Cell. Mol. Life Sci. (2010) 67:3573–3587
DOI 10.1007/s00018-010-0485-5

Cellular and Molecular Life Sciences

MULTI-AUTHOR REVIEW

Functions of disordered regions in mammalian early base excision repair proteins

Muralidhar L. Hegde · Tapas K. Hazra · Sankar Mitra

Received: 28 July 2010 / Accepted: 28 July 2010 / Published online: 17 August 2010

Abstract Reactive oxygen species, generated endogenously and induced as a toxic response, produce several dozen oxidized or modified bases and/or single-strand breaks in mammalian and other genomes. These lesions are predominantly repaired via the conserved base excision repair (BER) pathway. BER is initiated with excision of oxidized or modified bases by DNA glycosylases leading to formation of abasic (AP) site or strand break at the lesion site. Structural analysis by experimental and modeling approaches shows the presence of a disordered segment commonly localized at the N- or C-terminus as a characteristic signature of mammalian DNA glycosylases which is absent in their bacterial prototypes. Recent studies on unstructured regions in DNA metabolizing proteins have indicated their essential role in interaction with other proteins and target DNA recognition. In this review, we have discussed the unique presence of disordered segments in human DNA glycosylases, and AP endonuclease involved in the processing of glycosylase products, and their critical role in regulating repair functions. These disordered segments also include sites for posttranslational modifications and nuclear localization signal. The teleological basis for their structural flexibility is discussed.

M. L. Hegde · T. K. Hazra · S. Mitra (✉)
Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-1079, USA
e-mail: samitra@utmb.edu

T. K. Hazra
Department of Internal Medicine, University of Texas Medical Branch, Galveston, TX 77555-1079, USA

Keywords Base excision repair · DNA glycosylases · End processing proteins · Disordered terminal segments · Single strand breaks · Reactive oxygen species · Repair complex · Protein–protein and protein–DNA interactions

Abbreviations

BER	Base excision repair
SSBR	Single-strand break repair
AP	Abasic
APE	AP endonuclease
ROS	Reactive oxygen species
RNS	Reactive nitrogen species
SSB	Single-strand break
PONDR	Prediction of naturally disordered regions in proteins

Introduction

Damage to mammalian genomes, induced by a variety of reactive oxygen species (ROS) and reactive nitrogen species (RNS), that are generated either endogenously or by radiation and in response to other genotoxic and inflammatory agents, include a plethora of oxidatively damaged bases, abasic (AP) sites and DNA single-strand breaks (SSBs) that are often mutagenic and have etiological linkage to sporadic cancers and a variety of other pathophysiologies as well as aging [1, 2]. It is generally estimated that more than 10^4 base lesions and SSBs are induced daily in a mammalian cell genome [3]. These base lesions and SSBs are typically repaired via the evolutionarily conserved DNA base excision repair (BER) pathway. BER is initiated with excision of an oxidized, alkylated or aberrant base by a lesion-specific

DNA glycosylase, generating an AP site. In the case of oxidized bases, the AP sites are further cleaved by the intrinsic AP lyase activity of DNA glycosylases, while in other cases, the AP sites are cleaved by an AP endonuclease (APE). The resulting DNA strand breaks have blocking groups either at the 3′ or 5′ end which need to be removed in order to generate a single-stranded gap that could then be filled in by a DNA polymerase followed by the sealing of the nick by a DNA ligase [4]. The SSBs with blocked ends are also generated directly by ROS that could include 3′-phosphate (3′-P), 3′-deoxyribosephosphate (3′-dRP) or 5′-deoxyriobosephosphate (5′-dRP) and oxidized sugar fragments such as 3′-phosphoglycolate [5, 6]. While 5′-dRP is processed by the dRP lyase activity of DNA polymerase β (Polβ) in mammalian cells, 3′-P and 3′-dRP are removed by polynucleotide kinase (PNK) and APE1, respectively [7]. Thus, the early step in repair of modified/oxidized bases and SSBs involves DNA glycosylases and nick-end processing enzymes.

Recent studies have suggested that BER is highly complex, involving a network of integrated pathways which are likely to be genome sequence-specific as well as cell cycle-specific [8, 9]. Several studies by us and others have shown that the early BER proteins form a complex to initiate repair and binarily interact with most if not all of the downstream proteins presumably for efficient co-ordination and sequential handover [7, 10–13].

The major challenge for the early BER proteins in mammalian cells is lesion recognition via efficient scanning of the gigabase size genome. Furthermore, a unique issue in BER, in contrast to nucleotide or mismatch excision repair pathways, is that the substrate lesion does not significantly distort the DNA helix and could retain near normal base pairing. These lesions do not block transcription or replication and are invariably bypassed. Thus, lesion recognition, particularly in highly condensed chromatin, poses a serious challenge which has not been extensively addressed. Interestingly, most early BER proteins have a stretch of disordered peptide segment invariably at one of the termini or which could serve sometimes as a linker bridging two domains, suggesting that such a common structural feature could be important for their functions. The focus of this review is to address the complexity of early BER activity in mammals and explore the common and unique structural features among these proteins that enable them to perform such an exigent function proficiently.

Basic BER mechanism

BER, first elucidated in *E. coli* and subsequently found to be universally conserved, is initiated with the recognition and excision of altered base lesions by about a dozen of distinct DNA glycosylases, each of which acts on a limited number of damaged bases [14, 15]. Uracil DNA glycosylase (Udg) was the first DNA glycosylase to be discovered in *E. coli*, which removes the U from DNA. U is generated due to deamination of cytosine and is mutagenic [14]. Subsequently, similar enzyme activities were discovered in mammalian cells and nuclear (UNG2)-specific and mitochondria (UNG1)-specific UDG variants were characterized [16]. Thymine-(T·G)–DNA glycosylase (TDG) is another mammalian DNA glycosylase that excises T and U paired with G [17]. DNA glycosylases specific for repair of alkylated bases have been characterized. The methylated bases generated by chemical reactions with endogenous 5-adenosylmethionine and exogenous alkylating agents, including many chemotherapeutic drugs and *N*-methylnitrosamine, are repaired either by direct reversal without DNA repair synthesis or via the BER pathway [18]. Ada in *E. coli* was the first repair protein discovered that carries out direct reversal of O^6-alkylguanine [18, 19]. Its mammalian ortholog O^6-methylguanine methyltransferase (MGMT) was subsequently cloned [20]. AlkB (*E. coli*) and its mammalian ortholog ABH are the other direct reversal enzymes for several alkylbase adducts [21]. *N*-methylpurine-DNA glycosylase (MPG, also named MAG) and its *E. coli* ortholog AlkA repair *N*-alkylpurines via the conventional BER pathway [22]. While DNA glycosylases in general recognize only abnormal bases in DNA, MutY, a mismatch-specific glycosylase discovered in *E. coli* and its mammalian homolog MYH excise normal base A from A·G and A·8-oxoG mispairs [23, 24]. The glycosylases for the repair of U, alkyl base adducts or base mispairs are monofunctional as they excise the base lesion, leaving an AP site without generating a strand break.

Oxidized base-specific DNA glycosylases in mammalian cells have been categorized in two families based on their tertiary structure and AP lyase reaction and named after their bacterial prototypes, Nth (endonuclease III) and Nei (endonuclease VIII) [25, 26]. OGG1 and NTH1 belonging to Nth family were the first to be discovered. Subsequently, we and others discovered NEIL1 and NEIL2 which belong to Nei family that also includes bacterial Fpg (formamidopyrimidine–DNA glycosylase) [1, 26–30]. NEIL3 was recently added to the list of oxidized base-specific glycosylases [31]; however, its biological activity is not well characterized. The Nth family glycosylases perform β-elimination at the AP site generating a 3′-phospho-$\alpha\beta$-unsaturated aldehyde, a dehydration product of deoxyribosephospate (3′-PUA also named as 3′-dRP) at the strand break. The members of Nei family catalyze $\beta\delta$-elimination at the AP site to produce 3′-P [7, 32]. Thus, contrary to the APE-cleaved product of an AP site which contains 5′-dRP, and is removed by dRP lyase activity of

Polβ, the oxidized base-specific DNA glycosylase-mediated strand breaks (BER intermediate) generate 3′ blocking groups, which need to be cleaned by specific end-processing enzymes in the subsequent BER step.

Processing of blocked termini: common step in the repair of altered bases and SSBs

Single-strand interruption processing, the second step in the repair of oxidized bases and AP sites, is also required for repairing direct ROS-induced SSBs which invariably contain an unligatable blocking group, either 3′ or 5′ or both ends at the strand break [33]. Gap filling by a DNA polymerase and nick sealing by a ligase require 3′-OH and 5′-P at the break site. End processing is thus an obligatory step involving multiple essential enzymes, specific for the 5′ or 3′ blocking groups as well as their type.

5′ end processing at a strand break

APE cleavage of the AP site generates 3′-OH and 5′-dRP termini. DNA polymerase β in mammalian cells removes the unmodified 5′-dRP moiety via its lyase activity [34, 35]. However, Polβ cannot process 5′ blocking groups generated by the cleavage of oxidized AP site by APE, and such dirty ends are processed via the long-patch BER pathway which involves removal of displaced 2–8 nts along with the 5′ blocking group by FEN-1, as discussed later. A unique type of 5′ blocking groups are formed as intermediates during abortive DNA ligation, such as adenylate groups covalently linked to the 5′ P terminus at strand break, and such groups are processed by a protein named aprataxin [36, 37]. Aprataxin releases the adenylate moiety to form ligatable 5′-P.

3′ end processing at a strand break

The Nth family glycosylases with intrinsic β lyase activity generate 3′-dRP (3′-PUA) and 5′-P at the strand break as already mentioned. The 3′dRP is removed by the intrinsic phosphodiesterase activity of mammalian APE1 to generate the 3′-OH terminus. On the other hand, the $\beta\delta$ lyase activity of Nei family DNA glycosylases would generate the 3′-P. We had shown that 3′-P is a poor substrate of APE1 but is efficiently removed by PNK in mammalian cells [7]. In addition, the 3′ blocks at the ROS-induced SSBs, mainly 3′-phosphoglycolate and 3′-phosphoglycoladehyde, are processed by APE1 [38–40]. Tyrosylphosphodiesterase 1 (TDP1) is another 3′ end processing enzyme discovered in yeast and human cells, which removes the product of abortive topoisomerase 1 (Top1) reaction, namely Top1(Tyr)-linked 3′ termini to form 3′-P, a substrate for PNK [41–44].

While there are multiple APEs in prokaryotes and lower eukaryotes, APE1 is the predominant contributor of APE activity in mammalian cells. Its prototype in *E. coli* is exonuclease III (Xth). As AP sites and strand breaks are continuously generated in the genome, it is not surprising that multiple enzymatic processes involving APEs are evolved. We have recently reviewed the functions of multiple APEs in BER in Hegde et al. [1].

Furthermore, TREX1 and TREX2 DNA 3′ exonucleases are ubiquitous in mammalian tissues, whose primary function may be editing during replication by Polβ or Polα lacking constitutive 3′ exonuclease activity [45]. In addition, TREX1 has also been proposed to play a role in SSBR, and its deficiency has been linked to a severe brain disease [46].

XRCC1 and PARP are two other key proteins which play a role in early BER activity. While XRCC1 acts as a scaffold for recruiting BER proteins for excision or strand break repair, PARP acts a SSB sensor protein. We showed that XRCC1 physically interacts with NEILs suggesting its role in oxidized base repair [7, 47]. XRCC1 also interacts with end processing enzymes PNK and APE1, and other BER proteins, Polβ and LigIIIα [48]. Moreover, processing of 5′-OH and 3′-P termini at SSBs is reduced in XRCC1-deficient cells suggesting its role in end processing [49]. PARP present in mammalian cells and absent in *E. coli* is activated by SSBs and transfers ADP-ribose moiety from NAD to a variety of proteins including itself. PARP-1 and PARP-2, the two proteins of the PARP superfamily, have been shown to be important players in the repair of SSBs both as sensors and for recruiting other repair proteins to the strand break [50]. Thus, XRCC1 and PARP play an indirect but vital role in both base excision and end cleaning repair steps.

The early BER reactions for repair of modified bases and SSBs are schematically represented in Fig. 1.

Complexities in BER and SSBR: multiple repair subpathways

Until recently, BER was believed to be the simplest among the DNA repair pathways involving four or five reaction steps. However, recent studies reveal that BER is much more complex, involving a network of distinct cell cycle-dependent as well as genome region-specific repair subpathways and could also involve non-BER proteins.

Short-patch versus long-patch repair pathways

In the simple mammalian BER model, excision of the lesion leaves a 1-nt gap at the damage site, which is subsequently filled by Polβ and the resulting nick sealed by

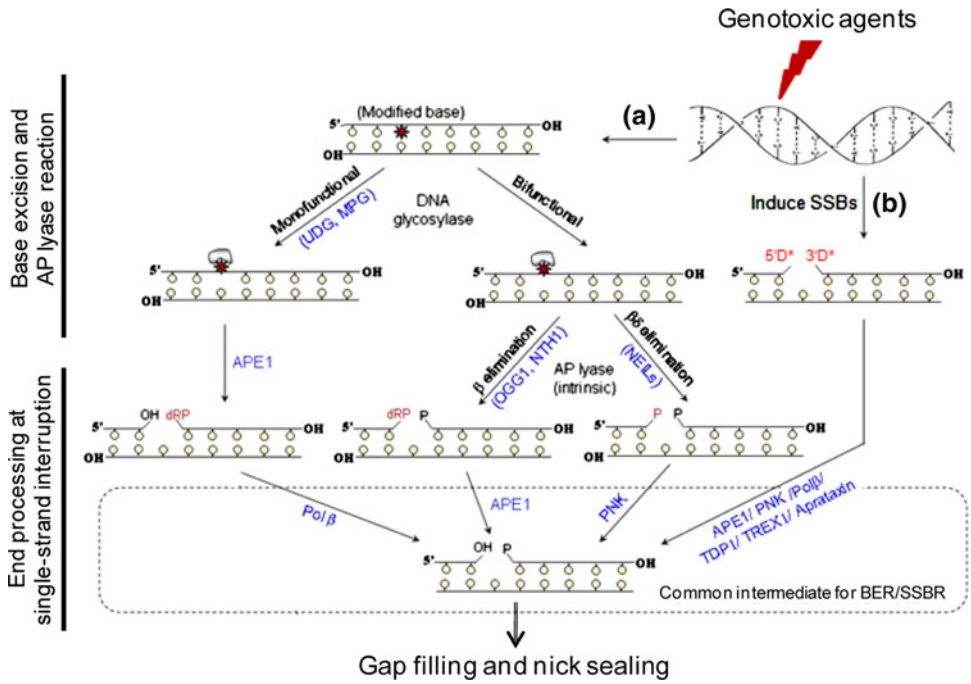

Fig. 1 Schematic representation of base excision (**a**) and single-strand break (**b**) repair steps in mammalian cells. Monofunctional DNA glycosylases (UDG, MPG) excise alkylated and modified bases from DNA to generate AP sites that are then cleaved by APE1. The 5′ blocking group at the break site is removed by Polβ to generate a single-nucleotide gap that is filled in by Polβ (and sealed by DNA ligase IIIα). Oxidized bases are excised by OGG1/NTH1 and NEILs which also cleave the DNA strand to generate 3′ blocking groups. DNA is directly cleaved by ROS/radiation and topoisomerases to generate 3′ or 5′ blocking groups (3′D* and 5′D*) which are removed by several end cleaning enzymes. Other details are described in the text

DNA ligase IIIα (LigIIIα) to complete single-nucleotide repair (SN-BER), also called short-patch repair (SP-BER) [51].

In contrast to this simple model, recent discoveries document several BER subpathways involving at least a dozen more proteins. The second mode of BER, characterized by Matsumuto, Dogliotti and coworkers [34, 52, 53], involves a repair patch size of 2–8 nucleotides at the lesion site and was named long-patch repair (LP-BER). LP-BER was first observed during repair of an AP site analog lacking the aldehyde group. An upstream segment containing 2–8 nts including the 5′-blocking group is displaced as a single-strand flap during repair synthesis which is subsequently cleaved by flap endonuclease 1 (FEN-1). FEN-1's normal function is to remove the 5′ RNA primers of Okazaki fragments during discontinuous replication of the lagging strand [54–56]. LP-BER is believed to utilize DNA replication proteins including DNA polymerase δ (Polδ), the sliding clamp PCNA, clamp loader replication factor-C (RF-C) and DNA ligase I (LigI) in addition to FEN-1 [57, 58]. However, Polβ has also been shown to participate in LP-BER via strand displacement, in collaboration with FEN-1 [59]. The choice of SN-BER versus LP-BER is a complex issue that is yet to be completely understood. Initial studies suggested that the nature of the 5′-phosphoribose terminus (normal vs oxidized) would be the deciding factor [60, 61]. However involvement of DNA replication proteins with LP-BER strongly suggests that LP-BER could be preferred in BER during DNA replication, irrespective of the 5′ terminal group.

DNA replication/transcription-specific BER subpathways

The mammalian genome at the replication fork or transcription bubble is relatively unfolded or nonchromatinized rendering it more prone to oxidative damage than non-replicating chromatin. Replication of unrepaired oxidized base lesions which do not block replication could be mutagenic. Thus, there is an urgency to repair these lesions prior to replication (pre-replicative repair) in order to maintain genome integrity. Furthermore, incorporation of abnormal or oxidized bases (e.g., uracil or 8-oxoG) from the nucleotide pool into nascent DNA could be mutagenic as well, which also warrants urgent repair. Such repair was earlier described as post-replicative repair [62].

Recent studies by us and others have suggested that there are distinct BER subpathways for transcriptionally active versus inactive genomes as well as for quiescent versus replicating genome. Our initial studies with human

NEIL1 and NEIL2 showed that both NEILs are active on bubble and fork-structured DNA substrates that mimic DNA replication or transcription intermediates [8]. At the same time, only NEIL1 is upregulated during the S-phase, based on which we had proposed that NEIL1 is preferentially involved in replication-associated repair (RA-BER) and NEIL2 in transcription-coupled BER (TC-BER) [8, 27, 28, 63]. Subsequently, our recent studies have shown NEIL1's preferential association with DNA replication proteins including PCNA [11], Replication protein A (RPA; [13]), FEN-1 [10], and Werner's helicase (WRN; [12]) that strongly support this hypothesis. Similarly, several other DNA glycosylases such as MYH and UNG (which are described later in this article) interact with replication proteins PCNA and RPA [64, 65]. Like NEIL1, UNG and MYH expression increases during the S-phase, and the UNG-PCNA-RPA complex co-localizes with replication foci suggesting preferential repair of nascent DNA [62, 66]. In contrast, our studies with NEIL2 which has cell cycle-independent expression suggested its association with transcription-associated BER. Although more studies are required to further characterize the genome region and cell-state-specific BER pathways, it is clear that additional complexities are associated with these distinct BER subpathways. For example, we have shown stable interaction between NEIL1 and 9-1-1 complex, a stress-activated sliding clamp implying a linkage between NEIL1-initiated BER and damage signaling pathways [67].

Involvement of non-BER proteins in the repair of base lesions/SSBs

While several non-BER proteins have been shown to be involved in BER, adding another dimension to the BER complexity, their precise in vivo role in repair has yet to be unraveled. We showed that NEIL2 interacts with YB-1, a Y-box binding protein, and it was suggested that YB-1 may be required for the fine-tuning of repair [68]. NTH1 was also shown to interact with and is stimulated by YB-1 [69]. The list of non-BER proteins interacting with BER proteins is still growing, which underscores the paradigm that the in vivo repair process is far more complex than in vitro repair demonstrated with minimal components.

Recent discoveries in BER

DNA glycosylases as hub proteins: binary interaction of DNA glycosylases with downstream proteins

Our initial characterization of NEIL1 and NEIL2 showed their similarity in in vitro repair of 5-OHU via SN-BER in a reconstituted system containing PNK, Polβ, LigIIIα and XRCC1. XRCC1 acts as a scaffold for recruiting BER proteins for excision or strand break repair. Both the NEIL immunoprecipitates from human cells contain these BER proteins [7, 47]. Furthermore, NEILs binarily interact in the absence of DNA with Polβ, LigIIIα and XRCC1, although not PNK. Direct interaction of NEILs with LigIIIα, the last enzyme in SN-BER, indicated that the repair is controlled or regulated by the initiating DNA glycosylase which acts as a hub protein. Similarly, as already discussed, NEIL1 binarily interacts with PCNA, RPA, FEN-1, and WRN in the absence of DNA, presumably for preferential BER during DNA replication. The stoichiometry of proteins in the repair complex and whether NEILs are present in distinct complexes or in a single complex are currently being investigated in our laboratories.

Repair complex versus sequential repair hand-off

The initial BER mechanism was proposed, based on co-crystal structure analysis of substrate-bound BER proteins, to involve 'hand-off' or 'passing the baton' process, wherein the repair product of each enzyme in the BER pathway is handed over to the next enzyme, primarily based on differential bending of DNA in each intermediate step [70, 71]. However, characterization of the BER interactome involving multiprotein interactions including stable complex of DNA glycosylase with DNA ligase, and the presence of multiprotein complexes has led to a new paradigm where complete repair occurs in the BER complex (BERosome). Although in vivo role of hand-off versus interactome modes of repair is not yet clear, we propose that preformed BER complexes predominantly repair endogenous base lesions, while repair via hand-off mechanism by sequential recruitment could occur with induced DNA damage. Further characterization of the dynamics of BERosomes is required to unravel the repair processes.

Common interaction interface of early BER proteins

For several years, our laboratory has focused on characterizing collaborations and mapping interactions among BER proteins. As already mentioned, NEIL1's stable interaction with downstream repair proteins utilizes a common interaction interface located near its C-terminus (residues 289–349) [7, 10–13]. The segment is absent in NEIL1's prototype Nei, and might have been acquired during evolution as a terminal addition [1]. We similarly identified the nonconserved N-terminal segment (65 residues) in APE1 which is involved in all its protein–protein interactions [72–74]. Although it is intriguing how NEIL1 or any other protein could simultaneously bind to so many proteins with high specificity via a small common

interface, recent studies have indicated that it is not uncommon for the mammalian hub proteins to have such an interaction surface, which invariably have a disordered structure. The flexibility of the disordered domain may facilitate interaction with diverse partners [75, 76].

Disordered structure of the interaction interface

The C-terminal region of human NEIL1 (hNEIL1) spanning about 100 residues contains the common interaction interface whose disordered conformation was first suggested from the fact that NEIL1's crystallization required deletion of 56 residues, while the proximal 44 residues did not form a defined structure [77]. We verified this conclusion using various protein structure prediction softwares.

Prediction of disordered structure: various softwares

Contrary to the concept about spontaneous formation of secondary and tertiary structure in properly folded proteins that prevailed before the turn of this century, recent experimental evidence as well as genome-wide prediction of intrinsic disorder in eukaryotic proteomes have indicated that a large percentage of proteins have long disordered (unfolded) segments under physiological conditions. Further, these disordered regions are essential for their biological functions. Recent advancement in the softwares for accurate prediction of protein secondary structure and their predisposition to remain intrinsically disordered has furthered our understanding of the role of disordered structures in functional hierarchy. Commonly used disorder prediction tools include PONDR [78–80], PrDOS [81, 82], RONN [82], FoldIndex [83], GlobPlot [84], IUPred [85, 86], FoldUnFold [87], etc. in the public domain, which evaluate intrinsic disorder on per residue basis. Among these, PONDR is most widely used, an advanced version of which contains a reference collection set of VSL predictors (trained on variously characterized, short and long disordered regions). The PONDR developers point out that short and long disordered regions might have differences in their amino acid characteristics because predictors based on short regions of disorder fare poorly for long regions of disorder and vice versa [88–90]. The VSL predictors in PONDR take advantage of such differences to yield more accurate predictions. We used PONDR, PrDOS and RONN software, which generated similar disorder prediction in early BER proteins.

The commonly used approaches to characterize protein disorder are NMR and circular dichroism spectroscopy, and also small angle X-ray scattering, while the structural information thus generated is often based on the crystal structure of truncated proteins [91].

Signature sequences of intrinsic disorder

A major sequence characteristic of intrinsic disorder is the low content of bulky hydrophobic amino acid residues such as Val, Leu, Ile, Met, Phe, Trp and Tyr, which would normally form the core of a folded globular protein. In contrast, a high proportion of polar and charged residues such as Gln, Ser, Pro, Glu, Lys, and Gly, and sometimes also Ala, are characteristically present in disordered regions [78, 92]. The presence of such charged residue-rich sequences was first discovered in transcriptional regulatory proteins about three decades ago, which are often classified based on their amino acid composition, for example Glu-rich, Pro-rich and acidic activation domains [93]. Later, NMR spectroscopy and other biophysical studies confirmed intrinsically disordered nature of such sequences [91, 94].

While disordered regions have been variously described as intrinsically disordered, intrinsically unstructured, natively unfolded, natively disordered and highly flexible [91], we believe that the term 'intrinsically disordered' would be more appropriate than 'unstructured' because many disordered regions have been shown to contain partial or transient secondary and/or tertiary structural organization [95]. Dunker and Obradovic [96] proposed that intrinsically disordered regions may exist in two different structural forms: molten globule-like (collapsed) and random coil-like (extended), whereas Uversky and co-workers suggested existence of another extended form, the pre-molten globule, which is distinguishable from fully extended and molten globular conformations by the presence of an unstable secondary structure [97]. The recently proposed protein quartet hypothesis suggests that the protein functions in eukaryotes could depend on any of the three disordered forms along with the ordered form or on transitions between them [97].

Disorder predictions have been extensively utilized by protein crystallographers to design crystallization targets after deletion of disordered segments. One first application of the disorder predictor was in crystallization of *Xeroderma pigmentosum* group A (XPA) protein involved in DNA repair [98]. Extended disordered regions at the N- and C-termini of XPA with ordered central core as predicted was confirmed with partial proteolysis and NMR spectroscopy [99].

The recently created databank of protein intrinsic disorder (http://www.disprot.org) suggests that proteins with long terminal extensions containing no or limited structure are common in eukaryotes and are involved in many key functions including cell cycle control, regulation, and signaling [99]. Such disordered tails were also shown to be more common in DNA binding proteins than in other proteins, particularly in the ones that are involved in target

sequence binding that include early repair proteins and transcription factors [100, 101]. It is likely that their structural flexibility and plasticity provides major functional advantage.

Predictions of disorder using accurate bioinformatic tools are in fact helping design experiments to characterize their biological functions, and this is one of the fastest growing topics of protein studies, humorously dubbed as 'the protein unfoldomics decade'. Although the number of proteins with experimentally determined disordered structure is still small, the behavior of the experimentally verified ones are highly consistent with the predictions [102].

Comparison of *E. coli* versus mammalian early BER proteins to identify disordered structures

We used PONDR (prediction of natively disordered regions) and PrDOS (prediction of disordered structures) softwares to compare the secondary structure of human and bacterial early BER proteins and correlated them with their available structural information. As already mentioned, hNEIL1 contains an extended disordered region spanning about 100 residues in the C-terminus which is absent in *E. coli* Nei protein (Fig. 2). Comparison of predicted structures of human DNA glycosylases NTH1, MYH and their *E. coli* prototypes endonuclease III and MutY, respectively, indicates that both hNTH1 and hMYH have extended disordered tails at the N-terminus that are absent in the *E. coli* enzymes. Similarly, the N-terminal disordered region present in hAPE1 is absent in Xth, its *E. coli* prototype (Fig. 3). PONDR modeling showed that such disordered terminal sequences may also be present in other human DNA glycosylases including UNG2 and TDG (Fig. 4). Although the unfolded sequence generally exists at the N- or C-terminus, this could exist internally in some proteins, acting as a linker joining two domains. HNEIL2 is such a protein with an internal disordered segment near the N-terminus (residues 45–130) as indicated from the PONDR plot (Fig. 4). The average size of disordered extensions in early BER proteins ranges from 50 to 100 residues, with few exceptions, e.g., hOGG1 and human Polβ, which have short ($\sim$10 residues) disordered tails at both termini, as predicted by PrDOS (Fig. 5). The early BER proteins are generally small (30–50 kDa) and monomeric, whereas other DNA transaction proteins such as PCNA are multimeric and typically possess disordered linkers bridging different subunits.

Among early BER proteins, the disordered N-terminal sequence of hNTH1 has been extensively characterized. HNTH1 has a lower specific activity than *E. coli* Nth that lacks the N-terminal extension [103, 104], deletion of which increases hNTH1's activity [105]. This suggests that this segment inhibits enzyme turnover in the absence of other BER proteins.

Role of disordered domain in function of early BER proteins

The presence of disordered extensions in proteins involved in transcriptional regulation, signal transduction, cell cycle control, DNA damage sensing and repair suggest their involvement in diverse functions. Furthermore, such disordered regions may regulate formation of large multiprotein complexes [106, 107]. An exhaustive discussion of this topic is beyond the scope of this review, focusing on the early BER proteins, in which these disordered segments include sites of posttranslational modifications, subcellular targeting, DNA scanning as well as common interface for protein–protein interactions, as already discussed (Fig. 6).

Posttranslational modifications in disordered segments

Posttranslational modifications of proteins such as phosphorylation, acetylation, ubiquitylation, ADP-ribosylation, sumoylation and methylation play a critical role in diverse cellular processes including DNA repair [108]. The modification sites are invariably localized in disordered regions, e.g., in the N-terminal segment in hAPE1 [74, 109], N- and C-terminus of p53 [110], and C-terminal region in hNEIL1 (Bhakat et al., unpublished). Our laboratory identified and characterized acetylation of hAPE1 at Lys6 or Lys7, and this modification plays an important role in APE1's transcriptional regulatory functions [74]. Recently, APE1 was also shown to be ubiquitynated at N-terminal Lys residues, which regulates its degradation as well as cellular functions [109]. Such covalent modifications may have multiple physiological effects on these proteins, including stability, interaction with DNA or other proteins, organelle targeting, and enzymatic activity [111]. We also showed that hNEIL2 is acetylated at Lys49 and Lys153 both in vitro and in cells [112]. Acetylation of Lys49 located in the disordered region (Fig. 4) inactivated NEIL2's base excision and AP lyase activity while acetylation of Lys150 had no effect on the activity. We have proposed that acetylation of Lys49 could act as a regulatory switch for NEIL2's activity [112]. TDG is acetylated in the N-terminal segment, Lys70, 94, 95 and 98 which is within the disordered segment of 100 residues [113]. PONDR modeling of TDG sequence indicates that the N-terminal 100 residues are in disordered conformation (Fig. 3). Strong acetylation sites in TDG were identified. Acetylation of TDG by CBP/p300 does not affect its

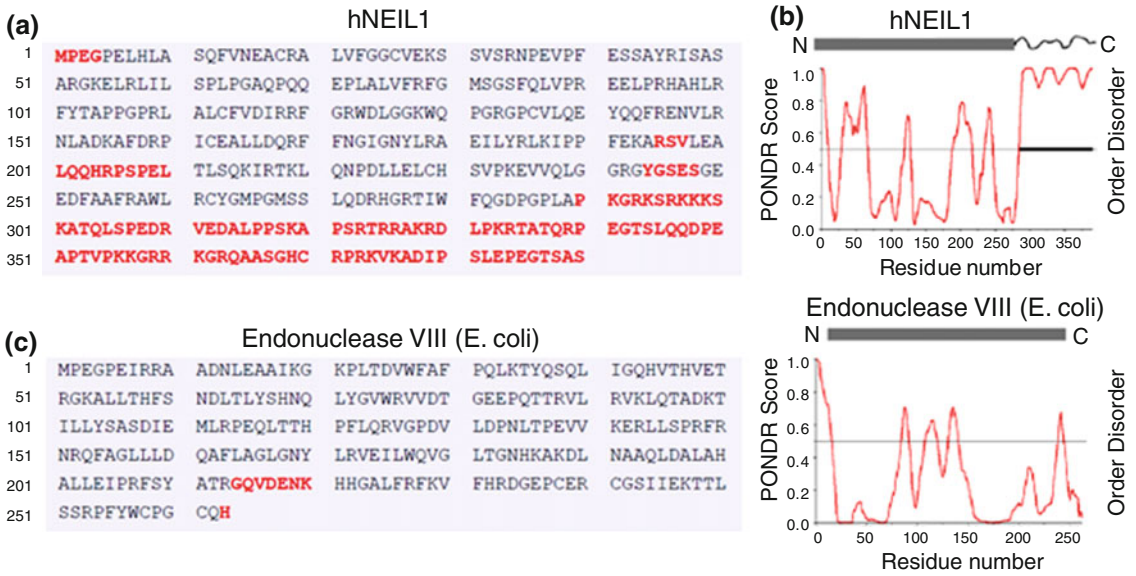

Fig. 2 Secondary structure prediction of hNEIL1 and its *E. coli* prototype endonuclease VIII (Nei) by PrDOS (**a**,**c**) and PONDR (**b**,**d**) softwares. The protein sequences were obtained from NCBI database. Sequences in *red* in PrDOS and a score of 0.5 and above in the PONDR plot indicate disordered structures. The disordered C-terminal segment (*wiggled line*) of hNEIL1 is absent in Nei

binding to G·T or G·U base mispairs but indirectly deregulate TDG-coupled repair by releasing CBP/p300 from DNA bound complex leading to reduced interaction with APE and in turn suppressing APE-dependent repair [113]. Thus, TDG acetylation could contribute to genomic instability and cancer susceptibility. We had earlier characterized acetylation of hOGG1 at Lys338 and Lys341 within its short disordered C-terminus, which increases its DNA glycosylase activity by reducing affinity for the product AP site [114].

The flexibility of disordered region appears to be a prerequisite for these modifications, presumably because the amino acid side chains in the flexible sequence are accessible for modifying enzymes, like kinases, phosphatases, acetyltransferases and deacetylases, methylases, and ubiquitin ligases, etc. [91].

Subcellular localization

Organelle localization signals such as nuclear localization signal (NLS) or mitochondrial transport signal (MTS) are contained in short segments (generally <20 residues) that mediate transport to the target organelle. Multiple types of mammalian NLS sequences have been identified, the major ones belonging to the classical type consisting of seven basic residues and the bipartite NLS with two strings of basic residues separated by a short intervening sequence [115]. Recent studies showed that almost all NLS sequences with overall basic nature are disordered [99, 116].

We mapped the NLS of hAPE1 to the disordered N-terminal 20 residues, the deletion of which markedly diminishes its translocation to the nucleus [117]. Our preliminary studies of GFP-fusion polypeptide of truncated NEIL1 suggest the presence of putative NLS at the disordered C-terminal region (unpublished observation). Similarly, the disordered N-terminal tails in hNTH1, UNG2 and TDG contain putative NLS and MTS [118–120]. Taken together, these studies show that subcellular distribution of many human repair proteins is mediated through signals localized in their disordered regions.

DNA scanning

Burg et al. have shown that target DNA search by proteins could be achieved via facilitated diffusion comprising four mechanisms, namely, one-dimensional (1D) sliding, hopping, 3D search and intersegmental transfer. An efficient search mechanism involves combination of these different modes [121].

Recent studies have shown that the most efficient and rapid scanning of the DNA for the target site involves 80% hopping and intersegmental transfer and 20% sliding by the DNA binding proteins, which invariably contain a disordered terminal extension, or a disordered linker for multidomain or multisubunit proteins [101]. Bioinformatics analysis has suggested that nearly 70% of DNA binding proteins have such disordered tails compared to about 25% for non-DNA binding proteins [100, 101]. In addition, the disordered segments are about seven residues longer on an average for DNA binding proteins compared to all proteins with disordered tails [100]. Another unique characteristic of the disordered tails in DNA binding proteins is

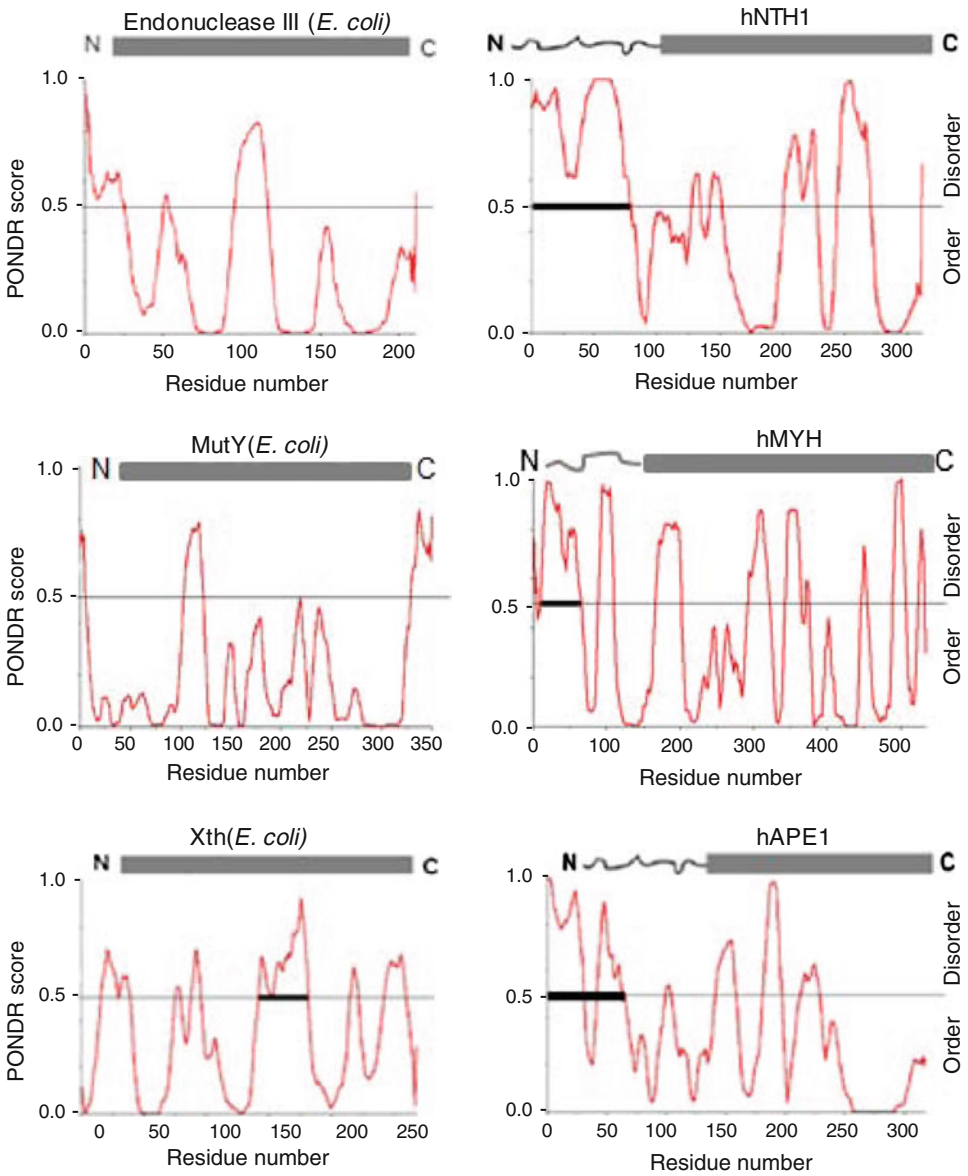

Fig. 3 PONDR plot of predicted secondary structures of hNTH1, hMYH, hAPE1 and their *E. coli* prototypes endonuclease III (Nth), MutY, Xth, respectively. *Wiggled lines* at the N-termini of human enzymes represent disordered segments

clustering of positively charged residues in the distal region, which turned out to be important for the scanning. Mutating such residues in HOXD9 and NK-2 markedly decreased scanning efficiency [101]. Similar results were obtained when the N-terminal segment in these proteins were deleted, suggesting that the initial scanning is mediated by a non-specific, mostly electrostatic, transient DNA binding via the basic disordered segment which is followed by target DNA sequence binding by the active site.

In light of the above studies, we examined hNEIL1's C-terminus, which possesses most characteristics required for DNA scanning as described above, including the clustered basic residues. Our recent biochemical studies using C-terminal deletion NEIL1 mutants showed that the C-terminus is important for NEIL1's substrate scanning and efficiency of damage recognition, via its non-specific DNA binding (Hegde et al., unpublished). Although limited studies are available on the role of disordered regions of other early BER proteins in such activity, we expect that all of them have similar functions.

Intrinsic disorder and hub proteins: dynamic repair complexes mediated by disorder/disorder interactions

As mentioned earlier, the 'hub' proteins like NEIL1 with several partners (usually >10) form a network of complexes [122]. Recent studies have indicated that 'hub' protein complexes are widely present in higher eukaryotes, whose formation mostly involves interaction with disordered regions [123]. The crucial role of intrinsic disorder in hub proteins was reviewed earlier [91, 124]. Bioinformatics

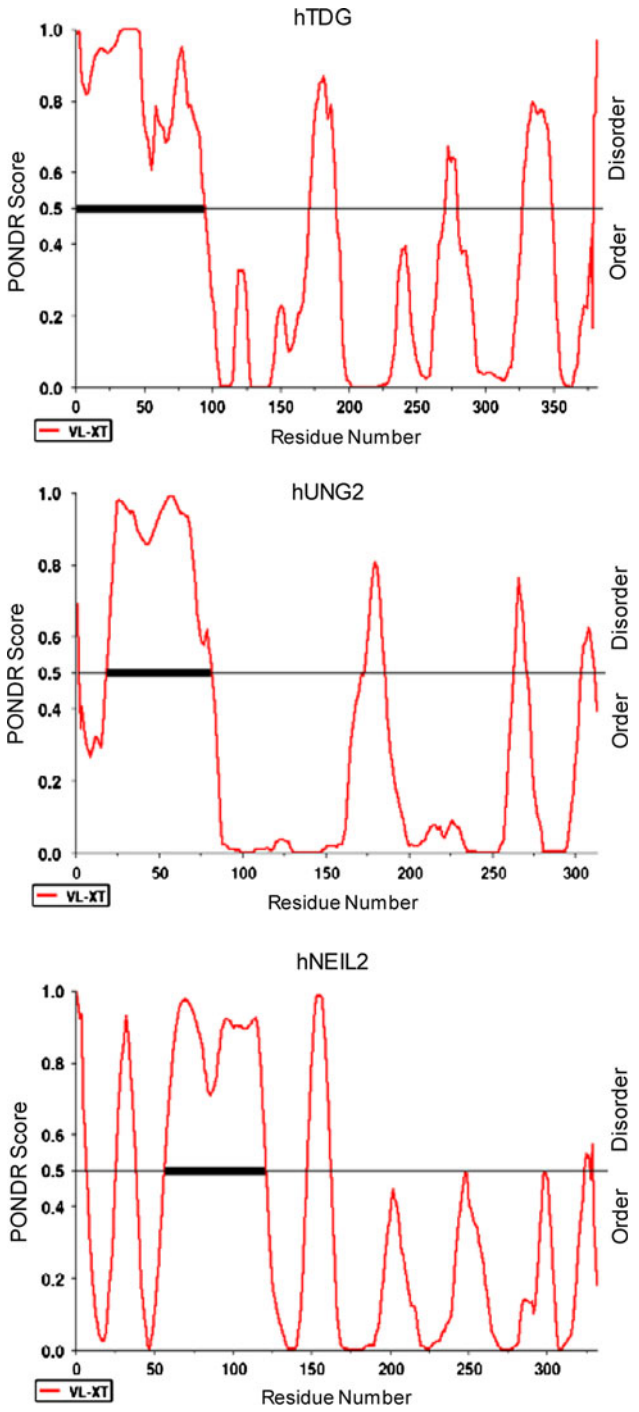

Fig. 4 PONDR plot of disordered conformation at the N-terminus of hTDG and UNG2. HNEIL2 has an internal disordered region

analysis of known protein interactions suggested that such interactions among disordered structures are significantly preferred among human proteins [123]. Disorder-mediated interactions could either involve disorder–disorder or disorder–order types. Both modes of interactions are prevalent in BER proteins, for example NEIL1–FEN-1 interaction involves disorder–disorder interfaces [10], while NEIL1–Polβ could involve disorder–order type of contacts [7]. Specific recognition and binding of hub proteins is achieved as a result of the flexibility of domain itself, because it facilitates conformational rearrangements and induced-fit with specific partners [99, 107]. Although it was inconceivable a few years ago that such specificity could be obtained via disorder–disorder interactions, these are quite common, particularly for hub proteins. Disorder-mediated interactions in fact confer advantages over order-mediated interactions because of the rapid and easy interconversion among diverse conformers, allowing formation of dynamic complexes [107]. The dynamics of such complexes could be further regulated by posttranslational modifications of one or more partners, as already discussed. Furthermore, disorder-mediated interactions have kinetic benefits, because the larger capture radius of disordered states facilitates faster on-rates for binding [125, 126]. Finally, disorder-mediated interactions have steric advantages by providing a large surface area for binding interface for wrapping around partners resulting in stronger specificity [127, 128].

Our recent studies using size fractionation chromatography of human cell nuclear extract suggest that the BER proteins indeed exist in large, stable complexes, presumably in the absence of DNA. Characterization of such complexes, the dynamics of their formation and regulation as well as their stoichiometry are warranted.

Evolutionary advantages of protein disorder

Disordered regions in proteins generally show higher rates of mutation, presumably because changes in their protein sequence may not affect protein stability and function as severely as that in ordered regions [107, 129, 130]. Although such an analysis of mutation distribution and mutation tolerance is yet to be carried out for BER proteins, a similar situation is likely to exist. The unique presence of disordered segments in eukaryotic proteins but not in the prokaryotic counterparts, with the highest degree of disorder in mammals, suggests its evolutionary development [107]. The disordered regions also enable alternative splicing in eukaryotic proteins without the risk of perturbing structured regions [131]. In addition, the disorder provides advantage of limiting molecular size as complexity increases, by providing common interface for multiprotein binding and sites of modifications. To achieve a similar goal, folded proteins need to be considerably larger, and thus disorder may help higher organisms to limit protein size and to reduce intracellular crowding [132].

Fig. 5 PrDOS secondary structure prediction of hOGG1 and Polβ indicate short disordered segments (sequence in *red*) at both termini

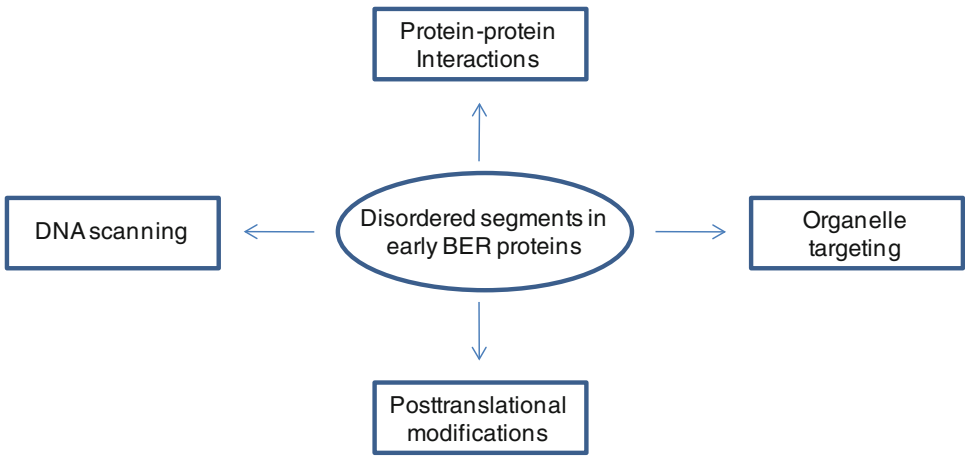

Fig. 6 Schematic of multiple regulatory functions of disordered regions in early BER enzymes

Disordered extensions in early BER proteins as potential targets of cancer therapy

The BER proteins have been explored as targets for cancer therapy which generally involved inactivation of key BER reactions such as ligation [133] or damage sensing proteins such as PARP [134]. DNA glycosylases turned out to be poor targets for such therapy, because of the nonessentiality of individual glycosylases due to back up functions of other glycosylases [1]. Although PARP inhibitors have been proven highly successful in recent therapeutic studies, inhibition of a key BER reaction poses the challenge of accurate targeting and dosage regulation to prevent their undesired effect on normal cells [134]. Based on the emerging evidence for a disorder-mediated repair regulatory switch in early BER hub proteins that controls the repair pathway, we propose that the disordered regions of early BER proteins could be targeted for cancer therapy which would disrupt repair regulation.

Conclusions and perspectives

A combination of experimental studies and structural predictions has revealed a critical role of disordered segments in many mammalian early BER enzyme functions including both protein–protein and protein–DNA interactions. Although few disordered regions have been experimentally characterized so far, we predict that such disordered segments are ubiquitous and essential for efficient repair.

Future studies should address the role of disordered sequences in other mammalian repair pathways and their evolutionary significance in complex repair regulation.

Acknowledgments The research in the authors' laboratory is supported by USPHS grants, R01 CA81063, R01 CA53791, P01 CA92586 and P30 ES06676 (S.M.) and R01 CA 102271, R21 ES017353 (T.K.H). Because of the limited focus of the article on protein disorder in early BER proteins, many appropriate references could not be included, for which the authors apologize. We thank Mitra lab members for various stimulating discussions during preparation of this review.

References

1. Hegde ML, Hazra TK, Mitra S (2008) Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. Cell Res 18:27–47
2. Ames BN, Shigenaga MK, Hagen TM (1993) Oxidants, antioxidants, and the degenerative diseases of aging. Proc Natl Acad Sci USA 90:7915–7922
3. Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715
4. Mitra S, Hazra TK, Roy R, Ikeda S, Biswas T, Lock J, Boldogh I, Izumi T (1997) Complexities of DNA base excision repair in mammalian cells. Mol Cells 7:305–312
5. Breen AP, Murphy JA (1995) Reactions of oxyl radicals with DNA. Free Radic Biol Med 18:1033–1077
6. Demple B, DeMott MS (2002) Dynamics and diversions in base excision DNA repair of oxidized abasic lesions. Oncogene 21:8926–8934
7. Wiederhold L, Leppard JB, Kedar P, Karimi-Busheri F, Rasouli-Nia A, Weinfeld M, Tomkinson AE, Izumi T, Prasad R, Wilson SH, Mitra S, Hazra TK (2004) Ap endonuclease-independent DNA base excision repair in human cells. Mol Cell 15:209–220
8. Dou H, Mitra S, Hazra TK (2003) Repair of oxidized bases in DNA bubble structures by human DNA glycosylases neil1 and neil2. J Biol Chem 278:49679–49684
9. Kavli B, Sundheim O, Akbari M, Otterlei M, Nilsen H, Skorpen F, Aas PA, Hagen L, Krokan HE, Slupphaug G (2002) Hung2 is the major repair enzyme for removal of uracil from u:A matches, u:G mismatches, and u in single-stranded DNA, with hsmug1 as a broad specificity backup. J Biol Chem 277:39926–39936
10. Hegde ML, Theriot CA, Das A, Hegde PM, Guo Z, Gary RK, Hazra TK, Shen B, Mitra S (2008) Physical and functional interaction between human oxidized base-specific DNA glycosylase neil1 and flap endonuclease 1. J Biol Chem 283:27028–27037
11. Dou H, Theriot CA, Das A, Hegde ML, Matsumoto Y, Boldogh I, Hazra TK, Bhakat KK, Mitra S (2008) Interaction of the human DNA glycosylase neil1 with proliferating cell nuclear antigen. The potential for replication-associated repair of oxidized bases in mammalian genomes. J Biol Chem 283:3130–3140
12. Das A, Boldogh I, Lee JW, Harrigan JA, Hegde ML, Piotrowski J, de Souza Pinto N, Ramos W, Greenberg MM, Hazra TK, Mitra S, Bohr VA (2007) The human werner syndrome protein stimulates repair of oxidative DNA base damage by the DNA glycosylase neil1. J Biol Chem 282: 26591–26602
13. Theriot CA, Hegde ML, Hazra TK, Mitra S (2010) Rpa physically interacts with the human DNA glycosylase neil1 to regulate excision of oxidative DNA base damage in primer-template structures. DNA Repair (Amst) 9:643–652
14. Lindahl T (1974) An n-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. Proc Natl Acad Sci USA 71:3649–3653
15. Friedberg EC, Aguilera A, Gellert M, Hanawalt PC, Hays JB, Lehmann AR, Lindahl T, Lowndes N, Sarasin A, Wood RD (2006) DNA repair: from molecular mechanism to human disease. DNA Repair (Amst) 5:986–996
16. Krokan HE, Drablos F, Slupphaug G (2002) Uracil in DNA—occurrence, consequences and repair. Oncogene 21:8935–8948
17. Hardeland U, Bentele M, Jiricny J, Schar P (2003) The versatile thymine DNA-glycosylase: a comparative characterization of the human, drosophila and fission yeast orthologs. Nucleic Acids Res 31:2261–2271
18. Mitra S (2007) Mgmt: a personal perspective. DNA Repair (Amst) 6:1064–1070
19. Demple B, Sedgwick B, Robins P, Totty N, Waterfield MD, Lindahl T (1985) Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis. Proc Natl Acad Sci USA 82:2688–2692
20. Tano K, Shiota S, Collier J, Foote RS, Mitra S (1990) Isolation and structural characterization of a cdna clone encoding the human DNA repair protein for O6-alkylguanine. Proc Natl Acad Sci USA 87:686–690
21. Sedgwick B, Bates PA, Paik J, Jacobs SC, Lindahl T (2007) Repair of alkylated DNA: recent advances. DNA Repair (Amst) 6:429–442
22. Mitra S, Kaina B (1993) Regulation of repair of alkylation damage in mammalian genomes. Prog Nucleic Acid Res Mol Biol 44:109–142
23. Slupska MM, Luther WM, Chiang JH, Yang H, Miller JH (1999) Functional expression of hMYH, a human homolog of the *Escherichia coli* MutY protein. J Bacteriol 181:6210–6213
24. Nghiem Y, Cabrera M, Cupples CG, Miller JH (1988) The muty gene: a mutator locus in *Escherichia coli* that generates g.C…t.A transversions. Proc Natl Acad Sci USA 85:2709–2713
25. Hazra TK, Izumi T, Kow YW, Mitra S (2003) The discovery of a new family of mammalian enzymes for repair of oxidatively damaged DNA, and its physiological implications. Carcinogenesis 24:155–157
26. McCullough AK, Dodson ML, Lloyd RS (1999) Initiation of base excision repair: glycosylase mechanisms and structures. Annu Rev Biochem 68:255–285
27. Hazra TK, Izumi T, Boldogh I, Imhoff B, Kow YW, Jaruga P, Dizdaroglu M, Mitra S (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. Proc Natl Acad Sci USA 99:3523–3528
28. Hazra TK, Kow YW, Hatahet Z, Imhoff B, Boldogh I, Mokkapati SK, Mitra S, Izumi T (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. J Biol Chem 277:30417–30420
29. Bandaru V, Sunkara S, Wallace SS, Bond JP (2002) A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease viii. DNA Repair (Amst) 1:517–529
30. Takao M, Kanno S, Shiromoto T, Hasegawa R, Ide H, Ikeda S, Sarker AH, Seki S, Xing JZ, Le XC, Weinfeld M, Kobayashi K, Miyazaki J, Muijtjens M, Hoeijmakers JH, van der Horst G, Yasui A (2002) Novel nuclear and mitochondrial glycosylases revealed by disruption of the mouse Nth1 gene encoding an endonuclease iii homolog for repair of thymine glycols. EMBO J 21:3486–3493
31. Liu M, Bandaru V, Bond JP, Jaruga P, Zhao X, Christov PP, Burrows CJ, Rizzo CJ, Dizdaroglu M, Wallace SS (2010) The

mouse ortholog of neil3 is a functional DNA glycosylase in vitro and in vivo. Proc Natl Acad Sci USA 107:4925–4930
32. Zharkov DO, Shoham G, Grollman AP (2003) Structural characterization of the fpg family of DNA glycosylases. DNA Repair (Amst) 2:839–862
33. Caldecott KW (2008) Single-strand break repair and genetic disease. Nat Rev Genet 9:619–631
34. Frosina G, Fortini P, Rossi O, Carrozzino F, Raspaglio G, Cox LS, Lane DP, Abbondandolo A, Dogliotti E (1996) Two pathways for base excision repair in mammalian cells. J Biol Chem 271:9573–9578
35. Sobol RW, Prasad R, Evenski A, Baker A, Yang XP, Horton JK, Wilson SH (2000) The lyase activity of the DNA repair protein beta-polymerase protects from DNA-damage-induced cytotoxicity. Nature 405:807–810
36. Ahel I, Rass U, El-Khamisy SF, Katyal S, Clements PM, McKinnon PJ, Caldecott KW, West SC (2006) The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. Nature 443:713–716
37. Rass U, Ahel I, West SC (2007) Actions of aprataxin in multiple DNA repair pathways. J Biol Chem 282:9469–9474
38. Caldecott KW (2007) Mammalian single-strand break repair: mechanisms and links with chromatin. DNA Repair (Amst) 6:443–453
39. Whitehouse CJ, Taylor RM, Thistlethwaite A, Zhang H, Karimi-Busheri F, Lasko DD, Weinfeld M, Caldecott KW (2001) Xrcc1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. Cell 104:107–117
40. Parsons JL, Dianova II, Dianov GL (2004) Ape1 is the major 3′-phosphoglycolate activity in human cell extracts. Nucleic Acids Res 32:3531–3536
41. Pouliot JJ, Robertson CA, Nash HA (2001) Pathways for repair of topoisomerase i covalent complexes in Saccharomyces cerevisiae. Genes Cells 6:677–687
42. El-Khamisy SF, Saifi GM, Weinfeld M, Johansson F, Helleday T, Lupski JR, Caldecott KW (2005) Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. Nature 434:108–113
43. Yang SW, Burgin AB Jr, Huizenga BN, Robertson CA, Yao KC, Nash HA (1996) A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc Natl Acad Sci USA 93:11534–11539
44. Pouliot JJ, Yao KC, Robertson CA, Nash HA (1999) Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase i complexes. Science 286:552–555
45. Mazur DJ, Perrino FW (2001) Structure and expression of the trex1 and trex2 3′ → 5′ exonuclease genes. J Biol Chem 276:14718–14727
46. Crow YJ, Hayward BE, Parmar R, Robins P, Leitch A, Ali M, Black DN, van Bokhoven H, Brunner HG, Hamel BC, Corry PC, Cowan FM, Frints SG, Klepper J, Livingston JH, Lynch SA, Massey RF, Meritet JF, Michaud JL, Ponsot G, Voit T, Lebon P, Bonthron DT, Jackson AP, Barnes DE, Lindahl T (2006) Mutations in the gene encoding the 3′–5′ DNA exonuclease trex1 cause aicardi-goutieres syndrome at the ags1 locus. Nat Genet 38:917–920
47. Das A, Wiederhold L, Leppard JB, Kedar P, Prasad R, Wang H, Boldogh I, Karimi-Busheri F, Weinfeld M, Tomkinson AE, Wilson SH, Mitra S, Hazra TK (2006) Neil2-initiated, ape-independent repair of oxidized bases in DNA: evidence for a repair complex in human cells. DNA Repair (Amst) 5:1439–1448
48. Vidal AE, Boiteux S, Hickson ID, Radicella JP (2001) Xrcc1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. EMBO J 20:6530–6539
49. Lan L, Nakajima S, Oohata Y, Takao M, Okano S, Masutani M, Wilson SH, Yasui A (2004) In situ analysis of repair processes for oxidative DNA damage in mammalian cells. Proc Natl Acad Sci USA 101:13738–13743
50. Schreiber V, Dantzer F, Ame JC, de Murcia G (2006) Poly(adp-ribose): novel functions for an old molecule. Nat Rev Mol Cell Biol 7:517–528
51. Kubota Y, Nash RA, Klungland A, Schar P, Barnes DE, Lindahl T (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the xrcc1 protein. EMBO J 15:6662–6670
52. Matsumoto Y (2001) Molecular mechanism of pcna-dependent base excision repair. Prog Nucleic Acid Res Mol Biol 68:129–138
53. Klungland A, Lindahl T (1997) Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for dnase iv (fen1). EMBO J 16:3341–3348
54. Maga G, Villani G, Tillement V, Stucki M, Locatelli GA, Frouin I, Spadari S, Hubscher U (2001) Okazaki fragment processing: modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein a, proliferating cell nuclear antigen, and flap endonuclease-1. Proc Natl Acad Sci USA 98:14298–14303
55. Singh P, Zheng L, Chavez V, Qiu J, Shen B (2007) Concerted action of exonuclease and gap-dependent endonuclease activities of fen-1 contributes to the resolution of triplet repeat sequences $(ctg)_n$- and $(gaa)_n$-derived secondary structures formed during maturation of okazaki fragments. J Biol Chem 282:3465–3477
56. Garg P, Stith CM, Sabouri N, Johansson E, Burgers PM (2004) Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. Genes Dev 18:2764–2773
57. Levin DS, McKenna AE, Motycka TA, Matsumoto Y, Tomkinson AE (2000) Interaction between pcna and DNA ligase i is critical for joining of okazaki fragments and long-patch base-excision repair. Curr Biol 10:919–922
58. Prasad R, Dianov GL, Bohr VA, Wilson SH (2000) Fen1 stimulation of DNA polymerase beta mediates an excision step in mammalian long patch base excision repair. J Biol Chem 275:4460–4466
59. Liu Y, Kao HI, Bambara RA (2004) Flap endonuclease 1: a central component of DNA metabolism. Annu Rev Biochem 73:589–615
60. Piersen CE, Prasad R, Wilson SH, Lloyd RS (1996) Evidence for an imino intermediate in the DNA polymerase beta deoxyribose phosphate excision reaction. J Biol Chem 271:17811–17815
61. Matsumoto Y, Kim K (1995) Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. Science 269:699–702
62. Otterlei M, Warbrick E, Nagelhus TA, Haug T, Slupphaug G, Akbari M, Aas PA, Steinsbekk K, Bakke O, Krokan HE (1999) Post-replicative base excision repair in replication foci. EMBO J 18:3834–3844
63. Hazra TK, Das A, Das S, Choudhury S, Kow YW, Roy R (2007) Oxidative DNA damage repair in mammalian cells: a new perspective. DNA Repair (Amst) 6:470–480
64. Nagelhus TA, Haug T, Singh KK, Keshav KF, Skorpen F, Otterlei M, Bharati S, Lindmo T, Benichou S, Benarous R, Krokan HE (1997) A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to xpa interacts with the C-terminal part of the 34-kDa subunit of replication protein a. J Biol Chem 272:6561–6566

65. Parker A, Gu Y, Mahoney W, Lee SH, Singh KK, Lu AL (2001) Human homolog of the muty repair protein (hmyh) physically interacts with proteins involved in long patch DNA base excision repair. J Biol Chem 276:5547–5555
66. Hagen L, Kavli B, Sousa MM, Torseth K, Liabakk NB, Sundheim O, Pena-Diaz J, Otterlei M, Horning O, Jensen ON, Krokan HE, Slupphaug G (2008) Cell cycle-specific ung2 phosphorylations regulate protein turnover, activity and association with rpa. EMBO J 27:51–61
67. Guan X, Bai H, Shi G, Theriot CA, Hazra TK, Mitra S, Lu AL (2007) The human checkpoint sensor rad9-rad1-hus1 interacts with and stimulates neil1 glycosylase. Nucleic Acids Res 35:2463–2472
68. Das S, Chattopadhyay R, Bhakat KK, Boldogh I, Kohno K, Prasad R, Wilson SH, Hazra TK (2007) Stimulation of neil2-mediated oxidized base excision repair via yb-1 interaction during oxidative stress. J Biol Chem 282:28474–28484
69. Marenstein DR, Ocampo MT, Chan MK, Altamirano A, Basu AK, Boorstein RJ, Cunningham RP, Teebor GW (2001) Stimulation of human endonuclease iii by y box-binding protein 1 (DNA-binding protein b). Interaction between a base excision repair enzyme and a transcription factor. J Biol Chem 276:21242–21249
70. Wilson SH, Kunkel TA (2000) Passing the baton in base excision repair. Nat Struct Biol 7:176–178
71. Parikh SS, Mol CD, Hosfield DJ, Tainer JA (1999) Envisioning the molecular choreography of DNA base excision repair. Curr Opin Struct Biol 9:37–47
72. Izumi T, Mitra S (1998) Deletion analysis of human ap-endonuclease: minimum sequence required for the endonuclease activity. Carcinogenesis 19:525–527
73. Izumi T, Wiederhold LR, Roy G, Roy R, Jaiswal A, Bhakat KK, Mitra S, Hazra TK (2003) Mammalian DNA base excision repair proteins: their interactions and role in repair of oxidative DNA damage. Toxicology 193:43–65
74. Bhakat KK, Izumi T, Yang SH, Hazra TK, Mitra S (2003) Role of acetylated human ap-endonuclease (ape1/ref-1) in regulation of the parathyroid hormone gene. EMBO J 22:6299–6309
75. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ (2008) Malleable machines take shape in eukaryotic transcriptional regulation. Nat Chem Biol 4:728–737
76. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends Biochem Sci 33:2–8
77. Doublie S, Bandaru V, Bond JP, Wallace SS (2004) The crystal structure of human endonuclease viii-like 1 (neil1) reveals a zincless finger motif required for glycosylase activity. Proc Natl Acad Sci USA 101:10284–10289
78. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. Proteins 42:38–48
79. Li X, Romero P, Rani M, Dunker AK, Obradovic Z (1999) Predicting protein disorder for N-, C-, and internal regions. Genome Inform Ser Workshop Genome Inform 10:30–40
80. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61(Suppl 7):176–182
81. Ishida T, Kinoshita K (2007) Prdos: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res 35:W460–W464
82. Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) Ronn: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21:3369–3376
83. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) Foldindex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435–3438
84. Linding R, Russell RB, Neduva V, Gibson TJ (2003) Globplot: exploring protein sequences for globularity and disorder. Nucleic Acids Res 31:3701–3708
85. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 347:827–839
86. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) Iupred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21:3433–3434
87. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY (2006) Foldunfold: web server for the prediction of disordered regions in protein chain. Bioinformatics 22:2948–2949
88. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics 7:208
89. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 3:35–60
90. Romero Obradovic, Dunker K (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inform Ser Workshop Genome Inform 8:110–124
91. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208
92. Vucetic S, Brown CJ, Dunker AK, Obradovic Z (2003) Flavors of protein disorder. Proteins 52:573–584
93. Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371–378
94. O'Hare P, Williams G (1992) Structural studies of the acidic transactivation domain of the vmw65 protein of herpes simplex virus using 1 h nmr. Biochemistry 31:4150–4156
95. Golovanov AP, Chuang TH, DerMardirossian C, Barsukov I, Hawkins D, Badii R, Bokoch GM, Lian LY, Roberts GC (2001) Structure-activity relationships in flexible protein domains: regulation of rho gtpases by rhogdi and d4 gdi. J Mol Biol 305:121–135
96. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59
97. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756
98. Iakoucheva LM, Kimzey AL, Masselon CD, Bruce JE, Garner EC, Brown CJ, Dunker AK, Smith RD, Ackerman EJ (2001) Identification of intrinsic order and disorder in the DNA repair protein xpa. Protein Sci 10:560–571
99. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456
100. Toth-Petroczy A, Simon I, Fuxreiter M, Levy Y (2009) Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. J Am Chem Soc 131:15084–15085
101. Vuzman D, Azia A, Levy Y (2010) Searching DNA via a "Monkey bar" mechanism: the significance of disordered tails. J Mol Biol 396:674–684
102. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. Cell Res 19:929–949

103. Thayer MM, Ahern H, Xing D, Cunningham RP, Tainer JA (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease iii crystal structure. EMBO J 14:4108–4120
104. Ikeda S, Biswas T, Roy R, Izumi T, Boldogh I, Kurosky A, Sarker AH, Seki S, Mitra S (1998) Purification and characterization of human nth1, a homolog of *Escherichia coli* endonuclease iii. Direct identification of lys-212 as the active nucleophilic residue. J Biol Chem 273:21585–21593
105. Liu X, Roy R (2002) Truncation of amino-terminal tail stimulates activity of human endonuclease III (hnth1). J Mol Biol 321:265–276
106. Stein A, Pache RA, Bernado P, Pons M, Aloy P (2009) Dynamic interactions of proteins in complex networks: a more structured view. FEBS J 276:5390–5405
107. Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit 23:105–116
108. Krueger KE, Srivastava S (2006) Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics. Mol Cell Proteomics 5:1799–1810
109. Busso CS, Iwakuma T, Izumi T (2009) Ubiquitination of mammalian ap endonuclease (ape1) regulated by the p53-mdm2 signaling pathway. Oncogene 28:1616–1625
110. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem 275:29426–29432
111. Seet BT, Dikic I, Zhou MM, Pawson T (2006) Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol 7:473–483
112. Bhakat KK, Hazra TK, Mitra S (2004) Acetylation of the human DNA glycosylase neil2 and inhibition of its activity. Nucleic Acids Res 32:3033–3039
113. Tini M, Benecke A, Um SJ, Torchia J, Evans RM, Chambon P (2002) Association of cbp/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol Cell 9:265–277
114. Bhakat KK, Mokkapati SK, Boldogh I, Hazra TK, Mitra S (2006) Acetylation of human 8-oxoguanine-DNA glycosylase by p300 and its role in 8-oxoguanine repair in vivo. Mol Cell Biol 26:1654–1665
115. Dingwall C, Laskey RA (1991) Nuclear targeting sequences—a consensus? Trends Biochem Sci 16:478–481
116. Lee BJ, Cansizoglu AE, Suel KE, Louis TH, Zhang Z, Chook YM (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. Cell 126:543–558
117. Jackson EB, Theriot CA, Chattopadhyay R, Mitra S, Izumi T (2005) Analysis of nuclear transport signals in the human apurinic/apyrimidinic endonuclease (ape1/ref1). Nucleic Acids Res 33:3303–3312
118. Sarker AH, Ikeda S, Nakano H, Terato H, Ide H, Imai K, Akiyama K, Tsutsui K, Bo Z, Kubo K, Yamamoto K, Yasui A, Yoshida MC, Seki S (1998) Cloning and characterization of a mouse homologue (mnthl1) of *Escherichia coli* endonuclease iii. J Mol Biol 282:761–774
119. Ikeda S, Kohmoto T, Tabata R, Seki Y (2002) Differential intracellular localization of the human and mouse endonuclease iii homologs and analysis of the sorting signals. DNA Repair (Amst) 1:847–854
120. Otterlei M, Haug T, Nagelhus TA, Slupphaug G, Lindmo T, Krokan HE (1998) Nuclear and mitochondrial splice forms of human uracil–DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. Nucleic Acids Res 26:4611–4617
121. Berg OG, Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry 20:6929–6948
122. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2:e100
123. Shimizu K, Toh H (2009) Interaction between intrinsically disordered proteins frequently occurs in a human protein–protein interaction network. J Mol Biol 392:1253–1265
124. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. Febs J 272:5129–5148
125. Pontius BW (1993) Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. Trends Biochem Sci 18:181–186
126. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA 97:8868–8873
127. Russo AA, Jeffrey PD, Patten AK, Massague J, Pavletich NP (1996) Crystal structure of the p27kip1 cyclin-dependent-kinase inhibitor bound to the cyclin a-cdk2 complex. Nature 382:325–331
128. Kiss R, Bozoky Z, Kovacs D, Rona G, Friedrich P, Dvortsak P, Weisemann R, Tompa P, Perczel A (2008) Calcium-induced tripartite binding of intrinsically disordered calpastatin to its cognate enzyme, calpain. FEBS Lett 582:2149–2154
129. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 55:104–110
130. Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. Science 324:203–207
131. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci USA 103:8390–8395
132. Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R (2003) Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 28:81–85
133. Chen X, Zhong S, Zhu X, Dziegielewska B, Ellenberger T, Wilson GM, MacKerell AD Jr, Tomkinson AE (2008) Rational design of human DNA ligase inhibitors that target cellular DNA replication and repair. Cancer Res 68:3169–3177
134. Drew Y, Plummer R (2009) Parp inhibitors in cancer therapy: two modes of attack on the cancer cell widening the clinical applications. Drug Resist Updat 12:153–156