An updated inventory of about 150 human DNA repair genes is described. The compilation includes genes encoding DNA repair enzymes, some genes associated with cellular responses to DNA damage, and other genes associated with genetic instability or sensitivity to DNA damaging agents. The updated human DNA repair genes table (http://www.cgal.icnet.uk/DNA_Repair_Genes.html) is a research and reference tool that directly links to several databases: Gene Cards, Online Mendelian Inheritance in Man, the NCBI MapViewer for chromosome position, and the NCBI Entrez database for the reference nucleotide sequence. This article discusses the approximately 25 genes added, since the original version of the table was first produced in 2001, and some other revisions.

Keywords: DNA; Base excision repair; Nucleotide excision repair; DNA repair gene; Damage response genes; Replication forks
Prologue

In 1993, the authors had the privilege of hosting Dr. Philip Hanawalt, together with his scientific colleague and spouse Dr. Graciela Spivak, for an extended visit to the UK. Many of us working at the Clare Hall Laboratories at the time took advantage of the opportunity to benefit from Phil Hanawalt’s broad perspective of the DNA repair field, his wide interest in many ongoing projects, and his generous willingness to provide advice and comments. Although his visit was technically a sabbatical, he stayed in daily contact with his laboratory and many other pressing duties in the USA, and was also able to visit many colleagues in Europe. In recognition of his pioneering work and his contribution towards the training of many European scientists, Phil Hanawalt was elected in 2002 as an Associate Member of the European Molecular Biology Organization.

1. Introduction

In the early 2001, as the “first complete draft” of the human genome sequence was published, we presented a listing and short analysis of the complement of human DNA repair genes, together with an online table of these genes[1]. The list included about 125 DNA repair enzymes and some gene products associated with cellular responses to DNA damage. The present article summarizes modifications to this list, as of early 2005. The updated table of human DNA repair genes is available on the World Wide Web and is not reproduced here. Since 2001, about 25 genes were added to the table, and a few deleted. This article discusses only the additional genes. An example of an entry and related links is shown in Fig. 1.

The definition of a “DNA repair gene” can be somewhat arbitrary. We attempt to be comprehensive regarding genes encoding actual DNA repair enzymes, as discussed below by pathway. Also included in the table are many “damage response genes” that modulate cellular sensitivity to DNA damage by, for example, altering cell cycle responses. A complete compilation of all DNA damage-related genes would include many more genes affecting cell cycle control, DNA replication and apoptosis, but is beyond the intended scope of this table. However, the table generally does include those genes that are defective in human diseases associated with sensitivity to DNA damage. More comprehensive information on genes and gene products associated with DNA-damage responses can be found in other articles in this issue, and elsewhere[2–5].

2. Base excision repair (BER)

Two previously unrecognized DNA glycosylases, NEIL1 and NEIL2, were found independently by at least three groups[6]. This brings up the number to 10 of clearly distinct DNA glycosylases recognising different types of base damage. The two new enzymes resemble the E. coli enzyme Nei (endonuclease eight), and catalyze the release of oxidized pyrimidine residues from DNA. In this regard, they show
considerable overlap in substrate specificity with the previously known NTHL1 glycosylase. Nevertheless, each of these three repair also enzymes appears to have unique specificity for certain subclasses of lesions. Consequently, suppression of NEIL1 expression by RNA interference resulted in a phenotype of hypersensitivity to ionizing radiation, possibly due to inability to excise a specific stereoisomer of thymine glycol which is recognized by NEIL1 but not by NEIL2 or NTHL1 [7].

DNA single-strand interruptions that cannot be directly rejoined by a DNA ligase are common lesions generated by reactive oxygen species. They are often corrected by a short-patch excision-repair process related to the later steps in BER. The mammalian polynucleotide kinase phosphatase (PNKP) can serve a useful initial role in such single-strand break repair, since this 57 kDa enzyme possesses both an activity for phosphorylation of DNA termini with a 5′-OH group, and an activity for dephosphorylation at 3′-termini. These activities are the same as those of the commonly used reagent PNK of phage T4 origin. RNA-mediated downregulation of human PNKP in tissue culture experiments [8] induces hypersensitivity to several DNA-damaging agents, delayed repair of /H9253-radiation induced single-strand breaks, and a seven-fold higher spontaneous mutation frequency, strongly indicating a direct role of human PNKP in DNA repair. Further, the NEIL1 and NEIL2 DNA glycosylases (discussed above) have associated AP lyase activities that can generate strand breaks with 3′ phosphate termini at damaged sites. PNKP is then required for processing of these blocked DNA termini [9].

One unusual form of DNA strand break can result from interruption of topoisomerase I-induced relaxation of a supercoiled DNA structure. If the intermediate is trapped as a cleavage complex, which could be a consequence of damage to either DNA or the protein, or enzyme inhibition, the topoisomerase remains covalently attached by a 3′-phosphotyrosine residue. A specific Tyr-DNA phosphodiesterase (TDP1) hydrolyzes the bond linking tyrosine to a 3′ DNA end. This enzyme was first discovered in yeast, and the human counterpart and its gene have been defined [10]. TDP1 associates with the BER enzyme DNA ligase III, which provides further evidence for a role in correction of certain types of DNA single-strand interruptions [11]. The neurodegenerative disease spinocerebellar ataxia with azonal neuropathy (SCAN1) is caused by mutations in the human TDP1 gene [12]. SCAN1-deficient individuals have defective single-strand break repair [11], although they do not appear to exhibit chromosome instability or increased cancer frequency.

A clinically different spinocerebellar ataxia syndrome, ataxia ocular apraxia (AOA1), is associated with mutations in APTX, a human gene not studied previously [13,14]. The gene product is referred to as aprataxin. The protein shows partial sequence homology with PNKP in a protein interaction domain, and may be involved in processing certain DNA single-strand interruptions [14]. This model is further supported by the recent finding that aprataxin interacts directly with, and stabilizes the BER protein XRCC1 [15,16].

3. Direct reversal of DNA damage

Single-stranded regions of DNA at transcription bubbles and replication forks are particularly susceptible to damage by alkylating agents, such as methyl methanesulphonate (MMS). The major lesions introduced are the cytotoxic bases 1-methyladenine and 3-methylcytosine, generated by alkylation at sites that are protected in double-stranded DNA. These lesions were recently found to be processed by an unanticipated strategy of DNA repair, oxidative demethylation with release of the methyl group as formaldehyde. The first example of this mode of repair was the E. coli AikB protein, and two human functional counterparts, ABH2 and ABH3, were then described [17,18]. These nuclear DNA dioxygenases employ Fe²⁺ and α-ketoglutarate as cofactors and directly revert DNA damage by a free-radical mechanism. The two human enzymes also act on the minor alkylation lesion 3-methylthymine [19], which differs from 3-methylcytosine and 1-methyladenine by being uncharged at neutral pH.

In addition to ABH2 and ABH3, genes encoding another six apparent homologs of E. coli AikB are present in human cells (for review, see [20]). It seems unlikely that these are all involved in DNA repair, since the various enzymatic strategies for dealing with the major lesions generated by simple alkylating agents are now known. Instead, some of them may be involved in enzymatic removal of methyl groups from nitrogen.
residues in other intracellular macromolecules. The human ABH1 protein has been studied in some detail. In contrast to ABH2 and ABH3, it completely fails to regenerate biological activity in MMS-treated single-stranded DNA used for transfection experiments [17]. Furthermore, ABH1 does not even bind to DNA [21], so at present, it is an unlikely candidate to participate in DNA repair.

Enzymatic monomerization of ultraviolet light-induced pyrimidine dimers by a photolyase is a widely distributed DNA repair mechanism in nature, but in contrast to other vertebrates, placental mammals have lost the function for photoreactivation of DNA during evolution. Thus, no photolyases appear in the table of human DNA repair genes. (Two mammalian proteins structurally related to photolyase, CRY1 and CRY2, are light-sensitive biological clock proteins needed for maintaining circadian rhythms [22]). The loss of DNA photolyases during early mammalian evolution appears unfortunate, because transgenic mice that express a photolyase for monomerization of cyclobutane pyrimidine dimers are highly resistant to sunlight-induced skin cancer [23].

4. Nucleotide excision repair (NER)

An additional entry regarding the NER pathway is the gene for a recently recognized 10th subunit of TFIH. Both yeast and mammalian TFIH have such a subunit, designated Tfb5 in S. cerevisiae and TFB5 or TTTA in human cells. The TTDA/GTF2H5 gene encodes a small protein of ~8 kDa that was overlooked for some time in TFIH preparations [24,25]. This subunit is not absolutely required for transcription or NER, but cells defective in TTDA have lower than normal amounts of TFIH, because TTDA functions to stabilize the TFIH complex [25,26]. Mutations in the gene occur in complementation group A of the sunlight-sensitive inherited disorder trichothiodystrophy [25]. A eukaryotic Tfb5 ortholog was first identified in the alga Chlamydomonas reinhardtii and designated REX1. Mutants in this algal gene are defective in the removal of UV radiation-induced cyclobutane pyrimidine dimers from DNA [27].

NER involves the accessory use of gene products that also function in DNA replication, such as DNA ligase I, PCNA, and RPA. These gene products also operate in several other DNA repair pathways and could be listed at several locations in the table, but are referred to in the NER section for convenience.

In the “NER-related” section of the table are several genes that participate in specialized aspects of NER, including the CSA and CSB genes that are needed for transcription-coupled nucleotide excision repair. The XAB2 (XPA-binding protein 2) gene is also included in this section because of its interaction with XPA, CSA, and CSB [28]. However, XAB2 may not participate directly in NER, although it does appear to be critical for normal transcription [28].

5. Homologous recombination

The mechanism of resolution of Holliday junctions formed during homologous recombination in human cells is a subject of continuing interest. For example, now there are indications that the "RAD51 paralogs" encoded by the RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3 genes participate in late stages of homologous recombination, including branch migration and Holliday junction resolution [29].

New entries have been added for the genes encoding MUS81 and its protein partner MMS4. Human MUS81-MMS4 is a nuclease (related to the NER nuclease XPF-ERCC1) that can cleave specific branched and forked structures in DNA, including Holliday junctions [30]. Investigation of the MUS81 ortholog in the fission yeast Schizosaccharomyces pombe indicates that the enzyme has many properties of a Holliday junction resolvase for homologous recombination [31]. On the other hand, the properties of the ortholog in the budding yeast Saccharomyces cerevisiae points to a different function in reactivation of DNA replication forks that become stalled at sites of DNA damage [32].

6. Non-homologous end joining (NHEJ)

The Artemis nuclease (SNM1C) has a special and important role in the hairpin cleavage of immunoglobulin genes during V(D)J recombination. SNM1C/Artemis may also play an accessory role during NHEJ due to its association with, and modification by, the DNA-dependent protein kinase in the cell nucleus [33]. In human cells exposed to ionizing ra-
radiation, Artemis is hyperphosphorylated by the ATM kinase. Joining of a significant minority of radiation-induced DNA double-strand breaks requires this ATM-activated form of the Artemis nuclease for processing of certain damaged or altered structures at DNA termini [34]. This could reflect Artemis-dependent cleavage of hairpins, generated by fold-back of frayed ends and subsequent accidental interstrand join by a nuclear DNA ligase. The results provide an intriguing model to help explain in molecular terms the marked radiosensitivity of ATM-deficient individuals.

7. DNA polymerases

Many or most of the approximately 14 different DNA polymerases found in human cells are specialized for operation in distinct DNA repair pathways, or for bypass of specific classes of adducts in DNA. These enzymes fall into four different families designated A, B, X, and Y.

POLQ (polH9258), recognized as a family A enzyme in human cells several years ago, is a larger polypeptide than originally appreciated. The full POLQ cDNA encodes a protein of 2590 amino acids including a DNA polymerase in the C-terminal region and a DNA helicase-like domain in the N-terminal region [35]. The purified enzyme efficiently inserts bases opposite AP sites and thymine glycol adducts, and efficiently extends from such misinsertions [36]. A mouse with a knockout of the PolQ gene shows spontaneous and radiation-induced chromosome instability in hematopoietic cells [37]. Simultaneous knockout of both mouse Atm and PolQ function severely prevents or impairs development [37], indicating that without the ATM checkpoint, PolQ is a particularly important enzyme that perhaps normally participates in the bypass of spontaneous DNA lesions in some cell types.

Two additional enzymes related to POLQ have also been found encoded in the human genome and added to the table. One of them is another DNA polymerase, named POLN (polH9263). This 900 amino acid family A polymerase appears to be less widely expressed than POLQ and has many alternatively spliced variants [38]. Little is known of its function. The second gene is a DNA helicase, designated HEL308, encoding an 1101 amino acid enzyme with a sequence similar to the helicase domain of POLQ [39]. The Drosophila ortholog of HEL308 is designated spn-C and is one of class of “spindle” mutants affecting embryonic development [40]. The other spindle genes spn-A, spn-B, and spn-D are the Drosophila orthologs of RAD51, XRCC3, and RAD51C, respectively, suggesting that HEL308 may be a helicase involved in a pathway of homologous recombination.

Functions are being suggested for several of the previously listed DNA polymerases. For example, the family X DNA polymerases POLM (polH9262) and POLL (polH9261) have properties and protein-protein interactions suitable for participation in non-homologous end-joining, specifically the ability to assist in forming ligatable ends from non-complementary DNA termini with “microhomology” [33,41].

Among the family Y DNA polymerases, POLK (polH9260) seems particularly suited to the bypass of benzo pyrene diol epoxide-guanine adducts [42,43]. The family Y enzyme POLI (polH9259) is a DNA polymerase similar to the xeroderma pigmentosum variant (XP-V) complementing polymerase POLH (polH9257). POLI associates closely with POLN at damaged sites in vivo [44], and can incorporate nucleotides opposite sites of UV radiation-induced [6-4] photoproducts and some cyclobutane pyrimidine dimers [45]. However, POLI-defective mouse cells do not seem to exhibit sensitivity to UV radiation or other obvious phenotypes [46], and so the in vivo function of this DNA polymerase is currently unclear.

8. Fanconi anemia

Most of the gene products of the Fanconi anemia pathway co-operate to carry out mono-ubiquitination of the FANCD2 gene product. The ubiquitination complex includes the FANC gene products of genetic complementation groups A–G, and L [47]. The reason for this elaborate targeting of FANCD2 for modification is currently not known. It appears to be related to improving some aspect of homology-directed DNA repair [48]. The BRCA2 gene product is identical to FANCD1, and is likely involved in a homologous recombination pathway. Information on the recently identified FANCB gene [49] has been added to the table, as well as the function for FANCL as an evolutionarily conserved ubiquitin E3 ligase component.
participating in the ubiquitination of FANCD2 [50].

9. Chromatin structure

Many gene products that alter chromatin structure could conceivably affect the operation or efficiency of various DNA repair pathways. We have not included various chromatin remodeling factors that might potentially be involved in DNA repair, but two gene products have been listed. One of these is H2AFX, encoding the histone H2A variant, H2AX. Phosphorylation of this histone variant is a rapid consequence of DNA strand breakage [51], and disruption of H2AX modulates chromosome instability phenotypes in the mouse [52,53].

The gene for the large (150 kDa) subunit of Chromatin Assembly Factor 1 (CAF1), designated CHAF1A, has also been added to the list as CAF1 can catalyze assembly of chromatin coupled to NER in vitro systems [54], and is localized to sites of NER in vivo [55].

10. Other conserved DNA damage response genes

Direct precursors for DNA synthesis during replication and repair are synthesised by the heterodimeric enzyme ribonucleotide reductase, comprised the R1 and R2 subunits. The R2 protein contains the tyrosine-derived stable free radical at the active site for reduction of ribose to deoxyribose. Human cells also encode a similar alternative form of the R2 subunit which is encoded by a p53-inducible gene. The P53R2 protein differs from the regular R2 protein by having a nuclear rather than cytoplasmic subcellular localization, but both proteins interact with the same R1 subunit. Expression of the P53R2 protein is induced by DNA damage after cellular exposure to UV light or ionizing radiation [56]. The results suggest that P53R2 might aid in the provision of precursors for DNA synthesis during damage-induced repair and recombination. A human cancer cell line, HCT116, carries an inactivating mutation in the P53R2 gene. This cell line is hypersensitive to ionizing radiation, with enhanced apoptotic cell death after DNA damage [57].

11. Editing and processing nucleases

A nuclease that can cleave a DNA strand at the second phosphodiester bond 3′ of a broad variety of structural alterations and perturbations has been investigated in E. coli for many years and called endonuclease V. Its physiological role remains uncertain, but it will incise DNA adjacent to a deaminated adenine (hypoxanthine) residue. Part of the protein sequence resembles that of the E. coli UvrC protein, which incises DNA during NER. Recently, the nfi gene encoding an endonuclease V-like enzyme was noted to be conserved during evolution and present in the human and mouse genomes. The cloned and overexpressed 37 kDa murine ENDOV enzyme has biochemical properties similar to the E. coli enzyme [58]. Ongoing studies with gene knockout mice deficient in this function may clarify its putative role in mammalian DNA repair.

12. Other genes that resemble DNA repair genes

Some previous entries in the table remain as genes that are homologous to known DNA repair genes, but for which no DNA repair function is yet known. ABHI is one such entry mentioned above in connection with the other ABH genes. POLN and HEL308 were also referred to above in the section on DNA polymerases. Other examples are two homologs of MUTL of unknown function (PMS2L3 and PMSL4/PMS6), an SNM1 family member called SNM1B, and a homolog of the middle subunit of Replication Protein A, designated RPA4.

RECQ-family DNA helicases are mutated in several human disorders associated with sensitivity to DNA damaging agents, including the products of the WRN (Werner syndrome), BLM (Bloom syndrome), and RECQL4 (Rothmund–Thompson syndrome) genes [59]. Two additional family members, RECQL (RECQ1) and RECQL5, have consequently been added to the table in the section of “other identified genes with a suspected DNA repair function”. A RECQL5 gene product isoform, known as RECQ5β, has both strand-annealing and helicase properties in vitro [60]. Studies with the chicken and C. elegans orthologs of RECQL5 suggest that the gene product serves as a backup to BLM function [61,62].
A protein with a specific recognition motif for binding to RNA and/or single-stranded DNA that exhibits partial sequence homology with the homologous recombination protein RAD52 is present in vertebrates. The novel protein is called RDM1, and it can assemble to form filament-like structures on DNA. Ablation of this function by a gene knockout in the chicken B cell line DT40 results in cellular hypersensitivity to cisplatin, suggesting that RDM1 may recognize cisplatin-DNA adducts and be involved in DNA repair [63].

The gene NEIL3 resembles the NEIL1 and NEIL2 genes encoding DNA glycosylases discussed above with respect to base excision repair. It has been added to the table, but because it has no currently known DNA repair activity, NEIL3 is presently relegated to the “suspected DNA repair function” section.

13. Omissions and deletions in this edition

As indicated earlier, we have not attempted to be comprehensive regarding damage-response genes that are tangential to enzymatic DNA repair functions. The p53 (TP53) gene is, however, included because of its central role in the damage-response in human cells, and we now also list ATRP (ATR-interacting protein), the ortholog of S. cerevisiae DDC2 and S. pombe rad26. Other genes that conceivably could have been included in the table, but are not, include MDM2 (HDM2), MDM4 (MDMX), TRF4 (TRF4-1, TRF4-2, and TRF2), subunits of replication factor C, and TRF2, encoding a telomeric protein.

Several genes have been deleted from the previous edition of this table. Two of them are TRF4-1 and TRF4-2 (PAPD5). TRF4-1, also known by the symbol POL5 (pol5), is homologous to the S. cerevisiae TRF4 gene. The S. cerevisiae gene is important for sister chromatid cohesion and is reported to have modest DNA polymerase activity [64]. On the other hand, a related gene in S. pombe, cdc13, encodes a cytoplasmic poly(A) polymerase, not a DNA polymerase [65]. If more information arises regarding DNA damage-related DNA polymerase activity for these gene products, they will be added to future editions. A gene encoding a homolog of poly(ADP-ribose) polymerase designated PARP3 (ADPRTL3) was also listed in our original table. The corresponding protein is localized to centrosomes and lacks a DNA binding domain, and so is omitted in the current version.

References


[42] N. Suzuki, E. Ohashi, A. Kolbasova, N.E. Geacint-


[46] J.P. McDonald, E.G. Frank, B.S. Plosky, I.B. Rogozin, C. Ma-


[48] K. Nakanishi, Y.G. Yang, A.J. Pierce, T. Taniguchi, M. Dig-

[49] A.R. Meetei, Z. Yan, W. Wang, FANCL replaces BRCA1 as the

[50] G. Almouzni, Initiation and bidirectional propagation of chromatin assembly from a target site for nu-

[51] O.A. Sedelnikova, D.R. Pilch, C. Redon, W.M. Bonner, Histone

[52] H. Tanaka, H. Arakawa, T. Yamaguchi, K. Matsui, S. Fukuda,

[53] C.H. Bassing, H. Suh, D.O. Ferguson, K.F. Chan, J. Manta,

[54] P-HL. Guillard, J.G. Mogg, D.M.J. Roche, J.-P. Quivy, P.B.


[56] T. Yamaguchi, K. Matsui, Y. Sugiy, M. Iwadate, M.A. Fujino,

[57] A. Moe, J. Ringvoll, L.M. Nordstrand, L. Eide, M. Bjoras, E.

[58] A. Moe, J. Ringvoll, L.M. Nordstrand, L. Eide, M. Bjoras, E.


[60] P.L. Garcia, Y. Liu, J. Jiricny, S.C. West, P. Janscak, Hu-

[61] P. Lecoq, J. van Dyck, RDM1, a Novel RNA-recognition motif (RRM)-


[64] Z. Wang, I.B. Castano, C. Adams, C. Vu, D. Fitzhugh, M.F.


[67] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[68] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[69] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[70] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[71] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[72] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[73] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[74] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[75] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[76] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[77] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[78] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[79] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[80] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[81] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[82] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[83] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[84] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[85] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[86] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[87] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[88] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[89] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[90] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[91] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.

[92] A. Celeste, S. Difilippantonio, M.J. Difilippantonio, O.