Catalytic Promiscuity and the Divergent Evolution of DNA Repair Enzymes

Patrick J. O'Brien*

Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109-0606

Received May 17, 2005 (Revised Manuscript Received November 17, 2005)

Contents

1. Introduction 720
2. Overview of Catalytic Promiscuity and the Evolutionary Diversification of Enzymes 722
3. The Chemical Landscape for DNA Damage and Repair 723
4. Overview of DNA Repair Pathways: A Common Theme of Enzymes that have Broad Substrate Specificities 726
   4.1. Base Excision Repair 726
   4.2. Nucleotide Excision and Mismatch Repair Pathways 731
   4.3. Nucleotide Incision Repair 731
   4.4. Direct Damage Reversal 731
   4.5. Gratuitous Repair: The Price of Broad Substrate Specificity 732
   4.6. Possible Role of Broad Substrate Specificity in Divergent Evolution of New Repair Activities 733
5. Examples of DNA Repair Enzymes that Exhibit Catalytic Promiscuity 733
   5.1. Exonuclease III/DNase I Superfamily 733
   5.2. Endonuclease IV and the Xylose Isomerase Superfamily 734
   5.3. Bifunctional DNA Glycosylases/Lyases 736
   5.4. Bifunctional dCTP Deaminase/dUTPase 737
   5.5. DNA Cytosine Methyltransferases: DNA Modification and Promiscuous DNA Damage 739
   5.6. Direct Repair (and DNA Damage) Catalyzed by Alkylguanine Alkyltransferases 739
   5.7. Promiscuity of DNA Ligases 741
   5.8. Challenges in Detecting Catalytic Promiscuity 741
6. Mechanistically Diverse DNA Repair Enzyme Superfamilies 741
   6.1. Helix–Hairpin–Helix DNA Glycosylase Superfamily 742
   6.2. Polynucleotide Kinase: Fusion of a 3′-Phosphatase from the Haloacid Dehalogenase Superfamily and a 5′-Kinase from the P-Loop-Containing Nucleotide Hydrolase Superfamily 744
   6.3. AlkB and the α-Ketoglutarate-Dependent Dioxygenase Superfamily 746
   6.4. The DNA Repair Nuclease, Artemis, and the Metallo-β-Lactamase Superfamily 747
7. Changes in Substrate and Reaction Specificity of DNA Repair Enzymes 747
8. Summary: DNA Repair, a Highly Evolvable System for Safeguarding the Genome 748
9. Acknowledgments 748
10. References 748

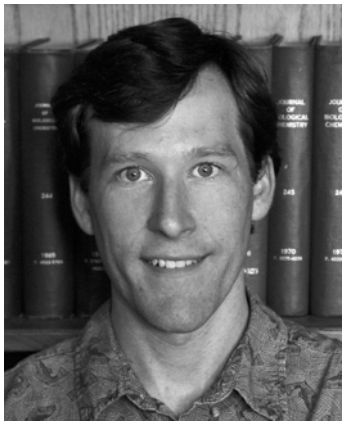
* To whom correspondence should be addressed. Phone: 743-647-5821. Fax: 734-763-4581. E-mail: pjobrien@umich.edu.

1. Introduction

DNA has a remarkably simple chemical structure, yet it encodes an amazing amount of diversity. It is likely that DNA was selected as the genetic molecule of life, at least in part, because of its exceptional stability. Nevertheless, there are many endogenous and exogenous sources of DNA damage. The alterations in the chemical structure of DNA that result from this damage have potentially dire cellular consequences because they can interfere with the normal DNA-templated processes of transcription and DNA replication and result in permanent mutation of the genome. On one hand, this damage-induced mutation can be beneficial because it provides the variation necessary for Darwinian selection. This cycle of variation and selection is the driving force for evolution, the process by which biological diversity is created. On the other hand, too much mutation is detrimental because beneficial sequences can be quickly lost. Therefore, it is not surprising that all cellular life forms and many viruses encode a multitude of proteins that function to repair damaged DNA. These repair pathways are remarkably complex and appear to be highly redundant. Individual cells have multiple pathways available for the repair of a given type of DNA damage and in many cases multiple proteins that can repair the same type of damage even within a single pathway.

How did this complex repertoire of DNA repair pathways evolve? It is generally accepted that mutation rates have changed over evolutionary time. Higher mutation rates would have provided selective advantage for short periods of time, and subsequently lower mutation rates proved to be more advantageous. Indeed mutation rates vary widely between organisms and even between different locations within the genome of a single organism.[1,2] The most obvious way to affect mutation rates is to alter the activity of the DNA repair and replication proteins themselves. Many cycles of selection for different mutation rates and exposure to fluctuating levels of DNA-damaging agents in different organisms and in different environments are likely to have shaped the complex and highly redundant pathways for DNA repair that are present in modern-day organisms. As most repair pathways require multiple gene products, this view of evolution suggests that it has been relatively easy for the function of DNA repair pathways to be altered and for new repair pathways to be assembled. Although the evolution of new catalytic activities plays a central role in creating and

10.1021/cr040481v CCC: $59.00
Published on Web 01/21/2006



Patrick O'Brien was born in Hollywood, California, and grew up in Portland, Oregon. He received his undergraduate degree in Biology and Chemistry from Santa Clara University in 1994 and his Ph.D. in Biochemistry from Stanford University in 2000. His graduate research was conducted with Professor Dan Herschlag, during which time he focused on understanding mechanisms of biological phosphoryl transfer reactions. He became interested in enzyme evolution during this time and will always be fascinated by the ingenuity and versatility of nature's catalysts. Pat was first exposed to the complexities of DNA repair while conducting postdoctoral work with Professor Tom Ellenberger at Harvard Medical School, and he was the recipient of a Ruth L. Kirchenstein postdoctoral fellowship from the National Institutes of Health. He moved to the University of Michigan in 2004 where he is currently an Assistant Professor of Biological Chemistry and Chemical Biology. His research group is currently studying the mechanisms by which enzymes recognize and repair DNA damage.

remodeling these pathways, protein−protein interactions, cellular localization, and regulatory mechanisms are also important factors. There are examples in which enzymes have lost enzymatic activity but have adopted different functions[3,4] and numerous examples of multifunctional proteins in which only one of the functions is catalytic activity. In other cases, a single polypeptide has been found to have different functions in different pathways or in different compartments of a given cell.[5,6] Throughout biology, and particularly in DNA repair, cellular processes are carried out by multiprotein complexes.[7−9] The evolution of these protein−protein interactions has no doubt been a critical component of evolving DNA repair pathways. Furthermore, regulation of these complexes via posttranslational modification provides an additional level of complexity.[10] Although these aspects of evolution are fascinating, this review will focus on the chemistry of the repair reactions themselves and on the processes by which new enzymatic activities have been recruited for DNA repair throughout evolution.

The de novo evolution of enzymes was clearly critical at the early stages of evolution and probably continued to play a role throughout evolution, otherwise all enzymes would share the same three-dimensional structure. However, the de novo creation of an enzyme is expected to be a low probability event because most random polypeptides are not expected to adopt a unique and stable structure, and enzymes require one or more specific substrate-binding pockets and the correct placement of multiple catalytic groups relative to the substrate(s). Divergent evolution from a preexisting enzyme via gene duplication is expected to provide a more favorable pathway to the creation of a new enzymatic activity, because a duplicated gene product already adopts a stable fold and the binding site and/or catalytic groups could be used in a new reaction with only minor changes. If the duplicated enzyme already was multifunctional and able to accept alternative substrates, then this process is expected to be more efficient because single beneficial mutations could confer a selective advantage.[11,12] The results from many studies suggest that the ability to accept alternative substrates that are closely related to the normal substrate (broad substrate specificity) and to catalyze different classes of reactions with the same or different substrates (catalytic promiscuity) are widespread and perhaps a fundamental property of biological catalysis.[11] Furthermore, the results from protein engineering experiments and in vitro evolution suggest that readily accessible pathways exist for creating a new enzyme from a preexisting one. In many cases, single mutations can greatly increase the activity of an enzyme toward a new substrate, even when the new reaction involves a change in catalytic mechanism.[11,13−16]

Consistent with this notion, there is abundant evidence that divergent evolution has played a central role in the evolutionary diversification of enzymatic function. Almost all enzymes have been conserved to some extent during evolution and are closely related to homologous enzymes from other organisms (orthologues).[17−19] These enzymes share the same biological function and catalyze the same biological reaction but oftentimes have significant differences in their ability to bind closely related substrates or inhibitors. Many enzymes also belong to enzyme families that are composed of closely related paralogues (homologous proteins within an organism that evolved from a common ancestor).[17−19] These enzymes catalyze a specific type of chemical transformation but can vary widely in their substrate specificity and often function in distinct biological processes or pathways.[12] Oftentimes these enzyme families are themselves homologous with one or more mechanistically distinct enzyme families that carry out different classes of chemical transformations, and together they constitute mechanistically diverse enzyme superfamilies thought to be related by divergent evolution.[20−26] Three examples of enzyme superfamilies are shown in Figure 1. These enzyme superfamilies are remarkable for the diversity of different reactions that are catalyzed: C−N, O═O, C−S, and P−O bond cleavage in the metallo-β-lactamase superfamily; C−Cl, C−P, and P−O bond cleavage and formation in the haloacid dehalogenase superfamily; C−S and C−N bond cleavage and C−N, C−O, and C−S heterocycle formation in the Fe(II)-dependent dioxygenase superfamily. Although the substrates and reactions vary greatly, key catalytic groups are conserved: two divalent metal ion binding sites in the β-lactamase superfamily; an aspartate nucleophile and a general base in the haloacid dehalogenase superfamily; a single divalent metal ion binding site in the Fe(II)-α-ketoglutarate-dependent enzymes. At least one DNA repair enzyme belongs to each superfamily, with the other enzymes functioning in a wide variety of biological roles. These examples showcase the evolutionary potential of biological catalysis and the probabilistic nature of the evolutionary process.

This review discusses the known evolutionary relationships among DNA repair enzymes and between DNA repair enzymes and other cellular enzymes, focusing on the divergence of enzymatic function. Our understanding of these relationships provides insight into the evolutionary past and current evolutionary potential of contemporary DNA repair processes. Consideration of the evolutionary relationships involving DNA repair enzymes, the multiplicity of DNA repair enzymes available to a given cell, and the current

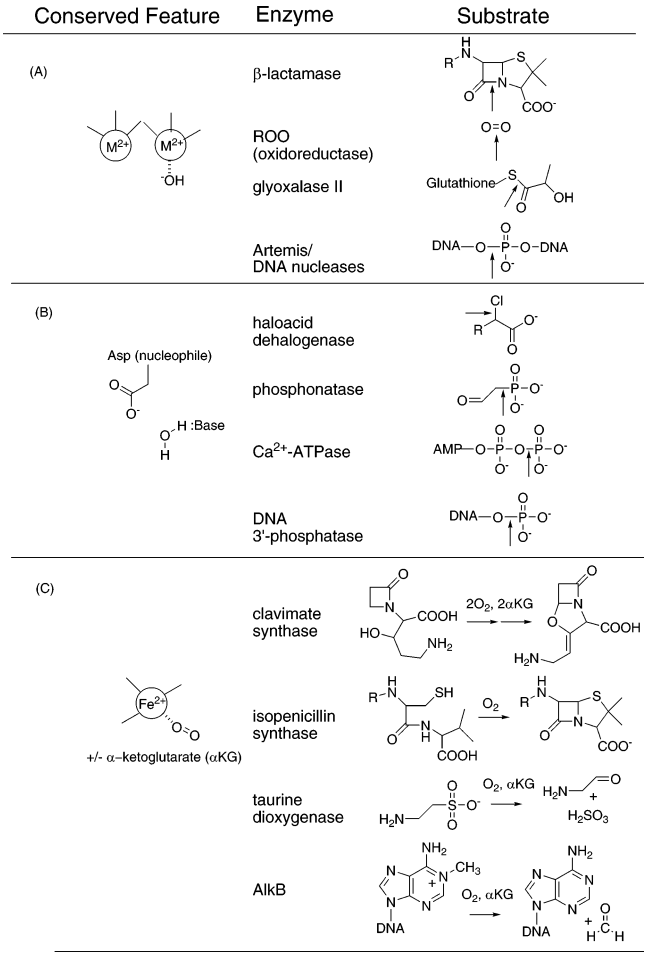


Figure 1. Representative superfamilies that contain DNA repair enzymes. Each protein has structural homology to the other members of the superfamily, and active site features have been conserved. The common mechanistic feature is shown for each superfamily, along with several enzymes and the reactions that they catalyze. Some of the reactions are not shown, but the bond that is broken is indicated with an arrow. (a) The metallo-β-lactamase superfamily members usually bind Zn^{2+}, but in a few cases the identity of the conserved metal ligands is altered to allow binding of Fe^{2+} instead. ROO is rubredoxin/oxygen oxidoreductase. Structures are not yet available for Artemis or any of the related DNA repair nucleases, but these enzymes are predicted to be homologous and to bind divalent metal ions in a similar manner.[290,292,296,329] (b) The haloacid dehalogenase (HAD) superfamily uses a highly conserved aspartate nucleophile. A general base subsequently activates a water molecule in the second step of the reaction to hydrolyze the covalent intermediate.[264,274] (c) The superfamily of Fe(II)/α-ketoglutarate-dependent dioxygenases.[330] Some members of this superfamily use Zn^{2+} instead of Fe^{2+}, and most, but not all, use α-ketoglutarate (αKG) as a cosubstrate. Although no structural information is yet available for the AlkB family, sequence homology and the confirmed Fe^{2+}/αKG-dependent mechanism strongly support this assignment.[139,141,284]

catalytic potential of these enzymes raises the possibility that DNA repair pathways themselves are under selective pressure to be evolvable. The ability to recognize and repair new types of DNA damage could provide a powerful selective force throughout evolution in response to changing intracellular and extracellular environments.

2. Overview of Catalytic Promiscuity and the Evolutionary Diversification of Enzymes

There are numerous examples of modern-day enzymes that appear to have evolved from preexisting enzymes via gene duplication. In the post-genomic era we know of many examples in which such gene duplication has taken place on a massive scale, such as the duplication of most of the *Saccharomyces cerevisiae* genome.[27] Divergent evolution has long been thought to require gene duplication as an initial step to free a gene from the previous functional constraints on its gene product.[28−30] However, once a gene is duplicated random drift will cause an accumulation of mutations in a functionally redundant gene. Because the majority of these mutations are expected to be deleterious to structure and/or function, most duplicated genes are expected to be relatively quickly lost. But, if a duplicated gene's product already had some low level of activity toward a new biological function, then evolutionary pressure has a higher probability of adapting and improving the new activity through Darwinian selection.

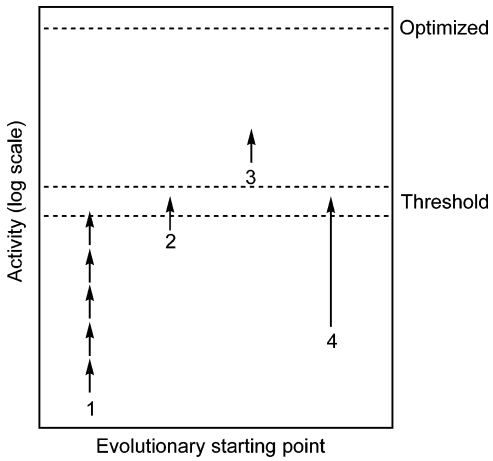


Figure 2. Threshold model for the evolution of a new activity. For any new activity there is a threshold below which the activity does not provide a selective advantage and the gene cannot benefit from Darwinian evolution. The selective pressure depends on the level of the activity, because greater activities can give larger selective advantage, so the threshold is not discrete. Three evolutionary starting points reflect the activities of newly duplicated genes. Very low activity (1) requires many advantageous mutations (arrows) to reach the threshold required for a selective advantage. Activity near (2) or above (3) the threshold can be immediately subject to selective pressure and single beneficial mutations could be selected for. Arrows indicate a typical energetic effect that a point mutation could have on activity, but the actual increase in activity obtained is completely dependent on the position and identity of the mutation. In rare cases, much larger increases might be observed (4), perhaps reflecting a domain fusion that brings together previously optimized catalytic or binding sites. The level of the threshold depends on the genetic background, the intra- and extracellular environment, and the biological function of the new activity. For example, new DNA repair activities are expected to require relatively lower reaction rates because single turnover would be sufficient to repair rare sites of damage. In contrast, a new metabolic or catabolic pathway is expected to require much higher reaction rates to generate sufficient product to affect viability. Adapted with permission from ref 11. Copyright 1999 Elsevier.

A threshold model for selective pressure based upon the biological function (e.g., enzymatic activity) illustrates some of these concepts for divergent evolution of any given gene product (Figure 2). A duplicated gene, or any other gene that is free from functional constraints, has many possible evolutionary pathways that can result in functional changes to the protein that it encodes. In most cases, a duplicated gene's product will not have any new desirable functions, and it would require many mutations to acquire a new function (Figure 2; gene 1). In other cases, a low level of

activity toward the new function could provide sufficient activity or be close to providing the necessary level of activity (Figure 2; genes 2 and 3). Once the activity reaches the threshold for which a selectable advantage is conferred, then subsequent rounds of mutation and selection can improve and ultimately optimize the new activity. Although a low level of activity toward the new function greatly increases the probability of divergent evolution, it does not guarantee that the function can necessarily be optimized. Even very low probability events can and do occur in the course of evolution. For example, domain fusion events, although rare, could provide much larger increases in activity than single point mutations because entire preformed binding/catalytic sites could be brought together (Figure 2; gene 4).[31−33] Nevertheless, the evolutionary potential to create new enzymatic activities is likely to be strongly influenced by the variety and frequency of alternative reactions that are carried out by the cellular complement of enzymes.

It has long been recognized that many enzymes have remarkably broad substrate specificities, and Jensen[12] proposed that the substrate ambiguity of such proteins could provide a starting point for divergent evolution after gene duplication. As our understanding of enzymatic catalysis increases, it has become clear that most if not all enzymes accept alternative substrates and quite commonly can catalyze transformations of remarkably different reactants utilizing distinct mechanistic features. The ability of a single active site to catalyze multiple types of reactions has been referred to as catalytic promiscuity.[11] Active sites abound with potentially catalytic groups, and any given active site contains several of the following groups: metal ions, general acids or bases, hydrogen-bond donors or acceptors, nucleophilic amino acids, bound cofactors. Furthermore, many side chains have versatile catalytic potential. For example, a carboxylate that binds a metal ion for a given reaction could function as a general base to abstract a proton from a substrate in an alternative reaction when the metal is not bound. Thus, it should come as no surprise that there are dozens of well-characterized examples of enzymes that exhibit catalytic promiscuity, utilizing a single active site to catalyze one or more alternative reactions that are substantially distinct from the normal, physiological reaction.[5,11,13] In many cases, single active sites have evolved to catalyze more than one type of chemical reaction at a single active site as part of their normal physiological function (see section 5).

Previous analyses of mechanistically diverse enzyme superfamilies and changes in reaction and substrate specificity have focused largely on metabolic enzymes. The recent evolution of new metabolic pathways in response to man-made chemicals has facilitated the identification of the evolutionary relationships and provides insight into the processes by which enzymes and pathways evolve.[5,34] Given the similar fundamental challenges of specificity and proficiency faced by all enzymes, it seems likely that these paradigms will apply to enzymes in other biological processes, including DNA repair. However, some fundamental features of DNA repair are expected to distinguish it from other metabolic pathways. Most notably, because the substrate is DNA, the source of genetic information, even low levels of the wrong activity or the correct activity toward the wrong sites could have profound effects on viability and genetic stability. Furthermore, once a protein has evolved the ability to bind to DNA nonspecifically, it has an increased likelihood of targeting any DNA sequence with a promiscuous activity. Consideration of these factors suggests that negative selection would operate against many alternative reactions, and this would serve to decrease the potential for divergent evolution. Alternatively, newly evolved enzymes that modify DNA substrates could have provided increased mutation rates during times of rapid change in the evolutionary past, and this could have provided a selective advantage. Perhaps the evolution of some repair activities paralleled the evolution of the organism as a whole so that there was ample time to evolve greater specificity even as the selective pressure to lower mutation rates increased. If so, then future evolution of repair pathways could differ considerably from past evolution.

The current wealth of genomic sequence data, protein structures, and mechanistic studies of DNA repair proteins provides extensive evidence for divergent evolution among proteins involved in DNA repair. This review will cover these evolutionary relationships from the perspective of the enzymologist and further evaluate the functional plasticity of DNA repair enzymes by examining examples of broad substrate specificity and catalytic promiscuity. The available evidence suggests that DNA repair proteins are not obviously different from other enzymes and that catalytic promiscuity and broad substrate specificity are likely to be common features of all enzymes. These features could have contributed to the diversity of DNA repair enzymes that are encoded within a single genome and are likely to influence the future evolution of DNA repair pathways.

3. The Chemical Landscape for DNA Damage and Repair

The early recognition that DNA is a remarkably stable molecule has been largely substantiated, but over the years we have gained a greater appreciation for the many spontaneous reactions that can alter the chemical structure of DNA.[35] Even very slow reactions can cause significant numbers of DNA-damaging events on a biologically relevant time scale given the large size of genomic DNA. Once damaged, the great chemical stability of DNA then acts to thwart the intentions of the DNA repair enzymes that must catalyze the cleavage and formation of these same stable bonds. As minor modifications of even a single atom can change the hydrogen-bonding potential of a base and cause errors during DNA replication or transcription, it is a considerable challenge for these enzymes to discriminate between sites of damage and the vast excess of normal DNA. In this section, the labile bonds in DNA are briefly introduced to present a context for understanding the chemistry behind DNA damage and repair. The reactivity of DNA is compared to other metabolic reactions, and finally the ways in which the chemistry of DNA damage and repair are expected to affect the evolution of DNA repair enzymes are discussed.

Efforts over the past three decades to characterize the intrinsic reactivity of DNA and of model compounds related to DNA have provided a wealth of information about its chemical stability (Figure 3).[35,36] As the phosphodiester backbone is remarkably stable under physiological conditions, it is difficult to measure the spontaneous hydrolysis reaction directly. However, estimates for the stability of the phosphodiester backbone of DNA, based upon the reactivity of the model phosphodiester dimethyl phosphate, suggest a half-life of ~140 000 years under physiological conditions.[37,38] As stable as this bond is, the carbon−carbon bonds of the deoxyribose sugar backbone are expected to be even

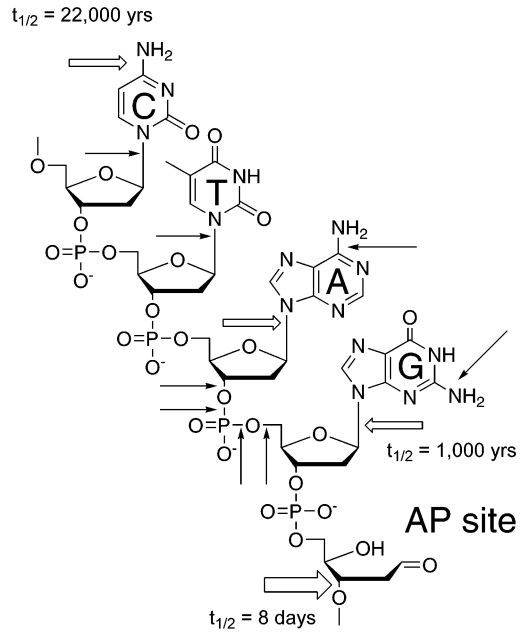


Figure 3. Hydrolytic reactions of DNA. Arrows indicate the most labile bonds in DNA; larger arrows indicate faster rates of spontaneous hydrolysis under physiological conditions. A single strand of a DNA duplex is shown, with the sequence from top to bottom C, T, A, G, and an AP site (resulting from spontaneous depurination).

more resistant to hydrolysis.[36] Under normal conditions the N-glycosidic bond that attaches the nucleobase moieties to the phospho-sugar backbone is the most hydrolytically sensitive (Figure 3). The purine nucleotides are especially susceptible to acid-catalyzed hydrolysis, but even at physiological pH guanine is hydrolytically released with a half-life of ~730 years.[39,40] Although pyrimidine nucleotides are also hydrolyzed, they are ~100-fold more stable than purine nucleotides under physiological conditions.[41,42] Once N-glycosidic bond hydrolysis occurs, the apurinic (AP) site that is produced is exceptionally unstable and undergoes a base-catalyzed elimination reaction to break the DNA backbone with a half-life of ~8 days in neutral solution.[43] Thus, depurination provides the lowest energy pathway to breaking the DNA polymer. Nature appears to have capitalized on this relative weakness of the N-glycosidic bond, because many different types of base damage are repaired by base excision repair that is initiated by enzymatic hydrolytic release of the damaged base (see below).

Nevertheless, the N-glycosidic bonds (and the other, more stable bonds in DNA) are all quite stable relative to many metabolic intermediates.[36,37] For example, spontaneous peptide bond hydrolysis, cytidine deamination, and isomerization of triose phosphate have half-lives of 450 years, 73 years, and 2 days, respectively.[36] The great chemical stability of DNA helps to preserve the integrity of the genetic information encoded by DNA but requires that the enzymes that repair DNA provide larger rate enhancements (rate enhancement = k_{cat}/k_{non}; in which k_{cat} is the enzymatic rate constant for multiple turnover and k_{non} is the nonenzymatic rate constant) to achieve a given rate constant. It appears to be generally true that enzymatic rate enhancements vary widely and they are largely dictated by the rate of the nonenzymatic reactions because enzymes appear to have been optimized for enzymatic turnover on a similar time scale.[36] This has previously been considered for mainly metabolic enzymes, but DNA repair pathways operate under a different kinetic regimen than primary metabolic pathways. Metabolic pathways usually require a high flux and the synthesis of relatively large amounts of products. In contrast, DNA damage is a rare event, and therefore many repair enzymes are only required to turn over a few times per cell cycle. It is critical that sites of damage be found, but there may not be a strong selective pressure to maximize the rate of reaction. If an enzyme can locate and tightly bind to a site of DNA damage, then this could prevent access by the DNA replication and RNA transcription machineries. Consistent with these notions, repair enzymes typically have k_{cat} values of $0.1-1\ s^{-1}$, and this differs from metabolic enzymes that often have rate constants that are 2−3 orders of magnitude larger.[36,37] Uracil DNA glycosylase (UNG) from *Escherichia coli* appears to be exceptional in its ability to turnover its substrate, with a single-turnover rate constant of more than $100\ s^{-1}$.[44] This fast reaction rate (and exceptional rate enhancement) likely reflects the abundance of this lesion relative to other types of base damage. As discussed below, UNG has an unusually narrow substrate specificity, and the tight uracil-specific pocket effectively excludes the other natural nucleobases. In contrast, most other DNA repair enzymes offer more modest rate enhancements. Presumably these lower rate enhancements reflect the ability to recognize a broader range of substrates.

In addition to DNA fragmentation, a large number of spontaneous reactions are known to alter the chemical structure of the nucleobases within DNA. Although most of these changes do not appreciably change the stability of the DNA, they can change the base-pairing properties and are therefore potentially mutagenic (Figure 4). The most common of these reactions is the hydrolytic deamination of the exocyclic amino groups of the nucleobases, of which the most unstable is the 4-amino group of C with a half-life of ~22 000 years in duplex DNA.[35,45−47] Deamination of C or m^5C, to generate either a dU·G or T·G mismatch, would be a mutagenic event if not repaired (Figure 4). A wide variety of additional chemical reactions involving the DNA bases are possible, including alkylation, oxidation, and damage caused by ionizing radiation or UV light (Figure 4), and the frequency of these events are dependent upon the level to which the DNA is exposed.[35] These base lesions pose a formidable challenge to the DNA repair machinery, because the damaged bases resemble the undamaged bases more closely than they resemble each other. This might suggest that DNA repair enzymes require exquisite sensitivity to distinguish damaged DNA from undamaged DNA. However, contrary to these initial expectations, many DNA repair enzymes exhibit broad substrate specificity, suggesting that other factors need to be considered.

To understand the landscape for the evolution of new DNA repair activities, it is important to recognize that very few biochemical reactions are required to accomplish repair: N-glycosidic bond hydrolysis, AP lyase, phosphodiester hydrolysis and formation, and phosphomonoester hydrolysis reactions can be used to completely repair any single nucleotide lesion or di-adduct (see sections 4.1 and 4.2 describing excision repair pathways). Although some repair enzymes catalyze different classes of reactions, such as the direct transfer and oxidative removal of alkyl groups catalyzed by the AGT and AlkB enzymes and the direct reversal of UV damage catalyzed by photolyase, alternative reactions such as those of excision repair could in principle be used (see section 4.4). The low chemical complexity of

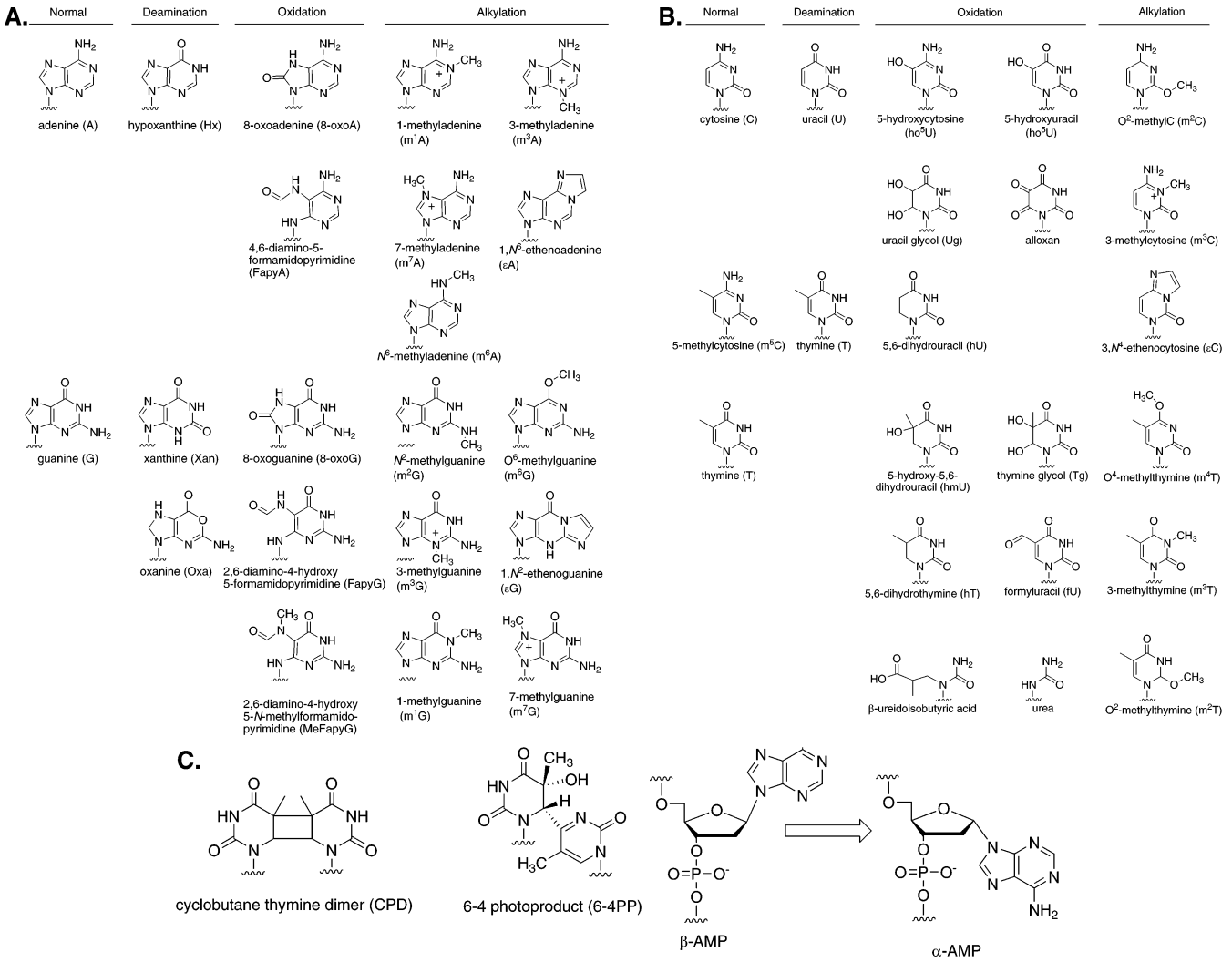


Figure 4. Normal and damaged bases found in DNA and their abbreviations. The base moieties for the normal nucleotides and for damaged (A) purine and (B) pyrimidine nucleotides resulting from deamination, oxidation, and alkylation. For alkylation by monofunctional agents, only methyl adducts are shown, but reactions with larger alkyl groups also occur, and many of these lesions are substrates for the same repair proteins that recognize the methyl adducts. Much larger alkylation adducts (not shown) are typically repaired by the nucleotide excision repair pathway. (C) Structure of pyrimidine dimers formed by exposure to UV light and conversion of a normal β-anomeric nucleotide into an α-anomeric nucleotide by exposure to ionizing radiation.

DNA suggests that simple changes in substrate specificity could provide abundant possibilities for the evolution of new repair activities. For example, the overwhelmingly dominant reaction catalyzed by DNA repair enzymes is the formation or cleavage of the phosphodiester bond. Phosphodiester bond hydrolysis is required for all excision repair pathways, including mismatch repair, nucleotide excision repair (NER), and base excision repair (BER). Phosphodiester bond formation is catalyzed by DNA polymerases and DNA ligases for the resynthesis of DNA, and topoisomerases and recombinases catalyze both the cleavage and the ligation of phosphodiester bonds to alter DNA topology and to carry out recombinational repair. Indeed, the functional interrelatedness of topoisomerases, ligases, and nucleases has long been appreciated, as topoisomerases are known to function as ligases and nucleases on different substrates.[48−51] Such broad substrate specificity could provide a favorable evolutionary starting point toward a variety of biological functions that utilize the same or similar chemistry (Figure 2). Other DNA repair pathways are considered below to address whether broad substrate specificity is a relatively common feature of DNA repair enzymes. If broad specificity is common, then this increases the number of possible genes that could be adapted for new enzymatic functions. The combination of large rate enhancements with broad substrate specificity is expected to result in considerable catalytic promiscuity and thereby expand evolutionary potential.

The physical properties of genomic DNA, namely, its large size and its homogeneous structure that sequesters the information-containing nucleobases, present additional challenges that must be met by any DNA repair pathway. Given the vast excess of undamaged DNA in the genome, it is not a trivial task to find rare sites of DNA damage. This task is further complicated by the fact that many damaged nucleobases involve subtle changes to the chemical structure of DNA, such as the addition of a methyl group or elimination of an amino group (Figure 4). Once detected, protein-driven conformational changes are typically required to gain access to the damaged site because the base-pairing and stacking interactions hide much of the DNA surface within the DNA duplex. A ubiquitous theme in the enzymatic modification/repair of duplex DNA substrates is the use of nucleotide flipping or base flipping, the process whereby a nucleotide is exposed by rotating it out of the duplex, and this ability

appears to have independently evolved many times. This strategy is critical for allowing the active site to access the bonds that are being transformed. The problem of finding one site of DNA damage among the sea of undamaged sites appears to be addressed in part by restricting the search to a two-dimensional one.[52−60] The ability to slide along DNA (linear diffusion) and sample multiple sites is expected to greatly decrease the search time relative to a three-dimensional search.[61,62] A number of DNA repair and modification enzymes and transcription factors have been shown to be capable of linear diffusion in vitro, suggesting that this is a general property of DNA-binding proteins.[63−67] The weak association with DNA that enables linear diffusion appears to be largely mediated by electrostatic interactions, because this process is very sensitive to ionic strength. Perhaps the best characterized example of linear diffusion by a DNA repair enzyme is the ability of T4 pyrimidine dimer glycosylase (PDG) to remove multiple lesions from a single molecule of DNA without dissociation.[66] Mutations have been generated that compromise the ability of PDG to diffuse along DNA, thereby changing the mechanism of repair from a processive one to a distributive one, but do not affect the kinetics of excision once a substrate is bound. Cells containing these distributive mutants are more sensitive to UV damage, providing compelling evidence that linear diffusion enhances the efficiency of repair in vivo.[68,69] Similar evidence has been obtained with EcoRV, demonstrating that linear diffusion is important for the in vivo restriction of bacteriophage DNA.[70]

These common requirements for enzyme-induced DNA conformational changes that provide access to the substrate, such as nucleotide flipping, and for employment of a two-dimensional search of genomic DNA that increases the efficiency of lesion detection provide additional barriers to the evolution of new DNA repair enzymes. This suggests that a repair enzyme with the ability to diffuse along DNA and flip out damaged nucleotides has a better chance of being adapted by divergent evolution than another protein that required simultaneous evolution of these properties along with the ability to catalyze the required chemical transformation.

4. Overview of DNA Repair Pathways: A Common Theme of Enzymes that have Broad Substrate Specificities

Damage to the nucleobases within DNA constitutes the most common type of DNA damage, accounting for thousands of damaged bases per human cell every day.[35] These spontaneous reactions include depurination, deamination, oxidation, alkylation, and UV-induced damage and create a structurally diverse array of base lesions (Figure 4). Most of these lesions are repaired by base excision repair (BER), a subset of these, particularly the bulky alkylation adducts, are repaired by nucleotide excision repair (NER), and a few specific lesions are repaired by direct damage reversal. A specialized DNA repair pathway, mismatch repair, exists for the repair of nucleotides that are misincorporated during DNA replication. A detailed description of each of these repair pathways is outside of the scope of this review, but a brief introduction is given below to emphasize the similarities and differences between these DNA repair pathways. The reader is referred to recent review articles and to the other reviews in this thematic issue on "DNA Damage and Repair". Remarkably each of these repair pathways is characterized by enzymatic activities that exhibit broad substrate specificity, with most enzymes recognizing multiple types of DNA damage. Although several repair pathways will be discussed, the focus will be on BER because there are many well-characterized examples of BER enzymes with broad substrate specificity, and there are several superfamilies of BER enzymes in which individual proteins have diverged to evolve different substrate specificity or even different reaction specificity (see below, Tables 1, 2, and 4).

4.1. Base Excision Repair

DNA glycosylases constitute one of the largest classes of repair enzymes that scan genomic DNA in search of damaged bases. DNA glycosylases initiate BER by flipping out the damaged nucleotide and catalyzing the cleavage of the N-glycosidic bond to release the damaged base. There are many different DNA glycosylases that have been discovered, and their substrate specificity varies such that different glycosylases are responsible for recognizing different lesions.[71−73] However, in many instances there is more than one glycosylase in a given cell that can recognize a particular lesion. Given the differences in size and intrinsic N-glycosidic bond reactivity of pyrimidine and purine nucleotides (Figure 3), it is not surprising that most glycosylases prefer either purine or pyrimidine substrates. Indeed, purine-specific but not pyrimidine-specific DNA glycosylases appear to use general acid catalysis to stabilize the purine leaving group (Figure 5).[73] Both purine- and pyrimidine-specific glycosylases are expected to benefit from positioning and activating the nucleophile and from providing a favorable electrostatic environment for the stabilization of the similar oxacarbenium-ion-like transition state that is expected in each case (Figure 5), and remarkably, several DNA glycosylases are able to efficiently catalyze the excision of both purine- and pyrimidine-derived base lesions.[73] DNA glycosylases are commonly classified as either monofunctional or bifunctional enzymes. The monofunctional glycosylases use water as the nucleophile to attack the anomeric carbon of the damaged nucleotide (Figure 5), and the bifunctional DNA glycosylases use an active site amine moiety to displace the damaged base and generate a Schiff base covalent enzyme intermediate (Figure 6). Subsequently the covalent intermediate is processed to generate either an AP-site-containing DNA product (hydrolysis), a 5′-deoxyribosephosphate (dRP) and a 3′-phosphate (β-elimination; Figure 6B), or a 3′-phosphate, 5′-phosphate, and oxo-2-pentenal (β,δ-elimination; Figure 6A). Both the AP site and the nicked DNA intermediates are potentially mutagenic or cytotoxic lesions, because DNA replication can result in either misincorporation or a double-stranded break. Therefore, it is critical that these intermediates be further processed and the repair pathway be completed, perhaps assisted by in vivo coordination of individual repair activities.[74] Although the specific pathway can vary considerably according to the specific repair enzymes that are used (Figure 7), all require DNA polymerization by a repair polymerase such as polymerase β and ligation by a DNA ligase. In the case of the monofunctional DNA glycosylases, the action of an AP endonuclease and an AP lyase are required to generate the free 3′-OH required for DNA synthesis and the 5′-phosphate required for DNA ligation.

Traditionally DNA repair glycosylases have been distinguished as having either narrow or broad substrate specificity. The first DNA glycosylase to be characterized

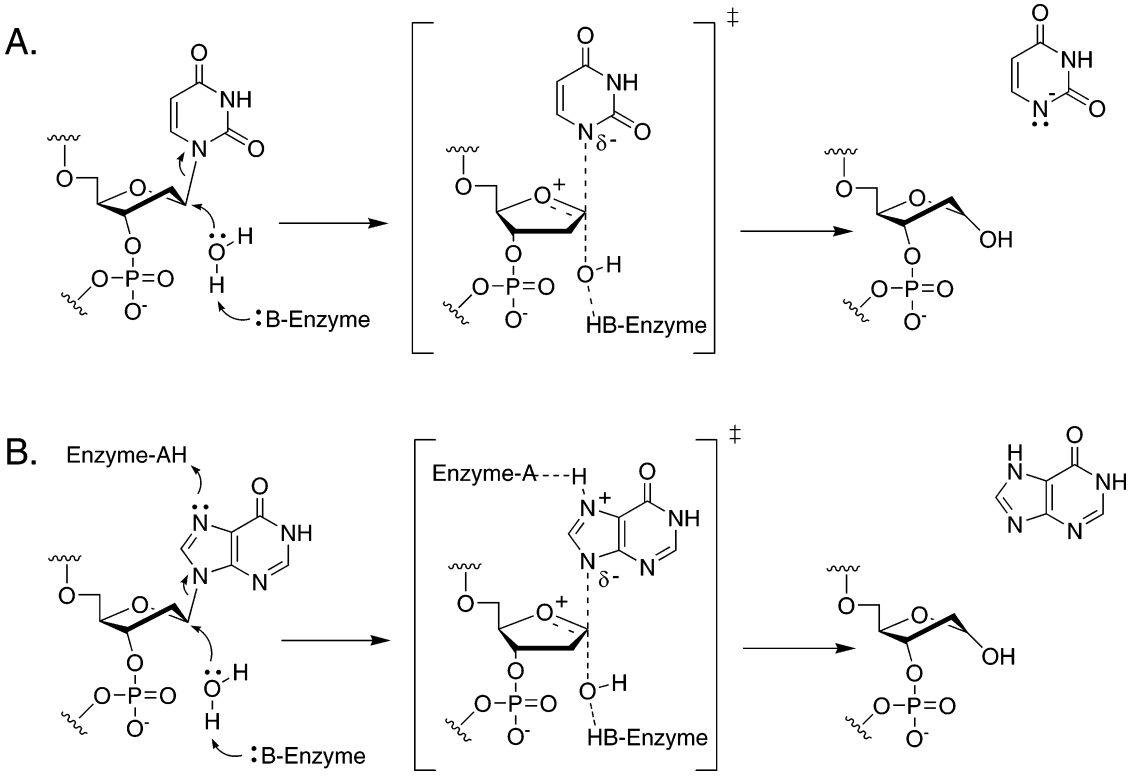


Figure 5. Comparison of the catalytic mechanisms for monofunctional DNA glycosylases acting on (A) pyrimidine and (B) purine substrates. The nucleotide substrate, proposed transition state (in brackets), free nucleobase, and AP site products are shown. Nonenzymatic and enzymatic hydrolysis of pyrimidine and purine nucleotides appear to proceed via highly dissociative transition states in which there is an accumulation of positive charge on the ribose group.[73] Most monofunctional glycosylases appear to utilize a carboxylate as a general base to deprotonate the nucleophilic water molecule. Even in a dissociative transition state, modest rate enhancement can be expected from activation of the nucleophile, and greater rate enhancement can be achieved by positioning of the nucleophile relative to the substrate. Additional stabilization could be achieved by preferential binding to the expected planar conformation of the sugar in the transition state and/or by electrostatic stabilization of the accumulating positive charge. Purine leaving groups can be stabilized by protonation either prior to or concurrent with cleavage of the N-glycosidic bond, and at least some DNA glycosylases appear to use this strategy to excise purine bases.[73,194,331−333] Additional differences in the binding pockets are expected given the different sizes and shapes of pyrimidine and purine bases.

was uracil DNA glycosylase encoded by the UNG gene of *E. coli*,[75,76] and orthologues have subsequently been found in prokaryotes, eukaryotes, and viruses. The glycosylase reaction mechanism has been thoroughly dissected, and there are several structures of the enzyme in complex with DNA.[73] UNG has exquisite specificity for dU in DNA, and it recognizes the base in either single or double-stranded DNA via specific hydrogen bonds (Asn123 in the *E. coli* enzyme accepts a hydrogen bond from N^3 and donates a hydrogen bond to the 4-oxo group of U). A phenylalanine side chain (Phe70, *E. coli*) makes a close contact with the bound substrate and is expected to cause a steric clash with the 5-methyl substituent of a bound T or m^5C base. Unnatural analogues such as 5-fluorouracil are accepted as substrates, but they closely mimic the natural substrate.[77,78] Many paralogues of UNG have been identified, and five distinct families of uracil DNA glycosylase (UDG) enzymes have been described that constitute a superfamily of structurally homologous enzymes that have diverged from a common ancestor.[79,80] In contrast to the UNG family, other families of UDGs are characterized by remarkably broad substrate specificity. As their name implies, all of the UDG superfamily members recognize uracil but additionally can recognize damaged purine and pyrimidine bases that have been deaminated, oxidized, or even alkylated. The substrate specificity of each family is distinct from one another, and the specificity of enzymes within a single family can differ considerably (Table 1; family 2). Some enzymes recognize uracil in either single-stranded or double-stranded DNA (families 1, 3, and 4) whereas the enzymes of families 2 and 5 have a strong preference for double-stranded DNA. Consistent with the requirement for nucleotide flipping, all of the members of this superfamily preferentially excise mismatched substrates in double-stranded DNA, with the double-strand-specific enzymes showing a preference for T·G or U·G, the expected product from the deamination of m^5C or C (Table 1).[54,59,78]

Although UNG is often cited as a prototypical highly specific DNA repair glycosylase, many closely related enzymes have quite broad substrate specificity. A survey of other known DNA glycosylases suggests that the narrow specificity of UNG is the exception and that most enzymes resemble the broadly specific UDG family members. For example, the monofunctional human 3-methyladenine DNA glycosylase AAG was originally identified as the enzyme responsible for the excision of cytotoxic 3-methyladenine lesions. The crystal structure revealed a unique fold that is different from other known glycosylases, including the 3-methyladenine DNA glycosylase AlkA from *E. coli*.[53] Subsequent experiments have established that it is the primary DNA glycosylase for the excision of more than a dozen different lesions, including the oxidative lesions hypoxanthine, xanthine, and oxanine and the lesions formed from alkylation such as 1,N^6-ethenoadenine, 7-methylguanine, 3-methylguanine, 7-methyladenine, and a variety of larger purine 3- and 7-alkyl adducts in human cells (Figure

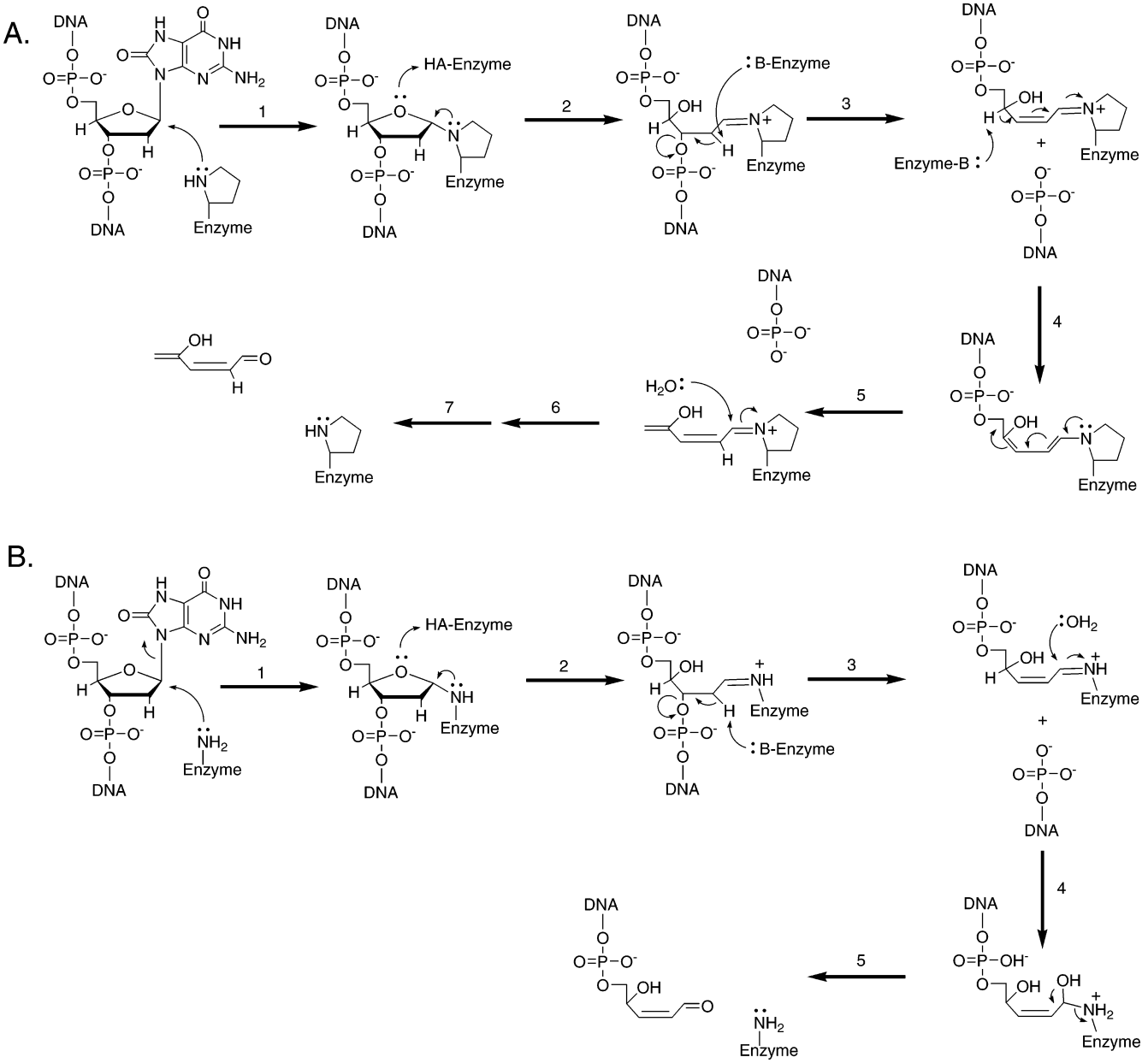


Figure 6. Two enzymatic mechanisms for the initiation of repair of 8-oxoguanine lesions. (A) The catalytic mechanism of FPG, a multifunctional DNA glycosylase that catalyzes N-glycosidic bond cleavage and β,δ-elimination.[83] The C1′ of the target nucleotide is attacked by the amino-terminal proline of FPG to excise the base and generate a covalent intermediate (1). General acid-catalyzed ring-opening (2) sets up the general base-catalyzed abstraction of a C2′ hydrogen (3) to drive β-elimination of the 3′ DNA strand. A second general base allows abstraction of the C4′ hydrogen (4) and δ-elimination of the 5′ strand (5). Subsequently, the enzyme catalyzes the hydrolysis of the Schiff base intermediate to complete the catalytic cycle (6 and 7). (B) Catalytic mechanism of OGG1, a bifunctional DNA glycosylase/AP lyase. A conserved active site lysine initiates nucleophilic attack at C1′ of the target nucleotide to form a covalent intermediate (1). The enzyme may use general acid catalysis to shift the equilibrium toward the ring-opened form of the sugar (2). The subsequent general base-catalyzed abstraction of a C2′ hydrogen allows β-elimination of the 3′ DNA strand (3). Alternatively, hydrolysis of this intermediate can release the AP-site-containing DNA (not shown; see the discussion of MutY, a related HhH DNA glycosylase in section 5.3). After β-elimination, hydrolysis of the Schiff base covalent intermediate generates the products of the reaction (4 and 5).

4).[72,81,82] The family of bifunctional glycosylases related to FPG is also characterized by an extremely broad substrate range, excising a wide variety of oxidative purine and pyrimidine lesions, including ring-opened formamidopyrimidine derivatives of guanine and adenine, 8-oxoguanine, formyluracil, thymine glycol, and dihydrouracil (Figure 4).[83]

The remaining DNA repair glycosylases belong to an extensive superfamily of related enzymes that contain a conserved helix−hairpin−helix (HhH) DNA-binding motif.[84−87] As is the case for the UDG superfamily, the HhH superfamily members have diverged to accept a wide variety of very different substrates, and the enzymes with narrow substrate specificity seem to be the exception rather than the rule (Table 2). Both Tag and MagIII are highly specific for 3-methyladenine, but both enzymes also excise 3-methylguanine, and MagIII also removes 7-methylguanine.[82,88,89] As N^3- and N^7-alkylated purines bear a positive charge (Figure 4), alkylation mimics the effect of protonation in acid-catalyzed depurination, and the spontaneous hydrolysis of these lesions is very rapid.[90] Therefore, the rate enhancements exhibited by MagIII and Tag are quite small and may preclude these enzymes from having reasonable excision

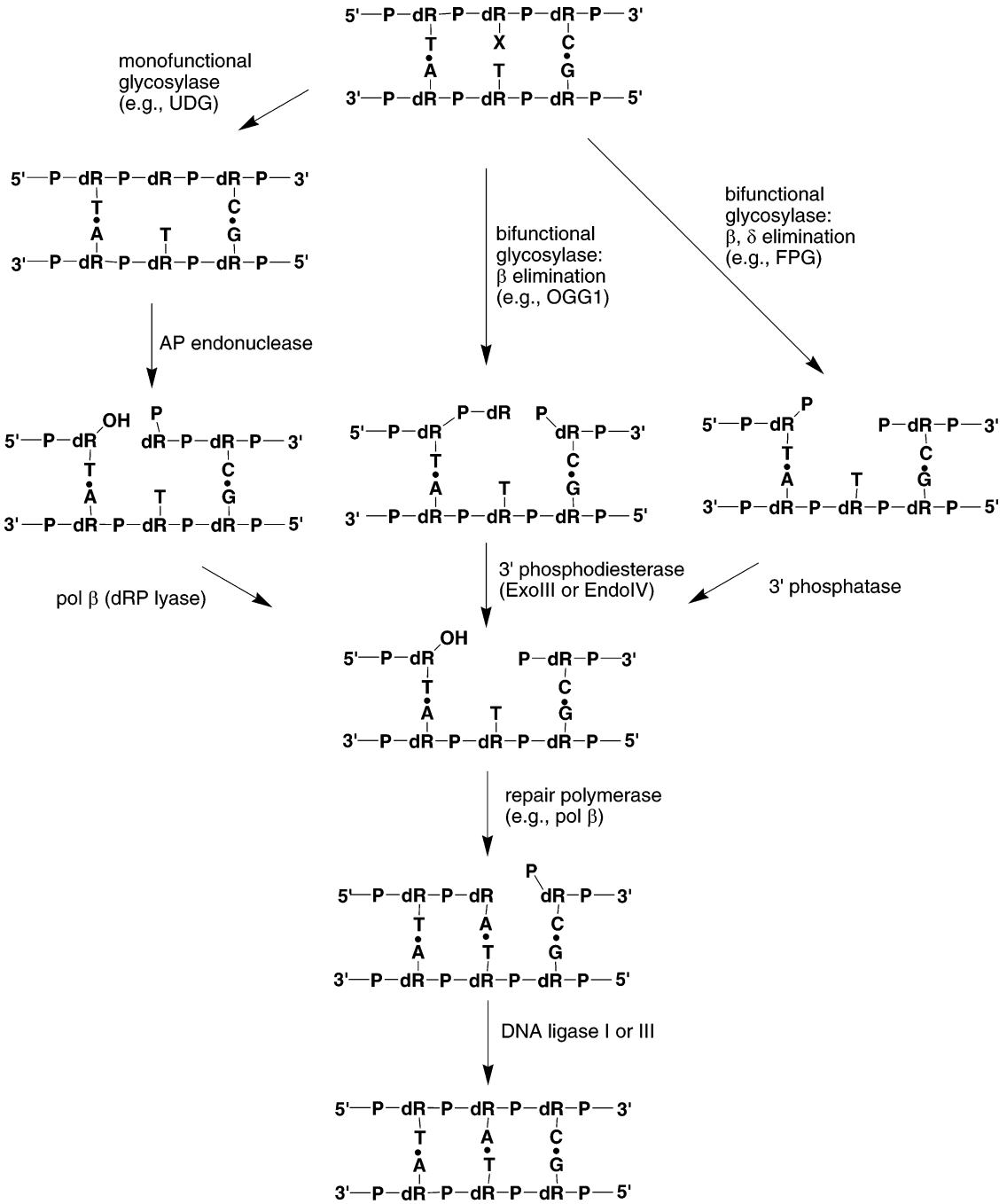


Figure 7. Base excision repair can follow one of several distinct biochemical pathways. If the base lesion (X) is recognized by a monofunctional DNA glycosylase (left pathway), then the subsequent actions of an AP endonuclease, 5′-deoxyribosephosphate (dRP) lyase, DNA polymerase, and DNA ligase are required to complete the repair pathway. If a bifunctional DNA glycosylase initiates repair by excising the base and catalyzing β-elimination (center pathway), then the AP endonuclease and dRP lyase are replaced by a 3′-exonuclease that generates a free 3′-OH. If a bifunctional enzyme catalyzes β,δ-elimination (right pathway), then a 3′-phosphatase is required. Remarkably, in many organisms the 3′-phosphatase reaction can be performed by the 3′-exonuclease (e.g., Xth in *E. coli* and APN1 in *S. cerevisiae*; Table 3). These pathways reflect the minimal biochemical pathways, and additional complexity is known to occur as this is not always a linear pathway. For example, repair synthesis by a DNA polymerase can cause strand displacement, in which case the endonuclease activity of FEN-1 is required to process the resulting 5′-overhang prior to ligation.

rates for stable lesions.[73,89] Remarkably, the HhH superfamily members contain both monofunctional and bifunctional enzymes, and these functional differences will be discussed below (see section 6.1).

Genetic and biochemical methods have been effective in discovering and characterizing the primary repair activities of DNA repair glycosylases, but activity toward alternative substrates can be difficult to detect and can often only be characterized by systematically examining the activity of each glycosylase toward each possible substrate. In many cases, the activity of additional enzymes in vivo can mask the more modest contribution of an enzyme toward its alternative substrates. A particularly powerful approach for the identification of alternative substrates involves the damage of DNA in situ, employing gas chromatography/mass spectrometry to isolate and identify the base lesions released by a purified enzyme.[91−95] Given sufficient levels of damage and sufficient repair activity, this approach has the potential to identify

Table 1. Substrate Specificity of Members of the Uracil DNA Glycosylase Superfamily[a]

enzyme	primary substrate(s)	additional substrates[b]	preferred DNA
	Narrow Specificity		
UNG (family 1)[77,78]	**U**·purine, **F^{5}U**·purine		ss/ds
Tth UDG (family 4)[314]	U	*c*	ss/ds
	Broad Specificity		
MUG (family 2)[78,315]	**U**·G, **ϵC**·G, **F^{5}U**·G, **ho^{5}h$_2$U**·G	**U**·A, **ϵA**·G, **Hx**·G, **hmU**·G, **T**·G	ds
TDG (family 2)[78,316,317]	**T**·G, **U**·G, **ϵC**·G, **Tg**·G	**T**·C/T, **F^{5}U**·G, Hx	ds
SMUG1 (family 3)[77,318−322]	**U**·purine, **hmU**·purine, **F^{5}U**·purine	**ho^{5}C**, **ho^{5}h$_2$U**, **f^{5}U**, **ϵC**	ss/ds
Pa UDGb (family 5)[317]	U, hmU, F^{5}U, ϵC, Hx		ds

[a] Many of these glycosylases prefer mismatched substrates. In the cases for which a strong preference for the identity of the opposing base is known, the base pair is shown with the excised base highlighted in bold-faced type. The abbreviations and structures of the bases are shown in Figure 4. [b] Excision of these lesions has been reported but at ≤1% the rate at which the best substrate is excised. [c] This enzyme has recently been discovered, and there are not yet extensive reports on additional substrates.

Table 2. Substrate and Reaction Specificity of the Helix−Hairpin−Helix Glycosylases[a]

enzyme	primary substrate	additional substrates	lyase?
	Narrow Specificity		
Tag	m^{3}A	m^{3}G	no
MagIII	m^{3}A	m^{3}G, m^{7}G, m^{7}A	no
PDG (EndoV)	thymine dimers	FapyA, FapyG	yes
MutY	**A**·8-oxoG	**Hx**·8-oxoG, **A**·G/C, **G**·8-oxoG	yes
OGG1	**8-oxoG**·C	FapyG, MeGapyG	yes
AGOG	**8-oxoG**·(T, C, G, A, ss)	*b*	yes
	Broad Specificity		
AlkA/Mag1	m^{3}A, m^{2}T, m^{2}C	m^{7}G, m^{7}A, Hx, ϵA, ϵC	no
EndoIII (Nth)	Tg, urea, ho^{5}C, hU, FapyG	ho^{5}U, MeFapyG	yes
MBD4	**T**·G, Tg	U, **T**·m^{6}G	no
MIG	**U**·G, **T**·G	**A**·G	no

[a] Most of these glycosylases prefer mismatched substrates. In the cases for which the enzyme is known to have a strong preference for the identity of the opposing base the base pair is shown with the substrate base highlighted in bold-faced type. The abbreviations and structures are shown in Figure 4. [b] This enzyme has recently been discovered, and additional substrates have not yet been tested.

novel substrates that are repaired via excision repair pathways. Even low levels of activity toward alternative substrates that may not be significant in the biological sense provide an opportunity for diversification of enzymatic function in the evolutionary sense. Contrary to initial expectations, it appears that broad substrate specificity is a relatively common feature of DNA repair glycosylases. Although most glycosylases appear to have a single preferred substrate and exhibit substantially lower rate enhancements for the other substrates (Tables 1 and 2), even low levels of activity can be effective in repairing DNA damage because most DNA-damaging events are infrequent.

The DNA repair intermediates generated by the action of monofunctional and some bifunctional DNA glycosylases must be further processed by AP endonucleases and/or phosphodiesterases to generate the 3′-hydroxyl required for replacement synthesis by a DNA polymerase and the 5′-phosphate required by DNA ligase (Figure 7). The AP endonucleases and repair phosphodiesterases have remarkably broad substrate specificity and many also exhibit catalytic promiscuity, processing multiple types of damaged DNA and repair intermediates. Several superfamilies of phosphodiesterases/phosphatases have been identified, and they are discussed below (Tables 3 and 4; sections 5.1 and 5.2). In addition to their roles in base excision repair, some of these enzymes make additional contributions to DNA repair capacity by acting as endonucleases at sites of DNA damage to initiate nucleotide incision repair.[96−100]

Exquisite specificity for Watson−Crick base pairs is a hallmark of the DNA polymerases involved in DNA replication, and this ability to recognize all four natural base pairs and discriminate against possible mismatches is critical to faithful DNA replication.[101] Polymerases are thought to achieve this specificity largely via shape recognition of the normal base pairs and steric exclusion of mismatches.[102] Consistent with this notion, a wide variety of synthetic base pairs can be tolerated,[102] and base lesions that interfere with normal hydrogen bonding cause stalling or misincorporation.[103] These principles appear to be generally true for replicative polymerases. However, several additional families of DNA polymerases have been identified that appear to function in DNA repair. These enzymes are characterized by greatly decreased fidelity for normal base pairs, and they have been shown to be able to bypass a wide variety of lesions by incorporating an appropriate or inappropriate nucleotide. These error-bypass polymerases belong to the same superfamily as replicative polymerases and share many mechanistic and structural features, but they evolved to have distinct patterns of substrate specificity.[101,104,105]

Similar to the replicative DNA polymerases, DNA ligases also show high specificity for the correct DNA ends, and the presence of damaged nucleotides, nucleotide gaps, or base mismatches all greatly decreases the catalytic efficiency for ligation.[106−112] However, human DNA ligase I ligates the 3′-OH of an RNA strand to the 5′-phosphate of a DNA strand with the same efficiency as it ligates a nicked DNA substrate,[112,113] and T4 DNA ligase is capable of joining RNA or DNA oligonucleotides in a variety of combinations.[106] Several DNA ligases are able to ligate DNA across a 1 nucleotide gap, albeit with decreased efficiency.[109,114] These observations reveal that the biological imperative for replicative DNA polymerases and ligases to exhibit high fidelity does not prevent them from accepting alternative substrates.

4.2. Nucleotide Excision and Mismatch Repair Pathways

Bulky DNA lesions, including UV-induced intrastrand di-adducts, are typically repaired via the nucleotide excision repair (NER) pathway. This repair occurs via remarkably similar mechanisms in prokaryotes and eukaryotes, given that the individual proteins are not conserved and appear to have independently evolved via convergent evolution.[115,116] In both prokaryotes and eukaryotes NER involves low-specificity recognition of the DNA damage, ATP-dependent unwinding of the DNA by one or more helicases that allow for increased specificity via kinetic proofreading, and removal of a damage-containing oligonucleotide by endonucleolytic incisions on both sides of the damaged nucleotide.[115−119] In prokaryotes this pathway is initiated by UvrA, UvrB, and UvrC, and in vitro reconstitution has established that these proteins are sufficient to carry out the dual excision of an oligonucleotide that is 12−13 nucleotides in length.[120] Subsequent action of UvrD, DNA polymerase I, and DNA ligase completes the repair. In eukaryotes, RPA, XPA, and XPC are responsible for recognizing the DNA damage, TFIIH unwinds the local DNA duplex in an ATP-dependent manner, and the incision nucleases XPG and XPF·ERCC1 are recruited to excise the damage-containing oligonucleotide that can be anywhere from 24 to 32 nucleotides in length.[115,116,121,122] After excision, the normal replicative machinery of RFC, PCNA, and polymerase δ/ϵ fill in the gap, and the nick is sealed by DNA ligase I. Given their independent origins, it is not surprising that there are mechanistic differences between prokaryotic and eukaryotic NER, but rather it is remarkable that there is so much in common.

The mismatch repair (MMR) pathway is responsible for the repair of single base mismatches or small nucleotide insertions/deletions, usually resulting from errors during DNA synthesis.[123−125] In prokaryotes, mismatched DNA is recognized and tightly bound by the MutS homodimer, a MutL homodimer is recruited, and this complex translocates away from the mismatch in an ATP-dependent manner until a hemimethylated GATC site is located, at which point the MutH endonuclease can be recruited and the newly synthesized (unmethylated) strand is nicked. UvrD helicase is employed to unwind the DNA, and the damaged strand is exonucleolytically degraded by exonuclease I if the nick was created on the 3′-side of the mismatch and RecJ or exonuclease VII if the nick was created on the 5′-side of the mismatch. The resulting gap (∼1 kb) is filled in by DNA polymerase III, and the nick is sealed by DNA ligase. In eukaryotes the MutS homodimer is replaced with one of two MutS-like heterodimers, MutSα and MutSβ, that have distinct damage specificities. There does not appear to be a eukaryotic homologue of MutH, and it is not clear how the newly synthesized strand is identified although it is possible that nicks associated with DNA replication are used as sites to load the exonucleases and helicases.

The NER and MMR pathways share some general mechanistic features, despite the fact that different complexes of proteins are involved and that the method of damage removal differs (oligonucleotide excision in the case of NER and exonucleolytic digestion in the case of MMR). Both of these pathways are initiated by endonucleolytic cleavage that is directed by a multiprotein damage recognition complex, both require ATP hydrolysis, and both involve excision and resynthesis of a relatively large region of undamaged DNA. Both pathways have an inherent broad substrate specificity built in, because damage recognition is physically separated from damage excision. Damage recognition relies on a change in the local conformation of damaged DNA, and the subsequent incision does not occur at the site of damage so the reaction can be optimized for normal DNA. The convergent evolution of these functionally similar repair pathways with broad substrate ranges suggests that such a strategy has been advantageous.

4.3. Nucleotide Incision Repair

In addition to the excision repair pathways described above, most cells have a distinct conserved pathway for repairing damaged nucleotides that has been called nucleotide incision repair.[97,126] This pathway uses some of the same enzymes that function in BER, but it differs in that the endonucleases are responsible for the initial recognition of DNA damage. Once bound to a site of damage they catalyze the hydrolysis of the phosphodiester backbone on the 5′-side of the damaged nucleotide. Both the APE1/Xth and the Nfo superfamilies that constitute the major endonucleases in mammalian and prokaryotic cells, respectively, have been shown to recognize a variety of oxidative lesions, bulky alkyl adducts, and some unusual sugar modifications such as α-anomeric nucleotides (Figure 4C).[97−100,127−129] This broad substrate specificity that has independently evolved in two families of functionally homologous, but structurally distinct enzymes, is discussed in more detail below (see sections 5.1 and 5.2). In many cases, these same damaged bases that are the substrates of nucleotide incision repair can also be recognized and excised by DNA glycosylases. For example, in human cells oxidative lesions such as dihydrothymidine and dihydrodeoxyuridine can be recognized by either APE1, the major endonuclease, or by TDG, a mismatch-specific monofunctional DNA glycosylase.[99]

4.4. Direct Damage Reversal

There are several known families of proteins that perform direct reversal of DNA damage without the synthesis of new DNA (direct reversal of damage). One of these families, the photolyase family, catalyzes the photoreversal of UV-induced intrastrand di-adducts and has narrow substrate specificity, repairing specific photoproducts. Two other families, the alkylguanine alkyltransferase (AGT) family and the AlkB family of DNA demethylases, function in the repair of alkylation damage and have broader substrate specificity, repairing a variety of different base adducts. Although both AGT and AlkB convert alkylated bases directly to the parent unmodified bases, they do so by very different mechanisms. DNA ligases directly repair single-strand and double-strand breaks that have a 5′-phosphate and a 3′-hydroxyl.

Photolyases repair UV-induced pyrimidine−pyrimidine cross-links such as a cyclobutane pyrimidine dimer (CPD) or a 6−4 photoproduct (6−4PP) by directly reversing the cross-link (Figure 4C).[119,130,131] This unusual mechanism relies on two chromophores, a flavin and a methenyltetrahydrofolate, to harvest blue light and it involves electron transfer from the flavin to break the DNA cross-link. Structural information is available for the interaction of photolyase with a CPD lesion, and the results from X-ray crystallography and NMR studies provide insight into the mechanism that this enzyme uses to flip out the sterically hindered base cross-link and place it in close proximity to

the flavin cofactor.[58,131,132] Given the steric restraints imposed by these unusual di-adducts and by the electron-transfer reaction that is catalyzed, it is not surprising that these enzymes repair either CPD or 6−4PP lesions but not both. Although humans have two proteins that are closely related to DNA photolyase, they do not have a functional photolyase. Instead these homologous proteins utilize the conserved cofactors as photoreceptors for blue light and appear to be involved in resetting the circadian clock.[130]

The biological reversal of DNA alkylation follows at least two distinct mechanisms, direct transfer via nucleophilic attack on the alkyl group and oxidation and elimination of the alkyl group. Human AGT and *E. coli* Ada are the best characterized examples of the alkyltransferases, and their primary biological function is the repair of O^6-alkylguanine lesions, because these lesions are highly mutagenic and they are not readily recognized by other repair pathways.[133,134] However, other O-alkylated bases, such as O^4-alkylthymine can also be recognized to a varying extent. These proteins bind to alkylated DNA and transfer the alkyl adduct from the nucleobase to an active site cysteine. Each protein can only carry out the reaction once, because the free active site cysteine cannot be regenerated from the alkylated form of the protein. However, removal of the cysteine by mutation allows the mutant Ada to use methanethiol as an exogenous nucleophile and to act in a truly enzymatic manner.[135] There is evidence that a similar direct transfer mechanism has evolved independently, because the amino terminus of the Ada protein contains a second domain that is structurally distinct and that recognizes alkyl phosphate triesters in DNA.[136] As is the case for photolyase, the amino-terminal domain of Ada is highly specific for a single type of DNA damage, catalyzing the transfer of an alkyl group from the S-stereoisomer to an active site cysteine. Structures of the enzyme in complex with both DNA substrate and product reveal the origin of this stereospecificity and suggest that it would be difficult for a single protein to detect both isomers.[136]

The family of enzymes related to AlkB of *E. coli* also catalyzes the removal of alkyl adducts from either single-stranded or double-stranded DNA without excision of the alkylated base. These enzymes can accept 1-alkyl adducts of A and G and 3-alkyl adducts of C and T, recently have been shown to reverse the etheno adducts ϵA and ϵC, and thus have a remarkably broad substrate specificity, reminiscent of the broadly specific DNA glycosylases.[134,137−144] This broad substrate specificity is particularly remarkable in light of the unusual reaction mechanism that involves the hydroxylation of the lesion's alkyl group (Figure 1C; see section 6.3) and suggests that narrow substrate specificity is not an absolute prerequisite of repair proteins that catalyze direct repair.

DNA ligases use the energy of either ATP or NAD+ hydrolysis to directly rejoin a DNA strand break.[106,111,145] Eukaryotic enzymes typically use ATP, whereas different prokaryotic enzymes utilize either ATP or NAD+. Irrespective of the cofactor, all ligases form a covalent enzyme−AMP intermediate via an active site lysine. In the second step of the reaction the AMP group is transferred to the 5′-phosphate to form a 5′−5′-phosphoanhydride linkage. Finally, attack by the 3′-OH displaces AMP to restore the phosphodiester bond. It is remarkable that these enzymes have evolved to carry out three distinct chemical transformations in a single active site. Eukaryotic cells typically have at least two DNA ligases, commonly referred to as DNA ligase I, which specializes in ligation of single-strand breaks (nicks), and DNA ligase IV, which specializes in ligation of double-strand breaks. All excision and incision repair mechanisms ultimately require ligation to complete the repair pathway. Although ligases are completely dependent upon a 5′-phosphate and 3′-OH and most can discriminate against mismatches at the DNA ends, several alternative reactions have been observed (see section 5.7).

4.5. Gratuitous Repair: The Price of Broad Substrate Specificity

Many DNA lesions differ from normal DNA by only one or a few atoms, and thus repair enzymes face the difficult task of distinguishing sites of damage from the vast excess of undamaged DNA. This has led to the notion that repair enzymes must have extremely narrow substrate specificity. However, the accumulated data regarding the specificity of repair enzymes suggests that this is the exception rather than the rule. As discussed above, DNA glycosylases show remarkably broad substrate ranges, and even the highly specific enzymes have some capacity for accepting alternative substrates (Tables 1 and 2). Most repair enzymes can recognize multiple substrates, and this property is at odds with exclusion of normal DNA. In fact, there is considerable evidence that DNA repair enzymes are not infallible and their specificity is not absolute. Several BER and NER enzymes have been shown to act on normal undamaged DNA,[40,146−149] and this phenomena has been termed gratuitous repair.[115,116,147,150] The existence of gratuitous repair may provide selective pressure to keep catalytic power in check. Larger rate enhancements would result in greater levels of gratuitous repair by broadly specific enzymes, and there may be no biological requirement for very fast rates of repair. Although gratuitous repair is not necessarily mutagenic and completion of the repair is likely to restore the original sequence, it is energetically wasteful and it does provide additional opportunity for mistakes to be made. Furthermore, gratuitous repair could target sites of damage in an undirected way. For example, a glycosylase capable of excising normal bases could cause mutations by removing the correct base from a damage-induced mismatch, leaving the repair polymerase to use the damaged base as a template. It is likely that the catalytic efficiency of repair enzymes reflects a compromise between sufficiently fast repair of damaged DNA and an acceptable level of gratuitous repair of undamaged DNA. Conversely, some level of gratuitous repair is likely to be an unavoidable consequence of a broadly specific repair system, and the widespread existence of such systems in nature suggests that this detrimental effect is counteracted by the efficiency and perhaps the adaptability of enabling a single repair enzyme to repair multiple types of damage.

The gratuitous repair catalyzed by the BER glycosylases, AlkA and AAG, provides insight into the possible biological and evolutionary significance of gratuitous repair. Presumably the broad substrate specificity of these enzymes is advantageous because alkylation damage results in a great structural diversity of adducts (Figure 4), and many enzymes with narrow specificity would be required to repair damage caused by a single alkylating agent. However, a consequence of this broad specificity is that normal bases are also excised to some extent.[40,146,148,149] Remarkably, *E. coli* AlkA catalyzes the excision of normal purines with the same rate enhance-

ment as for the excision of methylated purines, although the absolute rate constants favor excision of the damaged bases.[146,149] In *E. coli*, the adaption to alkylation response allows cells to detect alkylation damage and induce expression of the enzymes involved in the repair of alkylation damage, AlkA, AlkB, and Ada.[136] The high level of these repair proteins provides greater capacity to withstand exogenous alkylating agents. It is tempting to speculate that *E. coli* keeps the gratuitous repair catalyzed by AlkA in check by repressing its expression until its repair activity is needed. Consistent with this idea, the overexpression of either AlkA or its yeast homologue, Mag1, results in increased rates of mutation in both *E. coli* and yeast.[146,151] Interestingly, the human functional homologue AAG exhibits lower levels of gratuitous repair and is constitutively expressed. Thus, the factors that enter into establishing a baseline for acceptable gratuitous repair are complex. The actual threshold at which gratuitous repair could be selected against depends not only on the rate of initiation of gratuitous repair but also on the abundance and regulation of the repair protein, the accessibility of the protein to DNA, the fidelity of the repair pathway, and the full complement of other repair processes available in the cell. At some level gratuitous repair is an unavoidable consequence of a broadly specific repair system, and the widespread existence of such systems in nature suggests that this detrimental effect is counteracted by the greater economy or perhaps the adaptability of enabling repair of multiple types of damage by a single enzyme.

4.6. Possible Role of Broad Substrate Specificity in Divergent Evolution of New Repair Activities

It is evident that DNA repair enzymes have a great capacity for accepting alternative substrates that differ considerably in their size, net charge, and hydrogen-bonding ability. The pervasiveness of broad specificity suggests that this strategy may provide a selective advantage over an enzyme with a narrow specificity. Broad specificity allows a single enzyme to protect against multiple types of DNA damage but more importantly could provide some capacity to repair new types of damage and thereby provide a starting point for the evolution of a new repair activity. Indeed, repair enzymes with broad substrate specificity are able to recognize and repair a variety of lesions that natural evolution has not yet selected for, such as 5-fluorouracil[78,152] and 3′-azido-3′-deoxythymidine (AZT).[153] Even a very low level of activity toward a new type of DNA damage is expected to provide a favorable starting point for divergent evolution, because even single mutations can provide large changes in specificity. Indeed, the results from rational protein engineering and in vitro evolution suggest that mutations affecting specificity are relatively common (see section 7 and Table 5).

For enzymes that bind to and catalyze transformations of DNA, it is expected that many of the features of the reaction with the physiological substrate(s) will advance activity toward other substrates. For example, the presence of one or more DNA-binding sites and a mechanism for stabilizing an alternative conformation of DNA would each provide considerable advantage in binding of substrate. In addition, many repair enzymes appear to be able to detect changes in the regular structure of DNA, and hence new sources of damage that destabilize the DNA helix might be expected to elicit a similar binding response. This has long been recognized for the damage sensors that are involved in NER and MMR, and it also appears to be true for many BER enzymes. For example, many DNA glycosylases preferentially excise substrates from mismatched base pairs, and others can bind mismatched nucleotides that are not substrates.[40,54,59,60,78,149,154−158]

Broad substrate specificity, such as that exhibited by DNA repair enzymes, provides a relatively high probability pathway after gene duplication for the evolution of enzymes that catalyze new reactions that are mechanistically quite similar to already existing biological reactions. Catalytic promiscuity, as discussed below, expands the number of possible gene duplications that have the potential to evolve activity toward a new reaction (e.g., a new phosphatase could come from an already existing phosphatase, or it could come from any number of other enzymes that are not phosphatases but have latent or existing promiscuous phosphatase activity). Additionally, catalytic promiscuity increases the opportunity for the expansion of current catalytic potential with new types of chemical transformations and new catalytic mechanisms for already existing biological reactions.

5. Examples of DNA Repair Enzymes that Exhibit Catalytic Promiscuity

To highlight the possible role of catalytic promiscuity in the divergent evolution of DNA repair enzyme superfamilies, the individual examples are broken down into two sections. In the first section, examples of DNA repair enzymes that exhibit catalytic promiscuity are presented (Table 3), and the possible significance of this promiscuity in the diversification of the superfamily is discussed. In section 6, examples of functionally diverse enzyme superfamilies that are not necessarily known to exhibit catalytic promiscuity are presented (Table 4). In these cases, catalytic promiscuity by an ancestral enzyme could have facilitated the evolutionary diversification of enzymatic function even though promiscuity may no longer be detected. A few examples from each table have been selected for further discussion to illustrate the different ways that catalytic promiscuity can be manifested and the types of divergent evolution that have occurred.

5.1. Exonuclease III/DNase I Superfamily

Exonuclease III (Xth) from *E. coli* was one of the first enzymes found to carry out two distinct types of chemical reactions at a single active site, catalyzing the hydrolysis of both phosphomonoester and phosphodiester substrates.[159] As its name implies, it is an exonuclease capable of catalyzing the hydrolytic release of single 5′-phosphonucleotides from the 3′-end of double-stranded DNA (Figure 8). The exonuclease activity has a remarkably broad substrate specificity, and in addition to normal nucleotides, Xth also catalyzes the removal of damaged nucleotides and sugars, such as 3′-deoxyribose-5′-phosphate and 3′-phosphoglycolate that can result from free radical damage of DNA.[160−163] Xth also has a robust AP endonuclease activity, cleaving the DNA phosphodiester backbone 5′ of the AP site to generate a free 3′-hydroxyl and 5′-deoxyribosyl phosphate group, and it is the major AP endonuclease in *E. coli*.[164,165] Xth has a damage-specific endonuclease activity at other damaged sites, most notably at fragmented bases such as urea *N*-glycosides (Figure 4).[96,166] Xth has also been shown to have an RNaseH-like activity, preferentially cleaving the RNA strand of a DNA·RNA hybrid.[167,168] Although this RNaseH-like activity is conserved among some of the other proteins of the

Table 3. DNA Repair Enzymes Known to Exhibit Catalytic Promiscuity

enzyme	normal reaction	promiscuous reaction
AlkB	oxidative dealkylation of alkylated bases (C—N and C—O cleavage)	catalyzes the oxidation of Trp to form a blue chromophore[323]
AGT	direct transfer of an alkyl group from alkylated DNA bases (C—O cleavage)	1. reacts with bifunctional electrophiles, forms cross-link to DNA[219,221,324] 2. covalent modification by aldehydes[325] and a variety of electrophiles[217]
cytosine methyltransferases	methyl transfer (C—C bond formation)	hydrolytic deamination of C and m^5C (C—N cleavage)[202−204]
dCTP deaminase/dUTPase	1. phosphodiester hydrolysis (P—O; hydrolysis of α−β phosphoanhydride bond) 2. deamination of dCTP (C—N bond hydrolysis)	bifunctional enzyme[199,200]
DNA ligase (T4)	phosphodiester bond formation (P—O)	AP lyase activity (C—O cleavage)[232]
DNA ligases	phosphodiester bond formation (P—O)	covalent modification by pyridoxal phosphate via active site lysine[326]
EndoIV (Nfo), APN1, APN2	1. phosphodiester hydrolysis (P—O) 2. phosphomonoester hydrolysis (P—O)	bifunctional enzymes[163,242,327,328]
ExoIII (Xth), DNase I, APE1	1. phosphodiester hydrolysis (P—O) 2. phosphomonoester hydrolysis (P—O)	bifunctional enzymes[159,163]
FPG (MutM), EndoVIII (Nei), NEIL1, NEIL2	1. N-glycosidic bond hydrolysis (C—N cleavage) 2. lyase; β,δ-elimination (C—O cleavage)	bifunctional enzymes[72,73]
G3PD	glyceraldehyde 3-phosphate dehydrogenase (oxidative phosphorylation: C—O formation)	uracil DNA glycosylase (C—N cleavage)[239,240]
HhH glycosylases: OGG1, EndoIII, PDG (EndoV), (MutY)	1. N-glycosidic bond hydrolysis (C—N cleavage) 2. Lyase; β-elimination cleavage of DNA backbone (C—O cleavage)	bifunctional enzymes[72,73]

exonuclease III superfamily, the physiological significance of this activity is not known. In the biological sense, each of these biochemical activities serves a distinct function in dealing with different types of damaged DNA: DNA 3′−5′-exonuclease for resection of DNA ends at strand breaks, DNA endonuclease for cleaving at sites of damage, 3′-phosphoglycolate exonuclease and 3′-dRP exonuclease for processing damaged DNA ends, and DNA-directed RNA endonuclease for removing ribonucleotides from DNA. Although these reactions have important differences, all are classified as phosphodiester hydrolysis, and it is envisioned that each of these different polynucleotide substrates can be accommodated in the active site in a similar manner, with the same catalytic groups providing the same or very similar catalytic mechanism. However, in addition to these phosphodiesterase reactions, Xth also has a vigorous phosphatase activity releasing 3′-phosphates from double-stranded DNA to generate free 3′-OH ends. As phosphomonoesters and phosphodiesters differ in size, charge, and transition state structure, it was unexpected that a single active site could catalyze both types of reactions.[169−171] Several types of DNA damage can result in the formation of 3′-phosphates, including β,δ-elimination catalyzed by bifunctional DNA glycosylases and radiolytic cleavage of DNA (reviewed in ref 163).

Crystal structures of *E. coli* exonuclease III,[172] bovine DNase I,[173] and human APE1[174−176] revealed that these enzymes belong to a superfamily of nucleases that are likely to have evolved via divergent evolution from a common ancestor. Many of these enzymes have been extensively characterized and shown to also have broad substrate specificity and catalytic promiscuity (Table 3). In addition to these broadly specific exonucleases, the endonuclease domain of LINE elements (sequence-specific endonucleases),[177−180] the catalytic domain of structure-specific endonucleases including XPG and flap endonuclease I (FEN-1),[181−185] and sphingomyelin-specific phosphodiesterases[178,186] are also closely related (Table 4). These latter families of enzymes are not known to exhibit phosphatase activity but share conserved catalytic groups with the exonuclease-III-related nucleases and catalyze phosphodiester bond cleavage via a similar divalent metal-dependent reaction mechanism. A conserved metal binding motif[178,187] and extensive structural homology[188] strongly suggest that a family of inosine-5-phosphatases also belongs to this superfamily (Table 4). These phosphomonoester hydrolases are not known to catalyze phosphodiester hydrolysis, but they appear to have mechanistic similarities with the exonucleases. Thus, the exonuclease III superfamily of enzymes has diverged to recognize different substrates and to catalyze different types of reactions: phosphodiester hydrolysis of phospholipids, phosphomonoester hydrolysis of inositol polyphosphates, or both phosphomonoester and phosphodiester hydrolysis of DNA substrates. This divergent evolution is likely to have been facilitated by the catalytic promiscuity of a common ancestor of this superfamily.

5.2. Endonuclease IV and the Xylose Isomerase Superfamily

Convergent evolution has given rise to a functionally similar but structurally distinct family of broadly specific phosphodiesterases/phosphatases typified by *E. coli* endonuclease IV (Tables 3 and 4). Endonuclease IV and its homologues APN1/APN2 in *S. cerivisiae* exhibit phosphodi-

Table 4. Mechanistically and Functionally Diverse DNA Repair Superfamilies[a]

superfamily name	enzyme(s)	reaction	bond type
α-ketoglutarate-dependent dioxygenases (cupins)	**AlkB, ABH2, ABH3**	direct reversal of DNA alkylation	C—C
	clavaminate synthase	hydroxylation, ring formation, desaturation	C—O, C═C
	isopenicillin synthase	ring formation	C—N, C—S
	taurine dioxygenase	oxidative elimination of sulfite	C—O, C—S
DNA/RNA polymerases	**DNA Pol I family**	high-fidelity DNA synthesis	P—O
	Y-family lesion bypass	lower fidelity DNA synthesis	P—O
	reverse transcriptase	RNA-templated DNA synthesis	P—O
	T7 RNA polymerase	DNA-templated RNA synthesis	P—O
	RNA-dependent RNA Pol	RNA-templated RNA synthesis	P—O
DNA ligase	**DNA ligase (ATP-dependent)**	adenylation, ligation of DNA	P—O
	DNA ligase (NAD^+-dependent)	adenylation, ligation of DNA	P—O
	RNA ligase	adenylation, ligation of RNA	P—O
	mRNA capping enzyme	guanylation of mRNA	P—O
DNaseI metallo-phosphoesterases	**DNase I**	exonuclease, endonuclease, phosphatase	P—O
	ExoIII	exonuclease, endonuclease, phosphatase	P—O
	APE1	exonuclease, endonuclease, phosphatase	P—O
	L1 endonuclease (human)	sequence-specific endonuclease	P—O
	TRAS1 endonuclease	sequence-specific endonuclease	P—O
	sphingomyelinase	phosphodiesterase	P—O
	inositol polyphosphate-5-phosphatase	phosphatase	P—O
dUTPase/dCTP deaminase	**dCTP deaminase/dUTPase**	deamination of dCTP, dUTPase	C—N, P—O
	dCTP deaminase	deamination of dCTP	C—N
	dUTPase	dUTP pyrophosphatase	P—O
FPG glycosylases/lyases	**FPG/MutM**	DNA glycosylase/lyase	C—N, C—O
	Nei, NEIL1, NEIL2, NEIL3	DNA glycosylase/lyase	C—N, C—O
HAD/phosphohydrolases	haloacid dehalogenase (HAD)	dehalogenation	C—Cl, C—Br
	3′-phosphatase (Tpp1)	DNA phosphatase	P—O
	3′-phosphatase of PNKP	DNA phosphatase	P—O
	5′−3′-deoxyribonucleotidase	dUTP hydrolysis	P—O
	phosphonoacetaldehyde hydrolase	hydrolysis of phosphonate	C—P
	phosphoserine phosphatase	hydrolysis of phosphoserine	P—O
	β-phosphoglucomutase	mutase	P—O
	RNA Pol II CTP phosphatase	protein phosphatase	P—O
	P-type ATPases	ATPase	P—O
HhH DNA glycosylases	**EndoIII, PDG (EndoV), MutY,**	N-glycosylase, AP lyase	C—N, C—O
	OGG1, AlkA, MBD4, MIG, Tag	N-glycosylase	C—N
Hsp90 (GHKL ATPases/kinases)	Hsp90	ATPase, chaperone for protein folding	P—O
	DNA gyrase	ATPase, rearrange DNA topology	P—O
	MutL, PMS2	ATPases involved in mismatch repair	P—O
	CheA-like histidine kinases	uses ATP to phosphorylate histidine	P—O, P—N
metallo-β-lactamase	Zn-dependent β-lactamases	hydrolysis of β-lactam antibiotic	C—C
	Artemis, DNA repair nucleases	phosphodiesterase	P—O
	glyoxalase II	thiolesterase	C—S
	rubredoxin/oxidoreductase	reduction of molecular oxygen	O—O
nucleotidyl transferases	**DNA polymerase β**	DNA-templated DNA repair synthesis	P—O
	DNA polymerase λ	DNA-templated DNA repair synthesis	P—O
	terminal dNMP transferase	nontemplated DNA synthesis	P—O
	poly(A) polymerase	nontemplated RNA synthesis	P—O
	kanamycin nucleotidyltransferase	NMP transfer to kanamycin	P—O
	RelA/SpoT	ppGpp synthesis	P—O
Nudix hydrolases	**MutT**	8-oxoGTP pyrophosphatase, 8-oxoGDP	P—O
	ADP−ribose pyrophosphatase	phosphatase	P—O
	isopentenyl diphosphate isomerase	isomerization of double bond	C—C
P-loop nucleoside triphosphate hydrolases	**kinase doshain, PNKP**	phosphorylates 5′-OH of DNA	P—O
	Cam phosphotransferase	phosphorylates chloramphenicol antibiotic	P—O
	adenylate kinase	transfers phosphate group to AMP	P—O
	PAPS sulfotransferase	transfers sulfate group	S—O
	G-proteins	GTPase, diverse signaling functions	P—O
	motor proteins	ATPase, generation of mechanical force	P—O
	helicases	ATPase, unwinding of DNA or RNA	P—O
	RecA ATPase	ATPase, recombinational repair	P—O
	MutS ATPase	ATPase, mismatch repair	P—O
phospholipase D	phospholipase D	phosphodiester hydrolysis	P—O
	Tdp1	phosphodiester hydrolysis of covalent DNA	P—O
	bacterial nucleases	protein phosphodiester hydrolysis	P—O
photolyase	**DNA photolyase**	light-induced repair of pyrimidine dimers	C—C
	cryptochromes	light-induced signaling	
purple acid phosphatases	purple acid phosphatases	phosphatase (small molecule substrates)	P—O
	Mre11	DNA repair nuclease	P—O
	exonuclease SbcD	DNA repair nuclease	P—O
	calcineurin	protein phosphatase	P—O

Table 4 (Continued)

superfamily name	enzyme(s)	reaction	bond type
restriction endonucleases	restriction endonucleases class I/II	site-specific endonuclease	P—O
	MutH	endonuclease, mismatch repair	P—O
	EndoI (Holliday junction resolvase)	structure-specific endonuclease	P—O
	XPF nuclease	structure-specific nuclease (NER)	P—O
	VSR endonuclease	repair endonuclease	P—O
ribonuclease H	**ribonuclease H**	RNA hydrolysis, RNA—DNA hybrid	P—O
	flap endonuclease I (FEN-1)	structure-specific endonuclease	P—O
	MutS	ATPase	P—O
	retroviral integrase	DNA integration (phosphotransesterification)	P—O
	Mu transposase	integration	P—O
	DNAQ/exonucleases	3′—5′-exonuclease	P—O
	RuvC/RuvX	Holliday junction resolvases	P—O
SAM-dependent methyltransferases	**cytosine methyltransferases**	m^5C synthesis in DNA	C—C
	adenosine methyltransferases	m^6A synthesis in DNA	C—N
	thymine methyltransferases	m^4T synthesis in DNA	C—O
	CheR	methylation, two component signaling	C—O
	catechol *O*-methyltransferase	methylation of catechols	C—O
	histone Lys methyltransferases	methylation of histones	C—N
Toprim	**topoisomerase IA, II**	phosphodiester transesterification	P—O
	DNAG	primase	P—O
	old family nuclease	nuclease	P—O
xylose isomerase	xylose isomerase	sugar isomerase	C—C
	EndoIV, APN1, APN2	endonuclease, exonuclease, phosphatase	P—O

[a] DNA repair enzymes are indicated in bold-faced type. Superfamily names refer to the founding member (first to be structurally characterized) or to the family name used in the SCOP database.[254]

esterase activity as endonucleases at AP sites and as 3′—5′-exonucleases at single-strand breaks. They can also recognize damaged bases such as α-anomeric nucleotides and catalyze endonucleolytic cleavage to initiate base incision repair.[100,189] Furthermore, these enzymes exhibit catalytic promiscuity by catalyzing the hydrolysis of 3′-phosphates, a phosphomonoesterase activity (Table 3). Endonuclease IV adopts an α_8/β_8 TIM barrel structure and contains two divalent metal ion binding sites that are conserved across this superfamily.[56]

A family of sugar isomerases that are related to xylose isomerase appears to be distantly related to the endonuclease IV family, as they adopt very similar structures and have conserved active site residues.[56,190,191] The C—C isomerization reactions catalyzed by these enzymes appear to be very different from the P—O cleavage reactions catalyzed by the nucleases, but both activities require divalent metal ion cofactors. Neither family of enzymes is known to have promiscuous activity toward the other reaction, but it will be interesting to learn whether additional mechanistic features have been conserved between the isomerases and the endonucleases of this superfamily.

5.3. Bifunctional DNA Glycosylases/Lyases

Two distinct superfamilies of DNA glycosylases are capable of acting as bifunctional DNA glycosylases/lyases, those belonging to the helix—hairpin—helix superfamily (HhH) and those belonging to the FPG/MutM superfamily (Tables 3 and 4). There are distinct differences in their catalytic mechanisms, but the enzymes of both superfamilies use active site amines as nucleophiles to attack C1′ of the damaged nucleotide and displace a damaged nucleobase (Figure 6). The resulting covalent Schiff base intermediate can serve as an electron sink for either β-elimination in the HhH bifunctional glycosylases or β,δ-elimination in the FPG-related enzymes. This catalytic promiscuity that provides for both N-glycosidic (C—N) bond cleavage and β-elimination (C—O cleavage) at a single active site allows a single repair enzyme to carry out two sequential steps in the BER pathway (Figure 7).

FPG is known to exhibit relatively broad substrate specificity, recognizing damaged purine and pyrimidine nucleotides in DNA. The associated lyase activity also allows it to recognize AP sites as substrates. The amino-terminal proline acts as the nucleophile in the N-glycosidic bond cleavage step to form the initial covalent enzyme—DNA intermediate, and a universally conserved glutamic acid has been proposed to function as a general acid to protonate O4′ and favor opening of the sugar ring (Figure 6A).[83] Although the identity of the general bases are not yet known with certainty, two general bases are expected to be involved in the subsequent elimination reactions to abstract protons from C2′ and C4′. Presumably the ability to carry out two elimination reactions is advantageous, because all known members of this superfamily conserve the essential features of this mechanism.

In contrast, the HhH superfamily of DNA glycosylases includes both monofunctional and bifunctional DNA glycosylases. OGG1 is one of the most extensively studied bifunctional enzymes from this superfamily. Instead of the amino terminus of the protein, OGG1 uses an active site lysine to attack the damaged nucleotide and release the damaged base. The Schiff base covalent intermediate that is formed can subsequently undergo β-elimination (Figure 6B). Apparently OGG1 and the other bifunctional HhH glycosylases lack the second general base required for abstraction of the second proton, and so the β-elimination product is released via hydrolysis. MutY is another particularly interesting example, because it shares features of the monofunctional glycosylases and the bifunctional glycosylases.[72,73,192—194] MutY has an active site aspartic acid (Asp138) that acts as a general acid catalyst to protonate (presumably N^7) and to stabilize the leaving group in the N-glycosidic bond cleavage reaction. It also has an active site glutamate (Glu37) that appears to act as a general base to activate a nucleophilic water molecule. This mechanism closely matches that of other monofunctional glycosylases and results in the formation of the same AP-site-containing DNA intermediate. However, MutY has two active site lysine side chains (Lys20

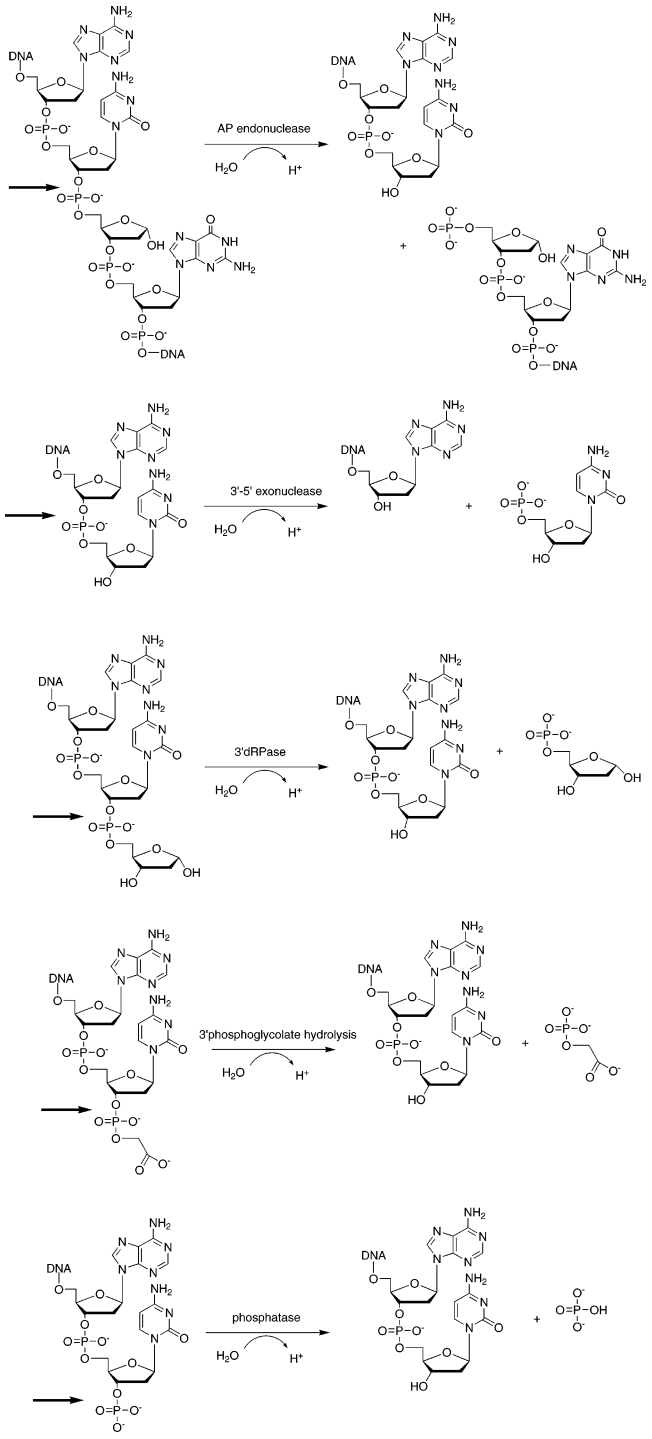


Figure 8. DNA repair reactions catalyzed by exonuclease III (Xth). Substrates and products are shown, and the arrows indicate the sites of attack by the metal-activated water nucleophile. Xth initiates repair of AP sites via endonucleolytic cleavage 5′ of the AP site. A similar endonucleolytic cleavage allows for the observed incision at damaged sites such as α-anomeric and base-fragmented nucleotides such as urea (Figure 4) and may account for the reported RNaseH activity (not shown). This enzyme accepts a broad range of substrates, processing a variety of common damaged DNA termini via an exonucleolytic cleavage reaction (intact nucleotides, dRP groups resulting from bifunctional glycosylase-catalyzed β-elimination, and 3′-phosphoglycolates produced from oxidative damage). In addition to these diverse phosphodiesterase reactions, Xth exhibits catalytic promiscuity by catalyzing a phosphomonoesterase (phosphatase) reaction.

and Lys142) that can act as nucleophiles in a subsequent reaction with an AP site.[194−196] The Schiff base that is formed by attack of an active site lysine can undergo β-elimination or proceed fully to β,δ-elimination (Figure 6).[194] Mutation of either lysine 20 or 142 to alanine showed that either residue is capable of forming a covalent species and permitting the elimination reaction to occur, with the Lys142-mediated pathway exhibiting substantially faster kinetics than the Lys20-mediated pathway. Although two pathways for elimination are possible, the AP DNA product is sufficiently long-lived that AP endonuclease preferentially processes this product when both AP endonuclease and MutY are present, demonstrating that hydrolysis of the intermediate to generate the AP site is favored over elimination.[197] This low level of promiscuous AP lyase activity of MutY could mimic an early stage in the evolution of a truly bifunctional DNA glycosylase. Remarkably, placement of a lysine side chain at a position analogous to the active site lysine in OGG1 via site-directed mutagenesis (S120K) results in a substantial increase in AP lyase activity (Table 5).[198] Selective pressure for an AP lyase activity could lead to improvements in this low level of activity if chance mutations allowed better positioning of the amine nucleophile and general base(s). Presumably, the presence of highly efficient AP endonucleases accounts for this apparent lack of selective pressure to improve the AP lyase activity of MutY. Additional aspects of HhH superfamily divergent evolution are discussed below, including the global changes in structure that have occurred during evolution of this superfamily (see section 6.1).

5.4. Bifunctional dCTP Deaminase/dUTPase

Many types of base damage can occur more readily in free nucleotides than in DNA, especially oxidation and deamination reactions. For certain types of damage, such as the oxidation of guanine and the deamination of cytosine nucleotides, the products have a high probability of being incorporated into DNA and creating a mutation. This is because 8-oxoGTP and dUTP can be readily incorporated by replicative polymerases opposite A, in place of TMP. In the case of 8-oxoGMP incorporation, the subsequent excision of A by MutY would be mutagenic; therefore all organisms appear to carry out a certain level of "DNA repair" by surveying the cellular pool of dNTPs and eliminating these damaged nucleotides. One of these enzymes, the bifunctional dCTP deaminase/dUTPase from *Methanococcus jannaschii*, exhibits catalytic promiscuity by catalyzing two chemically distinct reactions at a single active site (Table 3).[199,200] In two tightly coupled reactions, dCTP is deaminated to give dUTP and then rapidly hydrolyzed to give dUMP and inorganic pyrophosphate (Figure 9). This enzyme is presumably involved in pyrimidine biosynthesis, because dUMP is the precursor required for synthesis of dTTP. In most organisms that catalyze the deamination of dCTP, this reaction is carried out by a monofunctional dCTP deaminase that catalyzes the formation of dUTP. If dUTP is left in the nucleotide pool, then it can be misincorporated into DNA by the replicative polymerases that do not discriminate against this nucleotide, and thus these organisms also encode an efficient dUTPase to catalyze pyrophosphorolysis of dUTP to yield inorganic pyrophosphate and dUMP. In *M. jannaschii*, the deaminase carries out both reactions at a single active site. Organisms with a bifunctional dCTP deaminase/dUTPase can potentially benefit from the coupling of these reactions by avoiding dUTP release. Many of the monofunctional dUTPase and dCTP deaminases have detectable sequence homology and the three-dimensional structures

Table 5. Changes in Substrate and Reaction Specificity via Protein Engineering and In Vitro Evolution[a]

enzyme	change in specificity	mutations(s)
	Substrate Specificity	
UDG	excision of C instead of U	N204D (human)[299] or N123D (*E. coli*)[300]
	excision of T instead of U	Y147A, Y147C, or Y147S[299]
	excision of C·pyrene instead of U·G	N123D, L191A (*E. coli*)[301]
MIG	excision of A instead of T	A50V/L187Q[302]
AAG	uncreased excision of G	N169S or N169A[40,148]
AGT	increased activity toward m^4T	many mutants[298]
DNA polymerase I (Taq)	incorporation of rNTPs	A597T/E615G[303]
	incorporation of 2′-*O*-methyl NTPs	I614E/E615G[304]
(*E. coli*)	increased incorporation of rNTPs	E710A[305]
	increased incorporation of ddNTPs	F760Y[306]
DNase I	AP endonuclease conferred	14-amino-acid insertion[307]
NaeI	endonuclease converted to a topoisomerase	L43K[308]
	Reaction Specificity	
MutY	increased AP lyase activity	S120K[198]
Mth.TDG (EndoIII-like)	conferred AP lyase activity	Y126K[309]
EndoIV	decouples phosphatase and phosphodiesterase	G149D[242]

[a] There are many additional examples of successful engineering and directed evolution of metabolic enzymes that suggest substrate specificity and even reaction specificity can be readily changed by one or a few mutations.[11,13,14,16,310,311]

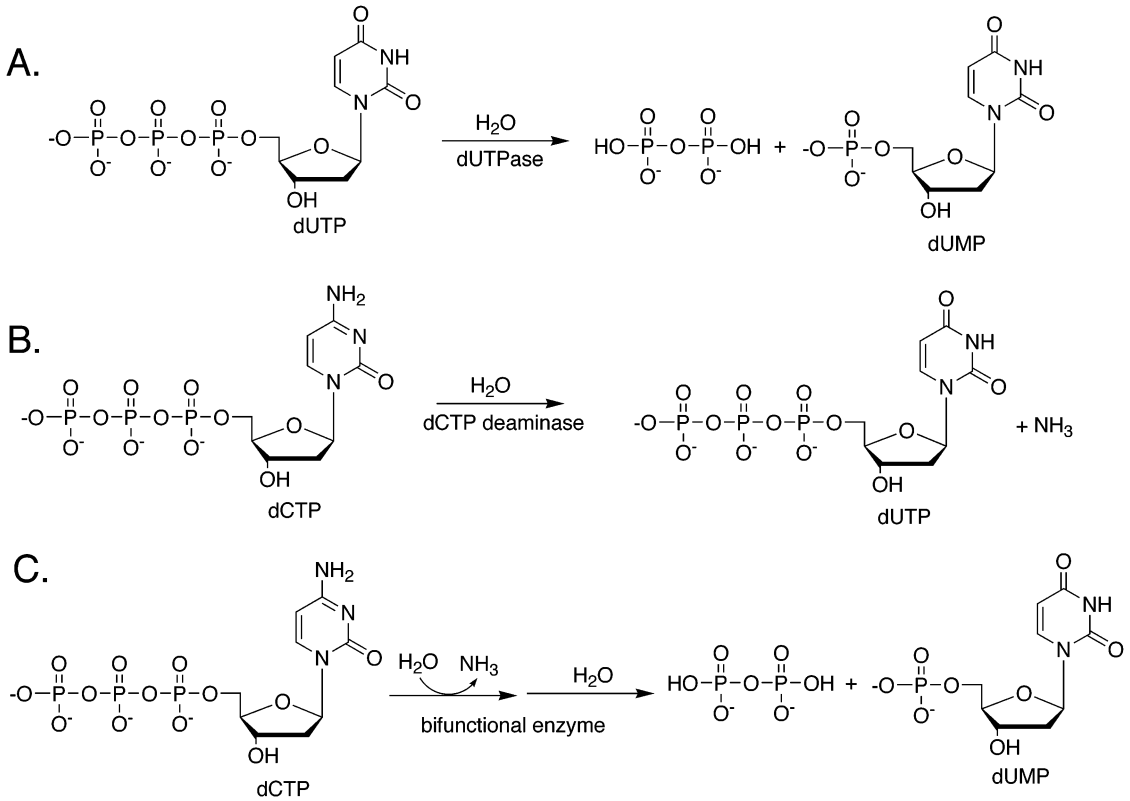


Figure 9. Reactions catalyzed by the dUTPase/dCTP deaminase superfamily of enzymes. (A) The dUTPase reaction converts dUTP to dUMP and pyrophosphate (hydrolysis of the phosphodiester bond). As replicative polymerases do not discriminate against dUTP, this reaction is critical to prevent the incorporation of dUMP into DNA, and thus all organisms appear to have a dUTPase to protect against spontaneous deamination of dCTP. (B) In some organisms, such as *E. coli*, the dUMP that is required for thymidine biosynthesis is generated from enzymatic deamination of excess dCTP. In contrast to other known deaminases, dCTP deaminase does not appear to use metal ion cofactors. The dCTP deaminase from *E. coli* has structural homology with dUTPase from the same organism, suggesting that they are related by divergent evolution. (C) The discovery of bifunctional dUTPase/dCTP deaminase enzymes such as that from *M. jannaschii* provides the missing piece of the puzzle. These enzymes exhibit catalytic promiscuity by catalyzing these two very different reactions at their active site. By tightly coupling these two steps of pyrimidine biosynthesis, organisms can avoid increasing the cellular pool of dUTP.

revealed that these enzymes belong to the same superfamily.[201] Many of the active site features of the deaminases and dUTPases are conserved, suggesting that relatively few changes separate these rather different chemical reactions (Figure 9). This remarkable divergence in function makes more sense given that the bifunctional enzyme from *M. jannaschii* also belongs to this same superfamily. Apparently these two biochemical activities have been kept together in the same enzyme in some organisms or separated into two enzymes in other organisms.

The examples of bifunctional DNA repair phosphodiesterases/phosphatases, glycosylases/lyases, and the deaminase/dUTPase provide evidence that bifunctionality can be selected for during evolution. More generally, many cellular enzymes are known to catalyze multiple mechanistically distinct reactions at a single active site as part of their normal

physiological function(s). Results from the in vitro evolution of new enzymatic activities provides compelling evidence that loss of catalytic activity is not a prerequisite to the gain of a new activity.[16] As long as the new catalytic activity does not confer negative selective pressure, then a given gene could expand or adapt the catalytic repertoire of its gene product prior to or without gene duplication.

5.5. DNA Cytosine Methyltransferases: DNA Modification and Promiscuous DNA Damage

DNA cytosine methyltransferases provide the first of two examples in which promiscuous reactions cause DNA damage, by catalyzing the deamination of C and m^5C (Table 3).[202−204] Although the DNA methyltransferases are not DNA repair enzymes by themselves, the DNA modification that they catalyze is critical for normal DNA metabolism.[125] This provides a clear example that natural selection has not been completely successful in ridding enzymes of potentially deleterious promiscuous reactions. The question remains whether this promiscuous deamination is an unavoidable side reaction of the physiological methyltransfer reaction or whether it might provide a latent source of increased mutations in times of stress.

DNA (cytosine-5-) methyltransferases catalyze carbon−carbon bond formation between the donor methyl group of *S*-adenosylmethionine (SAM) and the C5 of cytosine bases within DNA. The known DNA methyltransferases belong to an enzyme superfamily that conserves many features of the overall three-dimensional structure, the identity of catalytic residues, and a binding site for the SAM cofactor.[205] This catalytic domain is structurally homologous to a much larger superfamily of SAM-dependent methyltransferases that methylate small molecule substrates. Each enzyme has distinct sequence specificity, and the known sites of methylation are N^6 of adenine and either N^4 or C5 of cytosine. Some of the methyltransferases that catalyze the formation of m^6A have also been shown to catalyze methylation of N^4 of C, and in turn the N^4-specific methyltransferase M. PvuII can also methylate N^6 of A when present in the enzyme's recognition sequence.[205,206] This is somewhat surprising given the potential deleterious consequences of erroneous DNA methylation, and it suggests a favorable pathway for the divergent evolution of N^4- and N^6-specific families of methyltransferases from a common ancestor.

Although many methyltransferases are known to be highly sequence-specific, binding site recognition has clearly diverged over time because related enzymes recognize different DNA sequences. Intriguingly, the HaeIII methyltransferase (cytosine methyltransferase) has been documented to provide a low level of methyltransferase activity toward noncanonical sites, with a range of activities reported for sequences that closely resembled the canonical sequence.[207] This low level of activity was used as a starting point for in vitro evolution, and mutants were identified that were up to 1000-fold better at methylating alternative sites.[208] The ease with which substrate specificity could be altered demonstrates how new DNA modification patterns could evolve. It also raises the possibility that a mutation of a DNA modifying enzyme could lead to deleterious DNA modification.

The N-methylation reaction catalyzed by N^6-adenosine- and N^4-cytosine-specific methyltransferases is thought to proceed via a S_N2 mechanism, with direct nucleophilic attack of the exocyclic amine on the methyl group of SAM.[209,210] It is apparent that such a mechanism would be difficult to apply to methylation of C5, given the extremely poor nucleophilicity of this site, and indeed C5 cytosine methyltransferases use an elaborate and elegant mechanism to activate C5 for nucleophilic attack that is analogous to the strategy used by thymidylate synthetase.[211] The enzyme catalyzes a Michael addition of an active site thiolate on the C6 of cytosine to form a covalent intermediate (Figure 10).[211−213] Protonation of N^3 stabilizes the 4,5-enamine species that subsequently accepts the methyl group of SAM. The covalently bound DNA is released via the deprotonation of C5 and the β-elimination of the thiolate to generate the 5-methylcytosine product. This mechanism leaves these enzymes susceptible to hydrolytic deamination, particularly in the absence of the SAM cofactor.[202,203] Consistent with this notion, overexpression of cytosine methyltransferases[214] or mutations that disrupt SAM binding[215] lead to greatly increased levels of deamination of both cytosine and 5-methylcytosine. Thus, cytosine-5-methyltransferases exhibit deleterious catalytic promiscuity by catalyzing a hydrolytic deamination reaction in addition to their physiological carbon−carbon bond formation reaction (Figure 10).

5.6. Direct Repair (and DNA Damage) Catalyzed by Alkylguanine Alkyltransferases

Alkylguanine alkyltransferase (AGT; also known as methylguanine methyltransferase or O^6-methylguanine transferase) directly reverses alkylation damage by catalyzing the stoichiometric transfer of an alkyl group from an alkylated base to an active site thiolate.[57,133,216,217] This reaction has been called a suicide reaction because AGT can only accept a single alkyl group and is not able to regenerate the active site thiolate (Figure 11A). The O^6-alkylguanine adducts are the preferred substrates, but AGT has also been shown to recognize O^4-alkyl thymine adducts to a lesser extent.[133] The bacterial enzyme appears better able to tolerate bulky alkyl groups than the human enzyme, but several different alkyl adducts have been shown to be substrates of both classes of enzymes. The recent crystal structure of AGT in complex with a DNA substrate has provided considerable insight into the mechanisms of DNA binding and damage recognition and the catalytic mechanism of alkyl transfer.[216] Paradoxically, this DNA repair protein has been found to potentiate the toxicity of bifunctional electrophiles such as dibromoethane.[218−221] It was subsequently shown that the active site thiolate of Cys145, which has a relatively low pK_a, reacts readily with small electrophilic compounds such as dibromoethane.[217] When AGT binds to DNA, a normal guanosine nucleotide can be flipped out into the active site, placing the second electrophilic group in close proximity to the base. This facilitates the nucleophilic attack by N^7 of guanine, generating a covalent complex between AGT and the DNA (Figure 11B). Thus, the activation of an active site cysteine provides for the promiscuous reaction with electrophilic sites on small molecules. The subsequent DNA-binding and nucleotide-flipping reactions that are integral to the normal physiological reaction then bring the small molecule into close proximity to the DNA, allowing for a protein−DNA cross-link to be formed.

The DNA damage that is catalyzed by AGT results in a covalent protein−DNA complex that appears to be more difficult to repair than a small molecule adduct, because the toxicity of the bifunctional alkylating agent dibromoethane is greatly increased if active AGT is present in the cell.[218] As there are many enzymes that form covalent intermediates

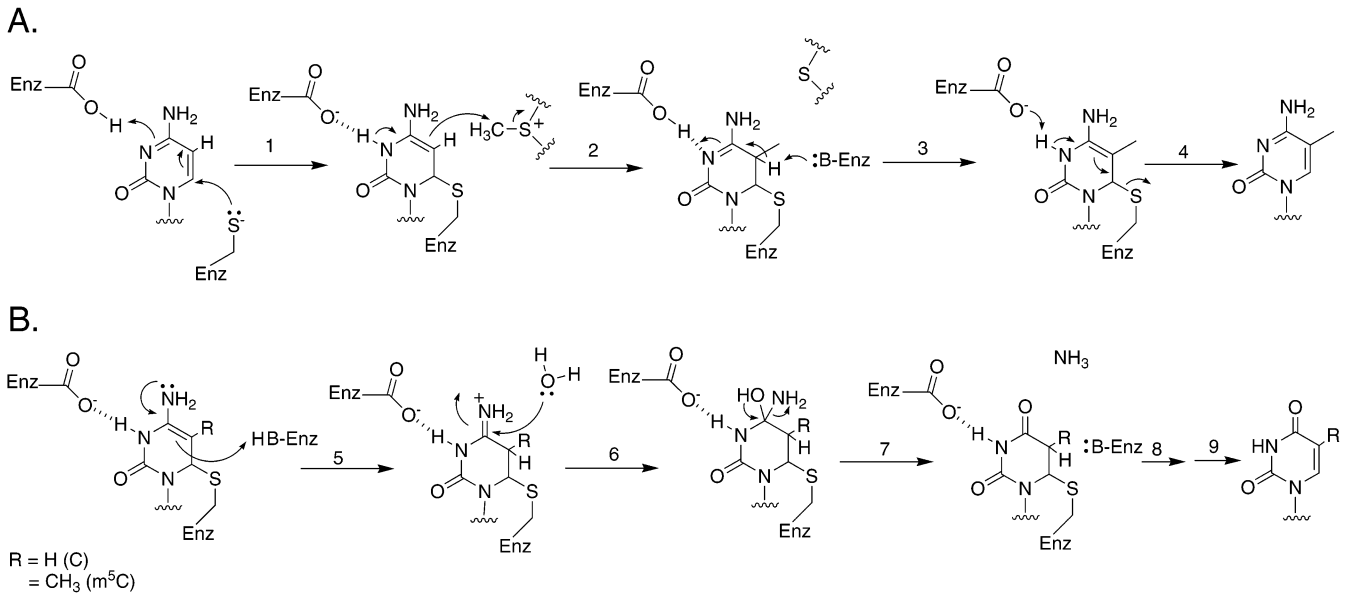


Figure 10. Mechanism of DNA methylation (A) and promiscuous deamination (B) catalyzed by cytosine methyltransferases. These are proposed to use an addition/elimination mechanism to methylate C5 of cytosine. In the first step of the reaction, the thiolate of the conserved active site cysteine undergoes a Michael addition at C6 (1). This step is assisted by a carboxylic acid that acts as a general acid to protonate N^3. In the presence of the methyl donor *S*-adenosylmethionine (SAM), the now activated C5 atom accepts the methyl group from SAM to generate a covalent m^5C intermediate (2). An active site base sets up the elimination of the thiolate by abstracting the C5 hydrogen (3). Finally, in a reverse of the addition step, the active site carboxylate deprotonates N^3, and the cysteine thiolate is eliminated to generate the m^5C-DNA product (4). In the absence of the SAM cofactor, the covalent C or m^5C intermediate can be protonated at C5, presumably by the acid form of the general base that assists in step 3 of the normal reaction (5). Subsequent hydrolytic deamination (6 and 7) generates either U from C or T from m^5C. The covalent, deaminated product is released by proton extraction and elimination of the active site thiolate (8 and 9).

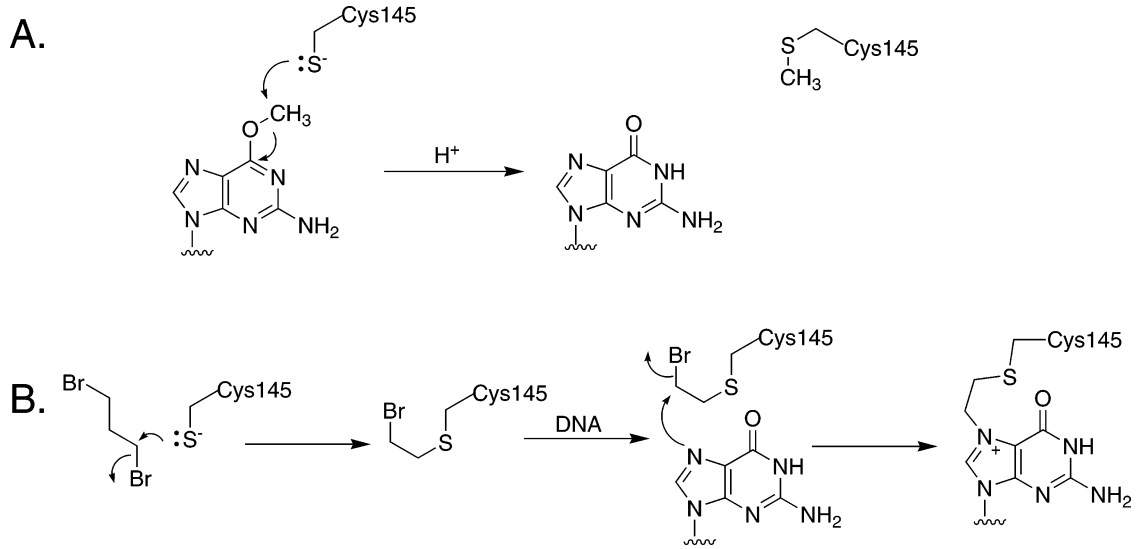


Figure 11. Alkylguanine methyltransferase accelerates rates of both repair and damage of DNA. (A) In the repair reaction AGT binds to alkylated DNA and flips out the damaged nucleotide (O^6-methylguanosine is shown). The highly reactive thiolate of a conserved active site cysteine, Cys145 in human AGT, initiates nucleophilic attack on the alkyl group to restore the normal base. The protein cannot regenerate the thiolate once it has been alkylated, so it is limited to a single turnover. (B) Due to its accessibility and high reactivity, Cys145 is readily modified by a wide variety of electrophiles (Table 3). If it reacts with a bifunctional electrophile such as dibromoethane, a potentially reactive covalent intermediate is created. Subsequent binding and nucleotide flipping by a covalently modified AGT assists a second nucleophilic attack by N^7 of G to form a covalent DNA−protein cross-link. This catalytic promiscuity is possible because the highly reactive active site nucleophile and the DNA-binding and nucleotide-flipping mechanisms are conserved in both reactions.

with DNA, including AP lyases, bifunctional glycosylases, cytosine methyltransferases, and topoisomerases, protein−DNA adducts may constitute a relatively common type of DNA damage. Indeed, it has been shown that DNA−peptide adducts are efficiently recognized by the prokaryotic nucleotide excision repair machinery.[222,223] The finding that DNA topoisomerases can become trapped as covalent intermediates via their phosphotyrosine linkages led to the identification and characterization of Tpp1 phosphodiesterase that catalyzes the hydrolytic excision of the enzyme from DNA.[224,225] Recently, redundant pathways have been identified in yeast that rely on the Mre11 or SLX4 nucleases.[226] It is not yet clear which DNA repair pathway or pathways processes protein−DNA cross-links that occur through the DNA bases,

but it seems likely that other small molecules are likely to be able to trap these covalent complexes in a manner analogous to the trapping of AGT with bis-electrophiles or topoisomerase I with campthothecin.[227] More generally, it is not yet known to what extent other enzymes involved in DNA repair also catalyze DNA-damaging reactions. However, the examples discussed above suggest that such reactions are possible and perhaps an unavoidable consequence of localizing proficient enzymes to genomic DNA.

5.7. Promiscuity of DNA Ligases

The reactivity of the active site cysteine of AGT is similar to the reactivity of the active site lysine of DNA ligases (Table 3). The physiological reaction of DNA ligases is phosphoryl transfer (P—O bond formation), but ligases are also known to exhibit catalytic promiscuity by catalyzing the formation of a covalent adduct with pyridoxal phosphate (C—N bond formation). In the phosphoryl transfer reactions catalyzed by DNA ligases, the universally conserved active site lysine forms a covalent intermediate with AMP and subsequently transfers the AMP group to the 5′-phosphate of a DNA substrate. The relatively low pK_a of this lysine (pK_a = 8.4 in T4 DNA ligase)[228] facilitates the nucleophilic attack on the α-phosphate of ATP in the first step of the ligase reaction but also allows for promiscuous nucleophilic attack on pyridoxal phosphate. Once covalently bound, it is easy to envision that the pyridoxal phosphate cofactor could be used to carry out chemistry that is unrelated to the normal reaction. Perhaps an analogous process led to the unconventional use of pyridoxal phosphate as a general acid/base in the reaction catalyzed by glycogen phosphorylase.[229−231]

T4 DNA ligase has also been reported to have AP lyase activity,[232] a promiscuous reaction that presumably would make similar use of the lysine nucleophile. Consistent with this model, the AP lyase activity could be inhibited by ATP, and a covalent complex between ligase and DNA could be trapped by borohydride, a treatment known to trap AP lyases by reducing the Schiff base intermediate.[232] More generally, the exceptional reactivity of active site residues, such as the active site lysine of DNA ligase and the cysteine of AGT, has long been exploited for the identification of catalytic groups in enzyme active sites in the absence of a structure, because catalytic residues are generally more reactive than other residues on a protein's surface. This greater reactivity of active site groups supports the idea that catalytic promiscuity is a common feature of enzyme active sites.[11]

5.8. Challenges in Detecting Catalytic Promiscuity

Additional examples of promiscuous DNA repair activities have been reported[233−235] but subsequently called into question.[236−238] And in other cases, alternative activities have been observed but not fully characterized.[232,239,240] These examples underscore the difficulty in detecting catalytic promiscuity. This is because promiscuous activities usually have much smaller rate enhancements than the normal activity and copurification of even low levels of a contaminating enzyme could give rise to this level of activity. Indeed, biochemical purification is particularly challenging for nucleic-acid-binding proteins, because they are likely to share physical properties with the other DNA repair proteins that are most likely to be problematic contaminants. Several methods can be used to establish that observed alternative activities are promiscuous activities of a given active site and not the result of a trace contaminant. Traditionally this has been done by biochemical copurification of the two activities, as a constant ratio of the two activities suggests that they arise from the same enzyme.[159] Although fractionation of two different biochemical activities provides direct evidence that distinct enzymes are responsible for the two activities, many proteins copurify over multiple columns and the failure to separate two activities cannot be taken as strong evidence for a multifunctional protein. Coinhibition of the two activities by a common inhibitor can be a useful technique, provided a specific inhibitor is known. If one or both of the activities can be saturated with substrate, then the competition between the substrates can be used to test whether the two substrates bind to the same site. Mutation of active site residues and evaluation of the effects on both reactions has also been used successfully to show that two reactions are carried out at the same active site. Such experiments have to be carefully analyzed, however, because promiscuous activities may only utilize a subset of the catalytic groups in an active site.[11,16,241,242] If possible, mutation of an active site residue that alters binding of a competitive inhibitor or substrate provides the most conclusive test of whether a reaction is carried out at a given active site. A change in the inhibition constant for inhibition of the alternative activity provides strong evidence that the mutant protein is responsible.[243]

It will be fascinating to learn to what extent other DNA repair enzymes exhibit catalytic promiscuity. It seems likely that many promiscuous reactions that utilize DNA as a substrate would be selected against, but the examples that are discussed above illustrate that such promiscuous activities do exist (Table 3). Furthermore, detectable (and potentially functionally significant) promiscuous catalysis may be latent within many more enzymes, as there are many examples of successful protein engineering in which a single mutation reveals a new enzymatic activity (Table 5; see section 7).[5,11,14,16,21,244,245] The characterization of the promiscuous reactions of DNA repair enzymes contributes to our understanding of the evolutionary history of DNA repair pathways and provides information about the current evolutionary potential for new DNA repair activities. Continuing progress in characterizing the three-dimensional structures of proteins is certain to identify additional evolutionary relationships and thereby suggest other types of catalytic promiscuity that may be possible.

6. Mechanistically Diverse DNA Repair Enzyme Superfamilies

Divergent evolution of ancestral enzymes has led to large superfamilies of modern-day enzymes that differ in their substrate specificity and, in many cases, their reaction specificity.[21] In many cases, the primary amino acid sequences have diverged beyond detectable sequence similarity, but characterization of their three-dimensional structures have allowed these ancient evolutionary connections to be identified. Not only is the overall three-dimensional fold conserved in each case, but the active site location and in many cases the identity of key catalytic residues have also been conserved. Once structures have been determined for several members of the same superfamily and the key catalytic residues have been identified, more sophisticated sequence searches can often be used to expand these superfamilies by predicting additional proteins that are likely to have conserved structural and mechanistic features.[19] The great

success of these approaches notwithstanding, this discussion will focus on enzymes for which the three-dimensional structures are known, given the difficulty in detecting distant evolutionary relationships by sequence comparisons. Comparison of three-dimensional protein structures can greatly facilitate the identification and functional characterization of conserved catalytic residues. Fortunately, the past decade has seen a rapid expansion in the number of DNA repair enzymes for which structural information is available,[246] and the coming decade is sure to yield similar increases in our knowledge of the structural biology of DNA repair.

The development of automated methods for the rapid and exhaustive comparison of new protein structures to all other known structures has greatly facilitated the classification of proteins into families and superfamilies of structurally homologous proteins.[247−254] The Structural Classification of Proteins (SCOP) Database[253,254] currently includes almost two dozen functionally diverse enzyme superfamilies that contain at least one enzyme involved in DNA repair (Table 4). Although there are a few superfamilies that are highly specialized in DNA repair, such as the HhH and FPG-related superfamilies of DNA glycosylases, many other superfamilies include enzymes that are involved in diverse aspects of metabolism and cell signaling. It is difficult to know with any certainty the actual evolutionary pathway that describes the divergence of any two paralogues, so we do not know which DNA repair enzymes diverged from enzymes with other cellular roles and which DNA repair enzymes diverged to function in other cellular roles. Nevertheless, it is likely that both paths have been taken, and the lack of any common biological function within most superfamilies is consistent with the probabilistic nature of gene duplication. A few examples are discussed below to illustrate some of the ways in which DNA repair enzymes have diverged to recognize different substrates and in some cases to catalyze markedly dissimilar reactions. In each case, key mechanistic features and/or substrate-binding features could have facilitated the evolutionary diversification of these enzymes.

6.1. Helix–Hairpin–Helix DNA Glycosylase Superfamily

Many of the DNA repair glycosylases that are responsible for recognition and repair of oxidative and alkylative base damage belong to a functionally diverse group of evolutionarily related enzymes known as the helix−hairpin−helix (HhH) superfamily. There is a great deal of structural diversity among these enzymes, but members of this superfamily can be identified by the conserved HhH motif that is invariably involved in DNA binding (Figure 12).[84,87] Many of these proteins have a second domain that also varies greatly but sometimes includes a four-cysteine Fe(II) center. Although some members of this superfamily have detectable sequence homology, others could not be assigned until the three-dimensional structures were solved.[84,255−257] Individual enzymes of this superfamily are capable of recognizing an extraordinary variety of modified bases in many different hydrogen-bonding contexts. As both substrate specificity and reaction specificity has varied (Table 2), it is not surprising that the amino acid sequences have diverged considerably. Nevertheless, a few residues critical for DNA binding, nucleotide flipping, and N-glycosidic bond cleavage are conserved (the amino acid side chains in parentheses reflect the numbering for *E. coli* EndoIII). For example, there is an aspartate residue (Asp138) just outside of the HhH motif that acts as a general base and is conserved in all of the glycosylases except for 3-methyladenine DNA glycosylase I (*E. coli* Tag). The absence of this catalytic residue may be the exception that proves the rule, because Tag is a very poor catalyst that is highly specific for destabilized 3-alkylpurine lesions that require little catalytic assistance.[73,255] Most of the bifunctional enzymes also conserve the location of the lysine nucleophile (Lys120) within the HhH motif (see above). Similarly, the position of the amino acid (Gln41) that is inserted into DNA in place of the extrahelical base is also conserved, although the identity appears to vary considerably with asparagine, leucine, and even arginine also occurring at this position.[89] Additional differences in the active site pocket can be rationalized in terms of the substrates that are recognized. However, it is difficult to understand why some enzymes have evolved to be monofunctional and require the subsequent action of a separate AP lyase and why others have evolved to be bifunctional. It could reflect differences in the repertoire of repair enzymes of the cell in which the glycosylase evolved (e.g., whether there was sufficient AP endonuclease activity already present in the cell, in which case there would not be selective pressure for a bifunctional mechanism), or it could simply reflect evolutionary chance. Indeed, there does not seem to be a strong selective pressure for either mono- or bifunctional glycosylases because there are numerous examples of each class. It is not yet known whether the extensive structural differences between HhH glycosylases reflect some additional functional difference, such as protein−protein interactions or, as recently proposed, long-distance DNA damage detection via an iron−sulfur center.[258,259]

This superfamily of DNA repair enzymes exemplifies how an active site can be altered through evolution to create enzymes with different catalytic mechanisms (bifunctional vs monofunctional) and very different substrate specificities. Is it possible that the diversification in the HhH superfamily was made possible by the broad substrate specificity and catalytic promiscuity of many of these proteins? Several factors could have contributed to the extensive diversification of the HhH superfamily and are expected to apply to the diversification of the other broadly specific DNA glycosylases related to FPG and to UDG that were discussed above (Tables 1 and 4). (i) The N-glycosidic bond is an attractive target for DNA repair because it is more labile than the phosphodiester backbone. Therefore an evolving DNA glycosylase is expected to require less catalytic power to achieve a given reaction rate, relative to a repair mechanism involving breaking another bond in DNA. (ii) All nucleotides share common structural features, such as the sugar moiety that is the site of nucleophilic attack by DNA glycosylases and the regular structure of the phosphate backbone that is used as a handle for sequence-independent DNA binding. (iii) Nucleotide flipping is an integral step for all DNA glycosylases, so a newly recruited DNA glycosylase would have an additional hurdle to overcome beyond the universal requirements for substrate binding and catalysis: stabilization of an unfavorable conformation of DNA with an extrahelical base lesion. This factor could serve as a powerful advantage for a duplicated DNA glycosylase relative to a glycosylase that acts on nucleotide substrates. Nevertheless, convergent evolution can occur, and the existence of four different structural families of DNA glycosylases demonstrates that that these enzymes have independently evolved at different times in the past (Table 4). One example of this functional

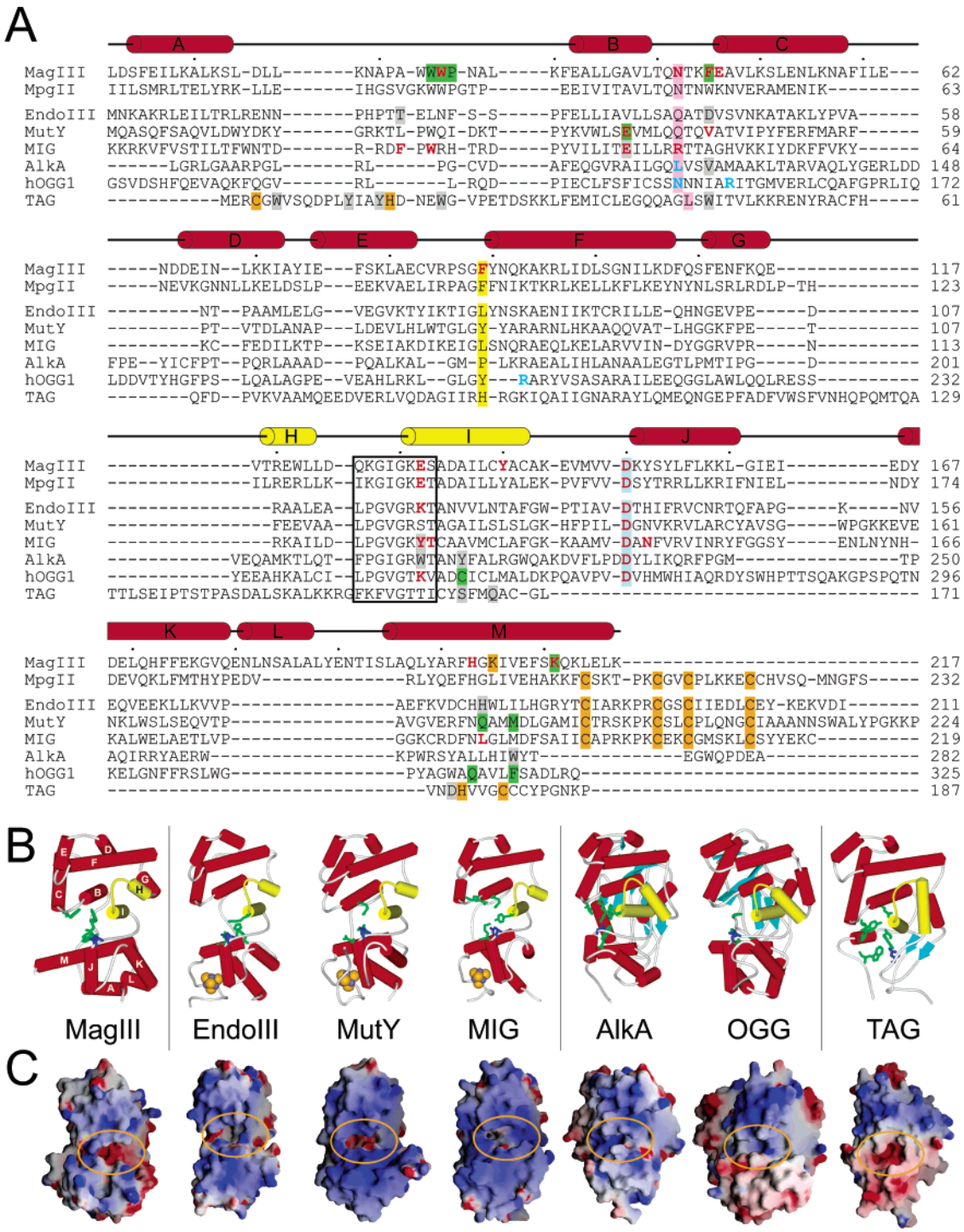


Figure 12. Structures of the DNA glycosylases belonging to the helix−hairpin−helix (HhH) superfamily. (A) The structure-based sequence alignment of some HhH glycosylases. The MagIII secondary structure elements are shown schematically, with the HhH motif highlighted as yellow cylinders (helices H−I of MagIII). The HhH residues that contact the DNA in the AlkA·DNA complex are boxed, and the conserved catalytic aspartic acid is shaded blue. Residues in the nucleobase binding pocket confirmed (green) or predicted (gray) to contact the target base are shaded, and the positions of the side chains that intercalate the DNA helix at the lesioned and nonlesioned strands are shaded pink and yellow, respectively. Residues that contact the orphaned DNA base opposite the modified base in AlkA and OGG1 are colored blue, while residues shown by mutagenesis to be important for either catalysis or DNA binding are colored red. Side chains that coordinate the Fe_4S_4 clusters (MpgII, EndoIII, MutY, and MIG) and Zn^{2+} ion (Tag) as well as the carbamylated lysine in MagIII are shaded orange. (B) Schematic representations of the HhH glycosylase structures. Helices are shown as red and yellow (HhH motif) cylinders, β-sheets as light blue arrows, and Fe_4S_4 clusters as golden CPK spheres. Side chains of functionally significant active site residues are rendered as sticks, with the conserved aspartic acid colored dark blue. (C) Solvent-accessible surfaces are colored according to electrostatic potential (blue, positive; red, negative). The substrate-binding pockets at the domain interface are circled. The structures have been rotated ~90° with respect to the views shown in part B. Reprinted with permission from ref 89 (http://embojournal.npgjournals.com). Copyright 2003 Macmillan Publishers Ltd.

convergence comes from the comparison of the human 3-methyladenine DNA glycosylase and that from *E. coli*. These enzymes share many of the same substrates but adopt different folds and are not evolutionarily related.[53] An even more remarkable example of functional convergence involves the evolution of the nucleotide excision repair pathways of

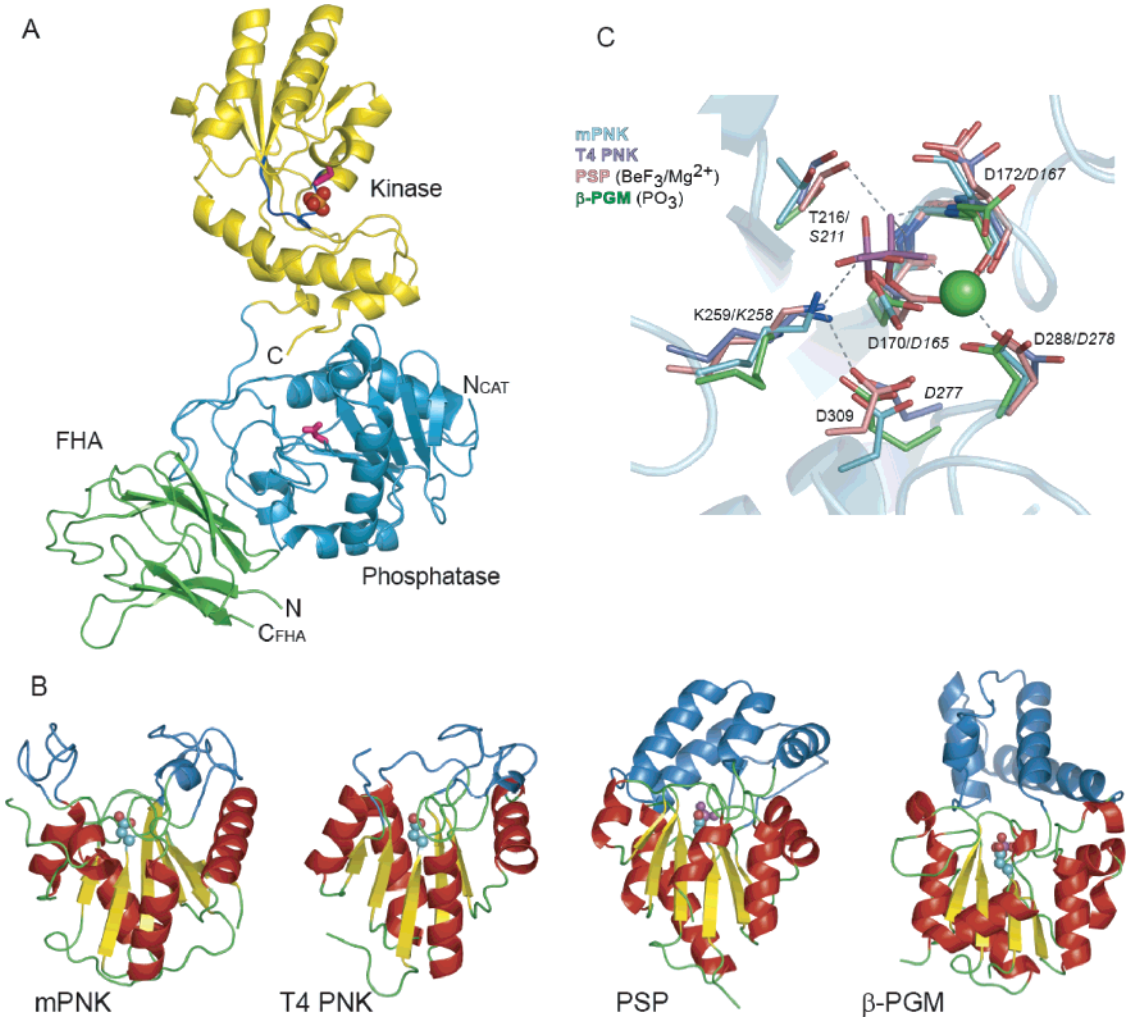


Figure 13. Structures of eukaryotic polynucleotide kinase/phosphatase (PNKP) and evolutionarily related enzymes of the HAD superfamily. (A) Ribbon diagram of mouse PNKP, with the kinase in yellow, the phosphatase in blue, and the FHA domain in green. Catalytic side chains (Asp170 and Asp396 in the phosphatase and kinase, respectively) are in pink, the ATP binding P-loop is in navy blue, and the sulfate bound at the P-loop is in orange and red spheres. (B) Comparison of mouse PNKP and T4 PNKP phosphatase domains, β-PGM and PSP in a common orientation, showing the variation in the active site capping structure (blue). The common α/β domain elements are colored yellow and red, and the catalytic aspartate is shown in a ball-and-stick representation. (C) Structure of the mouse PNKP phosphatase active site (cyan, PDB code 1YJ5) superimposed on the active sites of phosphorylated β-PGM (green, PDB code 1LVH), BeF_3-derivatized PSP (pink, Mg^{2+} in green, PDB code 1J97), and T4 PNKP (blue, PDB code 1LTQ). The hydrogen-bonding pattern for PSP is shown. Residue numbering shown is for mouse PNKP and in italics for T4 PNKP. The kinase domain is similarly homologous to a superfamily of nucleotide triphosphate hydrolases (not shown; Table 4). Reprinted with permission from ref 263. Copyright 2005 Elsevier.

prokaryotes and eukaryotes. These pathways share many common mechanistic features, but the individual proteins are not evolutionarily related.[115,116]

6.2. Polynucleotide Kinase: Fusion of a 3′-Phosphatase from the Haloacid Dehalogenase Superfamily and a 5′-Kinase from the P-Loop-Containing Nucleotide Hydrolase Superfamily

Polynucleotide kinase/phosphatase (PNKP) is a bifunctional DNA repair enzyme that is responsible for preparing nicked DNA sites for ligation. The 3′-phosphatase activity hydrolyzes a 3′-phosphate from DNA to generate a 3′-OH, and the kinase activity phosphorylates a 5′-OH DNA end to generate a 5′-phosphate, both of which are required by DNA ligases. This enzyme provides an interesting example of divergent evolution, because it is a multidomain protein presumably resulting from the fusion of a phosphatase and a kinase domain (Figure 13).[260−263] Similar multidomain assembly of proteins is commonplace, and the vast majority of multifunctional enzymes have separate domains for each catalytic activity. It is widely accepted that separate folding domains have considerable potential to be recombined via gene fusion events. Interestingly, both domains of PNKP belong to separate mechanistically diverse enzyme superfamilies that are widely represented by enzymes from a variety of cellular processes and thereby showcase the scope of divergent evolution.

A 3′-phosphate serves as an effective block of DNA replication and repair, and thus it is not surprising that most cells have multiple enzymes capable of recognizing this very stable lesion. The exonuclease-III- and endonuclease-IV-related enzymes comprise two evolutionarily distinct families of broadly specific bifunctional phosphodiesterase/phosphomonoesterases (see above). The phosphatase domain of PNKP is the third family of 3′-DNA phosphatases. In contrast to the broadly specific nucleases/phosphatases, these phosphatases appear to have a strong preference for nucleic acids with 3′-phosphates. As in the case of T4 polynucleotide

kinase, this phosphatase is usually found as one domain of a bifunctional PNKP, but in *S. cerevisiae* the 3′-phosphatase (TPP1) does not have an associated kinase domain, and yeast appear to lack a polynucleotide kinase homologue.[236] Crystal structures of PNKP enzymes from T4 phage[260−262] and from mouse[263] demonstrate that the phosphatase domain belongs to the haloacid dehalogenase (HAD) superfamily, despite low sequence homology (Figure 13).

The HAD superfamily is a large and functionally diverse enzyme superfamily.[264,265] Enzymes of this superfamily are found throughout metabolism and recognize a wide variety of substrates. Hydrolysis of phosphate monoesters is the most common reaction catalyzed by these enzymes, and representative examples include phosphoserine phosphatase,[266,267] P-type ATPase,[268,269] and RNA polymerase II CTD phosphatase.[265,270] Other enzymes form a similar phosphorylated aspartate intermediate, but instead of catalyzing hydrolysis (water acceptor) they favor other substrates as the acceptor. Examples of phosphoryl transfer include β-phosphoglucomutase that catalyzes the transfer of phosphoryl groups to and from glucose phosphates[271] and phosphoserine phosphatase (thrH) that catalyzes the transfer of phosphoryl groups between phosphoserine and homoserine.[272] Although phosphatase and phosphotransferase reactions (P—O cleavage/formation) are the most common reactions catalyzed by HAD superfamily members, additional reactions include dehalogenase (carbon−halogen bond hydrolysis) and phosphonatase (P—C bond hydrolysis; Figure 1B). The core of the HAD superfamily proteins consist of six parallel β-strands, sandwiched between two layers of α-helices (Figure 13). Many members of this superfamily have an additional domain, or cap structure, that is located adjacent to the active site and controls the substrate and reaction specificity of the enzyme.[263,273−275] The great variety of cap structures that are observed is consistent with the dramatic differences in substrate specificity across this superfamily.

Although the catalytic residues vary greatly among the different superfamily members, all appear to conserve an active site aspartate that serves as the nucleophile in the reaction (Figure 14). The position of this aspartate side chain is conserved in the three-dimensional structure, but the position and identity of nearby groups have diverged to allow for the different reactions and different substrates that are accommodated. Haloacid dehalogenase, the founding member of the HAD superfamily, recognizes halogenated compounds that contain a carboxylic acid. An unusual feature of the reaction mechanism is that the breakdown of the covalent intermediate occurs via water attack at the carbonyl carbon of the aspartate side chain (Figure 14A).[276] Although the phosphatases form an analogous covalent phosphorylated aspartate intermediate, the subsequent nucleophilic attack by a water molecule occurs at phosphorus.[277] Whereas haloacid dehalogenase uses an aspartate residue as a general base to position and activate the water nucleophile, this role is filled by a Mg^{2+} ion in the phosphatases (Figures 14A and 14B). The malleability of the HAD superfamily is perhaps best exemplified by the catalytic mechanism for phosphonate hydrolysis (C—P bond cleavage) by phosphonatase (Figure 14C).[274] The active site aspartate is conserved in this enzyme, and its role in nucleophilic attack on the phosphorus of the substrate is analogous to this step in the phosphatase mechanism, but clearly the carbanion of phosphonoacetaldehyde would be a poor leaving group. The enzyme has arrived at an ingenious solution to this problem; nucleophilic attack by the active site aspartate is facilitated by first forming a Schiff base between the carbonyl of the substrate and a lysine from the protein. Subsequent regeneration of the aspartate nucleophile can be accomplished in the same manner as in the phosphatases, and the Schiff base can be hydrolyzed to regenerate the active site lysine (Figure 14).[274] Thus, the identity and function of some active site groups have been changed (the Mg^{2+} ions of the phosphatases vs the general base of the haloacid dehalogenases), and others have remained the same (the aspartate nucleophile). Further changes in the three-dimensional structure, especially in the cap structure that forms the lid of the active site, have allowed for the recognition of either small molecule substrates (PSP and β-PGM) or macromolecular substrates such as nucleic acids (PNKP; Figure 13A).

The kinase domain of PNKP is structurally homologous to a very large superfamily of P-loop-containing NTP hydrolases/transferases. Nucleotide hydrolysis is arguably one of the largest classes of biological reactions, and the P-loop-containing enzymes constitute a large and functionally diverse superfamily that can be divided into 22 different structural families (see the SCOP database for a complete listing).[254] A few representative examples highlighting the different classes of reactions are listed in Table 4. Many of the same themes discussed above for the HAD superfamily appear to have influenced the evolutionary diversification of these nucleotide hydrolases. The most common reaction catalyzed by this superfamily is the hydrolysis of nucleotide triphosphates, but many enzymes also catalyze the transfer of the terminal phosphate from ATP to another acceptor. As is the case for the HAD superfamily, the substrate specificity varies for the hydrolases (ATP vs GTP) and for the transferases (small molecules such as AMP and chloramphenicol vs macromolecules such as DNA; Table 4). Phosphoryl transfer is by far the most common type of reaction catalyzed by the P-loop-containing enzymes, but several enzymes catalyzing sulfuryl transfer also adopt this same structure. Intriguingly, at least one of the phosphoryl transfer enzymes, adenylate kinase, has been shown to exhibit catalytic promiscuity for sulfate group transfer, suggesting a possible pathway for the evolutionary divergence of phosphoryl and sulfuryl transfer enzymes.[11,278]

PNKP is just one example out of a great many multifunctional proteins in which more than one catalytic domain has been fused. These multiple functions include bringing together two or more catalytic activities, as in the case of T4 PNKP, and bringing together catalytic modules with binding modules that are involved in nucleic acid binding or protein−protein interactions. Mammalian PNKP also serves as an example of the latter, because it contains a carboxy-terminal forkhead-associated (FHA) domain, a common phosphoprotein interaction domain that can allow for recruitment and additional regulation of higher-order protein complexes, in addition to the kinase and phosphatase domains (Figure 13C).[263] Such multifunctional polypeptides are conceptually indistinguishable from protein complexes composed of multiple proteins, differing only in the type of bonds that holds the domains together: covalent versus noncovalent. Indeed, there are many examples in evolution in which fused domains appear to have been separated into distinct polypeptides and other examples in which separate polypeptides have been fused together.[279,280] It is important to appreciate that this is a common theme during evolution and that it appears to be a very favorable way of creating new

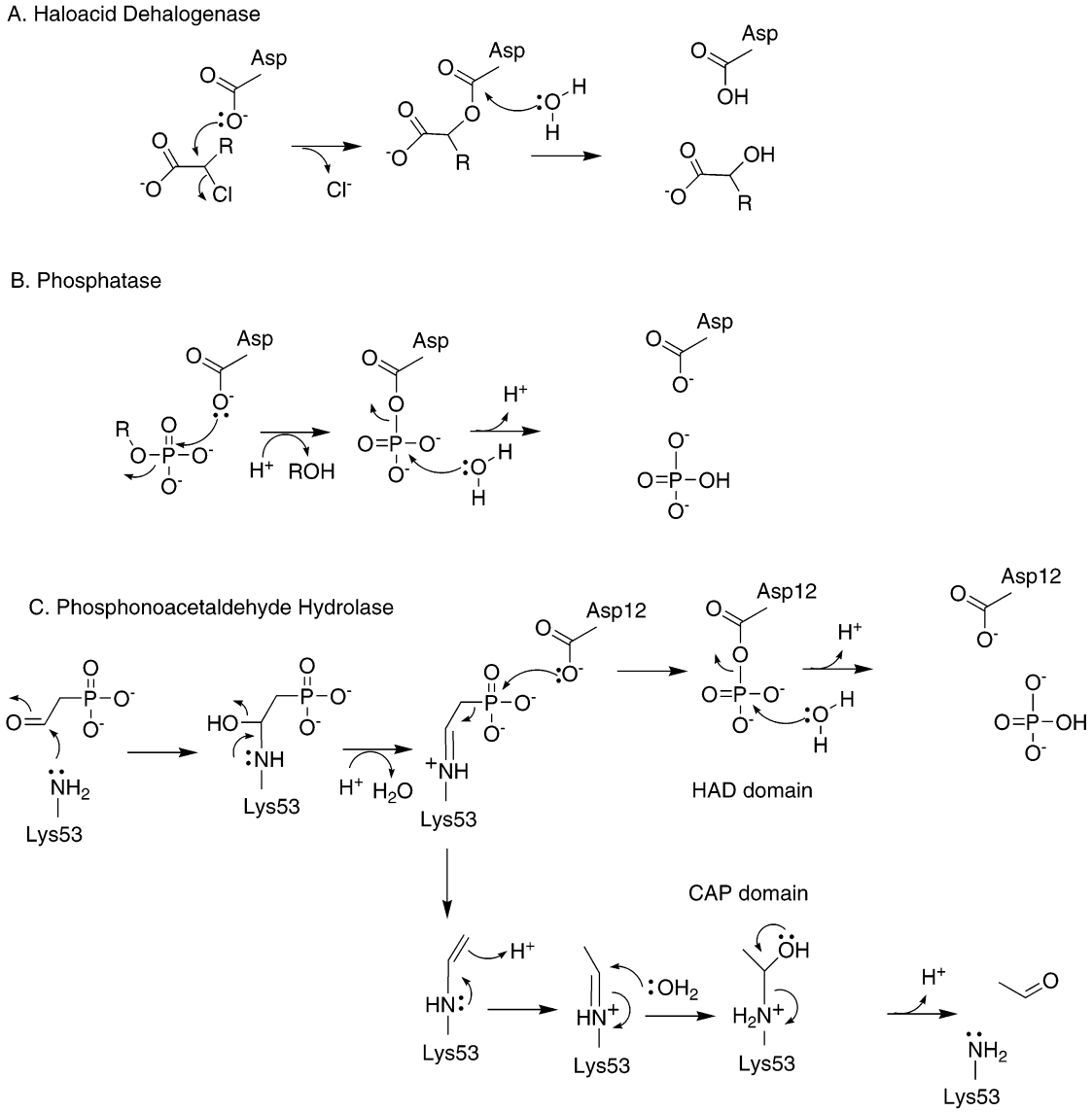


Figure 14. Reaction mechanisms of the mechanistically diverse enzymes of the haloacid dehalogenase (HAD) superfamily. An aspartate nucleophile in a common active site pocket appears to be universally conserved by the enzymes of this superfamily, but there are substantial mechanistic differences between different members. (A) Haloacid dehalogenases cleave a carbon−halogen to form a covalent aspartyl intermediate. In an unusual twist on the familiar covalent catalysis theme, the covalent enzyme intermediate is hydrolyzed via water attack on the carbonyl carbon of the aspartate side chain.[276] (B) The phosphatases of the HAD superfamily catalyze nucleophilic attack on phosphate monoesters to displace an alcohol and form a phosphoaspartate intermediate. In the phosphatase reactions, nucleophilic attack by water occurs at the phosphoryl group and not at the carbonyl oxygen. (C) Phosphonoacetaldehyde hydrolase forms a Schiff base intermediate between an active site lysine and the aldehyde of phosphonoacetaldehyde. This lysine is contributed from the cap structure that varies greatly across the superfamily in response to the different substrates that are recognized.[273−275] Schiff base formation activates the P—C bond for an elimination/nucleophilic attack by the conserved aspartate nucleophile. The phosphoasparate intermediate is hydrolyzed similarly to the phosphatases, and the Schiff base is also hydrolyzed to complete the reaction. Adapted with permission from ref 274. Copyright 2004 Elsevier.

proteins. Although interesting from a biological view, this review makes no attempt to comprehensively review these types of evolutionary connections. Other reviews have explored the general role of domain fusion in evolution[33,279,281] and more specifically as it pertains to the evolution of DNA repair proteins.[190]

6.3. AlkB and the α-Ketoglutarate-Dependent Dioxygenase Superfamily

The AlkB family of Fe(II)/αKG-dependent DNA demethylases provide an interesting example of how the identification of a protein's evolutionary relationships can provide incisive insight into the catalytic mechanism. Although AlkB was identified by genetic methods more than 20 years ago,[282] it was only in the past few years that its unusual catalytic mechanism has been revealed. The biological data strongly supports a role for AlkB in repairing alkylation-damaged single-stranded DNA, because a functional AlkB gene is required for the replication of single-stranded phage DNA.[282,283] However, attempts to reconstitute a biochemical repair activity were unsuccessful. The breakthrough came from the recognition that AlkB shares sequence homology with a functionally diverse enzyme superfamily of Fe(II)-dependent hydroxylases (Table 4).[284] This superfamily is remarkable both in the complexity of reactions that these enzymes catalyze and in the diversity of different types of reactions

that are catalyzed by family members (Figure 1C). The common mechanistic theme is the use of Fe(II) to form activated oxygen species, and a number of the enzymes in this superfamily are hydroxylases that couple the oxidative decomposition of αKG to the hydroxylation of a substrate. Analogy to these enzymes suggested the possibility that AlkB could function as a hydroxylase to oxidatively remove alkyl adducts.[284] Armed with this hypothesis, two groups were able to demonstrate that AlkB is indeed an Fe(II)/αKG-dependent hydroxylase that catalyzes the elimination of alkyl base adducts via direct oxidation (Figure 1C).[139,141] Subsequently a number of human analogues of AlkB have been found, and several of them also exhibit activity toward alkylated DNA and RNA substrates.[137,138,142,285,286] Remarkably, in addition to the expected repair of m^3C and m^1A lesions that are formed by the reaction of S_N2 alkylating agents with single-stranded nucleic acids, AlkB from *E. coli* can also directly repair the exocyclic adducts ϵA and ϵC that are formed upon exposure to chloroacetaldehyde or as a result of lipid peroxidation.[143,144] The molecular basis of this broad substrate specificity awaits high-resolution structures of AlkB in complex with its alkylated substrates. The versatility of the Fe(II)/αKG-dependent hydroxylases may be even wider than currently appreciated, because there is evidence that histone demethylases might also be evolutionarily related and could utilize a reaction mechanism directly analogous to the oxidative demethylation reaction of the AlkB family enzymes.[287]

6.4. The DNA Repair Nuclease, Artemis, and the Metallo-β-Lactamase Superfamily

Artemis is a nuclease that is involved in repairing the double-strand breaks that are formed during V(D)J recombination.[288−290] Mutations in the human protein are associated with a variety of immune deficiency disorders, and the gene was cloned and identified based upon weak sequence similarity with mouse SNM1 and yeast PSO2, enzymes that are involved in the repair of interstrand cross-links.[291] These DNA repair enzymes also have weak sequence homology with Zn^{2+}-dependent β-lactamases, and most importantly, the sequence motifs involved in metal binding are conserved.[291−293] The purified Artemis protein has 5′−3′-exonuclease activity on single-stranded DNA.[294] Remarkably, when reactions are reconstituted with DNA-PK, Artemis acts as a structure-specific endonuclease capable of opening the hairpin intermediates that are formed during V(D)J recombination.[294,295] This illustrates how protein−protein interactions influence the biochemical activities of DNA repair enzymes in vivo and suggests that changes in protein−protein interactions provide yet another way of influencing substrate specificity. Cells that lack Artemis have increased radiosensitivity, in addition to their recombinational defect, raising the possibility that the exonuclease activity of Artemis may be involved in a pathway for the repair of radiation damage.[288−290]

These DNA nucleases that function in DNA repair and β-lactamase that functions in degrading antibiotics belong to the metallo-hydrolase superfamily (Table 4). Members of this superfamily carry out a variety of different types of chemical transformations, including thioester hydrolysis catalyzed by glyoxalase II and reduction of molecular oxygen by ROO oxidoreductase in addition to the phosphodiester hydrolysis and β-lactam ring-opening reactions mentioned above (Figure 1A). Although the locations of two divalent metal binding sites are generally conserved across the superfamily, the identities of the metal ligands have been altered so that different metal ions can be bound. For example, metallo-β-lactamases bind either one or two Zn^{2+} ions, glyoxalase II prefers to bind one Zn^{2+} and one Fe^{2+} ion, ROO binds two Fe^{2+} ions, and Artemis is believed to bind one or two Mg^{2+} ions.[294,296] The different properties of these divalent metal ions presumably allows for the great catalytic diversity within the metallo-hydrolase superfamily. The di-iron center of ROO binds to molecular oxygen and catalyzes its reduction to water. In the hydrolyases, a Zn^{2+} or Mg^{2+} ion is thought to coordinate the water molecule and activate it for nucleophilic attack, and a second divalent metal ion, if present, is proposed to stabilize the development of negative charge during the reaction.[297] The mechanistic diversity that is displayed by the metallo-β-lactamase superfamily is consistent with the idea that a preformed active site can be adapted to bind different cofactors and substrates and catalyze very different types of chemical transformations (Figure 1A).

7. Changes in Substrate and Reaction Specificity of DNA Repair Enzymes

There is little doubt that divergent evolution has played an important role in the evolution of DNA repair enzymes, because many of these enzymes belong to enzyme superfamilies. The functional and mechanistic diversity of these superfamilies is particularly remarkable (Table 4), suggesting the role of catalytic promiscuity in the past evolution of many DNA repair enzymes. Despite the expectations that DNA repair enzymes should be more specific than other enzymes, because side reactions that alter the structure of DNA could be lethal or mutagenic, there are a number of examples of DNA repair enzymes that exhibit catalytic promiscuity (Table 3). An even larger number of DNA repair enzymes exhibit broad substrate specificity and catalyze a given reaction with a remarkable variety of DNA substrates. For example, the DNA glycosylases discussed in Tables 1 and 2 catalyze N-glycosidic bond hydrolysis but accept substrates that differ in size, charge, and hydrogen-bonding ability. These properties of broad substrate specificity and catalytic promiscuity suggest that many contemporary DNA repair enzymes retain considerable evolutionary potential. Consistent with this notion, there are a number of examples from rational protein engineering and in vitro evolution in which substrate specificity and reaction specificity have been dramatically altered by one or a few mutations (Table 5).

These examples not only provide insight into the catalytic mechanism and structural basis for substrate selection but also demonstrate how mutation and selection can alter both reaction mechanism and specificity. Although there have been relatively few attempts to change the specificity of repair enzymes, there are examples from each of the biochemical classes of repair reactions (Table 5). Random mutagenesis of human AGT identified many different mutations that broaden the substrate specificity of AGT to include m^4T.[298] Interestingly, most of these mutations did not affect the activity toward m^6G. Guided by structures of enzyme−DNA complexes, a number of DNA glycosylases have had their specificities altered to accept different bases.[40,148,299−302] Both rational engineering and in vitro evolution have demonstrated that the strong specificity of DNA polymerases for dNTPs can be dramatically changed so that rNTPs or even 2′-*O*-methyl NTPs are accepted.[303−306]

The evolutionary relationships between mammalian DNase I and the structurally related AP endonuclease Xth from *E. coli* were substantiated by the finding that insertion of a conserved 14-amino-acid helix from Xth into DNase I confers a dramatic increase in AP endonuclease activity.[307] Another interesting example is the conversion of NaeI endonuclease into a topoisomerase by a single mutation.[308] There have also been several examples of changes in reaction mechanism. MutY has only very limited AP lyase activity and so functions predominantly as a monofunctional DNA glycosylase. Mutation of Ser120, the position that is a conserved lysine in bifunctional members of the HhH superfamily, confers robust AP lyase activity.[198] Similarly, the thymine DNA glycosylase from *Methanobacterium thermoautotrophicum* (Mth TDG) is a monofunctional DNA glycosylase, but mutation of an active site tyrosine to lysine changes it into an AP lyase, albeit at the expense of the glycosylase reaction.[309] Finally, mutations in the active site of Nfo revealed that the phosphatase and phosphodiesterase activities toward 3′-DNA ends could be decoupled. In contrast to the examples of conferring a new activity, it was shown that an active site mutation in this enzyme greatly decreased the catalytically promiscuous activity toward phosphomonoester hydrolysis without altering phosphodiester hydrolysis activity.[242]

The relative ease with which DNA repair activities have been altered and the large effects of single mutants that are sometimes observed suggest that for at least some enzymes there are evolutionary pathways by which new activities could be incrementally optimized. Most protein engineering experiments have been directed toward metabolic enzymes, and there are many examples that span the range of structural enzyme classes and of chemical reaction types.[11,13,14,16,310,311] As relatively few experiments have been directed toward changing the specificity of DNA repair enzymes, it is still too early to say whether DNA repair enzymes differ fundamentally from other metabolic enzymes. Nevertheless, the early indications suggest that DNA repair enzymes are no different than other enzymes with respect to their evolutionary potential. This has implications for the evolution of DNA repair responses to novel types of DNA damage and for the way in which the balance between repair and mutation might be altered (see below).

8. Summary: DNA Repair, a Highly Evolvable System for Safeguarding the Genome

As we bring together the results from decades of studying the biochemical mechanisms of DNA repair and more recent results from structural biology of DNA repair proteins, we begin to understand the evolutionary relationships among DNA repair enzymes and between DNA repair enzymes and other cellular enzymes. The identification and dissection of these evolutionary relationships can contribute in many ways to our understanding of biology. Most practically, this information helps us to predict and to understand the mechanisms of specific enzymes as key mechanistic features are often conserved during the evolutionary divergence of enzymatic function.[11,21,244,312] This information can also help to guide comparisons of DNA repair pathways between organisms.[190,313] More generally, understanding these evolutionary pathways gives insight into the current evolutionary potential of DNA repair systems. Together with the results from protein engineering and directed evolution, the latter is important for understanding carcinogenesis and the ways that cells may adapt to counteract new sources of DNA damage.

The realization that DNA repair enzymes have a rich evolutionary past provides the opportunity to apply mechanistic insight from one enzyme to the study of another. The recent progress in understanding both the AlkB and Artemis families of enzymes, based upon their homology with well-characterized enzymes, provides examples of how successfully the information provided by an evolutionary relationship can be applied to inferring function and even catalytic mechanism. However, given the remarkable changes in reaction mechanisms that have occurred (Figure 1 and Table 4), there is the growing realization that such hypotheses need to be rigorously tested with the appropriate biochemical experiments. Indeed, homology with a mechanistically diverse enzyme superfamily provides very little help in elucidating the identity of substrate(s) and underscores the importance of obtaining functional information.

As a biological system, DNA repair exhibits several of the hallmarks of an evolvable system. As evolution takes place on the level of a population, it is reasonable to consider that evolution has shaped a system that is inherently amenable to evolution. First, many DNA repair activities are redundant. This could allow greater flexibility in recruiting any given enzyme to a new function because gene duplication would not necessarily be required, as mutation of any one enzyme would be buffered by the activity of a functionally redundant enzyme. Second, the majority of DNA repair enzymes appear to have broad substrate specificity and reasonably large rate enhancements. This provides an abundant source of enzymes with low levels of promiscuous activities that could be improved in response to a biological selection for new or increased DNA repair activity. Finally, the relatively simple structure of DNA ensures that few types of catalytic reactions are required to repair any damage. As evolutionary pathways for changes in substrate specificity are likely to be more favorable than the pathways for creation of new catalytic mechanisms, this is expected to provide increased capacity for new DNA repair pathways. Thus, the catalytic promiscuity and broad substrate specificity of DNA repair enzymes are consistent with their apparently complex evolutionary histories and further suggest that there is considerable potential for future evolutionary diversification of enzymatic function in response to changing levels and sources of DNA damage.

9. Acknowledgments

I thank N. Bernstein, J.N.M. Glover, B. Eichman, and T. Ellenberger for providing high resolution images, Suzanne Admiraal for her support and comments, Jesse Zalatan and Dan Herschlag for critical reading of this manuscript, and the University of Michigan for financial support during the writing of this review.

10. References

(1) Drake, J. W.; Charlesworth, B.; Charlesworth, D.; Crow, J. F. *Genetics* **1998**, *148*, 1667.
(2) Chuang, J. H.; Li, H. *PLoS Biol.* **2004**, *2*, E29.
(3) Piatigorsky, J.; Wistow, G. J. *Cell* **1989**, *57*, 197.
(4) Piatigorsky, J.; Kantorow, M.; Gopal-Srivastava, R.; Tomarev, S. I. *EXS* **1994**, *71*, 241.
(5) Copley, S. D. *Curr. Opin. Chem. Biol.* **2003**, *7*, 265.
(6) Jeffery, C. J. *Trends Genet.* **2003**, *19*, 415.
(7) Kohn, K. W. *Mol. Biol. Cell* **1999**, *10*, 2703.
(8) Noirot, P.; Noirot-Gros, M. F. *Curr. Opin. Microbiol.* **2004**, *7*, 505.

(9) Coulombe, B.; Jeronimo, C.; Langelier, M. F.; Cojocaru, M.; Bergeron, D. *Mol. Cell Proteomics* **2004**, *3*, 851.
(10) Fan, J.; Wilson, D. M., III. *Free Radical Biol. Med.* **2005**, *38*, 1121.
(11) O'Brien, P. J.; Herschlag, D. *Chem. Biol.* **1999**, *6*, R91.
(12) Jensen, R. A. *Annu. Rev. Microbiol.* **1976**, *30*, 409.
(13) Bornscheuer, U. T.; Kazlauskas, R. J. *Angew Chem., Int. Ed.* **2004**, *43*, 6032.
(14) Penning, T. M.; Jez, J. M. *Chem. Rev.* **2001**, *101*, 3027.
(15) Wackett, L. P. *Ann. N.Y. Acad. Sci.* **1998**, *864*, 142.
(16) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S. M.; Roodveldt, C.; Tawfik, D. S. *Nat. Genet.* **2005**, *37*, 73.
(17) Fitch, W. M. *Syst. Zool.* **1970**, *19*, 99.
(18) Fitch, W. M. *Trends Genet.* **2000**, *16*, 227.
(19) Koonin, E. V.; Galperin, M. Y. *Sequence−Evolution−Function*; Kluwer Academic Publishers: Boston, 2003.
(20) Wilmanns, M.; Hyde, C. C.; Davies, D. R.; Kirschner, K.; Jansonius, J. N. *Biochemistry* **1991**, *30*, 9161.
(21) Gerlt, J. A.; Babbitt, P. C. *Annu. Rev. Biochem.* **2001**, *70*, 209.
(22) Murzin, A. G. *Trends Biochem. Sci.* **1993**, *18*, 403.
(23) Murzin, A. G. *Curr. Opin. Struct. Biol.* **1998**, *8*, 380.
(24) Petsko, G. A.; Kenyon, G. L.; Gerlt, J. A.; Ringe, D.; Kozarich, J. W. *Trends Biochem. Sci.* **1993**, *18*, 372.
(25) Ollis, D. L.; Cheah, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Franken, S. M.; Harel, M.; Remington, S. J.; Silman, I.; Schrag, J.; Sussman, J. L.; Vershueren, K. H. G.; Goldman, A. *Protein Eng.* **1992**, *5*, 197.
(26) Babbitt, P. C.; Mrachko, G. T.; Hasson, M. S.; Huisman, G. W.; Kolter, R.; Ringe, D.; Petsko, G. A.; Kenyon, G. L.; Gerlt, J. A. *Science* **1995**, *267*, 1159.
(27) Wolfe, K. H.; Shields, D. C. *Nature* **1997**, *387*, 708.
(28) Ohno, S. *Evolution by Gene and Genome Duplication*; Springer: Berlin, 1970.
(29) Prince, V. E.; Pickett, F. B. *Nat. Rev. Genet.* **2002**, *3*, 827.
(30) Lynch, M.; Conery, J. S. *Science* **2000**, *290*, 1151.
(31) Holland, S. K.; Blake, C. C. *Biosystems* **1987**, *20*, 181.
(32) Doolittle, R. F. *Annu. Rev. Biochem.* **1995**, *64*, 287.
(33) Todd, A. E.; Orengo, C. A.; Thornton, J. M. *J. Mol. Biol.* **2001**, *307*, 1113.
(34) Wackett, L. P. *J. Biol. Chem.* **2004**, *279*, 41259.
(35) Lindahl, T. *Nature* **1993**, *362*, 709.
(36) Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938.
(37) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90.
(38) Wolfenden, R.; Ridgway, C.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 833.
(39) Lindahl, T.; Nyberg, B. *Biochemistry* **1972**, *11*, 3610.
(40) O'Brien, P. J.; Ellenberger, T. *J. Biol. Chem.* **2004**, *279*, 9750.
(41) Shapiro, R.; Kang, S. *Biochemistry* **1969**, *8*, 1806.
(42) Shapiro, R.; Danzig, M. *Biochemistry* **1972**, *11*, 23.
(43) Lindahl, T.; Andersson, A. *Biochemistry* **1972**, *11*, 3618.
(44) Drohat, A. C.; Jagadeesh, J.; Ferguson, E.; Stivers, J. T. *Biochemistry* **1999**, *38*, 11866.
(45) Griffiths, A. D.; Tawfik, D. S. *Curr. Opin. Biotechnol.* **2000**, *11*, 338.
(46) Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. *Biochemistry* **1990**, *29*, 2532.
(47) Lindahl, T.; Nyberg, B. *Biochemistry* **1974**, *13*, 3405.
(48) Jo, K.; Topal, M. D. *Science* **1995**, *267*, 1817.
(49) Krogh, B. O.; Shuman, S. *Biochemistry* **2000**, *39*, 6422.
(50) Kwon, K.; Nagarajan, R.; Stivers, J. T. *Biochemistry* **2004**, *43*, 14994.
(51) Shuman, S. *Mol. Cell* **1998**, *1*, 741.
(52) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. *Nature* **1996**, *384*, 87.
(53) Lau, A. Y.; Scharer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, T. *Cell* **1998**, *95*, 249.
(54) Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A. *EMBO J.* **1998**, *17*, 5214.
(55) Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67*, 181.
(56) Hosfield, D. J.; Guan, Y.; Haas, B. J.; Cunningham, R. P.; Tainer, J. A. *Cell* **1999**, *98*, 397.
(57) Daniels, D. S.; Mol, C. D.; Arvai, A. S.; Kanugula, S.; Pegg, A. E.; Tainer, J. A. *EMBO J.* **2000**, *19*, 1719.
(58) Mees, A.; Klar, T.; Gnau, P.; Hennecke, U.; Eker, A. P.; Carell, T.; Essen, L. O. *Science* **2004**, *306*, 1789.
(59) Krosky, D. J.; Schwarz, F. P.; Stivers, J. T. *Biochemistry* **2004**, *43*, 4188.
(60) Krosky, D. J.; Song, F.; Stivers, J. T. *Biochemistry* **2005**, *44*, 5949.
(61) Berg, O. G.; Winter, R. B.; von Hippel, P. H. *Biochemistry* **1981**, *20*, 6929.
(62) von Hippel, P. H.; Berg, O. G. *J. Biol. Chem.* **1989**, *264*, 675.
(63) Kampmann, M. *J. Biol. Chem.* **2004**, *279*, 38715.
(64) Kampmann, M. *Mol. Microbiol.* **2005**, *57*, 889.
(65) Stanford, N. P.; Szczelkun, M. D.; Marko, J. F.; Halford, S. E. *EMBO J.* **2000**, *19*, 6546.
(66) Lloyd, R. S. *Mutat. Res.* **2005**, *577*, 77.
(67) Zharkov, D. O.; Grollman, A. P. *Mutat. Res.* **2005**, *577*, 24.
(68) Dowd, D. R.; Lloyd, R. S. *J. Biol. Chem.* **1990**, *265*, 3424.
(69) Dowd, D. R.; Lloyd, R. S. *J. Mol. Biol.* **1989**, *208*, 701.
(70) Jeltsch, A.; Wenz, C.; Stahl, F.; Pingoud, A. *EMBO J.* **1996**, *15*, 5104.
(71) Krokan, H. E.; Standal, R.; Slupphaug, G. *Biochem. J.* **1997**, *325*, 1.
(72) David, S. S.; Williams, S. D. *Chem. Rev.* **1998**, *98*, 1221.
(73) Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729.
(74) Hosfield, D. J.; Daniels, D. S.; Mol, C. D.; Putnam, C. D.; Parikh, S. S.; Tainer, J. A. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *68*, 315.
(75) Lindahl, T. *Nature* **1976**, *259*, 64.
(76) Lindahl, T.; Ljungquist, S.; Siegert, W.; Nyberg, B.; Sperens, B. *J. Biol. Chem.* **1977**, *252*, 3286.
(77) Kavli, B.; Sundheim, O.; Akbari, M.; Otterlei, M.; Nilsen, H.; Skorpen, F.; Aas, P. A.; Hagen, L.; Krokan, H. E.; Slupphaug, G. *J. Biol. Chem.* **2002**, *277*, 39926.
(78) Liu, P.; Burdzy, A.; Sowers, L. C. *Chem. Res. Toxicol.* **2002**, *15*, 1001.
(79) Aravind, L.; Koonin, E. V. *Genome Biol.* **2000**, *1*, 0007.1.
(80) Pearl, L. H. *Mutat Res.* **2000**, *460*, 165.
(81) Hitchcock, T. M.; Dong, L.; Connor, E. E.; Meira, L. B.; Samson, L. D.; Wyatt, M. D.; Cao, W. *J. Biol. Chem.* **2004**, *279*, 38177.
(82) Wyatt, M. D.; Allan, J. M.; Lau, A. Y.; Ellenberger, T. E.; Samson, L. D. *Bioessays* **1999**, *21*, 668.
(83) Zharkov, D. O.; Shoham, G.; Grollman, A. P. *DNA Repair* **2003**, *2*, 839.
(84) Labahn, J.; Scharer, O. D.; Long, A.; Ezaz-Nikpay, K.; Verdine, G. L.; Ellenberger, T. E. *Cell* **1996**, *86*, 321.
(85) Kuo, C. F.; McRee, D. E.; Fisher, C. L.; O'Handley, S. F.; Cunningham, R. P.; Tainer, J. A. *Science* **1992**, *258*, 434.
(86) Kuo, C. F.; Mol, C. D.; Thayer, M. M.; Cunningham, R. P.; Tainer, J. A. *Ann. N.Y. Acad. Sci.* **1994**, *726*, 223.
(87) Nash, H. M.; Bruner, S. D.; Scharer, O. D.; Kawate, T.; Addona, T. A.; Spooner, E.; Lane, W. S.; Verdine, G. L. *Curr. Biol.* **1996**, *6*, 968.
(88) O'Rourke, E. J.; Chevalier, C.; Boiteux, S.; Labigne, A.; Ielpi, L.; Radicella, J. P. *J. Biol. Chem.* **2000**, *275*, 20077.
(89) Eichman, B. F.; O'Rourke, E. J.; Radicella, J. P.; Ellenberger, T. *EMBO J.* **2003**, *22*, 4898.
(90) Oivanen, M.; Hovinen, J.; Lehikoinen, P.; Lonnberg, H. *Trends Org. Chem.* **1993**, *4*, 397.
(91) Dizdaroglu, M.; Bauche, C.; Rodriguez, H.; Laval, J. *Biochemistry* **2000**, *39*, 5586.
(92) Dizdaroglu, M.; Karakaya, A.; Jaruga, P.; Slupphaug, G.; Krokan, H. E. *Nucleic Acids Res.* **1996**, *24*, 418.
(93) Dizdaroglu, M. *Anal. Biochem.* **1985**, *144*, 593.
(94) Dizdaroglu, M. *Mutat. Res.* **2003**, *531*, 109.
(95) Dizdaroglu, M.; Jaruga, P.; Birincioglu, M.; Rodriguez, H. *Free Radical Biol. Med.* **2002**, *32*, 1102.
(96) Kow, Y. W. *Biochemistry* **1989**, *28*, 3280.
(97) Ischenko, A. A.; Saparbaev, M. K. *Nature* **2002**, *415*, 183.
(98) Ishchenko, A. A.; Sanz, G.; Privezentzev, C. V.; Maksimenko, A. V.; Saparbaev, M. *Nucleic Acids Res.* **2003**, *31*, 6344.
(99) Gros, L.; Ishchenko, A. A.; Ide, H.; Elder, R. H.; Saparbaev, M. K. *Nucleic Acids Res.* **2004**, *32*, 73.
(100) Ishchenko, A. A.; Ide, H.; Ramotar, D.; Nevinsky, G.; Saparbaev, M. *Biochemistry* **2004**, *43*, 15210.
(101) Steitz, T. A.; Yin, Y. W. *Philos. Trans. R Soc. London, Ser. B* **2004**, *359*, 17.
(102) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191.
(103) Pages, V.; Fuchs, R. P. *Oncogene* **2002**, *21*, 8957.
(104) Baynton, K.; Fuchs, R. P. *Trends Biochem. Sci.* **2000**, *25*, 74.
(105) Prakash, S.; Johnson, R. E.; Prakash, L. *Annu. Rev. Biochem.* **2005**, *74*, 317.
(106) Lindahl, T.; Barnes, D. E. *Annu. Rev. Biochem.* **1992**, *61*, 251.
(107) Tomkinson, A. E.; Tappe, N. J.; Friedberg, E. C. *Biochemistry* **1992**, *31*, 1, 11762.
(108) Sriskanda, V.; Shuman, S. *Nucleic Acids Res.* **1998**, *26*, 3536.
(109) Shuman, S. *Biochemistry* **1995**, *34*, 16138.
(110) Bhagwat, A. S.; Sanderson, R. J.; Lindahl, T. *Nucleic Acids Res.* **1999**, *27*, 4028.
(111) Lehman, I. R. *Science* **1974**, *186*, 790.
(112) Pascal, J. M.; O'Brien, P. J.; Tomkinson, A. E.; Ellenberger, T. *Nature* **2004**, *432*, 473.
(113) Bedows, E.; Wachsman, J. T.; Gumport, R. I. *Biochemistry* **1977**, *16*, 2231.
(114) Ho, C. K.; Van Etten, J. L.; Shuman, S. *J. Virol.* **1997**, *71*, 1931.
(115) Reardon, J. T.; Sancar, A. *Prog. Nucleic Acid Res. Mol. Biol.* **2005**, *79*, 183.
(116) Sancar, A.; Reardon, J. T. *Adv. Protein Chem.* **2004**, *69*, 43.
(117) Hanawalt, P. C. *Oncogene* **2002**, *21*, 8949.

(118) Hanawalt, P. C.; Ford, J. M.; Lloyd, D. R. *Mutat. Res.* **2003**, *544*, 107.
(119) Sancar, A.; Lindsey-Boltz, L. A.; Unsal-Kacmaz, K.; Linn, S. *Annu. Rev. Biochem.* **2004**, *73*, 39.
(120) Sancar, A.; Rupp, W. D. *Cell* **1983**, *33*, 249.
(121) Huang, J. C.; Svoboda, D. L.; Reardon, J. T.; Sancar, A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3664.
(122) Mu, D.; Hsu, D. S.; Sancar, A. *J. Biol. Chem.* **1996**, *271*, 8285.
(123) Modrich, P. *J. Biol. Chem.* **1997**, *272*, 24727.
(124) Jiricny, J. *EMBO J.* **1998**, *17*, 6427.
(125) Kunkel, T. A.; Erie, D. A. *Annu. Rev. Biochem.* **2004**.
(126) Gros, L.; Saparbaev, M. K.; Laval, J. *Oncogene* **2002**, *21*, 8905.
(127) Hang, B.; Chenna, A.; Sagi, J.; Singer, B. *Carcinogenesis* **1998**, *19*, 1339.
(128) Hang, B.; Chenna, A.; Fraenkel-Conrat, H.; Singer, B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13737.
(129) Hang, B.; Rothwell, D. G.; Sagi, J.; Hickson, I. D.; Singer, B. *Biochemistry* **1997**, *36*, 15411.
(130) Sancar, A. *Chem. Rev.* **2003**, *103*, 2203.
(131) Weber, S. *Biochim. Biophys. Acta* **2005**, *1707*, 1.
(132) Torizawa, T.; Ueda, T.; Kuramitsu, S.; Hitomi, K.; Todo, T.; Iwai, S.; Morikawa, K.; Shimada, I. *J. Biol. Chem.* **2004**, *279*, 32950.
(133) Samson, L. *Mol. Microbiol.* **1992**, *6*, 825.
(134) Drablos, F.; Feyzi, E.; Aas, P. A.; Vaagbo, C. B.; Kavli, B.; Bratlie, M. S.; Pena-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. *DNA Repair* **2004**, *3*, 1389.
(135) He, C.; Wei, H.; Verdine, G. L. *J. Am. Chem. Soc.* **2003**, *125*, 1450.
(136) Sedgwick, B.; Lindahl, T. *Oncogene* **2002**, *21*, 8886.
(137) Falnes, P. O. *Nucleic Acids Res.* **2004**, *32*, 6260.
(138) Falnes, P. O.; Bjoras, M.; Aas, P. A.; Sundheim, O.; Seeberg, E. *Nucleic Acids Res.* **2004**, *32*, 3456.
(139) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. *Nature* **2002**, *419*, 174.
(140) Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16660.
(141) Falnes, P. O.; Johansen, R. F.; Seeberg, E. *Nature* **2002**, *419*, 178.
(142) Aas, P. A.; Otterlei, M.; Falnes, P. O.; Vagbo, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjoras, M.; Slupphaug, G.; Seeberg, E.; Krokan, H. E. *Nature* **2003**, *421*, 859.
(143) Mishina, Y.; Yang, C. G.; He, C. *J. Am. Chem. Soc.* **2005**, *127*, 14594.
(144) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 855.
(145) Doherty, A. J.; Suh, S. W. *Nucleic Acids Res.* **2000**, *28*, 4051.
(146) Berdal, K. G.; Johansen, R. F.; Seeberg, E. *EMBO J.* **1998**, *17*, 363.
(147) Branum, M. E.; Reardon, J. T.; Sancar, A. *J. Biol. Chem.* **2001**, *276*, 25421.
(148) Connor, E. E.; Wyatt, M. D. *Chem. Biol.* **2002**, *9*, 1033.
(149) O'Brien, P. J.; Ellenberger, T. *J. Biol. Chem.* **2004**, *279*, 26876.
(150) Hanawalt, P. C. *Mutat Res.* **2001**, *485*, 3.
(151) Posnick, L. M.; Samson, L. D. *J. Bacteriol.* **1999**, *181*, 6763.
(152) Warner, H. R.; Rockstroh, P. A. *J. Bacteriol.* **1980**, *141*, 680.
(153) Chou, K. M.; Cheng, Y. C. *Nature* **2002**, *415*, 655.
(154) Chepanoske, C. L.; Langelier, C. R.; Chmiel, N. H.; David, S. S. *Org. Lett.* **2000**, *2*, 1341.
(155) Biswas, T.; Clos, L. J., II; SantaLucia, J., Jr.; Mitra, S.; Roy, R. *J. Mol. Biol.* **2002**, *320*, 503.
(156) Vallur, A. C.; Feller, J. A.; Abner, C. W.; Tran, R. K.; Bloom, L. B. *J. Biol. Chem.* **2002**, *277*, 31673.
(157) Valinluck, V.; Liu, P.; Burdzy, A.; Ryu, J.; Sowers, L. C. *Chem. Res. Toxicol.* **2002**, *15*, 1595.
(158) Abner, C. W.; Lau, A. Y.; Ellenberger, T.; Bloom, L. B. *J. Biol. Chem.* **2001**, *276*, 13379.
(159) Richardson, C. C.; Kornberg, A. *J. Biol. Chem.* **1964**, *239*, 242.
(160) Sammartano, L. J.; Tuveson, R. W. *J. Bacteriol.* **1983**, *156*, 904.
(161) Demple, B.; Halbrook, J.; Linn, S. *J. Bacteriol.* **1983**, *153*, 1079.
(162) Demple, B.; Johnson, A.; Fung, D. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7731.
(163) Demple, B.; Harrison, L. *Annu. Rev. Biochem.* **1994**, *63*, 915.
(164) Yajko, D. M.; Weiss, B. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 688.
(165) Rogers, S. G.; Weiss, B. *Methods Enzymol.* **1980**, *65*, 201.
(166) Kow, Y. W.; Wallace, S. S. *Biochemistry* **1987**, *26*, 8200.
(167) Keller, W.; Crouch, R. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 3360.
(168) Weiss, B. *J. Biol. Chem.* **1976**, *251*, 1896.
(169) Thatcher, G.; Kluger, R. *Adv. Phys. Org. Chem.* **1989**, *25*, 99.
(170) Hengge, A. C. In *Comprehensive Biological Catalysis: A Mechanistic Reference*; Sinnott, M., Ed.; Academic: New York, 1998; Vol. 1.
(171) Cassano, A. G.; Anderson, V. E.; Harris, M. E. *Biopolymers* **2004**, *73*, 110.
(172) Mol, C. D.; Kuo, C. F.; Thayer, M. M.; Cunningham, R. P.; Tainer, J. A. *Nature* **1995**, *374*, 381.
(173) Suck, D. *J. Mol. Recognit.* **1994**, *7*, 65.
(174) Beernink, P. T.; Segelke, B. W.; Hadi, M. Z.; Erzberger, J. P.; Wilson, D. M., III; Rupp, B. *J. Mol. Biol.* **2001**, *307*, 1023.
(175) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A. *Nature* **2000**, *403*, 451.
(176) Gorman, M. A.; Morera, S.; Rothwell, D. G.; de La Fortelle, E.; Mol, C. D.; Tainer, J. A.; Hickson, I. D.; Freemont, P. S. *EMBO J.* **1997**, *16*, 6548.
(177) Feng, Q.; Moran, J. V.; Kazazian, H. H., Jr.; Boeke, J. D. *Cell* **1996**, *87*, 905.
(178) Dlakic, M. *Trends Biochem. Sci.* **2000**, *25*, 272.
(179) Maita, N.; Anzai, T.; Aoyagi, H.; Mizuno, H.; Fujiwara, H. *J. Biol. Chem.* **2004**, *279*, 41067.
(180) Weichenrieder, O.; Repanas, K.; Perrakis, A. *Structure* **2004**, *12*, 975.
(181) Hwang, K. Y.; Baek, K.; Kim, H. Y.; Cho, Y. *Nat. Struct. Biol.* **1998**, *5*, 707.
(182) Sakurai, S.; Kitano, K.; Yamaguchi, H.; Hamada, K.; Okada, K.; Fukuda, K.; Uchida, M.; Ohtsuka, E.; Morioka, H.; Hakoshima, T. *EMBO J.* **2005**, *24*, 683.
(183) Chapados, B. R.; Hosfield, D. J.; Han, S.; Qiu, J.; Yelent, B.; Shen, B.; Tainer, J. A. *Cell* **2004**, *116*, 39.
(184) Hosfield, D. J.; Mol, C. D.; Shen, B.; Tainer, J. A. *Cell* **1998**, *95*, 135.
(185) Lieber, M. R. *Bioessays* **1997**, *19*, 233.
(186) Hofmann, K.; Tomiuk, S.; Wolff, G.; Stoffel, W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5895.
(187) Whisstock, J. C.; Romero, S.; Gurung, R.; Nandurkar, H.; Ooms, L. M.; Bottomley, S. P.; Mitchell, C. A. *J. Biol. Chem.* **2000**, *275*, 37055.
(188) Tsujishita, Y.; Guo, S.; Stolz, L. E.; York, J. D.; Hurley, J. H. *Cell* **2001**, *105*, 379.
(189) Ide, H.; Tedzuka, K.; Shimzu, H.; Kimura, Y.; Purmal, A. A.; Wallace, S. S.; Kow, Y. W. *Biochemistry* **1994**, *33*, 7842.
(190) Aravind, L.; Walker, D. R.; Koonin, E. V. *Nucleic Acids Res.* **1999**, *27*, 1223.
(191) Carrell, H. L.; Glusker, J. P.; Burger, V.; Manfre, F.; Tritsch, D.; Biellmann, J. F. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4440.
(192) Guan, Y.; Manuel, R. C.; Arvai, A. S.; Parikh, S. S.; Mol, C. D.; Miller, J. H.; Lloyd, S.; Tainer, J. A. *Nat. Struct. Biol.* **1998**, *5*, 1058.
(193) Zharkov, D. O.; Grollman, A. P. *Biochemistry* **1998**, *37*, 12384.
(194) Manuel, R. C.; Hitomi, K.; Arvai, A. S.; House, P. G.; Kurtz, A. J.; Dodson, M. L.; McCullough, A. K.; Tainer, J. A.; Lloyd, R. S. *J. Biol. Chem.* **2004**, *279*, 46930.
(195) Williams, S. D.; David, S. S. *Nucleic Acids Res.* **1998**, *26*, 5123.
(196) Williams, S. D.; David, S. S. *Biochemistry* **1999**, *38*, 15417.
(197) Pope, M. A.; Porello, S. L.; David, S. S. *J. Biol. Chem.* **2002**, *277*, 22605.
(198) Williams, S. D.; David, S. S. *Biochemistry* **2000**, *39*, 10098.
(199) Li, H.; Xu, H.; Graham, D. E.; White, R. H. *J. Biol. Chem.* **2003**, *278*, 11100.
(200) Bjornberg, O.; Neuhard, J.; Nyman, P. O. *J. Biol. Chem.* **2003**, *278*, 20667.
(201) Johansson, E.; Fano, M.; Bynck, J. H.; Neuhard, J.; Larsen, S.; Sigurskjold, B. W.; Christensen, U.; Willemoes, M. *J. Biol. Chem.* **2005**, *280*, 3051.
(202) Shen, J. C.; Rideout, W. M., III; Jones, P. A. *Cell* **1992**, *71*, 1073.
(203) Yebra, M. J.; Bhagwat, A. S. *Biochemistry* **1995**, *34*, 14752.
(204) Zingg, J. M.; Shen, J. C.; Jones, P. A. *Biochem. J.* **1998**, *332*, 223.
(205) Jeltsch, A.; Christ, F.; Fatemi, M.; Roth, M. *J. Biol. Chem.* **1999**, *274*, 19538.
(206) Jeltsch, A. *Biol. Chem.* **2001**, *382*, 707.
(207) Cohen, H. M.; Tawfik, D. S.; Griffiths, A. D. *Nucleic Acids Res.* **2002**, *30*, 3880.
(208) Cohen, H. M.; Tawfik, D. S.; Griffiths, A. D. *Protein Eng. Des. Sel.* **2004**, *17*, 3.
(209) Pogolotti, A. L., Jr.; Ono, A.; Subramaniam, R.; Santi, D. V. *J. Biol. Chem.* **1988**, *263*, 7461.
(210) Reich, N. O.; Mashhoon, N. *Biochemistry* **1991**, *30*, 2933.
(211) Santi, D. V.; Garrett, C. E.; Barr, P. J. *Cell* **1983**, *33*, 9.
(212) Santi, D. V.; Norment, A.; Garrett, C. E. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6993.
(213) Bestor, T. H.; Verdine, G. L. *Curr. Opin. Cell Biol.* **1994**, *6*, 380.
(214) Bandaru, B.; Gopal, J.; Bhagwat, A. S. *J. Biol. Chem.* **1996**, *271*, 7851.
(215) Shen, J. C.; Zingg, J. M.; Yang, A. S.; Schmutte, C.; Jones, P. A. *Nucleic Acids Res.* **1995**, *23*, 4275.
(216) Daniels, D. S.; Woo, T. T.; Luu, K. X.; Noll, D. M.; Clarke, N. D.; Pegg, A. E.; Tainer, J. A. *Nat. Struct. Mol. Biol.* **2004**, *11*, 714.
(217) Guengerich, F. P.; Fang, Q.; Liu, L.; Hachey, D. L.; Pegg, A. E. *Biochemistry* **2003**, *42*, 10965.
(218) Liu, H.; Xu-Welliver, M.; Pegg, A. E. *Mutat. Res.* **2000**, *452*, 1.
(219) Liu, L.; Pegg, A. E.; Williams, K. M.; Guengerich, F. P. *J. Biol. Chem.* **2002**, *277*, 37920.
(220) Liu, L.; Williams, K. M.; Guengerich, F. P.; Pegg, A. E. *Chem. Res. Toxicol.* **2004**, *17*, 742.

(221) Valadez, J. G.; Liu, L.; Loktionova, N. A.; Pegg, A. E.; Guengerich, F. P. *Chem. Res. Toxicol.* **2004**, *17*, 972.
(222) Minko, I. G.; Kurtz, A. J.; Croteau, D. L.; Van Houten, B.; Harris, T. M.; Lloyd, R. S. *Biochemistry* **2005**, *44*, 3000.
(223) Minko, I. G.; Zou, Y.; Lloyd, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1905.
(224) Davies, D. R.; Interthal, H.; Champoux, J. J.; Hol, W. G. *Structure* **2002**, *10*, 237.
(225) Pouliot, J. J.; Robertson, C. A.; Nash, H. A. *Genes Cells* **2001**, *6*, 677.
(226) Deng, C.; Brown, J. A.; You, D.; Brown, J. M. *Genetics* **2005**.
(227) Guengerich, F. P. *Arch. Biochem. Biophys.* **2005**, *433*, 369.
(228) Arabshahi, A.; Frey, P. A. *J. Biol. Chem.* **1999**, *274*, 8586.
(229) Palm, D.; Klein, H. W.; Schinzel, R.; Buehner, M.; Helmreich, E. J. *Biochemistry* **1990**, *29*, 1099.
(230) Helmreich, E. J. *Biofactors* **1992**, *3*, 159.
(231) Johnson, L. N.; Hu, S. H.; Barford, D. *Faraday Discuss.* **1992**, 131.
(232) Bogenhagen, D. F.; Pinz, K. G. *J. Biol. Chem.* **1998**, *273*, 7888.
(233) Betti, M.; Petrucco, S.; Bolchi, A.; Dieci, G.; Ottonello, S. *J. Biol. Chem.* **2001**, *276*, 18038.
(234) Postel, E. H. *J. Bioenerg. Biomembr.* **2003**, *35*, 31.
(235) Postel, E. H.; Abramczyk, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13247.
(236) Vance, J. R.; Wilson, T. E. *J. Biol. Chem.* **2001**, *276*, 15073.
(237) Kumar, P.; Krishna, K.; Srinivasan, R.; Ajitkumar, P.; Varshney, U. *DNA Repair* **2004**, *3*, 1483.
(238) Bennett, S. E.; Chen, C. Y.; Mosbaugh, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6391.
(239) Meyer-Siegler, K.; Mauro, D. J.; Seal, G.; Wurzer, J.; deRiel, J. K.; Sirover, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8460.
(240) Wang, X.; Sirover, M. A.; Anderson, L. E. *Arch. Biochem. Biophys.* **1999**, *367*, 348.
(241) O'Brien, P. J.; Herschlag, D. *Biochemistry* **2001**, *40*, 5691.
(242) Izumi, T.; Ishizaki, K.; Ikenaga, M.; Yonei, S. *J. Bacteriol.* **1992**, *174*, 7711.
(243) O'Brien, P. J.; Herschlag, D. *Biochemistry* **2002**, *41*, 3207.
(244) Gerlt, J. A.; Babbitt, P. C.; Rayment, I. *Arch. Biochem. Biophys.* **2005**, *433*, 59.
(245) Schmidt, D. M.; Mundorff, E. C.; Dojka, M.; Bermudez, E.; Ness, J. E.; Govindarajan, S.; Babbitt, P. C.; Minshull, J.; Gerlt, J. A. *Biochemistry* **2003**, *42*, 8387.
(246) Huffman, J. L.; Sundheim, O.; Tainer, J. A. *Mutat. Res.* **2005**, *577*, 55.
(247) Day, R.; Beck, D. A.; Armen, R. S.; Daggett, V. *Protein Sci.* **2003**, *12*, 2150.
(248) Holm, L.; Sander, C. *J. Mol. Biol.* **1993**, *233*, 123.
(249) Holm, L.; Sander, C. *Nucleic Acids Res.* **1996**, *24*, 206.
(250) Dietmann, S.; Park, J.; Notredame, C.; Heger, A.; Lappe, M.; Holm, L. *Nucleic Acids Res.* **2001**, *29*, 55.
(251) Pearl, F.; Todd, A.; Sillitoe, I.; Dibley, M.; Redfern, O.; Lewis, T.; Bennett, C.; Marsden, R.; Grant, A.; Lee, D.; Akpor, A.; Maibaum, M.; Harrison, A.; Dallman, T.; Reeves, G.; Diboun, I.; Addou, S.; Lise, S.; Johnston, C.; Sillero, A.; Thornton, J.; Orengo, C. *Nucleic Acids Res.* **2005**, *33*, D247.
(252) Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure* **1997**, *5*, 1093.
(253) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536.
(254) Andreeva, A.; Howorth, D.; Brenner, S. E.; Hubbard, T. J.; Chothia, C.; Murzin, A. G. *Nucleic Acids Res.* **2004**, *32*, D226.
(255) Drohat, A. C.; Kwon, K.; Krosky, D. J.; Stivers, J. T. *Nat. Struct. Biol.* **2002**, *9*, 659.
(256) Lingaraju, G. M.; Sartori, A. A.; Kostrewa, D.; Prota, A. E.; Jiricny, J.; Winkler, F. K. *Structure* **2005**, *13*, 87.
(257) Sartori, A. A.; Lingaraju, G. M.; Hunziker, P.; Winkler, F. K.; Jiricny, J. *Nucleic Acids Res.* **2004**, *32*, 6531.
(258) Boon, E. M.; Livingston, A. L.; Chmiel, N. H.; David, S. S.; Barton, J. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12543.
(259) Lukianova, O. A.; David, S. S. *Curr. Opin. Chem. Biol.* **2005**, *9*, 145.
(260) Galburt, E. A.; Pelletier, J.; Wilson, G.; Stoddard, B. L. *Structure* **2002**, *10*, 1249.
(261) Wang, L. K.; Lima, C. D.; Shuman, S. *EMBO J.* **2002**, *21*, 3873.
(262) Eastberg, J. H.; Pelletier, J.; Stoddard, B. L. *Nucleic Acids Res.* **2004**, *32*, 653.
(263) Bernstein, N. K.; Williams, R. S.; Rakovszky, M. L.; Cui, D.; Green, R.; Karimi-Busheri, F.; Mani, R. S.; Galicia, S.; Koch, C. A.; Cass, C. E.; Durocher, D.; Weinfeld, M.; Glover, J. N. *Mol. Cell* **2005**, *17*, 657.
(264) Koonin, E. V.; Tatusov, R. L. *J. Mol. Biol.* **1994**, *244*, 125.
(265) Aravind, L.; Galperin, M. Y.; Koonin, E. V. *Trends Biochem. Sci.* **1998**, *23*, 127.
(266) Wang, W.; Cho, H. S.; Kim, R.; Jancarik, J.; Yokota, H.; Nguyen, H. H.; Grigoriev, I. V.; Wemmer, D. E.; Kim, S. H. *J. Mol. Biol.* **2002**, *319*, 421.
(267) Wang, W.; Kim, R.; Jancarik, J.; Yokota, H.; Kim, S. H. *Structure* **2001**, *9*, 65.
(268) Toyoshima, C.; Mizutani, T. *Nature* **2004**, *430*, 529.
(269) Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. *Nature* **2000**, *405*, 647.
(270) Kamenski, T.; Heilmeier, S.; Meinhart, A.; Cramer, P. *Mol. Cell* **2004**, *15*, 399.
(271) Lahiri, S. D.; Zhang, G.; Dunaway-Mariano, D.; Allen, K. N. *Biochemistry* **2002**, *41*, 8351.
(272) Singh, S. K.; Yang, K.; Karthikeyan, S.; Huynh, T.; Zhang, X.; Phillips, M. A.; Zhang, H. *J. Biol. Chem.* **2004**, *279*, 13166.
(273) Lahiri, S. D.; Zhang, G.; Dai, J.; Dunaway-Mariano, D.; Allen, K. N. *Biochemistry* **2004**, *43*, 2812.
(274) Allen, K. N.; Dunaway-Mariano, D. *Trends Biochem. Sci.* **2004**, *29*, 495.
(275) Morais, M. C.; Zhang, W.; Baker, A. S.; Zhang, G.; Dunaway-Mariano, D.; Allen, K. N. *Biochemistry* **2000**, *39*, 10385.
(276) Liu, J. Q.; Kurihara, T.; Miyagi, M.; Esaki, N.; Soda, K. *J. Biol. Chem.* **1995**, *270*, 18309.
(277) Dahms, A. S.; Kanazawa, T.; Boyer, P. D. *J. Biol. Chem.* **1973**, *248*, 6592.
(278) Peliska, J. A.; O'Leary, M. H. *Biochemistry* **1989**, *28*, 1604.
(279) Tatusov, R. L.; Galperin, M. Y.; Natale, D. A.; Koonin, E. V. *Nucleic Acids Res.* **2000**, *28*, 33.
(280) Tatusov, R. L.; Fedorova, N. D.; Jackson, J. D.; Jacobs, A. R.; Kiryutin, B.; Koonin, E. V.; Krylov, D. M.; Mazumder, R.; Mekhedov, S. L.; Nikolskaya, A. N.; Rao, B. S.; Smirnov, S.; Sverdlov, A. V.; Vasudevan, S.; Wolf, Y. I.; Yin, J. J.; Natale, D. A. *BMC Bioinf.* **2003**, *4*, 41.
(281) Heringa, J.; Taylor, W. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 416.
(282) Lindahl, T.; Sedgwick, B.; Sekiguchi, M.; Nakabeppu, Y. *Annu. Rev. Biochem.* **1988**, *57*, 133.
(283) Dinglay, S.; Trewick, S. C.; Lindahl, T.; Sedgwick, B. *Genes Dev.* **2000**, *14*, 2097.
(284) Aravind, L.; Koonin, E. V. *Genome Biol.* **2001**, *2*, 0007.1.
(285) Kurowski, M. A.; Bhagwat, A. S.; Papaj, G.; Bujnicki, J. M. *BMC Genomics* **2003**, *4*, 48.
(286) Ougland, R.; Zhang, C. M.; Liiv, A.; Johansen, R. F.; Seeberg, E.; Hou, Y. M.; Remme, J.; Falnes, P. O. *Mol. Cell* **2004**, *16*, 107.
(287) Trewick, S. C.; McLaughlin, P. J.; Allshire, R. C. *EMBO Rep.* **2005**, *6*, 315.
(288) Le Deist, F.; Poinsignon, C.; Moshous, D.; Fischer, A.; de Villartay, J. P. *Immunol. Rev.* **2004**, *200*, 142.
(289) Dudasova, Z.; Chovanec, M. *Neoplasma* **2003**, *50*, 311.
(290) Moshous, D.; Callebaut, I.; de Chasseval, R.; Poinsignon, C.; Villey, I.; Fischer, A.; de Villartay, J. P. *Ann. N.Y. Acad. Sci.* **2003**, *987*, 150.
(291) Moshous, D.; Callebaut, I.; de Chasseval, R.; Corneo, B.; Cavazzana-Calvo, M.; Le Deist, F.; Tezcan, I.; Sanal, O.; Bertrand, Y.; Philippe, N.; Fischer, A.; de Villartay, J. P. *Cell* **2001**, *105*, 177.
(292) Callebaut, I.; Moshous, D.; Mornon, J. P.; de Villartay, J. P. *Nucleic Acids Res.* **2002**, *30*, 3592.
(293) Aravind, L. *In Silico Biol.* **1999**, *1*, 69.
(294) Ma, Y.; Pannicke, U.; Schwarz, K.; Lieber, M. R. *Cell* **2002**, *108*, 781.
(295) Pannicke, U.; Ma, Y.; Hopfner, K. P.; Niewolik, D.; Lieber, M. R.; Schwarz, K. *EMBO J.* **2004**, *23*, 1987.
(296) Gomes, C. M.; Frazao, C.; Xavier, A. V.; Legall, J.; Teixeira, M. *Protein Sci.* **2002**, *11*, 707.
(297) Wang, Z.; Fast, W.; Valentine, A. M.; Benkovic, S. J. *Curr. Opin. Chem. Biol.* **1999**, *3*, 614.
(298) Encell, L. P.; Loeb, L. A. *Biochemistry* **1999**, *38*, 12097.
(299) Kavli, B.; Slupphaug, G.; Mol, C. D.; Arvai, A. S.; Peterson, S. B.; Tainer, J. A.; Krokan, H. E. *EMBO J.* **1996**, *15*, 3442.
(300) Handa, P.; Acharya, N.; Varshney, U. *Nucleic Acids Res.* **2002**, *30*, 3086.
(301) Kwon, K.; Jiang, Y. L.; Stivers, J. T. *Chem. Biol.* **2003**, *10*, 351.
(302) Fondufe-Mittendorf, Y. N.; Harer, C.; Kramer, W.; Fritz, H. J. *Nucleic Acids Res.* **2002**, *30*, 614.
(303) Xia, G.; Chen, L.; Sera, T.; Fa, M.; Schultz, P. G.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6597.
(304) Fa, M.; Radeghieri, A.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 1748.
(305) Astatke, M.; Ng, K.; Grindley, N. D.; Joyce, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3402.
(306) Joyce, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1619.
(307) Cal, S.; Tan, K. L.; McGregor, A.; Connolly, B. A. *EMBO J.* **1998**, *17*, 7128.
(308) Jo, K.; Topal, M. D. *Nucleic Acids Res.* **1996**, *24*, 4171.
(309) Begley, T. J.; Cunningham, R. P. *Protein Eng.* **1999**, *12*, 333.

(310) Shao, Z.; Arnold, F. H. *Curr. Opin. Struct. Biol.* **1996**, *6*, 513.
(311) Collins, C. H.; Yokobayashi, Y.; Umeno, D.; Arnold, F. H. *Curr. Opin. Biotechnol.* **2003**, *14*, 665.
(312) Gerlt, J. A.; Raushel, F. M. *Curr. Opin. Chem. Biol.* **2003**, *7*, 252.
(313) Eisen, J. A.; Hanawalt, P. C. *Mutat. Res.* **1999**, *435*, 171.
(314) Hoseki, J.; Okamoto, A.; Masui, R.; Shibata, T.; Inoue, Y.; Yokoyama, S.; Kuramitsu, S. *J. Mol. Biol.* **2003**, *333*, 515.
(315) O'Neill, R. J.; Vorob'eva, O. V.; Shahbakhti, H.; Zmuda, E.; Bhagwat, A. S.; Baldwin, G. S. *J. Biol. Chem.* **2003**, *278*, 20526.
(316) Yoon, J. H.; Iwai, S.; O'Connor, T. R.; Pfeifer, G. P. *Nucleic Acids Res.* **2003**, *31*, 5399.
(317) Sartori, A. A.; Fitz-Gibbon, S.; Yang, H.; Miller, J. H.; Jiricny, J. *EMBO J.* **2002**, *21*, 3182.
(318) Matsubara, M.; Masaoka, A.; Tanaka, T.; Miyano, T.; Kato, N.; Terato, H.; Ohyama, Y.; Iwai, S.; Ide, H. *Biochemistry* **2003**, *42*, 4993.
(319) Matsubara, M.; Masaoka, A.; Tanaka, T.; Terato, H.; Ohyama, Y.; Ide, H. *Nucleic Acids Res. Suppl.* **2003**, 233.
(320) Boorstein, R. J.; Cummings, A., Jr.; Marenstein, D. R.; Chan, M. K.; Ma, Y.; Neubert, T. A.; Brown, S. M.; Teebor, G. W. *J. Biol. Chem.* **2001**, *276*, 41991.
(321) Masaoka, A.; Matsubara, M.; Hasegawa, R.; Tanaka, T.; Kurisu, S.; Terato, H.; Ohyama, Y.; Karino, N.; Matsuda, A.; Ide, H. *Biochemistry* **2003**, *42*, 5003.
(322) Masaoka, A.; Matsubara, M.; Tanaka, T.; Terato, H.; Ohyama, Y.; Kubo, K.; Ide, H. *Nucleic Acids Res. Suppl.* **2003**, 263.
(323) Henshaw, T. F.; Feig, M.; Hausinger, R. P. *J. Inorg. Biochem.* **2004**, *98*, 856.
(324) Liu, L.; Hachey, D. L.; Valadez, G.; Williams, K. M.; Guengerich, F. P.; Loktionova, N. A.; Kanugula, S.; Pegg, A. E. *J. Biol. Chem.* **2004**, *279*, 4250.
(325) Krokan, H.; Grafstrom, R. C.; Sundqvist, K.; Esterbauer, H.; Harris, C. C. *Carcinogenesis* **1985**, *6*, 1755.
(326) Tomkinson, A. E.; Totty, N. F.; Ginsburg, M.; Lindahl, T. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 400.
(327) Jilani, A.; Vongsamphanh, R.; Leduc, A.; Gros, L.; Saparbaev, M.; Ramotar, D. *Biochemistry* **2003**, *42*, 6436.
(328) Vance, J. R.; Wilson, T. E. *Mol. Cell. Biol.* **2001**, *21*, 7191.
(329) Daiyasu, H.; Osaka, K.; Ishino, Y.; Toh, H. *FEBS Lett.* **2001**, *503*, 1.
(330) Hausinger, R. P. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21.
(331) Wright, P. M.; Yu, J.; Cillo, J.; Lu, A. L. *J. Biol. Chem.* **1999**, *274*, 29011.
(332) Francis, A. W.; Helquist, S. A.; Kool, E. T.; David, S. S. *J. Am. Chem. Soc.* **2003**, *125*, 16235.
(333) O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418.

CR040481V