the helical
56
57
58
59
60
61
62
63
64
65

axis and a slight widening of the minor groove. However, upon locating the recognition sequence, R.PabI transitions to a far more disruptive binding mode, in which one of the two dimers is displaced as the duplex is bent by nearly 90° and the minor groove is stretched until the GTAC base pairs are pulled apart (Figure 6C) [90]. This highly distorted DNA conformation is stabilized by insertion of Gln155 and Arg156 into the melted region of the duplex (Figure 6C). The unpaired nucleotides themselves form a large number of contacts with the remaining R.PabI dimer, including 13 sequence-specific hydrogen bonds with each strand in the palindromic sequence (Figure 6D). Such a large number of contacts is possible because both the 2'-deoxyguanosine nucleotide and the excised adenine nucleobase in the product complex are pulled into adjacent but separate nucleobase binding pockets (Figure 6C,E). Only the adenine binding pocket possesses the catalytic residues necessary for base excision. Asp214, which forms a hydrogen bond with the AP site in the product complex (Figure 6E), is positioned to stabilize the sugar as positive charge develops during cleavage of the glycosidic bond, while also pre-organizing the water nucleophile. The manner in which the nucleobase is activated for excision is less apparent (Box 1). No obvious general acid is present in the active site. Yet, N7 of adenine appears to form a hydrogen bond with the backbone carbonyl of Ile66, which would require N7 to be protonated (Figure 6E). Like the plant toxin ricin, which excises adenine from 28S rRNA [95, 96], R.PabI may select for substrates that are protonated at N7 prior to entering the nucleobase binding pocket. However, given the high acidity of N7-protonated adenosine ($pK_a = 2.2$) [97], the fraction of adenine likely to be activated for excision at neutral pH is exceedingly small. Conceivably, R.PabI—as well as ricin—could facilitate protonation prior to or during flipping of the substrate into the active site.

R.PabI and its homologs are the only DNA glycosylases known to function in a restriction-modification system, a role that requires sequence-specific excision of unmodified nucleobases [90]. The specificity that R.PabI exhibits for GTAC sequences comes from a large number of

sequence-specific contacts present in a homodimeric structure that are ideally suited for recognition of palindromic sequences. The residues involved in these contacts are highly conserved among homologs, suggesting all enzymes in the HALFPIPE superfamily recognize GTAC sequences. This is in stark contrast to other DNA glycosylases, which display little to no sequence selectivity, and only form sequence-specific contacts with the lesion and, in some instances, the nucleotide complementary to the lesion [9]. As with restriction endonucleases, the sequence-specific activity of R.PabI has potential biotechnology applications, which could be expanded by modulating the specificity of the enzyme through mutation of residues involved in DNA sequence recognition. However, it is currently unclear what features of the recognition sequence induce the tetrameric search complex to further remodel the DNA substrate and transition to the dimeric excision complex.

Concluding Remarks

Our understanding of how the BER pathway is involved in alternative biological roles has moved forward on two fronts, the discovery of new DNA glycosylases and the identification of new functions associated with previously known enzymes. In the past 12 years, three superfamilies of DNA glycosylase have been discovered, each with its own unique structural architecture and mechanism of lesion recognition and excision. While structurally unrelated, both YtkR2 and AlkZ employ non-base-flipping mechanisms to excise bulky adducts and ICLs, respectively, providing bacteria with self-resistance to genotoxic natural products. NEIL3, which is structurally distinct from both enzymes, likely also utilizes some form of a non-base-flipping mechanism to unhook ICLs, providing an incision-independent pathway for restart of stalled replication forks in vertebrates. Conversely, R.PabI flips not one but four nucleotides into nucleobase binding pockets, enabling recognition of a short palindromic sequence and sequence-specific excision of adenine, providing prokaryotes with protection from foreign DNA as part of a

1
2
3
4 unique restriction-modification system. Together, these enzymes illustrate how deviation from a
5
6 traditional base-flipping mechanism broadens the range of substrates amenable to excision by
7
8 DNA glycosylases, and ultimately how these enzymes expand the biological utility of the BER
9
10 pathway.
11

12
13
14 Despite our advances in understanding the structural and mechanistic features of these
15
16 unique DNA glycosylases, many questions remain unanswered (see Outstanding Questions). We
17
18 are only beginning to speculate how the products of these specialized enzymes are processed in
19
20 the downstream steps of the BER pathway, and how multiple repair pathways may be coordinated
21
22 to perform these alternative roles. As we work to answer these questions, additional DNA
23
24 glycosylases, potentially with new functions, are likely to be discovered. Without question, we still
25
26 have much to learn about a pathway once thought only to eliminate small nucleobase
27
28 modifications from the genome.
29
30
31
32
33

34 **Acknowledgments**

35
36
37 Research on DNA glycosylases in our laboratory is supported by the National Science Foundation
38
39 (MCB-1517695). A.A.R. and N.P.B. are supported by the National Science Foundation Graduate
40
41 Research Fellowship Program (DGE-1445197).
42
43
44
45
46

47 **REFERENCES**

- 48
49
50 1. Friedberg, E.C. et al. (2006) *DNA Repair and Mutagenesis*, 2nd edn., ASM Press.
51
52 2. Gates, K.S. (2009) An overview of chemical processes that damage cellular DNA: spontaneous
53 hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.* 22, 1747-1760.
54
55 3. Jackson, S.P. and Bartek, J. (2009) The DNA-damage response in human biology and disease.
56 *Nature* 461, 1071-1078.
57
58 4. Krokan, H.E. and Bjoras, M. (2013) Base excision repair. *Cold Spring Harb. Perspect. Biol.* 5,
59 a012583.
60
61
62
63
64
65

5. Zharkov, D.O. (2008) Base excision DNA repair. *Cell. Mol. Life Sci.* 65, 1544-1565.
6. Robertson, A.B. et al. (2009) Base excision repair: The long and short of it. *Cell. Mol. Life Sci.* 66, 981-993.
7. Stivers, J.T. (2004) Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog. Nucleic Acid Res. Mol. Biol.* 77, 37-65.
8. Hitomi, K. et al. (2007) The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst.)* 6, 410-428.
9. Brooks, S.C. et al. (2013) Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* 1834, 247-271.
10. Ramstein, J. and Lavery, R. (1988) Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7231-7235.
11. Roberts, R.J. and Cheng, X. (1998) Base flipping. *Annu. Rev. Biochem.* 67, 181-198.
12. Jiang, Y.L. et al. (2001) Turning on uracil-DNA glycosylase using a pyrene nucleotide switch. *J. Biol. Chem.* 276, 42347-42354.
13. Yang, W. (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst.)* 5, 654-666.
14. Tenconi, E. and Rigali, S. (2018) Self-resistance mechanisms to DNA-damaging antitumor antibiotics in actinobacteria. *Curr. Opin. Microbiol.* 45, 100-108.
15. Xu, H. et al. (2012) Self-resistance to an antitumor antibiotic: A DNA glycosylase triggers the base-excision repair system in yatakemycin biosynthesis. *Angew. Chem. Int. Ed.* 51, 10532-10536.
16. Wang, S. et al. (2016) Characterization of a novel DNA glycosylase from *S. sahachiroi* involved in the reduction and repair of azinomycin B induced DNA damage. *Nucleic Acids Res.* 44, 187-197.
17. Zhao, Q. et al. (2008) Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem. Biol.* 15, 693-705.
18. Robinson, E.H. et al. (2010) An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* 468, 406-411.
19. Wu, S. et al. (2017) Unified biosynthetic origin of the benzodipyrrole subunits in CC-1065. *ACS Chem. Biol.* 12, 1603-1610.
20. Shi, R. et al. (2018) Structural biology of the HEAT-like repeat family of DNA glycosylases. *Bioessays* 40, e1800133.
21. Jacob, C. and Weissman, K.J. (2017) Unpackaging the roles of *Streptomyces* natural products. *Cell Chem. Biol.* 24, 1194-1195.
22. Igarashi, Y. et al. (2003) Yatakemycin, a novel antifungal antibiotic produced by *Streptomyces* sp. TP-A0356. *J. Antibiot. (Tokyo)* 56, 107-113.
23. Tichenor, M.S. et al. (2007) Systematic exploration of the structural features of yatakemycin impacting DNA alkylation and biological activity. *J. Am. Chem. Soc.* 129, 10858-10869.

24. MacMillan, K.S. and Boger, D.L. (2009) Fundamental relationships between structure, reactivity, and biological activity for the duocarmycins and CC-1065. *J. Med. Chem.* 52, 5771-5780.
25. Gunz, D. et al. (1996) Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. *J. Biol. Chem.* 271, 25089-25098.
26. Mullins, E.A. et al. (2017) Toxicity and repair of DNA adducts produced by the natural product yatakemycin. *Nat. Chem. Biol.* 13, 1002-1008.
27. Yuan, H. et al. (2017) Gyrl-like proteins catalyze cyclopropanoid hydrolysis to confer cellular protection. *Nat. Commun.* 8, 1485.
28. Alseth, I. et al. (2006) A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol. Microbiol.* 59, 1602-1609.
29. Mullins, E.A. et al. (2015) The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature* 527, 254-258.
30. Dalhus, B. et al. (2007) Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res.* 35, 2451-2459.
31. Robinson, E.H. et al. (2008) A new protein architecture for processing alkylation damaged DNA: The crystal structure of DNA glycosylase AlkD. *J. Mol. Biol.* 381, 13-23.
32. Mullins, E.A. et al. (2014) The substrate binding interface of alkylpurine DNA glycosylase AlkD. *DNA Repair (Amst.)* 13, 50-54.
33. Parsons, Z.D. et al. (2016) A catalytic role for C-H/ π interactions in base excision repair by *Bacillus cereus* DNA glycosylase AlkD. *J. Am. Chem. Soc.* 138, 11485-11488.
34. Lawley, P.D., and D.H. Phillips (1995) DNA Adducts from Chemotherapeutic Agents. *Mutation Research* 355, 13-40.
35. Hata, T. et al. (1954) Carzinophilin, a new tumor inhibitory substance produced by streptomyces. I. *J. Antibiot. (Tokyo)* 7, 107-112.
36. Armstrong, R.W. et al. (1992) Novel interstrand cross-links induced by the antitumor antibiotic carzinophilin/azinomycin B. *J. Am. Chem. Soc.* 114, 3144-3145.
37. Fujiwara, T. et al. (1999) Highly efficient DNA interstrand crosslinking induced by an antitumor antibiotic, carzinophilin. *Tetrahedron Lett.* 40, 315-318.
38. Alcaro, S. and Coleman, R.S. (2000) A molecular model for DNA cross-linking by the antitumor agent azinomycin B. *J. Med. Chem.* 43, 2783-2788.
39. Coleman, R.S. et al. (2002) Studies on the mechanism of action of azinomycin B: definition of regioselectivity and sequence selectivity of DNA cross-link formation and clarification of the role of the naphthoate. *J. Am. Chem. Soc.* 124, 13008-13017.
40. McVey, M. (2010) Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds. *Environ. Mol. Mutagen.* 51, 646-658.
41. Deans, A.J. and West, S.C. (2011) DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* 11, 467-480.
42. Clauson, C. et al. (2013) Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb. Perspect. Biol.* 5, a012732.

- 1
2
3
4 43. Noll, D.M. et al. (2006) Formation and repair of interstrand cross-links in DNA. *Chem. Rev.*
5 106, 277-301.
6
7 44. Huang, Y. and Li, L. (2013) DNA crosslinking damage and cancer - a tale of friend and foe.
8 *Transl. Cancer Res.* 2, 144-154.
9
10 45. Zhang, J. and Walter, J.C. (2014) Mechanism and regulation of incisions during DNA
11 interstrand cross-link repair. *DNA Repair (Amst.)* 19, 135-142.
12
13 46. Wu, R.A. et al. (2019) TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature*
14 567, 267-272.
15
16 47. Zhang, J. et al. (2015) DNA interstrand cross-link repair requires replication-fork convergence.
17 *Nat. Struct. Mol. Biol.* 22, 242-247.
18
19 48. Amunugama, R. et al. (2018) Replication fork reversal during DNA interstrand crosslink repair
20 requires CMG unloading. *Cell Rep.* 23, 3419-3428.
21
22 49. Foulke-Abel, J. et al. (2011) Characterization of AziR, a resistance protein of the DNA cross-
23 linking agent azinomycin B. *Mol. Biosyst.* 7, 2563-2570.
24
25 50. Mullins, E.A. et al. (2017) Structure of a DNA glycosylase that unhooks interstrand cross-links.
26 *Proc. Natl. Acad. Sci. U.S.A.* 114, 4400-4405.
27
28 51. Moldovan, G.L. and D'Andrea, A.D. (2009) How the fanconi anemia pathway guards the
29 genome. *Annu. Rev. Genet.* 43, 223-249.
30
31 52. Kee, Y. and D'Andrea, A.D. (2010) Expanded roles of the Fanconi anemia pathway in
32 preserving genomic stability. *Genes Dev.* 24, 1680-1694.
33
34 53. Long, D.T. et al. (2014) BRCA1 promotes unloading of the CMG helicase from a stalled DNA
35 replication fork. *Mol. Cell* 56, 174-185.
36
37 54. Semlow, D.R. et al. (2016) Replication-dependent unhooking of DNA interstrand cross-links
38 by the NEIL3 glycosylase. *Cell* 167, 498-511.
39
40 55. Martin, P.R. et al. (2017) The human DNA glycosylases NEIL1 and NEIL3 excise psoralen-
41 induced DNA-DNA cross-links in a four-stranded DNA structure. *Sci. Rep.* 7, 17438.
42
43 56. Minko, I.G. et al. (2019) Processing of N(5)-substituted formamidopyrimidine DNA adducts by
44 DNA glycosylases NEIL1 and NEIL3. *DNA Repair (Amst.)* 73, 49-54.
45
46 57. Yang, Z. et al. (2017) A role for the base excision repair enzyme NEIL3 in replication-
47 dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites. *DNA*
48 *Repair (Amst.)* 52, 1-11.
49
50 58. Bandaru, V. et al. (2002) A novel human DNA glycosylase that removes oxidative DNA
51 damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair (Amst.)* 1, 517-
52 529.
53
54 59. Hazra, T.K. et al. (2002) Identification and characterization of a human DNA glycosylase for
55 repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3523-
56 3528.
57
58 60. Morland, I. et al. (2002) Human DNA glycosylases of the bacterial Fpg/MutM superfamily: An
59 alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA.
60 *Nucleic Acids Res.* 30, 4926-4936.
61
62
63
64
65

- 1
2
3
4 61. Takao, M. et al. (2002) Novel nuclear and mitochondrial glycosylases revealed by disruption
5 of the mouse Nth1 gene encoding an endonuclease III homolog for repair of thymine glycols.
6 *EMBO J.* 21, 3486-3493.
7
8 62. Liu, M. et al. (2013) Neil3, the final frontier for the DNA glycosylases that recognize oxidative
9 damage. *Mutat. Res.* 743-744, 4-11.
10
11 63. Hazra, T.K. et al. (2002) Identification and characterization of a novel human DNA glycosylase
12 for repair of cytosine-derived lesions. *J. Biol. Chem.* 277, 30417-30420.
13
14 64. Dou, H. et al. (2003) Repair of oxidized bases in DNA bubble structures by human DNA
15 glycosylases NEIL1 and NEIL2. *J. Biol. Chem.* 278, 49679-49684.
16
17 65. Banerjee, D. et al. (2011) Preferential repair of oxidized base damage in the transcribed genes
18 of mammalian cells. *J. Biol. Chem.* 286, 6006-6016.
19
20 66. Torisu, K. et al. (2005) Hematopoietic tissue-specific expression of mouse Neil3 for
21 endonuclease VIII-like protein. *J. Biochem.* 138, 763-772.
22
23 67. Hildrestrand, G.A. et al. (2009) Expression patterns of Neil3 during embryonic brain
24 development and neoplasia. *BMC Neurosci.* 10, 45.
25
26 68. Neurauter, C.G. et al. (2012) Release from quiescence stimulates the expression of human
27 NEIL3 under the control of the Ras dependent ERK-MAP kinase pathway. *DNA Repair (Amst.)*
28 11, 401-409.
29
30 69. Bandaru, V. et al. (2007) Human endonuclease VIII-like (NEIL) proteins in the giant DNA
31 Mimivirus. *DNA Repair (Amst.)* 6, 1629-1641.
32
33 70. Takao, M. et al. (2009) Human Nei-like protein NEIL3 has AP lyase activity specific for single-
34 stranded DNA and confers oxidative stress resistance in *Escherichia coli* mutant. *Genes Cells*
35 14, 261-270.
36
37 71. Liu, M. et al. (2010) The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and
38 in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4925-4930.
39
40 72. Liu, M. et al. (2012) Expression and purification of active mouse and human NEIL3 proteins.
41 *Protein Expr. Purif.* 84, 130-139.
42
43 73. Liu, M. et al. (2013) Structural characterization of a mouse ortholog of human NEIL3 with a
44 marked preference for single-stranded DNA. *Structure* 21, 247-256.
45
46 74. Couvé-Privat, S. et al. (2007) Psoralen-induced DNA adducts are substrates for the base
47 excision repair pathway in human cells. *Nucleic Acids Res.* 35, 5672-5682.
48
49 75. Couvé, S. et al. (2009) The human oxidative DNA glycosylase NEIL1 excises psoralen-
50 induced interstrand DNA cross-links in a three-stranded DNA structure. *J. Biol. Chem.* 284,
51 11963-11970.
52
53 76. Cheng, S. et al. (1988) Use of psoralen-modified oligonucleotides to trap three-stranded
54 RecA-DNA complexes and repair of these cross-linked complexes by ABC excinuclease. *J.*
55 *Biol. Chem.* 263, 15110-15117.
56
57 77. Kaushik Tiwari, M. et al. (2016) Triplex structures induce DNA double strand breaks via
58 replication fork collapse in NER deficient cells. *Nucleic Acids Res.* 44, 7742-7754.
59
60 78. Wilson, D.M., III and Seidman, M.M. (2010) A novel link to base excision repair? *Trends*
61 *Biochem. Sci.* 35, 247-252.
62
63
64
65

- 1
2
3
4 79. Doublié, S. et al. (2004) The crystal structure of human endonuclease VIII-like 1 (NEIL1)
5 reveals a zincless finger motif required for glycosylase activity. *Proc. Natl. Acad. Sci. U.S.A.*
6 101, 10284-10289.
7
8 80. Prakash, A. et al. (2014) Genome and cancer single nucleotide polymorphisms of the human
9 NEIL1 DNA glycosylase: Activity, structure, and the effect of editing. *DNA Repair (Amst.)* 14,
10 17-26.
11
12 81. Zhu, C. et al. (2016) Tautomerization-dependent recognition and excision of oxidation damage
13 in base-excision DNA repair. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7792-7797.
14
15 82. Qi, Y. et al. (2009) Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme.
16 *Nature* 462, 762-766.
17
18 83. Bickle, T.A. and Kruger, D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.* 57, 434-450.
19
20 84. Murray, N.E. (2002) Immigration control of DNA in bacteria: Self versus non-self. *Microbiology*
21 148, 3-20.
22
23 85. Samson, J.E. et al. (2013) Revenge of the phages: Defeating bacterial defences. *Nat. Rev.*
24 *Microbiol.* 11, 675-687.
25
26 86. Pingoud, A. et al. (2005) Type II restriction endonucleases: Structure and mechanism. *Cell.*
27 *Mol. Life Sci.* 62, 685-707.
28
29 87. Ishikawa, K. et al. (2005) Discovery of a novel restriction endonuclease by genome
30 comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-
31 GTA/C) from the hyperthermophilic archaeon *Pyrococcus abyssi*. *Nucleic Acids Res.* 33, e112.
32
33 88. Miyazono, K. et al. (2007) Novel protein fold discovered in the PabI family of restriction
34 enzymes. *Nucleic Acids Res.* 35, 1908-1918.
35
36 89. Humbert, O. and Salama, N.R. (2008) The *Helicobacter pylori* HpyAXII restriction-modification
37 system limits exogenous DNA uptake by targeting GTAC sites but shows asymmetric
38 conservation of the DNA methyltransferase and restriction endonuclease components.
39 *Nucleic Acids Res.* 36, 6893-6906.
40
41 90. Miyazono, K. et al. (2014) A sequence-specific DNA glycosylase mediates restriction-
42 modification in *Pyrococcus abyssi*. *Nat. Commun.* 5, 3178.
43
44 91. Fukuyo, M. et al. (2015) Restriction-modification system with methyl-inhibited base excision
45 and abasic-site cleavage activities. *Nucleic Acids Res.* 43, 2841-2852.
46
47 92. Zhang, Y. et al. (2017) Restriction glycosylases: Involvement of endonuclease activities in the
48 restriction process. *Nucleic Acids Res.* 45, 1392-1403.
49
50 93. Watanabe, M. et al. (2006) Hyperthermophilic DNA methyltransferase M.PabI from the
51 archaeon *Pyrococcus abyssi*. *Appl. Environ. Microbiol.* 72, 5367-5375.
52
53 94. Wang, D. et al. (2016) Tetrameric structure of the restriction DNA glycosylase R.PabI in
54 complex with nonspecific double-stranded DNA. *Sci. Rep.* 6, 35197.
55
56 95. Chen, X. et al. (2000) Transition-state analysis for depurination of DNA by ricin A-chain. *J.*
57 *Am. Chem. Soc.* 122, 6527-6534.
58
59 96. Ho, M.C. et al. (2009) Transition state analogues in structures of ricin and saporin ribosome-
60 inactivating proteins. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20276-20281.
61
62
63
64
65

- 1
2
3
4 97. Kapinos, L.E. et al. (2011) Understanding the acid–base properties of adenosine: the intrinsic
5 basicities of N1, N3 and N7. *Chem. Eur. J.* 17, 8156-8164.
6
7 98. Stivers, J.T. and Jiang, Y.L. (2003) A mechanistic perspective on the chemistry of DNA repair
8 glycosylases. *Chem. Rev.* 103, 2729-2759.
9
10 99. Drohat, A.C. and Maiti, A. (2014) Mechanisms for enzymatic cleavage of the N-glycosidic
11 bond in DNA. *Org. Biomol. Chem.* 12, 8367-8378.
12
13 100. Drohat, A.C. and Coey, C.T. (2016) Role of base excision "repair" enzymes in erasing
14 epigenetic marks from DNA. *Chem. Rev.* 116, 12711-12729.
15
16 101. Sadeghian, K. et al. (2014) Ribose-protonated DNA base excision repair: A combined
17 theoretical and experimental study. *Angew. Chem. Int. Ed.* 53, 10044-10048.
18
19 102. Sadeghian, K. and Ochsenfeld, C. (2015) Unraveling the base excision repair mechanism of
20 human DNA glycosylase. *J. Am. Chem. Soc.* 137, 9824-98231.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

FIGURE LEGENDS

Figure 1. Base Excision Repair of Damaged Nucleobases. (A) Common DNA lesions resulting from alkylation, oxidation, and deamination of nucleobases. 3mA, 3-methyl-2'-deoxyadenosine; 7mG, 7-methyl-2'-deoxyguanosine; 8oxoG, 8-oxo-2'-deoxyguanosine; ϵ A, 1, N^6 -etheno-2'-deoxyadenosine; mFapyG, methyl derivative of N^6 -(2'-deoxyribose)-2,6-diamino-4-oxo-5-formamidopyrimidine (FapyG); dU, 2'-deoxyuridine. (B) General steps in the base excision repair (BER) pathway. BER is initiated by lesion-specific DNA glycosylases, which remove damaged nucleobases to create an apurinic/apyrimidinic (AP) site. An AP endonuclease (or a bifunctional DNA glycosylase) then incises the modified strand, producing a single-strand break. As necessary, the break is processed by one of several enzymes to create a gap with a 3'-hydroxyl group and a 5'-phosphoryl group. A DNA polymerase fills the gap with new DNA, and a DNA ligase seals the strand to complete repair. In eukaryotes, if strand incision is performed by an AP endonuclease, repair synthesis occurs before end processing, displacing the AP site.

Figure 2. Initiation of Base Excision Repair by DNA Glycosylases. (A) Base excision and strand incision reactions performed by monofunctional and bifunctional DNA glycosylases. Monofunctional enzymes catalyze only base excision, wherein the glycosidic bond between the nucleobase and the phosphodeoxyribose backbone is hydrolyzed, removing the nucleobase and creating an AP site. Bifunctional enzymes catalyze both base excision and strand incision (lyase activity). During removal of the nucleobase, most bifunctional glycosylases form an iminium intermediate, which covalently links the protein and the DNA. Some bifunctional enzymes, however, initially hydrolyze the glycosidic bond to create an AP site before then converting the AP site to an iminium intermediate. Following base excision, all bifunctional DNA glycosylases incise the strand on the 3'-side of the AP site (β -elimination), generating a single-strand break with a 3'-phospho- α,β -unsaturated aldehyde (PUA) group and a 5'-phosphoryl group. Some bifunctional enzymes also subsequently incise the strand on the 5'-side of the PUA (δ -elimination),

1
2
3
4 leaving a 3'-phosphate, which must also be removed prior to repair synthesis, requiring the
5
6 phosphatase activity of a separate enzyme. Alternatively, if β -elimination occurs following strand
7
8 incision by an AP endonuclease, a gap is generated with the 3'-hydroxyl and 5'-phosphoryl groups
9
10 necessary for synthesis and ligation. (B) X-ray crystal structures of the bifunctional DNA
11
12 glycosylase Fpg (green) bound to DNA (orange and yellow) containing an 8oxoG lesion (red; PDB
13
14 ID: 3GO8, 3GPY) [82]. After base flipping, Met77, Arg112, and Phe114 (not shown) fill the void in
15
16 the duplex. (C) An extrahelical 8oxoG lesion in the nucleobase binding pocket of Fpg. Hydrogen-
17
18 bonding interactions are indicated with dashed lines.
19
20
21

22 **Figure 3. Removal of Bulky Adducts by AlkD.** (A) Bulky adducts excised by AlkD. 7POBG, 7-
23
24 pyridyloxobutyl-2'-deoxyguanosine; O²-POBC, O²-pyridyloxobutyl-2'-deoxycytidine; YTMA, 3-
25
26 yatakemycinyl-2'-deoxyadenosine. AlkD eliminates bulky lesions with modifications located in
27
28 either the major (7POBG) or the minor (O²-POBC and YTMA) groove. (B,D) X-ray crystal structure
29
30 of AlkD (cyan) in complex with free 3-methyladenine (3mAde) nucleobase (purple) and DNA
31
32 (orange and yellow) containing a tetrahydrofuran (THF) spacer (purple) to mimic an AP site (PDB
33
34 ID: 5CLE) [29]. Water molecules located in the large cavity between the protein and the DNA are
35
36 depicted as red spheres. (C,E) X-ray crystal structure of AlkD in complex with an excised
37
38 yatakeymycinyladenine (YTMAde) nucleobase (purple) and DNA containing an AP site (purple;
39
40 PDB ID: 5UUF) [26]. Hydrogen-bonding interactions are indicated with dashed lines. Unlike DNA
41
42 glycosylases that use a traditional base-flipping mechanism, the catalytic residues (Trp109,
43
44 Asp113, and Trp187) present in AlkD are located on the protein surface, not recessed in a
45
46 nucleobase binding pocket. Without a catalytic requirement for base flipping, the lack of protein–
47
48 DNA contacts in the major groove and the large solvent-filled cavity between the protein and the
49
50 minor groove allow AlkD to recognize and excise nucleobases with bulky modifications at any
51
52 position.
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 4. Unhooking of Interstrand Crosslinks by AlkZ. (A) Interstrand crosslink (ICL) formed between two 2'-deoxyguanosine nucleotides by azinomycin B (AZB). AZB preferentially reacts in GNC sequences to form {1-3} crosslinks. (B) Hypothetical models of AlkZ (green and yellow) bound to DNA (gray and blue) containing an AZB ICL (red). The models were constructed by rigid-body docking of an X-ray crystal structure of AlkZ (PDB ID: 5UUJ) [50] and a computationally derived model of DNA containing an AZB ICL [38]. The DNA could be docked equally well in either of two binding orientations by placing the β 11/ β 12 hairpin into the minor groove across from the ICL. Each orientation positions one of the two modified nucleotides (G_1 or G_2) near Gln39 in the putative active site of AlkZ. (C,D) Hypothetical model of an AlkZ dimer bound to DNA containing an AZB ICL. Docking two molecules of AlkZ with a single AZB ICL creates a network of salt bridges (Glu152 and Arg153), hydrogen bonds (Ser304 and Arg308), and hydrophobic contacts (Ala309 and Pro340) at the protein interface. (E) Alternate mechanisms of AZB ICL processing by the BER pathway. Concerted unhooking of both strands, consistent with the hypothetical dimeric complex, produces two closely spaced AP sites, potentially leading to a double-strand break (DSB). Sequential unhooking and repair of each strand avoids the concurrence of multiple AP sites and minimizes the possibility of a DSB.

Figure 5. Excision of Bulky Adducts and Interstrand Crosslinks by NEIL1 and NEIL3. (A) Bulky adducts and interstrand crosslinks removed by NEIL1 or NEIL3. NM-FapyG, nitrogen mustard derivative of FapyG; AFB₁-FapyG, aflatoxin B₁ derivative of FapyG; PSO, psoralen. Both psoralen and AP sites form {1-2} crosslinks by modifying nucleobases on opposing strands. (B) Triplex substrate for NEIL1. The crosslink is colored red and nascent DNA produced by translesion synthesis is gray. Triplex structures are generated during replication-associated ICL repair (Supplemental Figure S1). (C) Convergent fork substrate for NEIL3. Nascent DNA generated during replication is colored gray. NEIL3 is recruited to ICLs after fork convergence and replisome ubiquitinylation (Supplemental Figure S1). (D) X-ray crystal structure of NEIL1

(cyan) bound to DNA (orange and yellow) containing a thymine glycol (Tg) lesion (red; PDB ID: 5ITY) [81]. Three intercalating residues (Met81, Arg118, and Phe120) stabilize the extrahelical conformation of the DNA substrate, while a flexible capping loop allows for accommodation and excision of bulky lesions. (E) Crystal structure of the glycosylase domain of NEIL3 (PDB ID: 3W0F) [73]. NEIL3 lacks the two intercalating residues that interact with the undamaged strand in the NEIL1 complex, as well as the flexible capping loop that contacts the Tg lesion. Consistent with a preference for non-duplex substrates, NEIL3 also lacks the basic residue (Arg274) that forms a salt bridge with the undamaged strand in the NEIL1 complex.

Figure 6. Sequence-Specific Excision of Adenine by R.PabI. (A) Double-strand break resulting from dual base excision and subsequent incision of AP sites on opposing strands. (B) X-ray crystal structure of tetrameric R.PabI (blue, green, and white) bound to non-specific DNA (orange and yellow; PDB ID: 5IFF) [94]. Four salt bridges formed by Arg70 and Asp71 are the only interactions between the two R.PabI dimers in the tetrameric search complex. (C) Crystal structure of dimeric R.PabI bound to specific DNA after dual excisions of adenine (Ade) nucleobase (red) to create AP sites (red) on opposing strands (PDB ID: 3WAZ) [90]. Binding of the recognition sequence induces a transition to a dimeric excision complex in which the GTAC base pairs are pulled apart and the void created in the duplex is stabilized by insertion of Gln155 and Arg156. (D) Hydrogen-bonding interactions in the dimeric product complex. DNA binding residues are colored according to protein subunit. R.PabI forms 13 sequence-specific and 21 non-sequence-specific hydrogen-bonding interactions with each strand in the palindromic product. (E) Recognition of Ade in the active site of R.PabI (stereodiagram). Hydrogen-bonding interactions are indicated with dashed lines. *N*⁶-methyl-2'-deoxyadenosine, *N*⁶-mA. Hydrogen bonds between Ade and backbone atoms in Ile66 and Val164 select for an N7-protonated substrate to catalyze excision. Sequence-specific methylation of Ade by M.PabI introduces steric clashes with these same backbone atoms to prevent catalytically productive binding and excision of *N*⁶-mA.

TEXT BOXES

Box 1. The Chemistry of Base Excision. Two chemical mechanisms have been proposed for base excision, which differ primarily in the number of steps and the presence or absence of a chemical intermediate [98, 99]. By a concerted mechanism (one step), the nucleobase is displaced by an attacking nucleophile, and both groups simultaneously have some degree of bonding character. By a stepwise mechanism (two steps), the nucleobase departs before the nucleophile attacks, generating a short-lived but discrete oxocarbenium intermediate (Figure 1A). To date, all studies of glycosidic bond hydrolysis in DNA have been consistent with a stepwise mechanism, both for enzymatic and non-enzymatic reactions [99, 100]. As the glycosidic bond is broken, electron density is pulled away from the sugar and onto the nucleobase [29, 98, 99]. For neutral lesions, this creates a negatively charged and usually unstable nucleobase leaving group, as well as a positively charged and highly unstable oxocarbenium intermediate (Figure 1A). Stabilization of these developing charges is a key factor in how DNA glycosylases catalyze base excision. Depending on the enzyme and the lesion, transfer of electron density to the nucleobase is stabilized by one of two mechanisms [98, 100]. By acid catalysis, the nucleobase is protonated prior to excision, creating a positively charged substrate with a nucleobase that departs as a stable, neutral leaving group. Conversely, by electrostatic catalysis, the nucleobase is not protonated before excision, and departs as an anionic leaving group. Interactions within the nucleobase binding pocket stabilize the increase in electron density on the nucleobase. In both mechanisms, the corresponding loss of electron density on the sugar is stabilized through electrostatic interactions. Carboxylate (aspartate/glutamate) or carboxamide (asparagine/glutamate) side chains located in the active site and near the anomeric carbon (C1') provide electron donation that offsets the loss of electron density to the nucleobase (Figure 1B) [9, 100]. In monofunctional DNA glycosylases, these side chains also position the water nucleophile that attacks the oxocarbenium intermediate to produce an AP site. In bifunctional DNA

glycosylases, an amine functional group (lysine ϵ -NH₂ or an N-terminal valine or proline α -NH₂) positions the water nucleophile to generate an AP site, or more commonly, attacks the oxocarbenium intermediate to form an iminium crosslink. Recent studies have proposed an alternative mechanism for bifunctional glycosylases, whereby the amine group attacks the sugar before departure of the nucleobase [81, 101, 102]. In this scenario, instead of breaking the glycosidic bond, nucleophilic addition opens the deoxyribose ring between C1' and O4', requiring protonation of O4' by a catalytic carboxylic acid side chain. By this "ribose-protonated mechanism", the glycosidic bond remains intact until the iminium crosslink is fully formed.

Figure I. Base Excision by DNA Glycosylases. (A) Putative mechanism of glycosidic bond cleavage. Addition of a water molecule or an amine group to the oxocarbenium intermediate would produce an AP site or an iminium crosslink, respectively. (B) Key catalytic residues required for base excision. Carboxylate (aspartate/glutamate) and carboxamide (asparagine/glutamine) functional groups are depicted both as lowest-energy resonance forms (left) and as resonance hybrids (right).

OUTSTANDING QUESTIONS

- Three new families of DNA glycosylase, each with a unique structural architecture and a novel function, have been discovered in the past ten years. Are there more DNA glycosylases, particularly with new functions, that have yet to be identified?
- Two of the three recently discovered families of DNA glycosylase play roles in self-resistance to genotoxic secondary metabolites. Are YtkR2 and AlkZ representative of a more widespread mechanism of bacterial self-resistance to natural products? That is, do other bacteria that produce genotoxins also employ specialized DNA repair enzymes that have not been discovered? If so, are these enzymes, like YtkR2 and AlkZ, encoded within biosynthetic gene clusters?
- Base excision of non-traditional substrates (e.g., ICLs) generates non-traditional products. It is unclear if these products can be processed by downstream enzymes in the BER pathway. If not, have specialized enzymes evolved, or are other canonical DNA repair pathways involved?
- Many DNA glycosylases remove a range of nucleobase substrates. Moreover, every DNA glycosylase that has been shown to excise bulky adducts or interstrand crosslinks also acts upon nucleobases with small modifications. Do any of the currently known DNA glycosylases that have been shown to excise traditional substrates also process unidentified substrates or perform unknown functions?

HIGHLIGHTS

- Recent studies revealed new and unexpected functions for diverse DNA glycosylases, expanding the known roles of the base excision repair pathway.
- Two specialized bacterial DNA glycosylases excise bulky adducts and interstrand crosslinks to provide self-resistance to genotoxic secondary metabolites.
- A vertebrate DNA glycosylase previously associated with excision of oxidative lesions performs replication-coupled unhooking of interstrand crosslinks to restart stalled replication forks.
- An archaeal DNA glycosylase originally annotated as a restriction endonuclease uses sequence-specific excision of adenine to induce double-strand breaks in foreign DNA.

(A) Figure 1

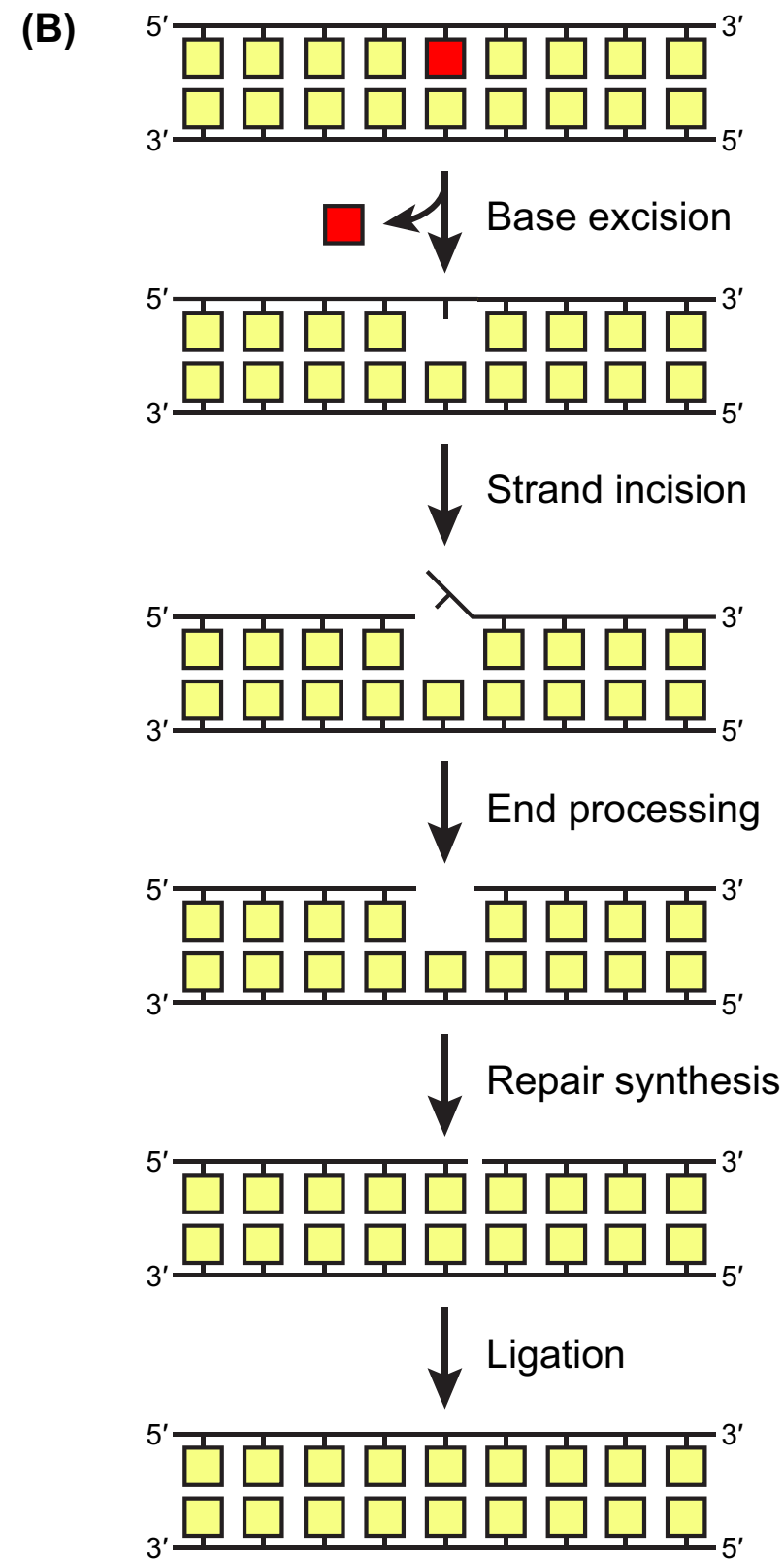
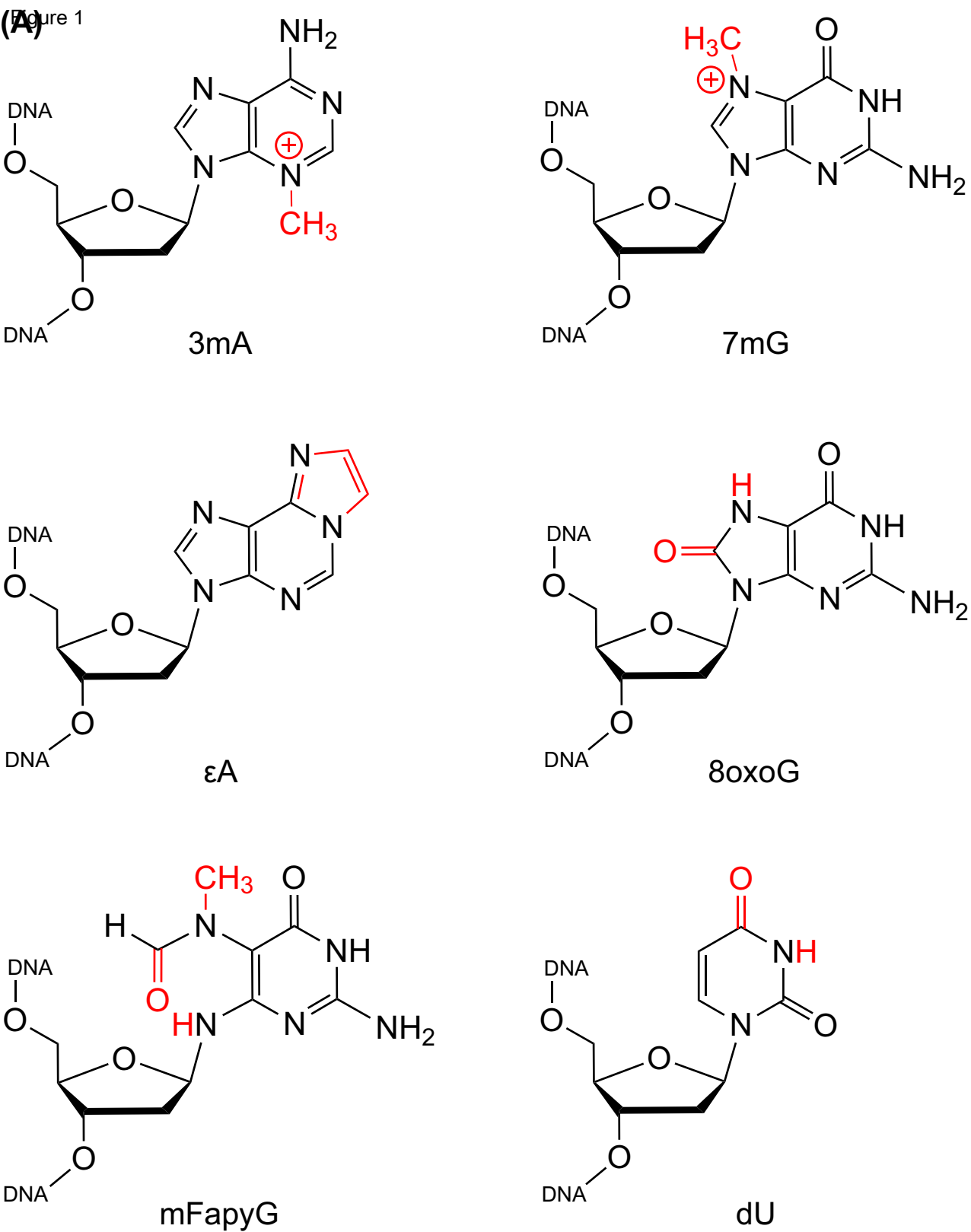
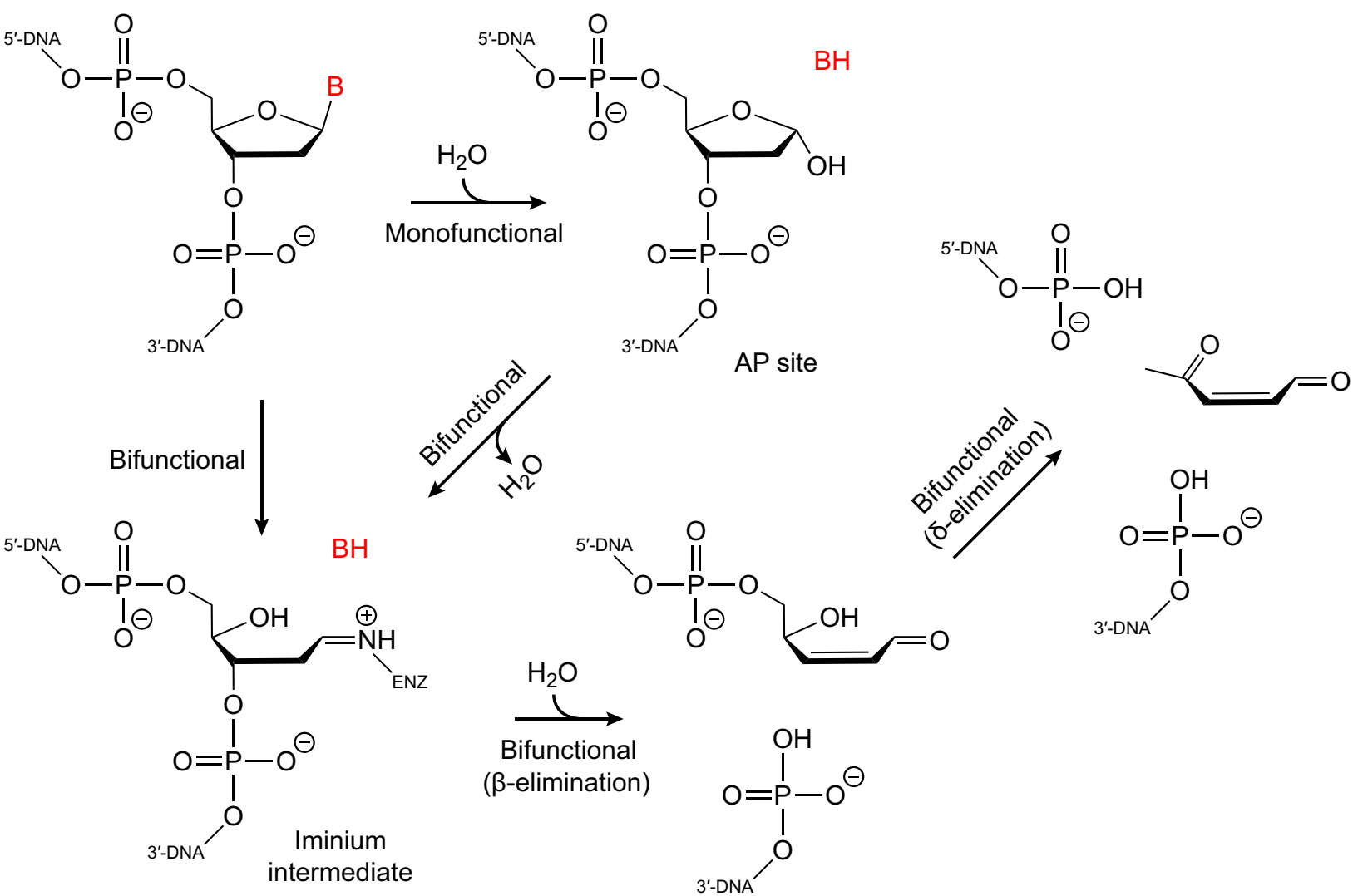
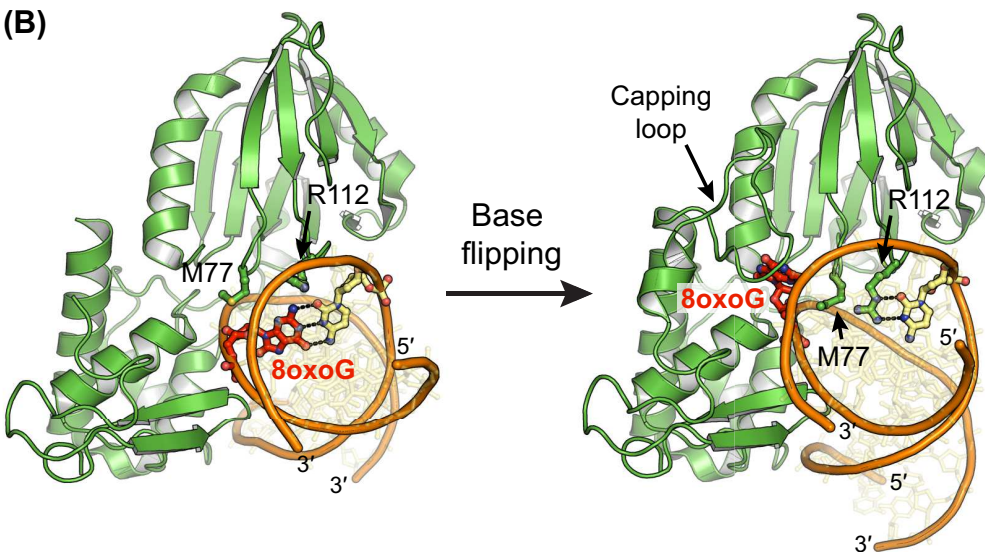


Figure 2

(A)



(B)



(C)

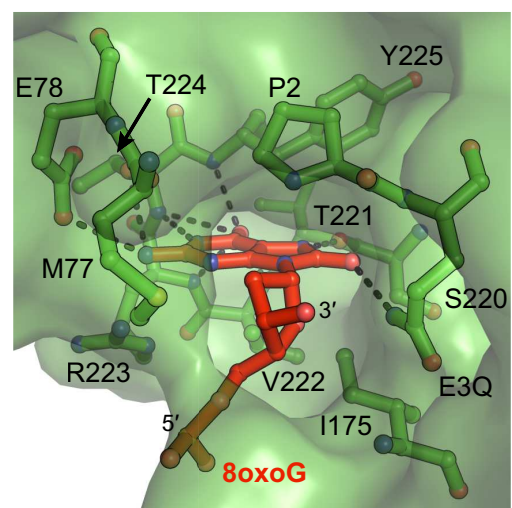
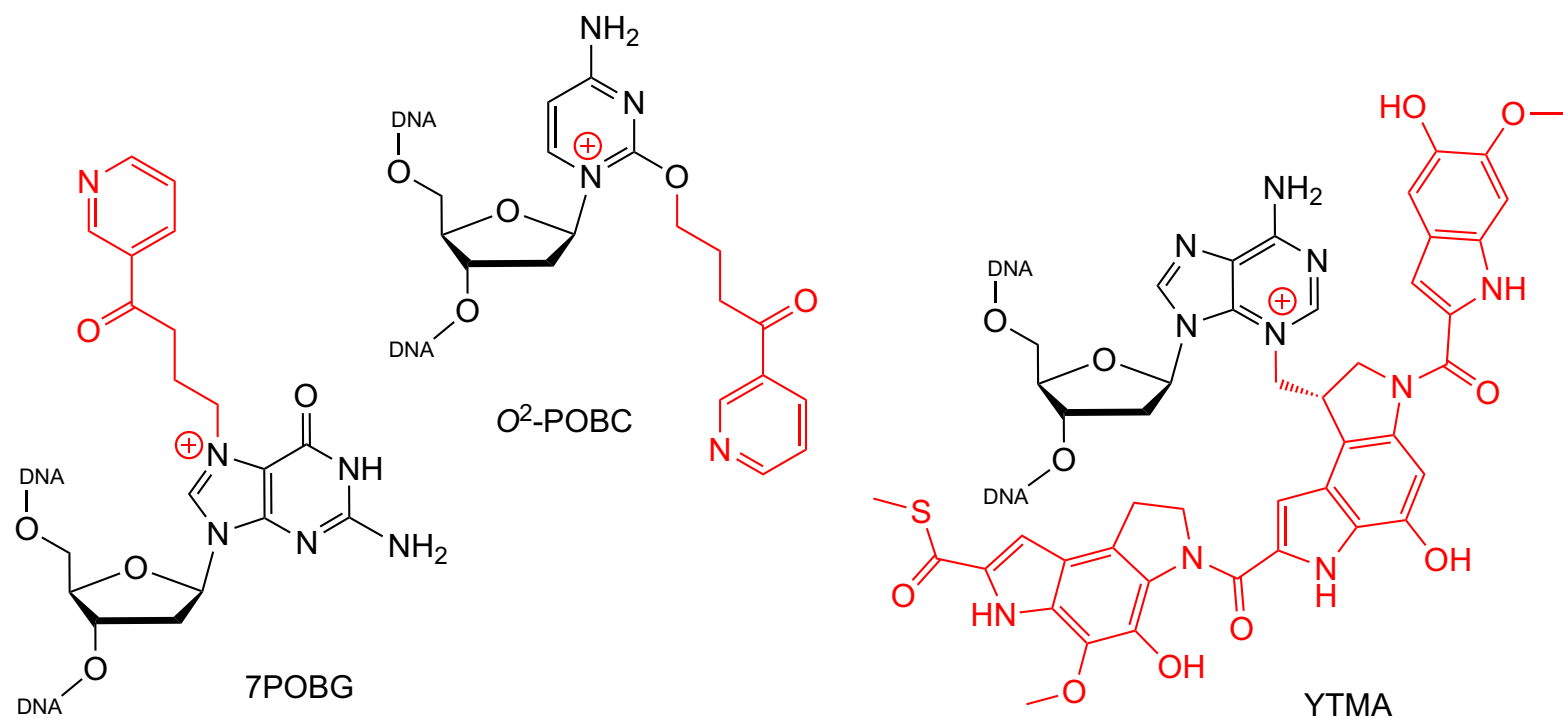
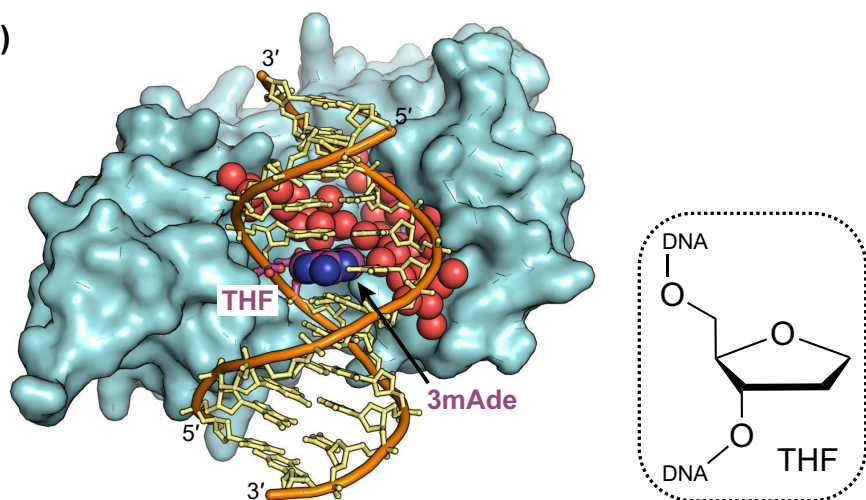


Figure 3

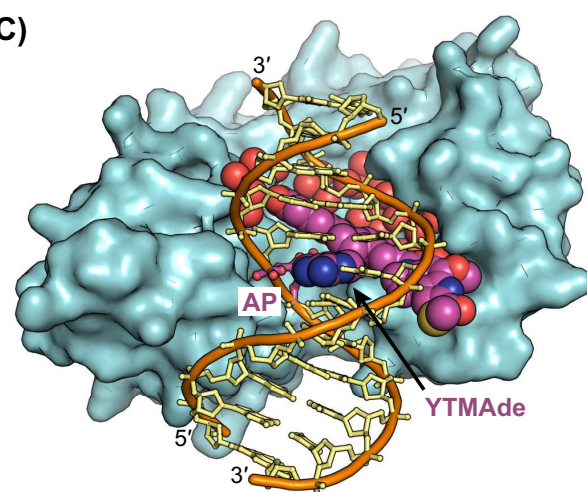
(A)



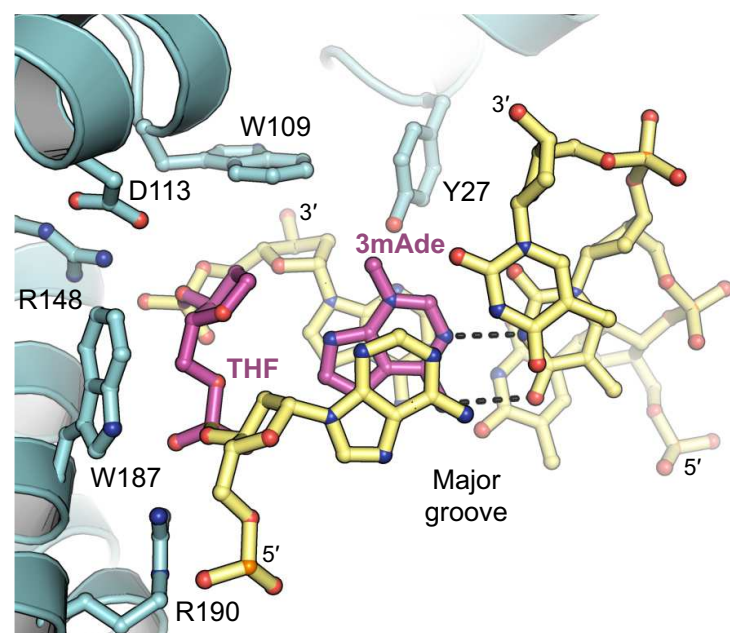
(B)



(C)



(D)



(E)

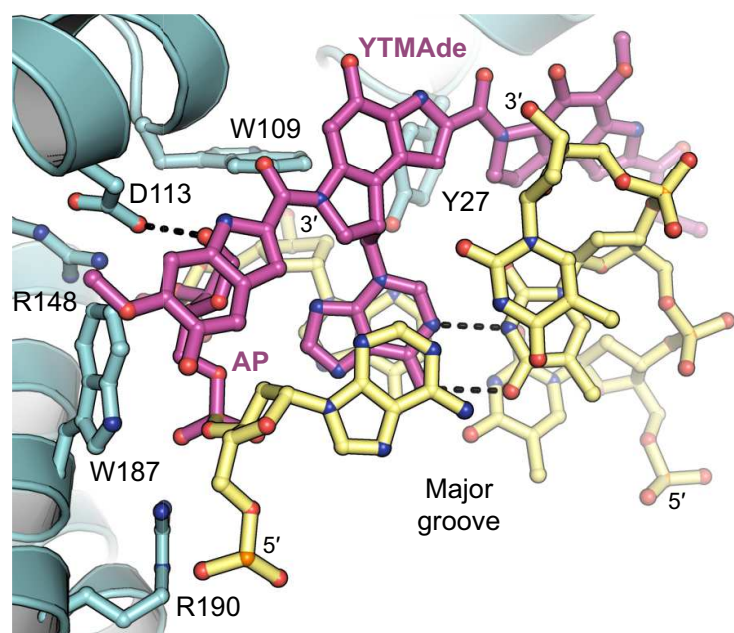
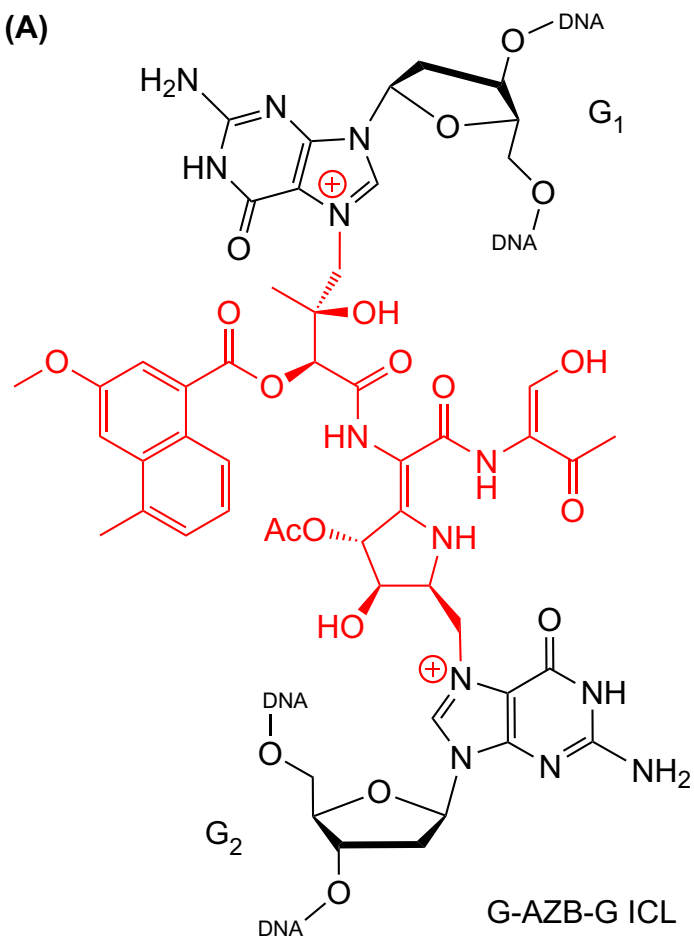
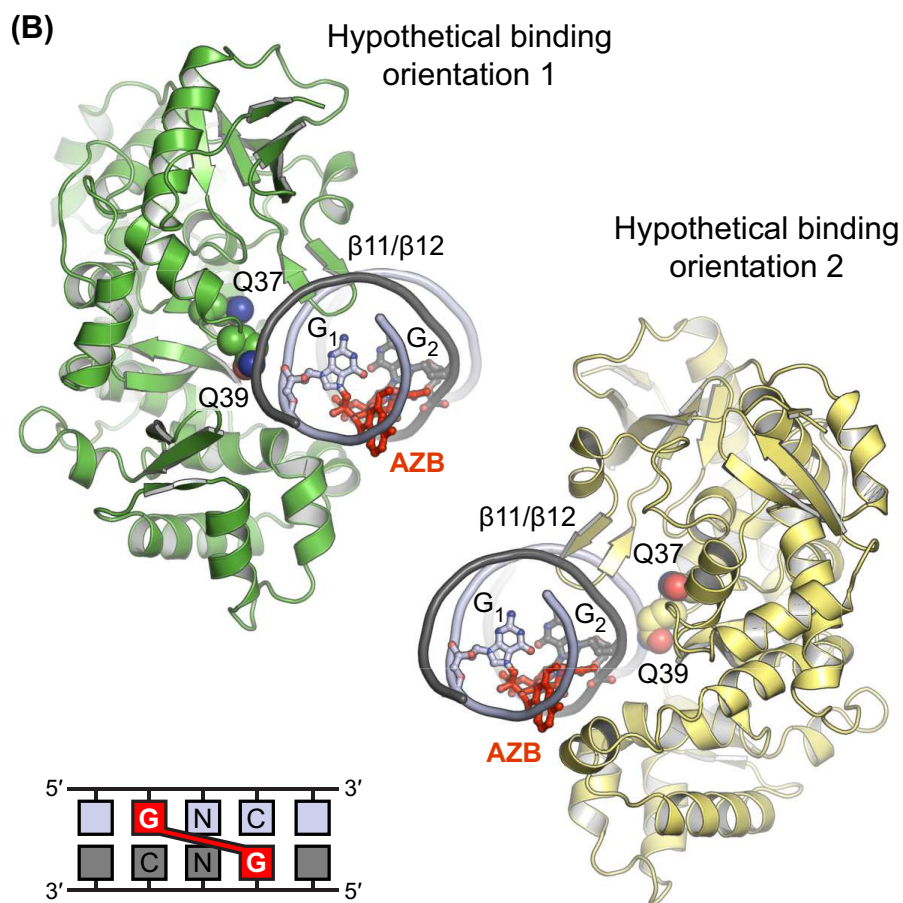


Figure 4

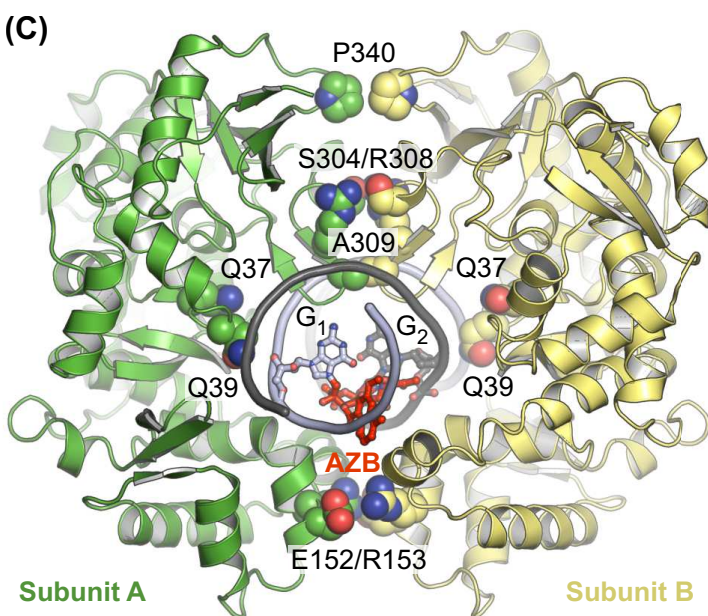
(A)



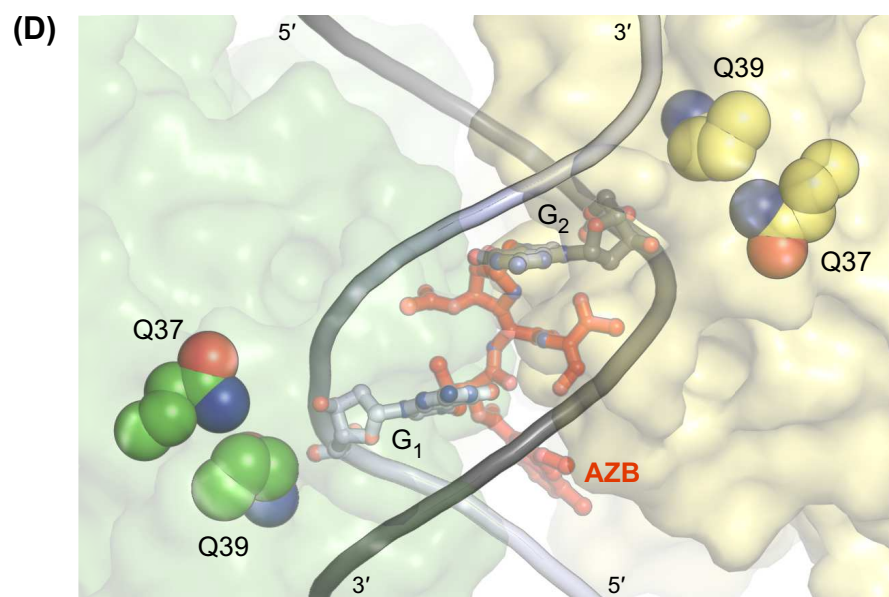
(B)



(C)



(D)



(E)

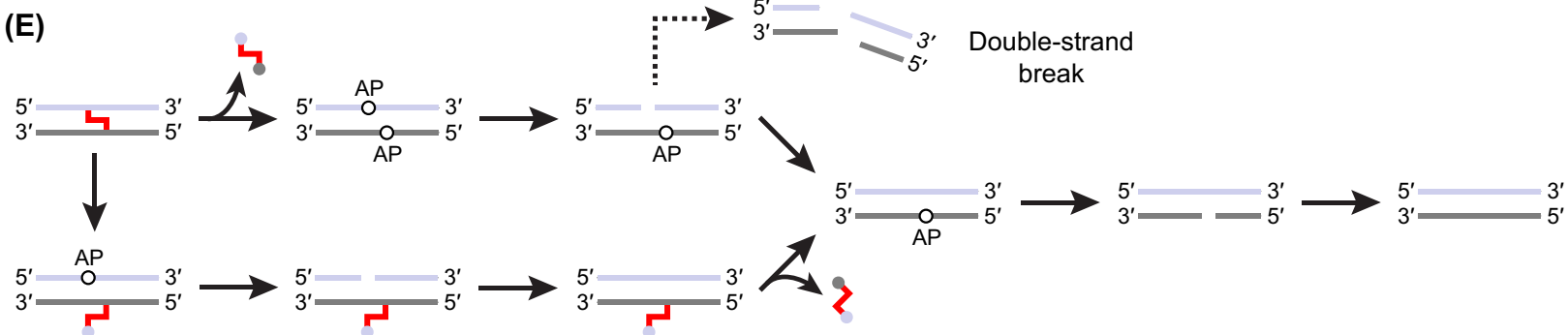


Figure 5

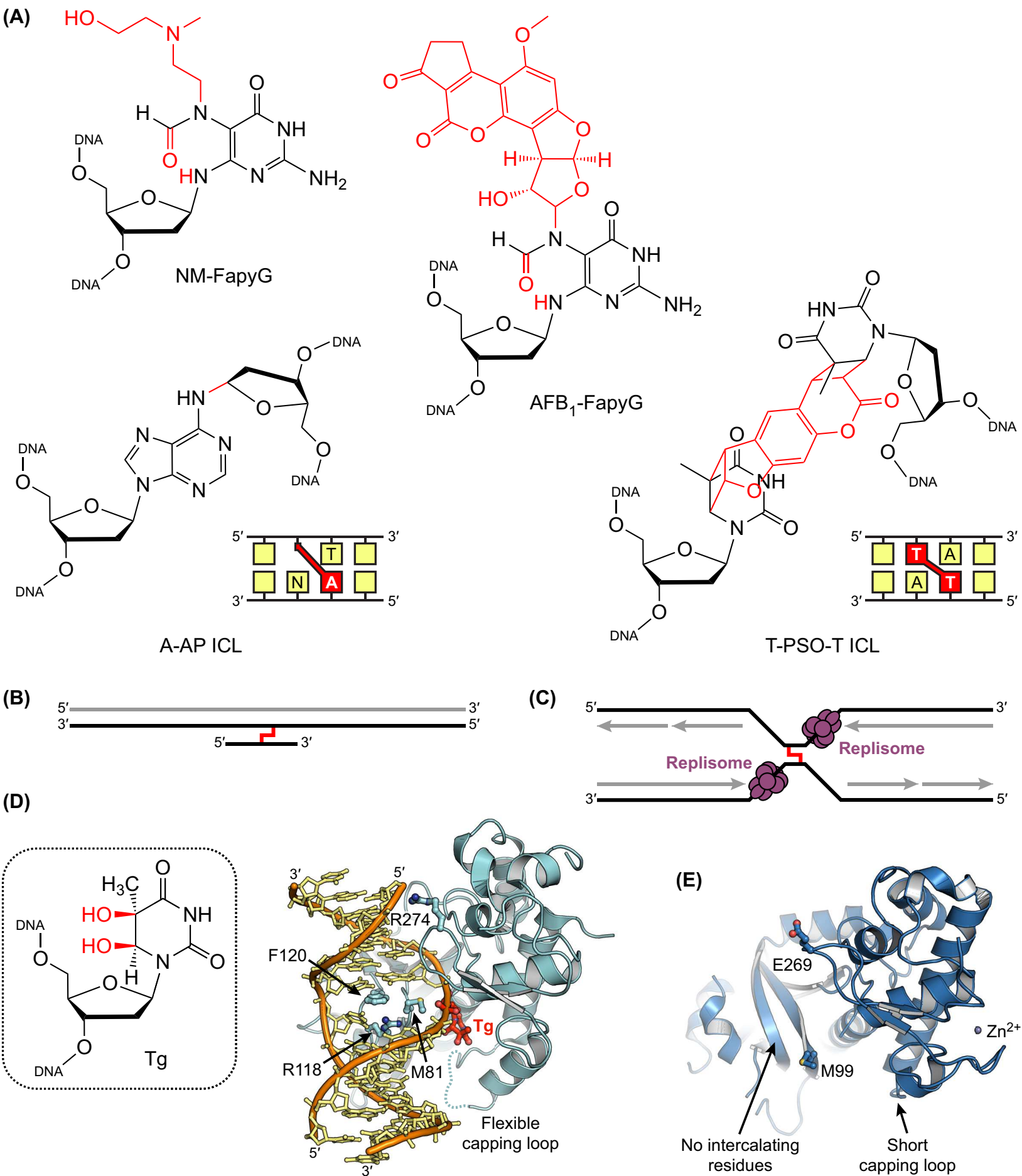
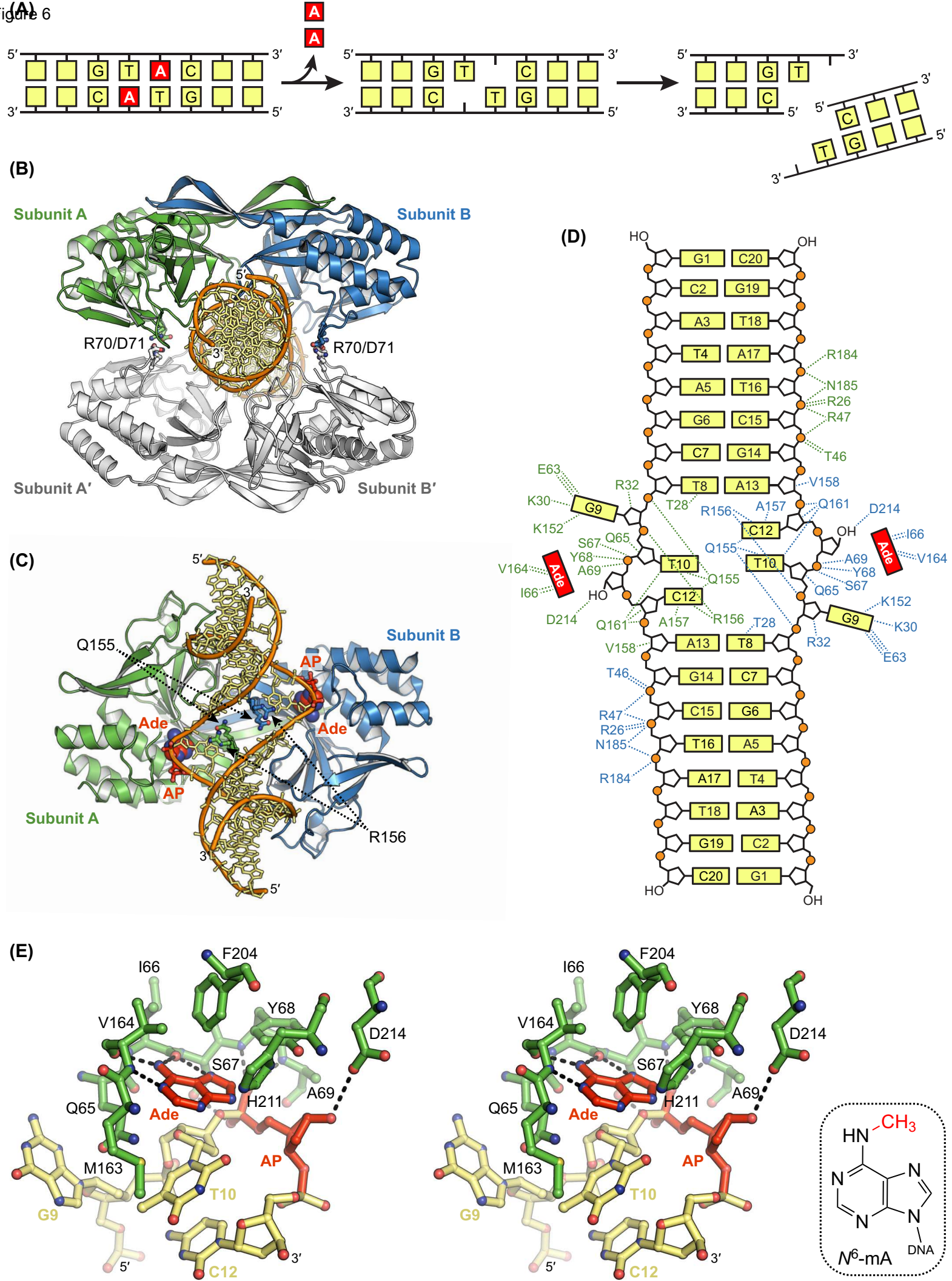
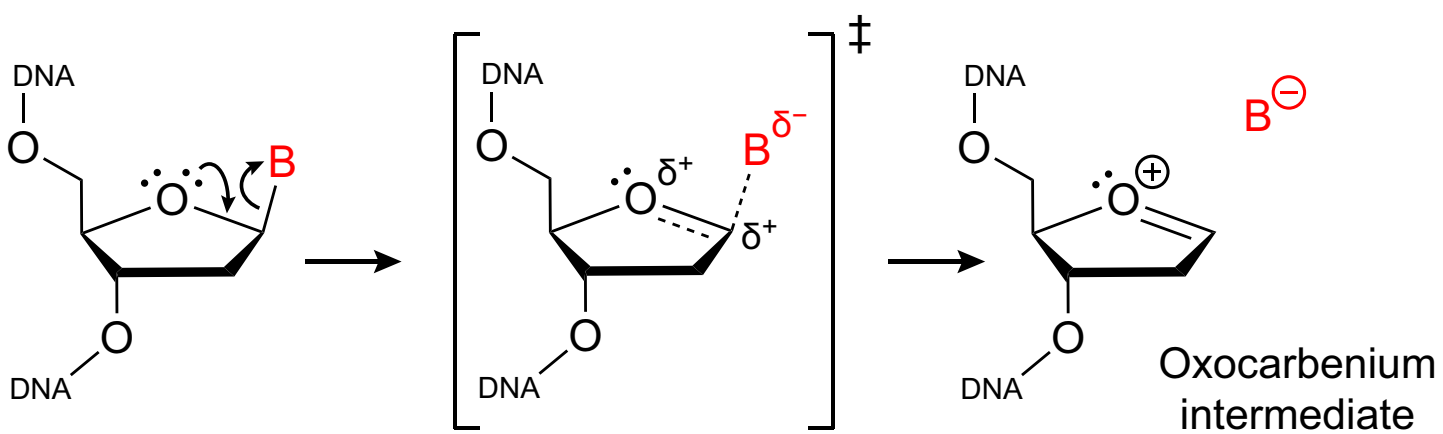
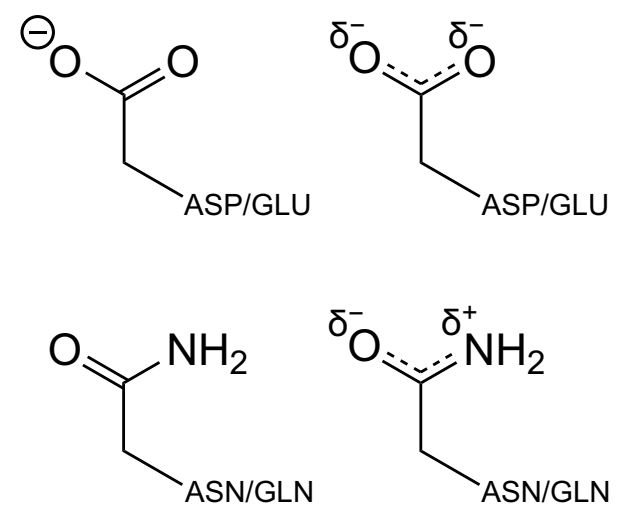


Figure 6



(A)**(B)**



Click here to access/download
Supplementary Material
TIBS_review_SI.pdf