REVIEW ARTICLE

The DNA damage-recognition problem in human and other eukaryotic cells: the XPA damage binding protein

James E. CLEAVER*[1] and J. Christopher STATES†

*Laboratory of Radiobiology and Environmental Health, University of California, San Francisco, CA 94143-0750, U.S.A., and †Center for Molecular Medicine and Genetics, Wayne State University, 2727 Second Avenue, 4th Floor, Detroit, MI 48201, U.S.A.

The capacity of human and other eukaryotic cells to recognize a disparate variety of damaged sites in DNA, and selectively excise and repair them, resides in a deceptively small simple protein, a 38–42 kDa zinc-finger binding protein, XPA (xeroderma pigmentosum group A), that has no inherent catalytic properties. One key to its damage-recognition ability resides in a DNA-binding domain which combines a zinc finger and a single-strand binding region which may infiltrate small single-stranded regions caused by helix-destabilizing lesions. Another is the augmentation of its binding capacity by interactions with other single-stranded binding proteins and helicases which co-operate in the binding and are unloaded at the binding site to facilitate further unwinding of the DNA and subsequent catalysis. The properties of these reactions suggest there must be considerable conformational changes in XPA and associated proteins to provide a flexible fit to a wide variety of damaged structures in the DNA.

INTRODUCTION

The capacity to recognize and respond to DNA damage is a universal property of living organisms. The capacity to repair DNA is found in such minimal organisms as the mycoplasma through to complex multicellular vertebrates and plants. DNA damage was once the province of radiobiologists, who had interests ranging from the identification of the size and shape of sensitive targets to the physiology of lethally irradiated individuals. Inherent in these interests was a fundamental mystery of how small amounts of energy could cause such major cellular disruption, when delivered in the form of ionizations rather than as thermal energy. Now, the cellular capacity to respond to damage from radiation, free radicals and reactive chemicals is seen as a central feature of the homoeostasis of living organisms.

At the core of all cellular responses to DNA damage, whether it comes from external ionizing radiation or UV light, or from internal sources such as oxidative metabolism, is the question of the mechanism by which damaged DNA is recognized. This is a subset of the questions of how proteins find their cognate target sites in DNA, in which the recognition signal is not a particular sequence or secondary structure, but a departure from normality. This can include loss of normal base-pairing, photochemical modification of bases and chemical addition and modification to various parts of the DNA bases and polynucleotide chain.

DAMAGE RECOGNITION BY THE BASE AND NUCLEOTIDE REPAIR SYSTEMS

The DNA repair systems now known include a wide diversity of solutions to the damage-recognition problem. Two major solutions have evolved according to the nature of the damage: base excision-repair with a narrow and highly specific recognition mechanism and nucleotide excision-repair (NER) with a much wider and versatile recognition capacity [1,2]. These two mechanisms are further distinguished by their separation of damage recognition from catalytic activity. In base excision-repair, the protein which recognizes and binds to the damaged base is catalytically active in carrying out cleavage of the glycosyl bond. In NER the recognition event involves a protein which is not catalytically active but acts as a binding site or anchor on which catalytically active proteins then congregate.

The separation of recognition from catalysis has occurred also in mismatch repair [3,4]. This pathway mainly acts on singly mismatched bases that occur during DNA replication errors and as a result of deamination of 5-methylcytosine to form uracil, and on small DNA loops. A mismatched base-pair or loop acts as a binding site for a heteroduplex protein that recruits a long stretch of DNA containing a nick in newly replicated DNA, from which the nascent strand is degraded and resynthesized. The similarity in principle is reflected in the overlap observed between NER and mismatch repair in some of their substrate specificities [5,6] and in their dual involvement in regulation of the strand selectivity in transcription-coupled NER [7]. These overlaps lead to the unexpected results that cells defective in mismatch repair can exhibit increased resistance to DNA-damaging agents [6] and do not show the increased rates of repair of UV damage usually observed in the transcribed strands of expressed genes [7].

Base repair mainly operates on simple, single-base modifications, many of which may arise from endogenous damage such as oxidation. The repair involves systems of apparent high specificity by which damaged bases are removed by glycosylase cleavage of the sugar–base linkage [8,9]. Recent crystal structures of typical enzymes in this, the base excision-repair pathway, indicate that they recognize damage embedded in DNA by swinging the damaged base into an extrahelical configuration and then acting in a typical enzyme–substrate fashion with a

Abbreviations used: UVB and UVC, UV light of medium wavelengths (≈ 280–320 nm) and short wavelengths (≈ 260 nm) respectively; NER, nucleotide excision-repair; XPA, XPB etc., xeroderma pigmentosum group A, B etc.; CSA, CSB etc., Cockayne syndrome group A, B etc.; TTD, trichothiodystrophy; UVRA, UVRB and UVRC, UV-resistance genes A, B and C of *Escherichia coli*; RPA(HSSB), replication protein A (human single-strand binding protein); TFIIH, transcription factor IIH; ERCC1, excision-repair cross-complementing group 1; RFC, replication factor C; PCNA, proliferating-cell nuclear antigen; NLS, nuclear localization signal.

[1] To whom correspondence should be addressed.

binding pocket that fits the damaged base with exquisite specificity [8,9]. Uracil-N-glycosylase, endonuclease III and related base-excision enzymes appear to act in this fashion to remove uracil and other bases. This apparent specificity may be relaxed with other enzymes such as 3-methyladenine glycosylase and formamidopyrimidine-N-glycosylase, where the substrate range may be wider than the original nomenclature would suggest [10]. These enzymes appear to recognize damage of a wider range of structures that would make a simple match between an extrahelical damaged base and a specific binding pocket less easy to envisage.

The more complex NER system further resolves the conflicting requirements that a repair system have both specificity and versatility, in the face of DNA damage that can be an essentially unlimited number of different modifications. This system solves the recognition and excision problem by separating the recognition of damage from most of the catalytic activity and cleaves the polynucleotide chain distal from the site of damage, removing a short oligonucleotide containing the damage with little regard for the chemical modification itself [11]. Even here, however, some form of specific recognition is required and there is a hierarchy of efficiencies by which some damaged sites are more readily acted upon compared with others. Influences of the damage itself, the flanking bases, its chromatin context, and more general metabolic activities all influence the rates, specificity and efficiency of damage recognition and repair.

TWO MECHANISMS OF NER – OR MORE?

At least two major mechanisms of NER have evolved, approximately corresponding to the Prokaryotic and Eukaryotic Kingdoms. The typical prokaryotic mechanism is highlighted by the UVRA, UVRB and UVRC proteins of *Escherichia coli* [1] (products of the UV-resistance genes A, B and C). The UVRA protein acts as a non-catalytic damage-sensing protein, which serves to deliver the catalytically active nucleases UVRB and UVRC to the site of damage. UVRB and UVRC then make single-strand cleavages in the DNA at defined sites 5′ and 3′ to the damage, 12–13 nucleotides apart. This 'matchmaker' function of UVRA is also found in the more complex eukaryotic NER system. The product of the human xeroderma pigmentosum gene (XPA) [12], and its homologues, is the principal DNA damage binding protein and appears to fulfil the matchmaker function in eukaryotes. A similar principle holds, that the initial damage binding protein attracts two nucleases which cleave either side of the damage. In both prokaryotic and eukaryotic cells the NER system is also linked to gene transcription either for functional reasons to deliver repair systems to sites of particular importance in the hierarchy of repair or as an example of cellular economy in which repair enzymes that have infrequent use exercise a second function [13]. Understanding the structure and function of the damage-recognition XPA protein will be an important step in discovering how higher cells have the capacity to employ a single system for the repair of a range of damage.

A particularly interesting question in repair is how archaeal cell types manage their extreme environments which subject them to reactive chemicals, heat, pressure, high salt, and many other adverse conditions which would be expected to damage DNA. These organisms appear to have a transcription apparatus based on an RNA polymerase II-like mechanism with its associated transcription factors that may more closely resemble those of eukaryotes [14]. A mammalian-type transcription-coupled NER might therefore be anticipated. But a search of the DNA sequence databases for currently sequenced archaeal genomes does not locate genes which correspond to a eukaryotic-like NER system. This raises the possibility that the Archaeal Kingdom may have evolved a third variation on the NER principle.

DISTRIBUTION OF XPA-DEPENDENT NER

An NER system based on a protein homologous with XPA is ubiquitous in eukaryotes. Homologous NER proteins with closely similar amino acid sequences have been identified in *Saccharomyces cerevisiae* and *S. pombe*, *Drosophila*, fish, amphibia, chicken, rodents and humans. Thus DNA repair proteins represent a class of proteins with a very high degree of homology across species. The third Kingdom, the Archaea, has similarities to eukaryotes in several respects, but in a homology search we carried out among the archaeal genomes that have been partly or fully sequenced (e.g. *Methanococcus jannaschii*) XPA-like homologues have not been detected. *M. jannaschii* also lacks DNA sequences corresponding to the the UVRABC mechanism, raising the possibility that some of the archaeal species may have developed different strategies for dealing with DNA adducts. A study of repair of UV damage in halophilic bacteria has shown that they do have the capacity to repair the two major photoproducts, cyclobutane dimers and [6-4] pyrimidine pyrimidinones, both by photoreactivation and by excision-repair pathways [15]. One possible candidate for an alternative excision mechanism is the nuclease which can cleave the phosphodiester bond adjacent, 5′, to a photoproduct; a UV endonuclease with such a mechanism is also found in *Neurospora crassa*, *S. pombe* and *Bacillus subtilis*, and its distribution within the Archaea and other Kingdoms has only recently begun to be investigated[16,17].

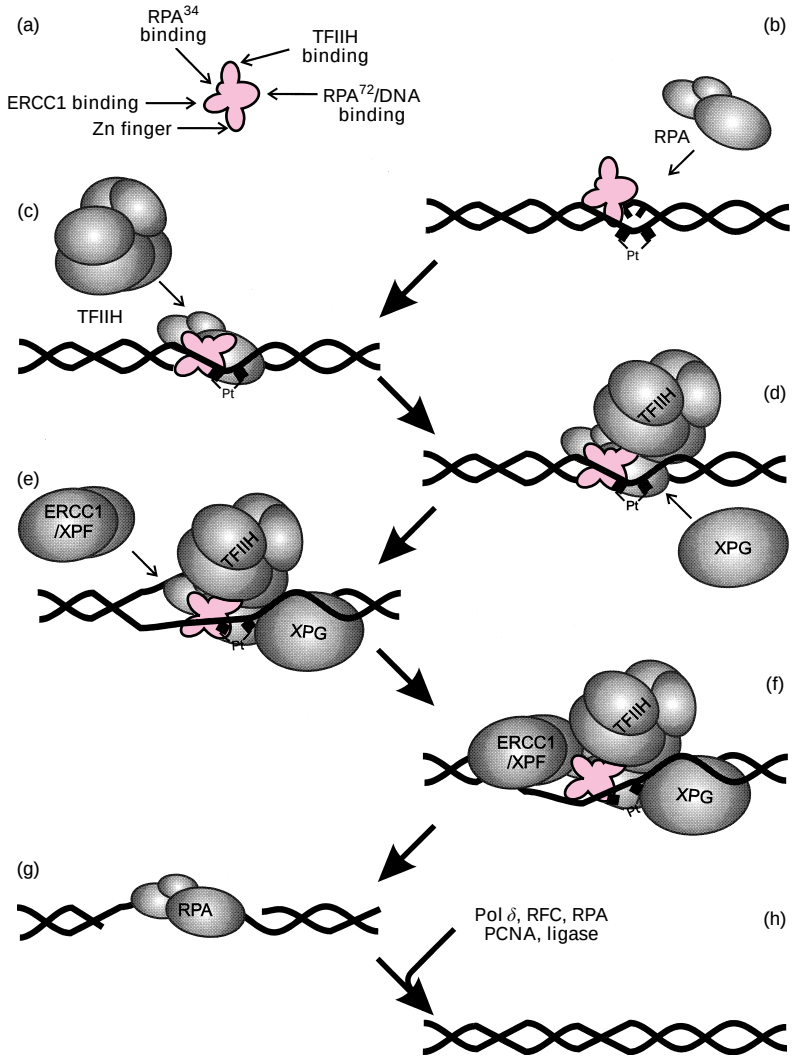
The question of repair in the mitochondria of mammalian cells also raises interesting evolutionary questions in relation to the origin of these cellular organelles and their exposure to high levels of oxidative damage. The mitochondrial genome appears to have few, if any, genes which specify DNA repair systems, and may therefore rely on systems imported from the nucleus. Mitochondria appear unable to remove pyrimidine dimers from their DNA and have been considered to lack a NER capacity, even in cells which have a competent nuclear system [18]. They do, however, have considerable capacity to repair oxidative and alkylation damage to their DNA, and may therefore import the components for base excision-repair from the cell nucleus [19]. Surprisingly, deficiencies in nuclear NER capacity by loss of function of the XPA DNA binding protein decreases the capacity of mitochondrial repair toward oxidative damage [19]. This indicates that the XPA binding protein can stimulate repair of oxidative damage, but not pyrimidine dimers, in the mitochondria, suggesting that XPA has a role in a mechanism of repair that may not require the RNA polymerase-II transcription coupled repair of the nucleus.

THE MECHANISM OF XPA-DEPENDENT NER

The sequence of events involved in damage recognition and repair can be summarized by the following sequence of events (Scheme 1), which represent an approximate temporal sequence [11,20].

Step 1 (Scheme 1b)

The damage in DNA is recognized and bound by XPA, the primary recognition protein [12]. This protein can also form a complex with human single-strand binding protein [HSSB, or RPA (replication protein A)], which may bind on the opposite undamaged strand [21]. Two other DNA-binding proteins coded



Scheme 1 Model of XPA and general scheme of NER

(**a**) Functional domains of XPA. XPA, shown in pink throughout this Scheme, is modelled as having lobed domains for interaction with DNA and other components of the DNA damage recognition/excision complex. (**b**) XPA bound at the site of a cisplatin intra-strand cross-link. XPA is shown inserting in the helix and recruiting RPA. (**c**) XPA/RPA inserted into the helix and recruitment of TFIIH to the complex. (**d**) XPG is recruited by binding to TFIIH and RPA. (**e**) ERCC1-XPF is recruited by binding to XPA. (**f**) The major components of the complex are assembled and ERCC1-XPF, and XPG make their incisions. (**g**) The oligonucleotide containing the lesion has been removed, leaving a gapped duplex with RPA bound to the single-stranded (ss) DNA region. (**h**) Repair replication enzymes synthesize new DNA and seal the remaining nick to complete repair. Other components are known to participate in recognition and excision (i.e. XPC/HHR23A, XPE), but their roles are less clearly defined at present. Therefore only principal components are shown for clarity in presentation of the model focusing on XPA.

for by the *XPE* and the *XPC* genes also appear to be involved, but their precise role is unclear. XPC may augment binding to lesions which make slight helix distortions, such as cyclobutane pyrimidine dimers [22]. XPE deficiencies can be complemented *in vitro* by RPA [23].The binding region of RPA defines the limits of a 30-nucleotide patch which will eventually replace the damaged strand and stretches five nucleotides 3′ to the dimer and 24 nucleotides 5′ to the dimer.

Step II (Scheme 1c)

The XPA–RPA complex recruits the multicomponent transcription factor TFIIH which contributes the helicases, XPB and XPD, which act in opposing directions to unwind the DNA in the damaged region. This recruitment appears to be mediated by HHR23B, a partner to the two-protein XPC–HHR23B complex [24,25].

Step III (Scheme 1d)

The XPG nuclease binds to the RPA protein on the 3′ side of the dimer. This nuclease then cuts the DNA on the strand containing the dimer at the the single-strand/double-strand junction on the 3′ side [26].

Step IV (Schemes 1a and 1f)

The XPA protein then acts as a binding site for the 5′-nuclease,

which is a multimeric protein containing the ERCC1 and XPF proteins [27,28]. This cuts at the single-strand/double-strand junction on the 5′ side of the damaged site. The binding of XPA to the damaged site appears to be enhanced by the interaction with ERCC1 [29,30]. The nucleases which carry out both cleavages appear to be activated by interaction with RPA [31]. The incision sites lie at a 90° angle to the site of the lesion, so that steric hindrance is avoided between the cleavage enzyme complexes and a bulky adduct on which XPA is bound [32].

Step V (Scheme 1g)

The 30-nucleotide oligonucleotide which contains the damaged site and is still bound with some of the components of the excision complex is displaced, leaving a single-strand gap which still contains RPA.

Step VI (Scheme 1h)

The gapped DNA is then patched by the combined action of DNA polymerase δ and its cofactors, proliferating-cell nuclear antigen (PCNA) and replication factor C (RFC). Repair is then completed by DNA ligase, which closes the remaining strand gap.

There may be an important difference between the mechanism and efficiency of XPA binding to damaged DNA in isolation, and the capacity of the system as a whole. Although a major part of the specificity and the efficiency of the repair process is dictated by the properties of XPA, its binding appears to require augmentation for lesions that represent minimal deviations from the normal DNA structure [33].

DISCOVERY OF XPA IN A HUMAN DISEASE PREDISPOSING TO CANCER

XPA (Table 2) was recognized as one of eight genes underlying the human disorder xeroderma pigmentosum [34] and was associated with some of the most severely affected patients. It is now known that all of the genes play a role in the response of human cells to UV damage; seven of the genes (*XPA–XPG*) represent components of the NER system (Scheme 1) and the eighth, which is presently uncloned, is likely a cofactor involved in replication fork progression on damaged parental DNA.

Mutations in the *XPA* gene result in loss of the capacity to repair UV damage from solar or artificial sources, in the UVC and UVB ranges (approximately up to 280 or 300 nm), which makes cultured XP-A cells extremely UV-sensitive. Patients exhibit a 10^3–10^4 increase in the age-specific incidence of skin cancer on sun exposed areas of the skin [35], and cells in culture have a much elevated UV-induced mutation rate [36]. The evidence that the gene product is responsible for the initial recognition of DNA damage came from a variety of experiments which became possible once the gene was cloned [12]. XPA has a sequence that indicated it is a DNA binding protein with a single Cys_2-Cys_2 zinc finger [12,37]. Its binding capacity on UV-damaged DNA was demonstrated by gel-shift experiments [38]. The binding of UV-damaged DNA by purified XPA was shown to be principally a result of binding to [6-4] pyrimidine-pyrimidone photoproducts and not to the more prevalent cyclobutane pyrimidine dimers. The binding of XPA to the latter lesion is significantly increased by formation of a heterodimer between XPA and RPA [21,39]. Thus the functional form of XPA in the cell may be as a heterodimer with RPA.

THE *XPA* GENE

XPA is located on chromosome 9 (9q34.1), contains six exons and spans approx. 23 kb. XPA expression is extraordinarily low in fibroblasts, with only five to eight molecules of *XPA* mRNA present in each cell, putting it among the class of genes with the lowest expression levels [40]. Consistent with the steady-state mRNA data, is that the intact promoter is extraordinarily weak in transient expression assays in fibroblasts [41]. Like several other NER genes [42,43], the *XPA* promoter lacks common basal transcription signals (TATA, CCAAT, G/C box). *XPA* also lacks the sequence motif conserved between mouse and human in ERCC1 hypothesized as a NER regulatory element [43]. A region containing a negative control element has been identified in both the human [41,44] and mouse [45] *XPA* promoters, but deletion of the negative element results in only a two-fold increase in expression [41]. However, intracellular XPA levels can be rate-limiting for NER, and modest increases have dramatic effects on the efficiency of removal of different DNA lesions [46]. Tissue-specific variation in the levels of *XPA* mRNAs were recently demonstrated, suggesting that *XPA* expression may be transcriptionally regulated in a cell-type-specific manner [40]. It is possible that positive elements active only in other cell types are present that were not identified in the fibroblast experiments.

THE XPA PROTEIN

The XPA protein is relatively small, 38–42 kDa, and functions as a monomer, despite its ability to be involved in an extensive network of protein–protein interactions (Scheme 1, Table 1). The protein shows two bands on SDS/PAGE gels for reasons which have not been fully explained. Southwestern-blot analysis of the bovine XPA indicated that only the fastest migrating species bound DNA, whether single- or double-stranded, or UV-irradiated double-stranded [47]. This observation suggests that XPA may exist in multiple forms and that conversion of one form into another may regulate the availability of functional XPA.

Sequences critical to the interaction of XPA with other proteins or DNA have been identified in all exons except exon V (Figures 1 and 2). Although no proteins interacting with exon V sequences have been identified, exon V is highly conserved among vertebrates [48] and its deletion inactivates XPA [49].

The normal subcellular localization of XPA is in the nucleus, but its specific sequestration in the nucleus is not essential for function. When XPA lacking the nuclear localization signal encoded in exon I (Figure 1) is overexpressed by a strong viral promoter [49] the protein can still carry out repair of nuclear DNA. DNA binding is dependent on the presence of the zinc finger (Figure 1) containing four cysteine residues encoded by exon III [50]. Mutation of any of the four cysteine residues of the zinc finger to a serine greatly reduces XPA function [49], and deletion of exon III eliminates function.

XPA binds UV-irradiated DNA modestly better than it does unirradiated DNA [38]. It appears that XPA alone recognizes only the [6-4] photoproduct and has little affinity for cyclobutane pyrimidine dimers, which are the predominant UV-induced lesions in DNA. However, XPA forms a complex with RPA [21,39]. The XPA–RPA complex has a greater affinity for UV-damaged DNA than either protein alone and also preferentially binds cyclobutane pyrimidine dimers [21]. Thus the lesion specificity of XPA is enhanced and broadened by its association with RPA, and possibly with other components of the repair system. The degree of association may also influence the specificity of the repair system, as for example the requirement for the XPC single-strand binding protein for repair of cyclobutane pyrimidine dimers but not cholesterol adducts *in vitro* [22].

Table 1 Genes and gene products involved in the excision steps of NER in eukaryotic cells*

The complete excision-repair process, including resynthesis requires also RPA, PCNA, RFC, DNA polymerase δ or ϵ, and ligase. Additional genes associated with Cockaye syndrome, *CSA* and *CSB*, are involved with regulating transcription-coupled repair, and there is evidence that another repair gene, *TTDA*, specifically involved in trichothiodystrophy, may be among the TFIIH components.

Gene*	Size (kb)†	Site	Protein (no. of amino acids; molecular mass)	Functions‡
XPA(*Rad14*)	25	19q13.2	273; 31 kDa	Damage recognition
XPB(*Rad25*)	45	2q21	782; 89 kDa	3′–5′ helicase (TFIIH)§
XPC(*Rad4*)	24	3p25.1	940; 106 kDa	Single-stranded DNA binding
XPD(*Rad3*)	19	19q13.2	760; 87 kDa	5′–3′ helicase(TFIIH)§
XPE‖	?	11q12-13	1140; 127 kDa	Damage binding
XPF(*Rad1*)	30	16p13.13	916; 104 kDa	Cleavage 5′ to damage
XPG(*Rad2*)	32	13q33	1186: 133 kDa	Cleavage 3′ to damage
ERCC1(*Rad10*)	15	19q13.2	297; 33 kDa	Cleavage 5′ to damage
HHR23B(*Rad23*)	–	3p25.1	409: 43 kDa	XPC complex

* The human gene nomenclature is given first and the corresponding gene in *S. cerevisiae* is given in parentheses.

† Genomic size.

‡ For fuller description of functions, see the text and Scheme 1.

§ The basal transcription factor TFIIH includes a suite of five to nine proteins, most of which, in the corresponding *S. cerevisiae* transcription complex, are associated with UV-sensitivity. But only the two, XPB and D, have been identified in human disorders.

‖ XPE is extremely enigmatic. The gene listed is missing from some XPE cell lines, and is found associated with a smaller 48 kDa protein; mutations have been found in this subunit in some XPE cell lines [109]; the protein has specificity for binding to damaged DNA and resembles XPA in function, but its precise role in NER is still unclear.

Table 2 General properties of the *XPA* gene and gene product in human and mouse

Property	Human	Mouse
Genomic size	25 kb	21 kb
Exon number	6	6
Chromosomal location	9q34.1	4C2
Promoter location	−156 nt upstream	−313 nt upstream
Acceptor/donor junctions	AG/GT	AG/GT
mRNA	1.3–1.4 kb	1.0–1.1 kb
Open reading frame	273 amino acids	303 amino acids
Molecular size (SDS/PAGE)	40, 42 kDa	
Solubility	Hydrophilic	Hydrophobic
Isoelectric point (pI)	7.5	
Motifs	Zn finger, Glu^7	Zn finger
Thermostability	< 40 °C	

XPA binding of RPA is essential for NER, and the regions involved in binding have been characterized (Figures 1 and 2). RPA is a heteromeric protein containing three subunits of 70 (RPA70), 34 (RPA34) and 14 kDa. XPA interacts with the 70 kDa and the 34 kDa subunits [39]. XPA binding to RPA70 is dependent on two oligopeptide sequences conserved in exon IV from *Drosophila*, *Xenopus*, chicken, mouse and human. An ill-defined region of exon I is responsible for XPA binding to RPA34. Binding to RPA70 is essential for XPA function in complementation of XPA cells in a UV survival assay or of XP-A cell extracts in an *in vitro* DNA repair assay. In contrast, binding to RPA34 enhances, but is not essential to, complementation of XP-A cells.

XPA is critical to assembly of the NER complex (Figure 2). XPA interacts with ERCC1 which, in a complex with XPF, functions as a single-strand endonuclease. The ERCC1/XPF nuclease cleaves 5′ to a lesion contained in a denaturation bubble [31]. The affinity of XPA for damaged DNA is enhanced by the interaction with ERCC1 [29]. XPA sequences essential to the interaction of XPA and ERCC1 are the polyglutamic acid sequence (Glu_{78}–Glu_{84}) and a nearby tetrapeptide (Gly_{72}–Phe_{75}) in exon II [51]. Deletion of the Gly_{72}–Phe_{75} tetrapeptide destroys XPA function, whereas deletion of the polyglutamate severely depresses, but does not completely eliminate, XPA function. In addition, XPA with the tetrapeptide deletion acts as a dominant negative in an *in vitro* repair assay with cell-free extract from wild-type cells because the deletion does not affect DNA binding by XPA. Thus the interaction of XPA with ERCC1 is critical for recruitment of the 5′ endonuclease activity to the NER complex. An intriguing observation was made on the importance of the order of binding of RPA and ERCC1 to XPA [30]. It appears that ERCC1 can bind to a preformed XPA-RPA heterodimer, but that RPA cannot bind to a preformed complex of ERCC1 and XPA. This suggests that formation of these binding complexes involves conformational changes which affect the protein–protein recognition motifs. The 3′ nuclease, XPG [26,31], is recruited to the NER complex via its binding to the RPA–XPA complex [21]. Thus both nuclease components are dependent on XPA interactions for recruitment to the site of DNA damage, in a similar fashion to the role of the $UVRA_2$ homodimer in recruiting the UVRB and UVRC nucleases in *E. coli*. Furthermore, the form of XPA that recruits the nucleases appears to be the heterodimer with RPA.

In addition to its interactions with RPA and ERCC1, XPA also interacts with the TFIIH complex. This is a bifunctional complex containing both XPB and XPD helicases and acting in both RNA polymerase II transcription and NER. It is believed that TFIIH may open up the denaturation bubble containing the lesion recognized by XPA–RPA. The interaction of XPA with TFIIH is dependent on sequences encoded in exon VI (Figure 1) and is mediated by the action of the protein HHR23B, which is tightly bound and co-isolates with XPC. Deletion of exon VI in a cDNA expression construct greatly decreases, but does not eliminate, the ability of the cDNA to complement XP-A cells [49]. More modest decreases in the ability to complement XP-A cells is induced by mutation of either of the two cysteines (Cys_{261}, Cys_{264}) in exon VI to a serine. It is possible that a structure dependent on a disulphide bridge is involved in the interaction [49].

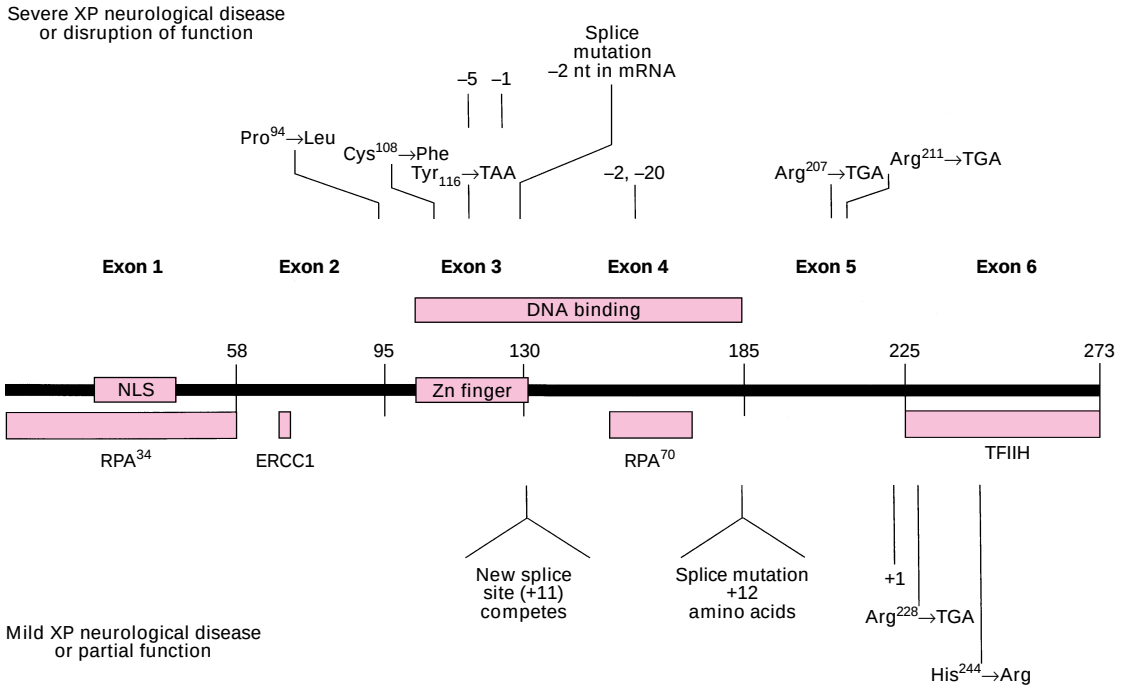


Figure 1 Map of the *XPA* gene, mutations and clinical phenotype

The linear sequence of *XPA* is represented by a heavy horizontal line (N terminal to the left, C terminal to the right). The amino acid numbers at the ends of each exon are indicated above vertical lines. Structural features are indicated by boxes. The DNA-binding region is indicated by a pink box above exons 3 and 4, although its precise 3′ terminus might extend into exon 5. The regions involved in binding other components of the NER complex (RPA, ERCC1 and TFIIH) are indicated by pink boxes below the heavy line. The types and positions of mutations associated with either severe XP disorders or complete loss of function are indicated at the top. The mutations associated with mild XP disease or retention of partial function are indicated on the bottom. Nucleotide insertions are indicated by +, deletions by —. NLS, nuclear localization signal (pink box); Zn, the Cys_4 zinc-finger motif (pink box).

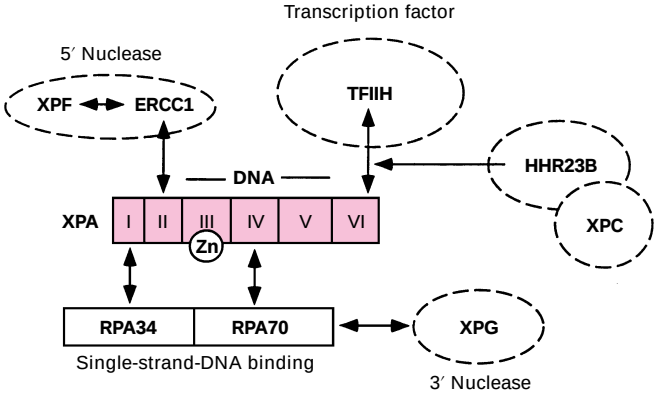


Figure 2 Representation of the protein–protein interactions within the NER complex, emphasizing the central role of the XPA–RPA DNA binding heterodimer in acting as a recruitment centre for all of the major components

Exons I–VI are shown as pink squares.

QUANTITATIVE STUDIES OF *XPA* GENE EXPRESSION

In early studies the rates of photoproduct excision clearly indicated that, under experimental conditions, the rate of repair was limited by the capacity of the repair system. Typical doses of UV radiation that correspond to about 50% lethality in a human cell population produce of the order of 2×10^5 pyrimidine dimers per nucleus. *In vivo*, however, only about 500 sites are involved in repair per nucleus at any one time, indicating that only about 2% of the damaged sites are capable of being repaired at the maximal rate *in vivo* [52]. In contrast, imaginative early studies on the effective quantity of XPA in a cell indicated that this protein was in excess [53,54]. Heterozygous XP-A cells were fused with homozygous cells to produce an increasing number of nuclei in heterokaryons which then were required to repair UV damage from the product of one *XPA* gene. Under these conditions, the product of one *XPA* gene could support repair at normal rates in multiple nuclei. These observations are consistent with others in patients and in heterozygous knockout mice, which indicate that the heterozygous state is sufficient to support completely normal levels of repair and is without significant clinical symptoms [55].

The apparent presence of excess of XPA protein in cells contrasts with the apparent limitation of the overall repair process by XPA binding to DNA damage and the extremely low transcription rate of the gene [40]. The promoter is extraordinarily weak [41]. These observations indicate the the *XPA* gene is at the extreme of low levels of transcription, but the protein must be long-lived to sustain an excess over minimal cellular requirements. The limitation on the total number of sites of repair per nucleus must therefore involve either some other rate-limiting component of the overall system, or a limit set by the need to assemble and disassemble a complex with multiple components for each site to be repaired.

XPA expression at below heterozygous levels is associated with less than normal levels of repair. Sub-heterozygote XPA expression occurs in some XP-A patients with delayed onset of neurological disease. These patients have either one or both

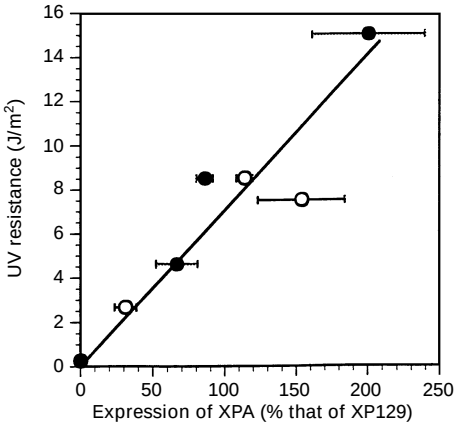


Figure 3 Relationship between expression levels of *XPA* cDNA driven by a conditional promoter and the UV resistance of cells with that level of expression, measured by the 37% surviving dose (D_{37}) in J/m^2

The cDNA cassette was inducible by isopropyl β-D-thiogalactoside using the Stratagene LacSwitch^{TM} system; the expression level was relative to a cell line XP129, which carried a single functional XPA gene. ○, Uninduced; ●, induced. Reproduced from [108] by permission of Oxford University Press.

alleles with leaky splice-site mutations [56,57]. Cells from these patients have intermediate resistance to UV which correlates directly with mild clinical symptoms of their disease. Detailed lesion-specific repair studies have not been performed with cells from these patients to determine if preferential repair of different types of lesions or in different regions of the genome occurs. Greater repair of [6-4] photoproducts and specific repair of transcriptionally active regions could be predicted, on the basis of studies done with conditional expression of XPA [46].

In experiments in which the *XPA* cDNA was expressed from a conditional promoter, the intracellular protein level could be regulated artificially and, under these conditions, *XPA* expression could be made rate-limiting [46]. The UV-sensitivity of human cells was then a linear function of the expression level of *XPA* (Figure 3). There was also a hierarchy in functions that could be brought into play with increasing *XPA* expression: at very low levels of expression, only [6-4] photoproduct excision occurred; at intermediate levels cyclobutane dimers were excised, but only from transcriptionally active genes. Cyclobutane dimers were excised from the overall genome only at the highest expression levels. The difference in [6-4] and cyclobutane pyrimidine dimer excision is consistent with the known binding constants of the XPA protein to these two substrates, but not with the affinities of the XPA–RPA heterodimer for the different lesions. Thus it is not clear that the heterodimer is the principal functional state of XPA. The assembly of the heterodimer at a lesion site may, instead, be the rate-limiting event.

The difference in cyclobutane dimer repair in active genes versus the overall genome as a function of *XPA* expression level suggests that there must be intracellular competition between XPA protein and chromatin proteins for photoproducts in unexpressed regions of the genome. XPA therefore also plays a role in setting the priorities of which sites are to be repaired in the nucleus, and contributes to the enhanced rates of repair seen in actively expressed regions of the genome [58].

Differences in the efficiency of dimer and [6-4] excision are seen in numerous experimental situations; apparently any reduction in repair due to mutations, loss of function in associated proteins or reduced expression levels are seen first in loss of cyclobutane pyrimidine dimer repair rather than [6-4] repair. This may be because the cyclobutane dimers represent such minimal distortions of the DNA that their repair is the most sensitive to reductions in the efficiency of damage recognition.

NATURALLY OCCURRING MUTATIONS

Studies of naturally occurring mutations in patients have provided a detailed understanding of the functional domains of XPA (Figure 1). Since the cloning of the *XPA* gene, many of the mutations in XP-A patients have been characterized [56,57, 59–67]. One of the common symptoms of XPA patients is a progressive neurological degeneration [68–70]. The manifestations of neurological disease are often severe, with onset in early childhood. However, in some cases the neurological disease is moderate and onset is delayed into the teens or early adulthood. The mutational data indicate that the patients with early onset of neurological disease have mutations that totally inactivate XPA. These mutations are commonly in the DNA-binding region of the protein. Patients with delayed onset of neurological disease have mutations that allow expression of partially functional XPA or are point mutations in exon VI. Leaky mutations include a point mutation that introduces a new splice acceptor site near the end of intron III that competes with the normal site and causes a frameshift in the DNA-binding region [57]. The mutations that allow partial function cause either small insertions or deletions at the end of exon IV [56,60] or loss of part or all of exon VI [66]. There are several naturally occurring mutations in or near exon VI in patients with delayed onset of neuological disease. Three of these mutations result in premature termination thus effectively deleting exon VI [64–67]. Mild XP disease also is associated with a missense mutation in exon VI (His_{244} to Arg) [60]. This histidine may be essential for proper interaction with TFIIH and its recruitment to damaged sites already bound by XPA. These observations suggest that the interaction between XPA exon VI sequences and TFIIH (Figures 1, 2 and 4) is not essential for NER because interactions between other components of the excision complex (e.g. XPG) and TFIIH can still occur. Consistent with this inference is that transfection of XPA mutant cDNAs with either a deletion of exon VI or point

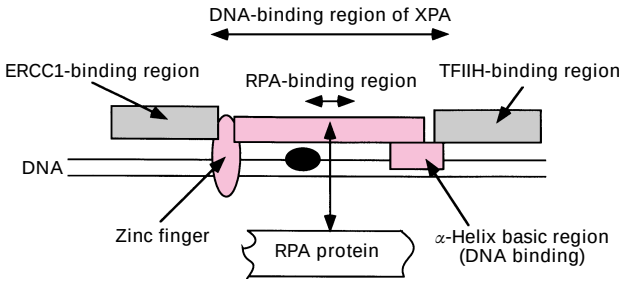


Figure 4 Schematic diagram of the XPA protein binding to a double-stranded region of DNA containing a damaged site (solid oval on the upper DNA strand)

The core DNA-binding region, exons III, IV and V, of the XPA protein (pink) is indicated in three sections: the zinc finger which binds tightly to DNA near to the damaged site; a centre region which is hypothesised to have a flexible association with DNA to accommodate different kinds of DNA damage, and the α-helix basic region which will also bind to DNA, and may in fact be internal to the helix and bind to single-strand DNA. A short region of the center of XPA also binds to the single-strand binding protein (RPA). The N-terminal end of the protein to the left (grey box) interacts with the ERCC1 component of the 5′ nuclease of the repair process. The C-terminal end of the protein to the right (grey box) interacts with the TFIIH transcription factor which unloads the XPB and XPD helicases on to the damaged region to locally unwind the DNA.

mutations in exon VI partially complements the repair defect in XP-A cells [49].

SYNTHETICALLY PRODUCED MUTATIONS

Three classes of mutations have been made synthetically in the XPA gene: specific deletions, reversions, and transgenic knockouts.

Deletion mapping was first used to identify that the exon I of XPA was dispensible for function [49], and more recently truncated genes have been used to define the DNA-binding domain of the protein in exons III–V [71]. These have been useful in isolating the minimal DNA-binding region of the protein for structural analysis by NMR and eventual solution of the crystal structure. Small deletions in exon II have also highlighted the functional interaction of XPA with ERCC1 [51]. Deletion of the Gly_{72}–Phe_{75} tetrapeptide in exon II results in XPA functioning as a dominant negative in an *in vitro* repair assay with cell-free extract from wild-type cells. The presumed mechanism is that the deletion prevents association between XPA and ERCC1, even though XPA can still bind to DNA. Recruitment of the ERCC1-XPF 5′ nuclease to the site of the lesion is therefore prevented.

Revertants of the *XPA* gene were initially identified as artefacts during early attempts to clone the *XPA* gene by transfection of XP-A cells with wild-type genomic DNA. The yield of cells with increased UV resistance was enhanced during the transfection and UV-selection process because the recipient XP-A cell lines had elevated mutation rates associated with their repair deficiency. Direct analysis of a series of these UV-resistant cells showed that they had arisen by reversion of the original mutation in the *XPA* gene, and were therefore more likely to be observed when the original XP-A cell line carried point mutations rather than deletions. Reversion of a TGA (stop) codon to GGA (glycine) or CTA (leucine) in a site that normally codes for CGA (arginine) was observed in one series of experiments [72]. The particular interest in these revertants was that the cell lines carrying these *XPA* genes had normal UV-resistance, but were unable to excise pyrimidine dimers from inactive regions of the genome; instead they had a more restricted repair capacity confined to normal excision of [6-4] photoproducts but excision of pyrimidine dimers from active genes [73,74]. The site of the reversion in these cell lines was within the DNA-binding region of the XPA protein, and it is conceivable that this amino acid change influences the photoproduct-binding capacity, a hypothesis we will return to below.

Transgenic animal models for patients with mutations in the XPA gene have been made by inserting a neomycin-resistance gene into a coding region of the mouse *XPA* gene and transfecting this mutant gene into embryonic stem cells followed by selective breeding to generate homozygote defective animals. Two strains have now been developed. In one, exons III and IV were removed and replaced with the neomycin-resistance gene [75]; in the other, a slightly different construct was made [76]. These strains both exhibited reduced NER and increased susceptibility to UV and chemical-carcinogen-induced skin carcinogenesis, but neither showed the progressive neurological degeneration which is characteristic of human patients. This is particularly striking, because both transgenic knockouts deleted the DNA-binding region of the XPA protein, which is the region in which mutations are associated with neurological symptoms in man. When these knockout animals were crossed with p53 knockout animals, the resulting double knockouts were normal and viable, but carcinogenesis from benzo[*a*]pyrene was greatly accelerated compared with either single-knockout animal [77]. This is in contrast with crosses between XPC knockout animals, which are similar to XPA knockouts in having elevated UV- and chemically induced carcinogenesis [78], and p53 knockouts. The double knockouts are embryo lethals and die of neural-tube defects [79], but since these morphological changes in the neural tube also occur in a significant proportion of p53 knockout animals, the significance of this observation in the XPC × p53 hybrid animals remains to be evaluated [80]. Clearly there are multiple unexpected interactions between the products of repair genes and of other regulatory genes which have yet to be fully understood.

These knockout animals have also been useful in beginning to understand the role of immunosuppression in skin carcinogenesis, which has been a vexed and controversial issue [81], and early experiments appear to demonstrate an important role for the effects of UV damage and repair deficiency in enhancing immunological responses [82].

SUBSTRATE RANGE OF XPA BINDING

The most important question underlying the NER process is the mechanism of damage recognition carried out by the XPA protein. This is especially difficult to understand predictively because of the wide variety of different kinds of DNA damage that can be recognized by the repair process. There may, however, be a difference between the substrate range for the overall repair process, in which a large number of different proteins interact to recognize and repair damage, and the substrate range of the XPA protein alone. The repair process excises cyclobutane pyrimidine dimers with a half-time *in vivo* of about 15 h, but the XPA protein has very weak or negligible binding to a pyrimidine dimer alone [38]. The yeast homologue, rad14, also has also been shown to have little affinity for cyclobutane pyrimidine dimers [83].

The substrate range of the repair process as a whole has been extensively investigated *in vivo* and *in vitro*, and indicates a very versatile repair capacity. The substrates range from large helix-distorting DNA adducts like AAAF adducts and [6-4] photoproducts, benzo[*a*]pyrene, cisplatin adducts, aflatoxin adducts, synthetic cholesterol pseudonucleotides, and even DNA triplexes, to small or less distorting lesions such as dimerized pyrimidines, psoralen adducts, apurinic sites, oxidative damage and G:G mismatches [84–86]. In the case of the synthetic triplex structure created by covalent cross-linking of a homologous oligonucleotide to a site in a plasmid, a functional XPA was required for the structure to be processed intracellularly to produce targeted mutations [86]. The rate of repair in a competition assay is clearly very dependent on the lesion chemistry: *N*-acetylaminofluorene adducts, for example, are repaired with about 100-fold greater efficiency than pyrimidine dimers [33,87]. Damage recognition has been suggested to depend on the degree of helix distortion and associated single-strandedness caused by the damage to DNA [33,87].

The correlation with helix destabilization is consistent with the reported preferential binding of single-stranded DNA versus double-stranded DNA by both bovine XPA in South-Western-blot analysis [47] and human XPA in competition assays in solution [38]. It is also likely that, *in vivo*, XPA does not bind DNA alone, but rather as a heterodimeric complex with RPA [21,39]. Association of XPA with RPA is likely to increase further the preferential binding of single-stranded DNA by XPA. There is, therefore, a distinct possibility that the driving mechanism of XPA binding to damaged DNA is a preference for single-stranded DNA. Thus it is unlikely that recognition would be identical with the kind of mechanism recently identified for photolyase and uracil-N-glycosylase in which the damaged base

swings out and fits into a pocket in the protein [8,9,88,89]. One observation that makes this mechanism less likely is the property of XPA revertants in which a single amino acid change (arginine to glycine in codon 207 of the XPA protein) changes the substrate specificity. This missense mutation in XPA incapacitates the cell's ability to initiate repair on psoralen–DNA but psoralen-DNA cross-links are still repaired [90]. The DNA-binding region of the protein may therefore need to be flexible to accommodate different forms of DNA damage and might undergo conformational changes when bound to DNA and to the other proteins involved in the repair process.

The Cys_2-Cys_2 zinc finger contains a 17-amino-acid loop which in many similar structures constitutes a region which binds in the major groove of DNA and makes contacts with a small core site of three to six bases and other flanking sites [91,92]. The XPA DNA-binding region has superficial similarities to the GATA-1 transcription factor, which is also a single finger structure [91], rather than TFIIA and Zif268, which are multi-finger complexes [92]. GATA-1 requires the zinc finger itself and a long C-terminal region which wraps around the DNA, conferring part of the specificity. The structure wraps around the DNA like a 'hand gripping a rope' [91]. The total site occupied by the GATA-1-binding domain is only about eight base-pairs. It is interesting at this time to speculate about the corresponding binding region of the XPA protein, in the current absence of a detailed structure directly. The C-terminal tail of the DNA-binding region of the XPA protein is considerably longer that that of GATA-1, about 90 amino acids as compared with about 34, so the 'footprint' could be larger. Since the strength of XPA binding to damaged DNA is greater for helix-distorting lesions and is augmented by RPA, it is conceivable that the C-terminal end of the XPA DNA-binding region may penetrate between the DNA strands in a locally denatured bubble in the DNA. Although the speculative model shown has put XPA binding on one strand and RPA on the other (Figure 4), this has no current experimental basis, and the precise orientations remain to be determined.

One technical puzzle is that although recombinant XPA has been available for a considerable time, and a DNA footprint of the larger damage binding protein XPE has been available for some considerable length of time [93], no corresponding footprint made by DNase I protection or other techniques has been described for XPA. This may because its binding is actually relatively weak and clear resolution may require specific cross-linking by photochemical or other means, such as used to define the binding of some transcription factors such as NFATc [94]. There may additionally be the possibility of considerable conformational mobility because of the need to accommodate a range of possible substrates. For comparison, the single-strand binding protein RPA, which plays an important part in XPA binding and can complement the the missing activity associated with loss of the XPE binding protein [23], has a footprint that can be as large as 30–90 nucleotides, corresponding to the actual excision patch size *in vivo* [95] and *in vitro* [96]. But in a crystal structure it appears to bind only to a core of eight nucleotides [97].

The relationship of XPA domains to the other components of the repair system, especially the 5′ and 3′ nucleases would predict that the zinc finger and the N-terminal domain would lie on the 5′ side of a damaged site to position ERCC1/XPF correctly. The region to the carboxy side of the zinc finger and the RPA70 domain cannot bind to the 3′ side of the lesion, but must be available for binding TFIIH. Because XPG cleaves occurs four to six nucleotides 3′ to the damaged site, this region of the damaged strand must be accessible and not covered by XPA or RPA. XPE masks this site, indicating that it is unlikely to act as a damage-recognition protein, but instead plays a lesser role than RPA [23,93].

CONCEPTS IN DAMAGE RECOGNITION

In consequence, our current knowledge of the XPA DNA-binding reaction described above leads to the following concepts.

(1) The binding reaction must be very flexible to accommodate a series of structurally diverse chemical modifications to DNA. Although the binding strength may differ according to the lesion, most significant digressions from normal DNA structure can be recognized and bound to initiate repair. The preference for single-stranded DNA may drive the binding to damaged DNA, and the ease with which XPA can insert itself into the helix may determine the apparent affinity differences.

(2) XPA contains several discrete functional and structural units that can be recognized by chymotrypsin digestion and from the distribution of mutations in severely and mildly affected XPA patients [62,64,71]. These units can therefore be analysed individually to build up a picture of the overall process of damage recognition.

(3) There are two classes of mild XP-A cases: splice-site mutations that are leaky and allow a small amount of normal translation to occur, and mutations deleting or altering exon VI sequences. The intermediate repair levels in cells expressing leaky mutant alleles indicate that even very small amounts of normal XPA protein can have a disproportionate effect on UV-sensitivity and repair. The mutations that result in small insertions or deletions in the DNA-binding region provide evidence of the flexibility of this region. The other class of mildly affected XP-A patients have mutations in the TFIIH-binding region of exon VI. The intermediate repair levels in cells from these patients indicate that the XPA–TFIIH interaction is not as essential for repair as is DNA damage binding by XPA.

(4) There is a striking absence of point mutations which would result in amino acid changes within the DNA-binding region among naturally occurring mutations in XPA patients. This is conspicuous, despite the prevalance of point mutations within the splice sites and insertions and deletions that cause frameshifts. Mis-sense mutations in this region can, however, be generated in cell culture that alter amino acids in the DNA-binding region, but yet express an XPA protein which retains high activity [72],

These considerations lead us to a key concept which we wish to advocate to explain the versatility of XPA binding in the NER process, which can be called the mechanism of 'flexible fit'.

THE FLEXIBLE-FIT HYPOTHESIS: A SUGGESTED MECHANISM OF DAMAGE RECOGNITION

The current information which has identified several functional regions of the XPA protein, and general properties of XPA in repair of different kinds of damage, can now be used to predict a mechanistic and structural model for the binding reaction of XPA to a damaged site (Figure 4).

The binding region of XPA, which consists of the 122-amino-acid core region of exons III, IV and V [71], contains the zinc finger (exon III) and a large region with high helical density that includes the RPA interaction domain. The model we propose involves the concept that the DNA-binding domain of XPA has two components: the zinc finger for generalized binding to DNA and a flexible region that contains the a single-strand-binding region that works in concert with RPA. We hypothesize that XPA may be anchored to the DNA by its zinc finger, but that this finger is not involved in direct binding to the damaged bases. Rather, lesion binding is more likely mediated by the region encoded by exons IV and V. We suggest that this region is

flexible around the DNA lesion and contains a single-strand-DNA-binding domain. Thus this region may actually insert itself into the DNA helix in quest of single-stranded DNA with which to bind (Scheme 1 and Figure 4), thus allowing it to recognize and to bind to a large variety of damaged sites. We envisage that other components (RPA, TFIIH, XPG, ERCC1-XPF) are recruited in succession subsequent to the initiation of XPA insertion. This model predicts that lesion recognition and insertion into the helix by XPA would be facilitated both by helix-destabilizing lesions and by the presence of RPA. Complete opening of the region containing the lesion would be facilitated by the binding of TFIIH and the subsequent action of the XPB and XPD helicases. This model also accommodates the positioning of the ERCC1 interaction region, which is located on the amino side of the zinc finger, 5′ to the lesion such that ERCC1-XPF will be in the vicinity of the 5′ cleavage site. Recruitment of XPG to the 3′ cleavage site by binding TFIIH and RPA is also accommodated.

XPA and RPA may therefore both be quite plastic and adopt varied conformations on damaged DNA to ensure secure recognition and binding. One possibility would be that XPA and RPA initially bind short regions around the damaged site within the region of hydrogen-bond disruption caused by the damage itself. Hence there will be a hierarchy of binding efficiency according to the helix-distorting capacity of the damage. These proteins then act as matchmakers [98] to guide the XPB and XPD helicases into the region to create a larger single-stranded DNA 'bubble'. This may then result in RPA changing conformation to coat the 30-nucleotide single strand complementary to the damaged strand, which then defines the excision sites. Evidence that RPA does change conformation is found in alterations in the proteolytic-digestion patterns of the 32 and 70 kDa components upon binding to single-stranded DNA [99]. Interestingly, no similar changes in proteolytic digestion of XPA upon binding to DNA was detected [71]. However, structural changes in XPA when binding damaged DNA have not been investigated. XPA may therefore adopt several structural conformations depending on its binding partners. Consequently, solution of a single structure of XPA may only be the beginning of a long series of investigations.

ACTION OF REPAIR SYSTEMS ON UNUSUAL DNA STRUCTURES

The extensive studies of DNA repair systems over the past 30 years has resulted in formal descriptions of discrete systems such as base excision–repair, NER, recombination repair etc, mostly based on classification of the nature of the DNA damage. For example, base excision–repair was defined from its action on single base lesions and NER from its action on extended large DNA adducts. While there was recognition that there was some overlap in substrate specificity, these definitions could still be used profitably to make progress in understanding the repair systems individually. Recently there has been the emergence of molecular studies of more complex and unusual DNA structures, with resulting interest in how different repair systems may be recruited to act on these.

Some complex structures occur naturally as a result of the metabolic activities of DNA, such as the recombination intermediates represented by three- and four-armed Holliday junctions, replication intermediates, immunoglobulin rearrangements and chromosomal exchanges. Others may arise from DNA damage or replication errors such as loops around DNA damaged sites, blocked replication forks, microsatellite instability and chromosome breakages. A third set, potentially of great practical importance, are artificially synthesized molecules such as the targeting vectors used for embryonic cell targeting in transgenic animal development, triple-stranded structures with reactive termini ('warheads') that have been used for gene targeting [86,100,101], and recA-coated DNA structures used for site-directed homologous recombination [102].

These varied structures are likely to recruit individual components of repair and recombinational systems and use them in novel ways depending on the particular structure. Loops containing damage, or undamaged loops, seem to employ a combination of mismatch and excision-repair [5,103]. Holliday junctions and recombination intermediates may employ the nucleases of NER, which cleave at junctions between single- and double-stranded DNA, together with being strong binding sites for the p53 protein [104]. Triple-stranded DNA structures appear to require a functional XPA protein and active transcription coupled repair, but a defective XP variant protein, for the structures to be processed into site-directed mutations at high frequency [86,100]. Repair of DNA–DNA cross-links seem particularly complex, requiring a functional XPA and NER system and some combination of recombinational and double-strand-break repair with the possibility of both accurate repair involving homologous recombination and less accurate non-homologous mechanisms [105,107].

With the increasing need to establish functions for the many genes that are emerging from large-scale cloning and sequencing studies, improved methods to manipulate and direct mutagenesis to genes, on the basis of their sequences, will become valuable technical approaches. Knowledge of how cells see and process DNA structures involved in these approaches will surely be of value.

CONCLUDING REMARKS

With an understanding of the structure and mechanism of XPA binding to DNA, there are numerous practical applications that can be envisaged. One possibility would be to engineer the protein into a direct DNA-cleavage enzyme, by the addition of a peptide or a photoactivatable group which can cut the DNA at sites of damage. Alternatively the DNA-binding region of the protein could be covalently linked to a biotin moiety or to a fluorescent protein, to further quantify sites and levels of DNA damage *in vivo* or in isolated DNA. This would be useful in enabling quantification to be made of a large variety of DNA lesions on DNA from experimental or field sources. Specific amino acid changes could also be engineered into the XPA protein to change its specificity, and it can be speculated whether the protein as currently known is at its optimum, or could actually be improved. The protein could also be encapsulated into a therapeutic agent for delivery to patients for protection and therapy after sun exposure, or exposure to other carcinogens.

Our understanding of the complex eukaryotic NER system has increased enormously in recent years with the cloning of the majority of the genes involved, their expression as recombinant proteins, the development of *in vitro* cell-free systems to dissect the enzyme activities, and the establishment of the minimal set of components. It would be premature, however, to believe that we are approaching the 'end of history' in DNA repair. The next steps will, at one level, be to develop a three-dimensional picture of the proteins in their functional orientations in the damage-recognition and excision complex. On other levels, the manner in which NER fits into the networks of cellular signalling by which cells orchestrate their overall responses to DNA damage and convert damage into mutations, genomic instability and cancer is yet to be understood fully. A full understanding of the XPA

protein is likely to emerge in the next few years, which we hope will make a major contribution to these future developments.

J.E.C. was supported by the Cohn Fund of the Academic Senate of the University of California, San Francisco, the March of Dimes Birth Defects Foundation (grant FY96-1032) and the American Cancer Society (grant CN-156), and additional support, to J.C.S., was from the National Institutes of Health (grant ES06460).

REFERENCES

1 Sancar, A. and Sancar, G. B. (1988) Annu. Rev. Biochem. **57**, 29–67
2 Wood, R. D. (1996) Annu. Rev. Biochem. **65**, 135–167
3 Jiricny, J. (1996) Cancer Surv. **28**, 47–68
4 Kolodner, R. (1996) Genes Dev. **10**, 1433–1442
5 Mu, D., Tursun, M., Duckett, D. R., Drummond, J. T., Modrich, P. and Sancar, A. (1997) Mol. Cell. Biol. **17**, 760–769
6 Mello, J. A., Acharya, S., Fishel, R. and Essigman, J. M. (1996) Chem. Biol. **3**, 579–589
7 Mellon, I., Spivak, G. and Hanawalt, P. C. (1987) Cell **51**, 241–249
8 Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. and Tainer, J. A. (1995) Cell **82**, 701–708
9 Savva, R., MacAuley-Hecht, K., Brown, T. and Pearl, L. (1995) Nature (London) **373**, 487–493
10 Singer, B. and Hang, B. (1997) Chem. Res. Toxicol. **10**, 713–732
11 Sancar, A. (1996) Annu. Rev. Biochem. **65**, 43–81
12 Tanaka, K., Satokata, I., Ogita, Z., Uchida, T. and Okada, Y. (1989) Proc. Natl. Acad. Sci. U.S.A **86**, 5512–5516
13 Hanawalt, P. C. (1994) Science **266**, 1957–1958
14 Thomm, M. (1996) FEMS Microbiol. Rev. **18**, 159–171
15 McCready, S. (1996) Mutat. Res. **364**, 25–32
16 Yajima, H., Takao, M., Yasuhira, S., Zhao, J. H., Ishii, C., Inoue, H. and Yasui, A. (1995) EMBO J. **14**, 2393–2399
17 Takao, M., Yonemasu, R., Yamamoto, K. and Yasui, A. (1996) Nucleic Acids Res. **24**, 1267–1271
18 Clayton, D. A., Doda, J. N. and Friedberg, E. C. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 2777–2781
19 Driggers, W. J., Grishko, V. I., LeDoux, S. P. and Wilson, G. L. (1996) Cancer Res. **56**, 1262–1266
20 Aboussekhra, A., Biggerstaff, M., Shivji, M. K. K., Vilpo, J. A., Moncollin, V., Podust, V. N., Protic, M., Hubscher, U., Egly, J. M. and Wood, R. D. (1995) Cell **80**, 859–868
21 He, Z., Henrickson, L. A., Wold, M. S. and Ingles, C. J. (1995) Nature (London) **374**, 566–569
22 Reardon, J. T., Mu, D. and Sancar, A. (1996) J. Biol. Chem. **271**, 19451–19456
23 Kazantsev, A., Mu, D., Nichols, A. F., Zhao, X., Linn, S. and Sancar, A. (1997) Proc. Natl. Acad. Sci. U.S.A. **93**, 5014–5018
24 Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J., Bootsma, D. J. et al. (1994) EMBO J. **13**, 1831–1843
25 Guzder, S. N., Bailler, V., Sung, P., Prakash, L. and Prakash, S. (1995) J. Biol. Chem. **270**, 8385–8388
26 O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. and Wood, R. D. (1994) Nature (London) **371**, 432–435
27 Biggerstaff, M., Szymkowski, D. E. and Wood, R. D. (1993) EMBO J. **12**, 3685–3692
28 van Vuuren, A. J. V., Appeldoorn, E., Odijk, H., Yasui, A., Jaspers, N. G. J., Bootsma, D. and Hoeijmakers, J. H. J. (1993) EMBO J. **12**, 3693–3701
29 Nagai, A., Saijo, M., Kuraoka, I., Matsuda, T., Kodo, N., Nakatsu, Y., Mimaki, T., Mino, M., Biggerstaff, M. and Wood, R. D. (1995) Biochem. Biophys. Res. Commun. **211**, 960–966
30 Saijo, M., Kuraoka, I., Masutani, C., Hanoaka, F. and Tanaka, K. (1996) Nucleic Acids Res. **24**, 4719–4724
31 Matsunaga, T., Park, C. H., Bessho, T., Mu, D. and Sancar, A. (1996) J. Biol. Chem. **271**, 11047–11050
32 Reardon, J. T., Thompson, L. H. and Sancar, A. (1993) Cold Spring Harbor Symp. Quant. Biol. **58**, 605–617
33 Gunz, D., Hess, M. T. and Naegeli, H. (1996) J. Biol. Chem. **271**, 25089–25098
34 Cleaver, J. E. and Kraemer, K. H. (1995) in The Metabolic and Molecular Bases of Inherited Disease, vol. 3, (Scriver, C. R., Beaudet, A. L., Sly, W. S. and Valle, D., eds.), pp. 4393–4419, McGraw-Hill, New York
35 Kraemer, K. H., Lee, M. M. and Scotto, J. (1987) Arch. Dermatol. **123**, 241–250
36 Maher, V. M., Dorney, D. J., Mendrake, A. L., Konze-Thomas, B. and McCormick, J. J. (1979) Mutat. Res. **62**, 311–323
37 Asahara, H., Wistort, P. M., Bank, J. F., Bakerian, R. H. and Cunningham, R. P. (1989) Biochemistry **28**, 4444–4449
38 Jones, C. J. and Wood, R. D. (1993) Biochemistry **32**, 12096–12116
39 Li, L., Lu, X., Peterson, C. A. and Legerski, R. J. (1995) Mol. Cell. Biol. **15**, 5396–5402
40 Layher, S. K. and Cleaver, J. E. (1997) Mutat. Res. **383**, 9–19
41 Topping, R. S., Myrand, S. P., Williams, B. L., Albert, J. C. and States, J. C. (1995) Gene **166**, 341–342
42 Weeda, G., Ma, L., van Ham, R. C. A., van der Eb, A. J. and Hoeijmakers, J. H. J. (1991) Carcinogenesis **12**, 2361–2368
43 van Duin, M., van den Tol, J., Warmerdam, P., Odijk, H., Meijer, D., Westeveld, A., Bootsma, D. and Hoeijmakers, J. H. J. (1988) Nucleic Acids Res. **16**, 5305–5322
44 Satokata, I., Iwa, K., Matsuda, T., Okada, Y. and Tanaka, K. (1993) Gene **136**, 345–348
45 van Oostrom, C. T., de Vries, A., Verbeek, S. J., van Kreijl, C. F. and van Steeg, H. (1994) Nucleic Acids Res. **22**, 11–14
46 Cleaver, J. E., Charles, W. C., McDowell, M. L., Sadinski, W. J. and Mitchell, D. L. (1995) Cancer Res. **55**, 6152–6160
47 Eker, A. P. M., Vermeulen, W., Miura, N., Tanaka, K., Jaspers, N. G. J., Hoeijmakers, J. H. J. and Bootsma, D. (1992) Mutat. Res. **274**, 211–224
48 Shimamoto, T., Kohno, K., Tanaka, K. and Okada, Y. (1991) Biochem. Biophys. Res. Commun. **181**, 1231–1237
49 Miyamoto, I., Miura, N., Niwa, H., Miyazaki, J. and Tanaka, K. (1992) J. Biol. Chem. **267**, 19786–19789
50 Asahina, H., Kuraoka, I., Shirakawa, M., Morita, E. H., Miura, N., Miyamoto, I., Ohtsuka, E., Okada, Y. and Tanaka, K. (1994) Mutat. Res. **315**, 229–237
51 Li, L., Peterson, C. A., Lu, X. and Legerski, R. J. (1995) Mol. Cell. Biol. **15**, 1993–1998
52 Cleaver, J. E. (1975) in Methods in Cancer Research, vol. 11 (Busch, H., ed.), pp. 123–165, Academic Press, New York
53 Gianelli, F., Croll, P. M. and Lewin, S. A. (1973) Exp. Cell Res. **78**, 175–185
54 Gianelli, F. and Pawsey, S. A. (1974) J. Cell Sci. **15**, 163–176
55 Berg, R. J. W., de Vries, A., van Steeg, H. and de Gruijl, F. R. (1997) Cancer Res. **57**, 581–584
56 States, J. C. and Myrand, S. P. (1996) Mutat. Res. **363**, 171–177
57 States, J. C., McDuffie, E. R., Myrand, S. P. and Cleaver, J. E. (1997) Hum. Mutat. in the press
58 Bohr, V. (1995) Carcinogenesis **16**, 2885–2892
59 Satokata, I., Tanaka, K., Miura, N., Miyamoto, I., Satoh, Y., Kondo, S. and Okada, Y. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 9908–9912
60 Satokata, I., Tanaka, K., Yuba, S. and Okada, Y. (1992) Mutat. Res. **273**, 203–212
61 Satokata, I., Tanaka, K. and Okada, Y. (1992) Hum. Genet. **88**, 603–607
62 Satokata, I., Tanaka, K., Miura, N., Narita, M., Mimaki, T., Satoh, Y., Kondo, S. and Okada, Y. (1992) Mutat. Res. **273**, 193–202
63 Satokata, I., Uchiyama, M. and Tanaka, K. (1995) Hum. Mol. Genet. **4**, 1993–1994
64 Nishigori, C., Zghal, M., Yagi, T., Imamura, S., Komoun, M. R. and Takebe, H. (1993) Am. J. Hum. Genet. **53**, 1001–1006
65 Cleaver, J. E., Charles, W. C., Thomas, G. H. and McDowell, M. L. (1995) Hum. Mol. Genet. **4**, 1685–1687
66 Kondoh, M., Ueda, M. and Ichihashi, M. (1995) Br. J. Dermatol. **133**, 579–585
67 Sato, M., Nishigori, C., Yagi, T. and Takebe, H. (1996) Mutat. Res. **362**, 199–208
68 Kenyon, G. S., Booth, J. B., Prasher, D. K. and Rudge, P. (1985) Brain **108**, 771–784
69 Robbins, J. H., Kraemer, K. H., Lutzner, M. A., Festoff, B. W. and Coon, H. G. (1974) Ann. Intern. Med. **80**, 221–248
70 Robbins, J. H., Brumback, R. A., Mendiones, M., Rarrett, S. F., Carl, J. R., Cho, S., Dencla, M. B., Ganges, M. B., Gerber, L. H., Guthrie, R. A. et al. (1991) Brain **114**, 1335–1361
71 Kuraoka, I., Monta, E. H., Saijo, M., Matsuda, T., Monkawa, K., Shirakawa, M. and Tanaka, K. (1996) Mutat. Res. **362**, 87–95
72 McDowell, M., Nguyen, T. and Cleaver, J. E. (1993) Mutagenesis **8**, 155–161
73 Cleaver, J. E., Cortes, F., Lutze, L. H., Morgan, W. F., Player, A. N. and Mitchell, D. L. (1987) Mol. Cell. Biol. **7**, 3353–3357
74 Lommel, L. and Hanawalt, P. C. (1993) Mol. Cell. Biol. **13**, 970–976
75 De Vries, A., van Ostrom, C. T. M., Hofhuis, F. M. A., Dortant, P. M., Berg, R. J. W., de Gruijl, F. R. P., Wester, W., van Kreijl, C. F., Capel, P. J. A., van Steeg, H. and Verbeek, S. J. (1995) Nature (London) **377**, 169–173
76 Nakane, H., Takeuchi, S., Yuba, S., Saijo, M., Nakatsu, Y., Murai, H., Nakatsura, Y., Ishikawa, T., Hirota, S., Kitamura, Y. et al. (1995) Nature (London) **377**, 165–168
77 De Vries, A. (1997) Ph.D. Dissertation, University of Utrecht
78 Sands, A. T., Abuin, A., Sanchez, A., Conti, C. J. and Bradley, A. (1995) Nature (London) **377**, 162–165
79 Cheo, D. L., Meira, L. B., Hammer, R. E., Burns, D. K., Doughty, A. T. and Frieberg, E. C. (1996) Curr. Biol. **6**, 1691–1694
80 Tebbs, R. S., Bishop, J. and Pedersen, R. A. (1997) Transgenic and Knockout Mice: Genetic Models for Environmental Stress, Marcel Dekker, New York

81 Kripke, M. L. (1991) J. Dermatol. **18**, 429–433
82 Miyauchi-Hashimoto, H., Tanaka, K. and Horio, T. (1996) J. Invest. Dermatol. **107**, 343–348
83 Guzder, S. N., Sung, P., Prakash, L. and Prakash, S. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 5433–5437
84 Huang, J. C., Hsu, D. S., Kazantsev, A. and Sancar, A. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 12213–12217
85 Cleaver, J. E. (1973) Cancer Res. **33**, 362–369
86 Wang, G., Seidman, M. M. and Glazer, P. M. (1996) Science **271**, 802–805
87 Hess, M. T., Gunz, D. and Naegeli, H. (1996) Nucleic Acids Res. **24**, 824–828
88 Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. and Tainer, J. A. (1995) Cell **80**, 869–878
89 Park, H. W., Kim, S. T., Sancar, A. and Deisenhofer, J. (1995) Science **268**, 1866–1872
90 Vuksanovic, L. and Cleaver, J. E. (1987) Mutat. Res. **184**, 255–263
91 Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J. and Gronenborn, A. M. (1993) Science **261**, 438–446
92 Greisman, H. A. and Pabo, C. O. (1997) Science **275**, 657–661
93 Reardon, J. T., Nichols, A. F., Keeney, S., Smith, C. A., Taylor, J. S., Linn, S. and Sancar, A. (1993) J. Biol. Chem. **268**, 21301–21308
94 Wolfe, S. A., Zhou, P., Dotsch, V., Chen, L., You, A., Ho, S. N., Crabtree, G. R., Wagner, G. and Verdine, G. L. (1997) Nature (London) **285**, 172–176
95 Cleaver, J. E., Jen, J., Charles, W. C. and Mitchell, D. L. (1991) Photochem. Photobiol. **54**, 393–402
96 Huang, J. C., Svoboda, D. L., Reardon, J. T. and Sancar, A. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 3664–3668
97 Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. and Frappier, L. (1997) Nature (London) **385**, 176–181
98 Sancar, A. and Hearst, J. E. (1993) Science **259**, 1415–1420
99 Gomes, X. V., Henricksen, L. A. and Wold, M. S. (1996) Biochemistry **35**, 5586–5595
100 Raha, M., Wang, G., Seidman, M. M. and Glazer, P. M. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 2941–2946
101 Kutyavin, I. V., Gamper, H., Gall, A. A. and Meyer, R. B. (1993) J. Am. Chem. Soc. **115**, 9303–9304
102 Revet, B. M. J., Sena, E. P. and Zarling, D. A. (1993) J. Mol. Biol. **232**, 779–791
103 Kirkpatrick, D. T. and Petes, T. D. (1997) Nature (London) **387**, 929–931
104 Lee, S., Cavallo, L. and Griffith, J. (1997) J. Biol. Chem. **272**, 7532–7539
105 Hoy, C. A., Thompson, L. H., Mooney, C. L. and Salazar, E. P. (1985) Cancer Res. **45**, 1737–1743
106 Thompson, L. H. (1996) Mutat. Res. **363**, 77–88
107 Jeggo, P. A., Taccioli, G. E. and Jackson, S. P. (1995) BioEssays **17**, 949–957
108 Afzal, A., Feeney, L., Thomas, G. H., Volpe, J. P. G. and Cleaver, J. E. (1995) Mutagenesis **10**, 457–452
109 Nichols, A. F., Ong, P. and Linn, S. (1996) J. Biol. Chem. **271**, 24317–24320