Structural and functional studies of the human ERCC1/XPF DNA repair complex

Konstantinos Tripsianes

cover picture by Alexis Vassilikos

ISBN: 978-90-393-4721-8

Het werk beschreven in dit proefschrift werd gefinancierd door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek, gebied Chemische Wetenschappen (NWO-CW)

Structural and functional studies of the human ERCC1/XPF DNA repair complex

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 11 februari 2008 des ochtends te 10:30 uur

door

Konstantinos Tripsianes

geboren op 5 februari 1980, te Edessa, Griekenland

Promotoren: Prof. Dr. R. Boelens
Prof. Dr. R. Kaptein

Co-promotor: Dr. G.E. Folkers

verbonden aan het Bijvoet Centrum voor Biomoleculair Onderzoek,
Faculteit Scheikunde van de Unversiteit Utrecht

Paranimfen: Mark Daniels
Efi Petritsi

Vormgeving: AV-dienst Faculteit Scheikunde, Universiteit Utrecht

Contents

Chapter 1 Introduction to Nucleotide Excision Repair 7
Evolution of XPF function through formation of dimers

Chapter 2 The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair 31

Chapter 3 Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization 47

Chapter 4 Structural insight to the first ERCC1 deficiency 63

References 75

Summary 87

Samenvatting 89

Acknowledgements 93

Curriculum Vitae 95

List of publications 97

Full color figures 99

Chapter one

Introduction to nucleotide excision repair

Evolution of XPF function through formation of dimers

The maintenance of DNA integrity is vital for the viability of cells and the health of organisms. The genome is under constant attack from endogenous metabolic byproducts and environmental factors that can alter its chemical structure and corrupt its encoded message. DNA lesions are of many types including single- and double-strand breaks, inter- and intra-strand crosslinks and different kinds of base modifications. These assaults are potentially mutagenic and a clear example in man is the strong correlation between UV (sunlight) exposure or smoking cigarettes and the development of skin and lung cancer, respectively (Freedberg et al., 2003). For tolerating and repairing various types of damage that seriously impede replication and transcription at the cellular level, and are implicated in cancer and ageing at the organismal level, cells have an arsenal of ways of responding to such injury (Bootsma et al., 2002).

Multiple genome surveillance mechanisms, which are in part redundant, collectively counteract the detrimental consequences of genomic insults (Lindahl and Wood, 1999). These embrace the restoration of both single- and double-strand breaks in the genome (SSB and DSB repair) (Lisby et al., 2003), as well as the excision of base damage, base modification or incorrectly incorporated nucleotides during replication carried out by the excision repair pathways: base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR) (Figure 1) (Hoeijmakers, 2001). The excision pathways all involve recognition of the DNA damage, removal of a single-stranded section of DNA that contains the lesion, repair replication across the gap and restoration of the DNA

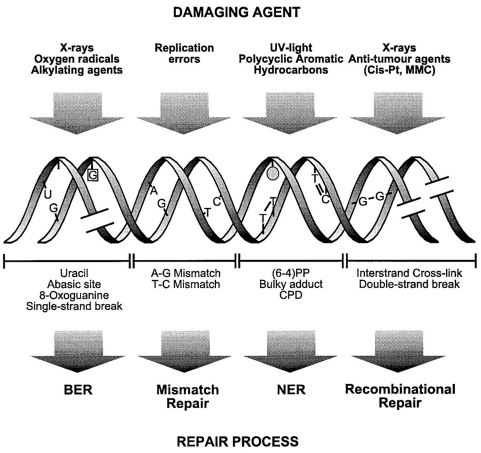


Figure 1. DNA lesions and repair mechanisms. Common DNA damaging agents (top); examples of lesions that can be introduced by these agents into the DNA double helix (middle); and most frequently used repair mechanisms for such lesions (bottom). Reprinted with permission from de Boer and Hoeijmakers. Copyright 2000 Carcinogenesis.

continuity by ligation of the repaired strand. Among them, NER is the most versatile pathway (Balajee and Bohr, 2000; Wood, 1997) because of its ability to recognise and repair a bewildering array of physically dissimilar lesions, such as DNA damage formed upon exposure to the UV radiation from sunlight and numerous bulky DNA adducts induced by mutagenic chemicals. NER comprises two subpathways: global genome NER (GG-NER), which eliminates DNA damage throughout the chromatin, and transcription coupled NER (TC-NER) which removes DNA damage in actively transcribed regions of the genome and in general is the faster of the two.

Syndromes associated with NER deficiency

NER appears to be a unique caretaker pathway because defects lead to two distinct phenotypes (de Boer and Hoeijmakers, 2000): sun-induced skin

cancer risk xeroderma pigmentosum (XP) or to segmental progerias without an increase in cancer incidence Cockayne syndrome (CS) and trichothiodystrophy (TTD). Parchment skin (xeroderma) and freckles (pigmentosum) are the prominent cutaneous hallmarks of XP patients, which are strikingly limited to sun-exposed areas of the skin (Kraemer, 1997). Furthermore, XP is associated with a >1000-fold elevated risk to develop skin cancer, especially in regions exposed to sunlight. The symptoms of XP are readily explained by a defect in the repair of DNA damage, particularly the UV lesions. Cockayne syndrome is a very pleiotropic disorder with physical and mental retardation (Nance and Berry, 1992). CS patients exhibit progeroid features including cachectic dwarfism, retinopathy, microcephaly, ganglial calcification, deafness, neural defects, and retardation of growth and development after birth. However, CS is not characterised by cancer (Nance and Berry, 1992). CS cells are deficient in TC-NER but proficient in GG-NER (Venema et al., 1990). In fact, the two subpathways share the same set of core factors but differ in the way the DNA damage is initially detected. The second segmental progeria, TTD, is characterized by sulpur-deficient brittle hair and nails, growth retardation and neurogolical abnormalities, but not pigmentation or cancer (Itin et al., 2001). TTD is now known to correlate with mutations in genes involved in NER (XPB, XPD and TTDA; all belong to the transcription/repair factor TFIIH) (Dubaele et al., 2003; Giglia-Mari et al., 2004). In retrospect, the spectrum of clinical symptoms differs considerably between the three syndromes and many of the peculiar abnormalities of CS and TTD are difficult to comprehend as a consequence of defective NER only (Bootsma et al., 2002).

Determination of the different XP complementation groups and identification of the corresponding genes

The first link between XP clinical picture and DNA repair deficiency was demonstrated by Cleaver (Cleaver, 1968) who observed impaired unscheduled DNA synthesis (UDS) in XP cultured fibroblast cells after UV irradiation. Moreover, it became evident that fibroblasts from different XP patients exhibit different survival capabilities upon UV stress. These data implied a genetic heterogeneity for this syndrome and the probable involvement of several genes in the repair process (De Weerd-Kastelein et al., 1972). Systematic cell fusion analysis confirmed this scenario and led to identification of seven complementation groups, ranging from XP-A to XP-G. A variant form was classified as XP-V and recently has been shown to be deficient in a "sloppy" DNA polymerase (Pol η) (Masutani et al., 1999), that is capable of overcoming replication blocks at DNA lesions.

The identification of the XP genes was facilitated by the use of chinese hamster ovary (CHO) cells that display UV sensitivity analogous to that of human XP-deficient cells (Bootsma, 2001). Human genomic DNA was transfered into rodent cell lines, which were subsequently screened for the recovery of UV survival. This strategy successfully resulted in the cloning of the first excision repair cross-complementing (ERCC) human genes: ERCC1 (Westerveld et al., 1984), ERCC2 (Weber et al., 1988), ERCC3 (Weeda et al., 1990), ERCC4 (Thompson et al., 1994), and

ERCC6 (Troelstra et al., 1990). The genes ERCC2, ERCC3, ERCC4 and ERCC6 were found to complement the human cells XP-D, XP-B, XP-F, and CS-B respectively, and their names were reassigned. The first cloned human DNA repair gene ERCC1 failed to complement any known XP complementation group and thus retained its rodent name. XPA gene was cloned by direct complementation of human cells transfected with mouse genomic DNA (Tanaka et al., 1990; Tanaka et al., 1989), while in the cases of the XPC and CSA genes, the clones were obtained directly from human cells after transfection of cDNA libraries (Henning et al., 1995; Legerski and Peterson, 1992). Other genes, such as hHR23B (heterodimeric partner of XPC) (Masutani et al., 1994), TTDA (the 10th subunit of TFIIH) (Giglia-Mari et al., 2004) and DDB heterodimer (missing in XP-E cell lines) (Chu and Chang, 1988; Dualan et al., 1995; Takao et al., 1993) were identified by chance or by sequence homology with repair genes discovered in other organisms. Finally XPG gene (MacInnes et al., 1993; Scherly et al., 1993) was found to be equivalent to the ERCC5 rodent line (Table 1) (O'Donovan and Wood, 1993). CSA and CSB are specific TC-NER factors and their function will not be discussed further (Laine and Egly, 2006).

From genes to NER proteins

The comprehension of the NER reaction and the roles of the individual proteins participating in this, was enhanced by the development of two in vitro assays which monitored the repair of DNA damage in whole cell extracts. The first assay, developed by Wood and co-workers, was designed to follow the specific incorporation of radioactive dNTPs into damaged plasmids by the polymerase carrying out the repair synthesis step (Wood et al., 1988). The second one, developed by Sancar and co-workers, monitored the excision step of the NER by analyzing the release of DNA fragments containing a radiolabeled phosphate close to the lesion (Huang and Sancar, 1994; Huang et al., 1992). Both assays led to the identification of additional factors required for the NER reaction such as RPA for the incision step (Coverley et al., 1991) and PCNA for the repair synthesis (Shivji et al., 1992). RPA is a core NER

Table 1. Xeroderma pigmentosum genes

gene	rodent	protein	pathway	disorder	function
XPC		XPC-HR23B	GG-NER	XP	damage DNA binding
XPE(DDB2)		DDB2-DDB1	GG-NER	XP	facilitates XPC binding
XPB	ERCC3	TFIIH	GG-NER/TC-NER	TTD, XP/CS	helicase 3′-5′
XPD	ERCC2	TFIIH	GG-NER/TC-NER	XP, TTD, XP/CS	helicase 5′-3′
XPA		XPA	GG-NER/TC-NER	XP	damage verification
XPG	ERCC5	XPG	GG-NER/TC-NER	XP, XP/CS	3′ endonuclease
XPF	ERCC4	ERCC1-XPF	GG-NER/TC-NER	XP	5′ endonuclease

Abbreviations: GG-NER, global genome nucleotide excision repair; TC-NER, transcription-coupled nucleotide excision repair; XP, xeroderma pigmentosum; TTD, trichothiodystrophy; CS, Cockayne syndrome.

component, but with additional functions in DNA replication and recombination (Wold, 1997). Deficiency of this protein has not been found because it is essential for other vital aspects of DNA metabolism.

Subsequent characterization of each protein led to preliminary assumption of their roles in NER. XPA (He et al., 1995; Li et al., 1995), RPA (Burns et al., 1996; He et al., 1995), XPC-HR23B (Masutani et al., 1994; Shivji et al., 1994), and DDB (Keeney et al., 1993; Reardon et al., 1993) were found to exhibit affinity for damaged DNA and were thus proposed to be involved in the damage recognition step of NER. XPB (Schaeffer et al., 1993) and XPD (Schaeffer et al., 1994) were identified as two ATP-dependent helicases, subunits of TFIIH complex involved in basal transcription. The fact that NER in eukaryotes requires elements of the basic transcription apparatus has yielded insight into the complex relationships between deficient DNA repair, defective transcription and hereditary human diseases (Friedberg et al., 1995). ERCC1-XPF (Bardwell et al., 1994; Sijbers et al., 1996) and XPG (O'Donovan et al., 1994) were found to be structure-specific endonucleases and exhibited preference for single-stranded/double-stranded junctions, which were able to incise with specific, and opposite, polarities. As such, ERCC1-XPF and XPG were proposed to perform the dual incision step by cleaving the damaged DNA strand 5′ and 3′ of the lesion, respectively. Eventually, the in vitro reconstitution of the NER reaction became possible by using all purified proteins in recombinant form (Aboussekhra et al., 1995; Araujo et al., 2000; Mu et al., 1995). It is now known that no less than 20 polypeptides accomplish the excision step and another 13 the repair synthesis step, indicating the complexity of the mechanism (Figure 2).

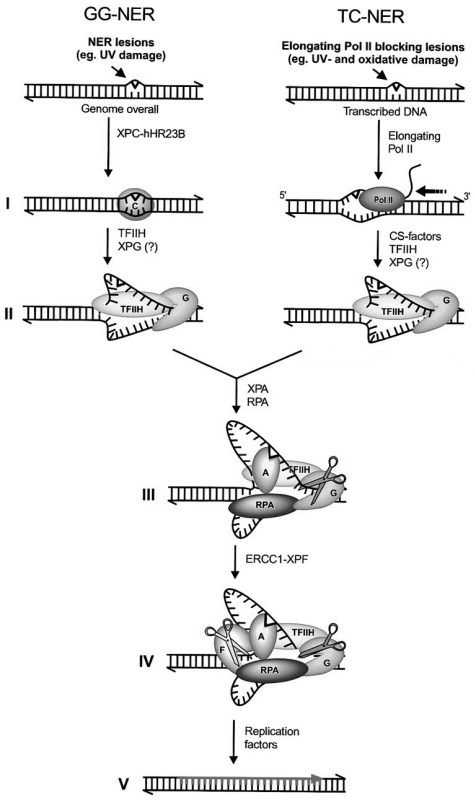


Figure 2. Model for the sequential assembly of the NER incision complex. (I) In global genome NER (GG-NER) the helix distortion is recognized by XPC-HR23B, whereas in transcription coupled repair the elongating RNA Pol II is arrested in front of a damage lesion on the transcribed DNA strand. (II) XPC-HR23B attracts TFIIH in GG-NER, while during TC-NER Pol II is displaced by specific factors that may facilitate the recruitment of TFIIH and XPG. The subsequent stages may be identical. The XPB and XPD helicases of TFIIH unwind the DNA helix in an ATP dependent manner (III) RPA, XPA and XPG are assembled to form the pre-incision complex. XPA probably verifies the damage and RPA stabilizes the open intermediate by binding to the non-damaged strand. (IV) ERCC1/XPF is the last factor to join and the dual incision occurs (3′ by XPG and 5′ by ERCC1/XPF). (V) Dual incision is followed by gap-filling DNA synthesis and ligation. Reprinted with permission from de Laat et al. Copyright 1999 Genes and Development.

NER in vivo

The NER process becomes much more complex in the chromatin environment of a eukaryotic nucleus (Smerdon, 1991). The compact nucleosomal structure restricts the access of trans-acting factors involved in various aspects of DNA metabolism (Wolffe, 1997). In NER particularly, the rates of removal of lesions differ substantially between TC- and GG-NER, because in transcription coupled repair, chromatin accessibility appears to be granted by the transcription process (Friedberg, 2001). Indeed, the efficiency of damage excision carried out by purified NER factors is lower on both UV-irradiated SV40 minichromosomes and reconstituted mononucleosomes than on naked DNA (Araki et al., 2000; Hara et al., 2000). In short, insights from chromatin restructuring upon transcription initiation, have gained ground for the detection of DNA lesions on the chromatin level as well. In the 'Access Repair Restore' model (Green and Almouzni, 2002; Moggs and Almouzni, 1999), initial repair steps include reactions that permit access of the repair machinery to DNA damage and later steps restore the canonical nucleosomal organization of the repaired DNA. These include ATP-dependent chromatin remodeling factors, such as ACF (Ura et al., 2001), and assembly of nucleosomes around a repair site by CAF-1 in a PCNA-dependent manner after repair synthesis (Green and Almouzni, 2003; Ridgway and Almouzni, 2000). In addition, p53 seems to fulfil a chromatin accessibility factor for NER that relaxes chromatin globally after UV irradiation, perhaps via the histone acetylase p300 (Rubbi and Milner, 2003). Finally, accumulating evidence points to a transcriptional regulatory role for p53 in NER, mediating expression of the global genomic repair (GGR)-specific damage recognition genes, *DDB2* and *XPC* (Adimoolam and Ford, 2002; Hwang et al., 1999).

Lesions addressed by NER

The diversity of DNA damage to be recognised and excised by NER is astonishing. Lesions are formed mainly by UV-irradiation or upon exposure to chemically reactive molecules. Prokaryotes and many eukaryotes can remove UV lesions from DNA by NER as well as through the action of specific enzymes, which use energy from light to revert pyrimidine dimers to the original monomers (Sancar, 2003; Thoma, 1999). In placental mammals, however, NER is the only pathway to remove UV-lesions from DNA, and the fact that humans have no backup mechanism for the repair of UV-lesions is evidenced by the XP phenotype (Sarasin, 2003).

The two major classes of DNA lesions induced by ultraviolet light (UV-C, 200-280 nm, and UV-B, 280-320 nm) are cis-syn cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts (6-4PPs) (Friedberg et al., 1995). The majority of CPDs are formed between adjacent thymine residues (TT) but can also occur between adjacent TC, CT, or CC. The 6-4PPs are induced preferentially at TC, CC, and TT sites and are formed at 20-30% of the yields of CPDs (Mitchell, 1988). While both types are formed at similar wavelengths, they produce different types of DNA distortion. CPDs induce a minor bent or kink in the DNA helix (7-9°) and do not impair Watson-Crick base pairing (Lee et al., 2004; McAteer et al., 1998). 6-4PPs, on the other hand, introduce large

bends (44°) and distort pairing with the complementary bases (Figure 3) (Kim and Choi, 1995). These differences are reflected in the processing of these lesions by NER and it has been observed that 6-4PPs are removed faster than CPDs (Kusumoto et al., 2001; Szymkowski et al., 1993). The UV-lesions, as well as other types of modification induced by carcinogens, can be accommodated in nucleosomes without disruption of the fundamental organization of the particle.

Many other bulky DNA adducts that are subject to NER, are caused by electrophilic compounds that exhibit some affinity for double-stranded DNA and react with the nucleophilic atoms of the DNA. Cisplatin (cis-diaminodichloroplatinum) is a chemotherapeutic drug for the treatment of tumours and forms monoadducts, intra-strand, or inter-strand cross-links by nucleophilic attack on the DNA (Jamieson and Lippard, 1999). Platinum drugs produce a high proportion of intra-strand cross-links between neighbouring pyrimidine residues: 1,2-d(GpG) and 1,3-d(GpNpG) where one nucleotide (N) separates the cross-linked guanines (Jamieson and Lippard, 1999; Kartalou and Essigmann, 2001). Since they involve differently spaced bases, the 1,2- and 1,3- cross-links exhibit different structural constraints on the DNA double helix. Again, as for the UV lesions, different DNA distortions correspond to different repair rates and 1,3- adducts are removed more efficiently than 1,2- adducts (Zamble et al., 1996). The inter-strand cross-links, on the other hand, are critical cytotoxic lesions but they are not addressed by NER (in fact they are inaccessible to any process that requires separation of the strands, like replication and transcripton). Intriguingly,

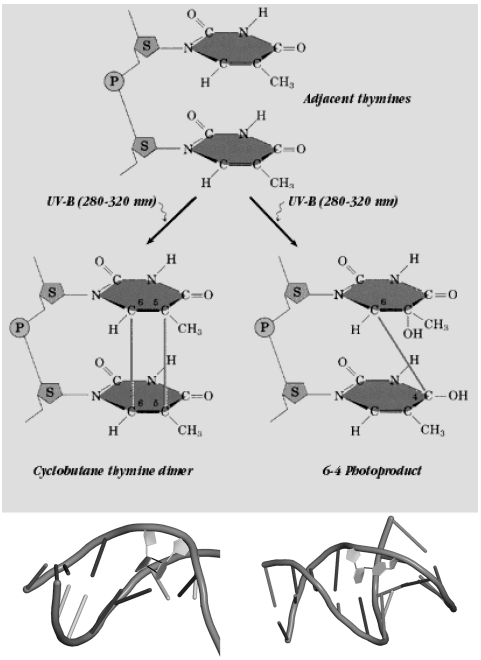


Figure 3. Formation of the most common UV photoproducts between two adjacent thymine residues on the same strand of DNA (top) and their stereochemistry in the double helix of the DNA (bottom) as revealed by the solution structures. Notice that CPDs do not alter the Watson-Crick base pairing, in contrast to 6-4PPs that result in loss of Watson-Crick hydrogen bonds.

repair of interstrand cross-links (ICL repair) by homologous recombination employs the function of ERCC1-XPF endonuclease, a core NER factor, to uncouple the cross-link (Niedernhofer et al., 2004).

Nonpolar carcinogens, such as Benzo[a]-pyrene, which is abundant in cigarette smoke and automobile exhausts, can intercalate into the DNA helix and therefore present a NER substrate. Similar is the case for the aromatic amines, e.g. aminofluorene (AF), which is used as a model compound to study carcinogenesis and DNA repair (Heflich and Neft, 1994). Dozens of other base modifications fill in the repertoire of DNA damage recognised

and processed subsequently in a "cut-and-patch" manner by NER.

Damage detection and sequential assembly of the NER factors

The broad DNA damage specificity suggests that NER does not recognise the lesion itself but rather specific conformational features of the DNA caused by the lesion. In order to be well-repaired by NER, a lesion must both distort the structure and covalently modify the DNA (Hess et al., 1997). These observations were formulated in the "bipartite substrate discrimination" model (Hess et al., 1997), where in a first step, DNA helix distortion is recognised, and in a second step, damaged strand and chemical alteration are located. Consistent with this model, the higher repair rates of 6-4PPs versus CPD lesions and of 1,3-d(GpNpG) versus 1,2-d(GpG) intra-strand cisplatin cross-links agree well with the amount of base pairing disruption and unwinding induced by the adducts (Huang et al., 1994; Szymkowski et al., 1992). The same relationship seems to hold true for the damages induced by other carcinogens. Apart from the flexibility in damage detection, the NER multiprotein repair machinery must as well have the capacity to find lesions anywhere in the chromatin condensed genome of living cells.

A long-standing problem in understanding the detailed mechanism of NER was whether it operates by sequential assembly of the core factors at the site of DNA damage or through the action of a preassembled 'repairosome' that would continuously scan the genome for the lesions. Pre-assembled factories are indeed involved in many cellular processes, as replication, RNA splicing, nuclear transport and translation (Alberts, 1998). Co-purification of several NER factors from yeast or human cell extracts (Rodriguez et al., 1998; Svejstrup et al., 1995), or even of an active repair complex (He and Ingles, 1997; Kong and Svejstrup, 2002), supported an analogous mode of action for NER. On the other hand, reconstitution of the incision reaction with individually purified factors and existence of stable protein-DNA intermediates (pre-incision complexes) favoured the sequential mode of action (Guzder et al., 1996; Wakasugi and Sancar, 1998).

The controversy was resolved by elegant in vivo studies using fluorescently tagged proteins (GFP-ERCC1 (Hoogstraten et al., 2002; Houtsmuller et al., 1999), GFP-XPB (Hoogstraten et al., 2002) and GFP-XPA (Rademakers et al., 2003)) combined with photobleaching techniques. These experiments demonstrated that the NER proteins move freely in the nuclei of nonirradiated cells with a diffusion coefficient proportional to the molecular weight of the individual factors (ERCC1-XPF, TFIIH, or XPA, respectively) and are transiently immobilized to perform repair on damage sites. Thus, the dynamic nature of the protein-DNA complexes that serve as the intermediates in NER illustrates how mutually exclusive, and competitive, reactions can be harnessed through temporally ordered formation and protein exchange to produce new nucleoprotein complexes that are kinetically competent for the subsequent step in the reaction mechanism (Riedl et al., 2003). Finally, since the sites of DNA damage are randomly distributed around the chromosomes, the sequential assembly at the site of damage facilitates the relocation of the individual

NER components (Kowalczykowski, 2000).

Consistent with the sequential assembly mechanism is the existence of factors responsible for the initial detection of DNA damage and the subsequent recruitment of all other NER proteins. Measurements of binding affinities and specificities however, proved inconclusive in this respect because none of the potential candidates (DDB (Fujiwara et al., 1999; Keeney et al., 1993; Reardon et al., 1993), XPC-HR23B (Batty et al., 2000; Hey et al., 2002; Sugasawa et al., 1998; Sugasawa et al., 2001), XPA (Buschta-Hedayat et al., 1999; Jones and Wood, 1993; Li et al., 1995), RPA (Burns et al., 1996; He et al., 1995; Schweizer et al., 1999)) exhibited a clear preference for binding damaged DNA over nondamaged DNA in the context of the genome. The matter was complicated further by the in vitro reconstitution of NER with different order of the addition of the factors, where either XPC-HR23B or XPA/RPA was proposed as the initial damage recognition factor (Sugasawa et al., 1998; Wakasugi and Sancar, 1999). The controversy was finally resolved by the use of in vivo experiments. The breakthrough came from a novel technique of local UV irradiation at well-defined regions in the nuclei of living cells through the pores of a polycarbonate filter, combined with fluorescent antibody labelling (Volker et al., 2001). Both the UV lesions inflicted at specific sites and the recruitment of the NER factors to these sites can be detected by staining with an appropriate antibody. It was demonstrated that, while the requirement of XPA to DNA damage requires functional XPC, XPA is not needed for the accumulation of XPC at sites of damage and therefore XPC-HR23B is the principal DNA damage binding protein for GG-NER. In agreement with the very early role of XPC in the process, XPC is not required for TC-NER where the damage is recognised by a stalled RNA polymerase (Laine and Egly, 2006).

The core NER reaction

I) Damage recognition by XPC-HR23B

The XPC protein forms in vivo a heterotrimeric complex involving one of the two human homologs of *S. cerevisiae* Rad23p, hHR23B and hHR23A (Masutani et al., 1994), and centrin 2 (Araki et al., 2001), a centrosomal protein. Both partners stabilize the XPC protein in the heterotrimer by inhibiting polyubiquitylation of XPC (Ng et al., 2003) and preventing its degradation by the 26S proteasome (Ortolan et al., 2000). Furthermore, HR23B (Sugasawa et al., 1996) and centrin2 (Nishi et al., 2005) stimulate NER via their interactions with the C-terminal half of XPC by enhancing its damaged DNA binding activities. Finally, expression of XPC is upregulated by p53-induced transcription in response to DNA damage (Adimoolam and Ford, 2002; Amundson et al., 2002; Fitch et al., 2003a), probably to avoid potential toxic effects caused by the overexpression of this factor.

The XPC-HR23B complex can bind efficiently to artificial substrates such as undamaged ssDNA, short 3′-overhang dsDNA, and DNA with 3-nucleotide bulge (Sugasawa et al., 2002). All these DNA structures mimic aspects of the flipped-out nucleotides of the damaged DNA conformation, revealed by the crystal structure of the *S. cerevisiae* XPC ortholog, Rad4, when bound to damaged DNA containing a CPD photolesion (Min

and Pavletich, 2007). Rad4 mostly covers the 3′ side of the lesion, leaving the 5′side almost completely free. A β-hairpin is inserted through the DNA duplex from the major groove at the damaged site, forming a highly kinked DNA conformation (Figure 4A). Importantly, the two undamaged bases opposite the CPD lesion make extensive interactions with Rad4, whereas the lesion itself is flipped out and structurally disordered (Min and Pavletich, 2007). This represents strong support for a model in which XPC/Rad4 recognize and bind non-hydrogen-bonded bases within the DNA duplex, rather than the lesions themselves, and explains the broad substrate specificity of GG-NER.

Furthermore, it has been shown that the N-terminal half of XPC interacts physically with XPA (Bunick et al., 2006), suggesting an additional role for XPC in later steps of NER. Finally, CPDs are poorly recognized by XPC-HR23B, but recent in vivo studies have demonstrated an important role for DDB (Fitch et al., 2003b; Wakasugi et al., 2002) in recruiting XPC-HR23B at the sites of CPDs. Once XPC-HR23B has been localized to aberrant DNA sites, it most likely produces a well-defined DNA conformation (Tapias et al., 2004) and recruits TFIIH, the next NER factor, by protein-protein interactions (Araujo et al., 2001; Riedl et al., 2003).

II) Open complex formation by TFIIH

TFIIH is a basal transcription factor that is composed of a core complex (XPB, XPD, p62, p44, p34, p52, TTDA) and a cdk activating kinase (CAK) subunit (Mat1, Cdk7, CyclinH). The CAK subunit is required for optimal transcriptional activity but is dispensable for NER (Araujo et al., 2001; Coin et al., 1999). TFIIH is normally involved in RNA II and I polymerase transcription but can switch readily to repair sites upon UV irradiation (Hoogstraten et al., 2002). Of high interest for NER are the two ATP-depedent helicases of TFIIH, XPB and XPD, responsible for 3′->5′ and 5′->3′ opening of the DNA around the lesion, respectively (Hwang et al., 1996; Winkler et al., 2000). While both activities are required for NER, and are crucial for the recruitment of the two endonucleases (Mone et al., 2004; Zotter et al., 2006), XPG and ERCC1-XPF, the XPD helicase activity is dispensable for in vitro transcription (Tirode et al., 1999). XPD, however, has an additional architectural role within the complex by connecting the core TFIIH with the CAK subunit (Schultz et al., 2000).

The assembly of TFIIH to the damaged DNA can be distinguished in two distinct steps. First, TFIIH is recruited to the site of damage by protein-protein interactions with XPC-HR23B in an ATP-independent manner without modifying significantly the DNA configuration stabilized by XPC-HR23B (Tapias et al., 2004). In a second step TFIIH unwinds the DNA further in both directions utilizing the energy of ATP and yielding a new nucleoprotein intermediate, able to attract the subsequent NER factors. This step can be characterized as "kinetic proofreading" because one of the two helicases will get stalled at the site of damage as it encounters the chemical modification of the DNA (Dip et al., 2004). Such a proofreading corroborates perfectly with the bipartite damage recognition process described above (Hess et al., 1997). The base-pairing disruption is sensed by XPC-HR23B and the chemical modification is confirmed subsequently by the blockage of a TFIIH helicase at the lesion. This bipartite

discrimination mechanism confers high levels of selectivity to avoid futile repair events on undamaged DNA and also protects the intact complementary strand from inappropriate cleavage. Furthermore, immobilization of TFIIH after unwinding is critical because the XPC-HR23B/TFIIH/DNA intermediate forms a platform for the following steps of NER.

III) Pre-incision complex formation by RPA, XPA and XPG

Three factors join subsequently the open NER substrate, RPA XPA and XPG, with the order of assembly remaining unclear (Riedl et al., 2003; Volker et al., 2001; Wakasugi and Sancar, 1998). However it is known that they form a stable "pre-incision complex" (Mu et al., 1996) missing only ERCC1-XPF, the last factor to arrive and trigger the dual incision. In addition, it is thought that XPC-HR23B leaves the nucleoprotein complex upon arrival of XPG (Riedl et al., 2003; Wakasugi and Sancar, 1998). It seems that the three proteins, RPA, XPA, and XPG, do not require the presence of each other in order to join the damage sites. It is rather that TFIIH plays a central role beyond unwinding the DNA whereby protein/protein interactions recruits these factors (Araujo et al., 2001).

RPA consists of three subunits (RPA70, RPA32, and RPA14) (Henricksen et al., 1994). The trimerization core comprises the C-terminal DNA-binding domain of RPA70 (DBD-C), the central DNA binding

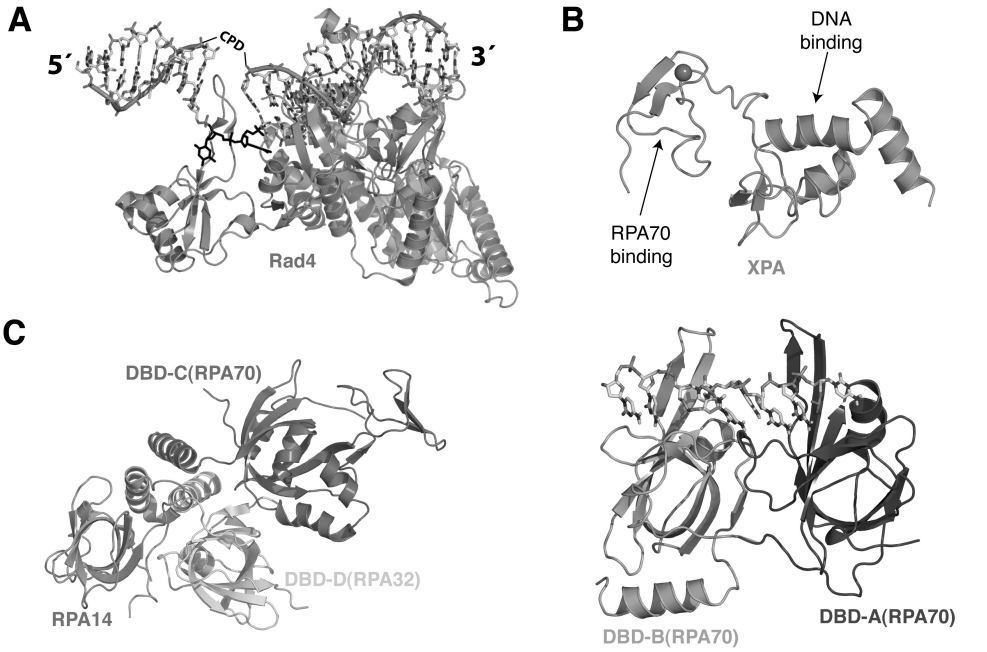


Figure 4. Available structures of NER proteins. (A) Structure of the Rad4-DNA complex (yeast XPC ortholog). The damaged strand is indicated by a worm representation through the phosphate backbone. The disordered CPD photoproduct is shown schematically, and the two flipped-out thymidines opposite to them are colored black. (B) Solution structure of the central region of XPA. Binding sites are indicated. (C) Structures of the trimerization core of RPA (left) and the DBD-A and DBD-B domains of the RPA70 subunit bound to ssDNA (right).

domain of RPA32 (DBD-D) and the entire RPA14 subunit (Figure 4C) (Bochkareva et al., 2002). RPA binds ubiquitously to ssDNA by using four so-called OB-fold motifs (Bochkarev and Bochkareva, 2004), DBD-A, DBD-B, DBD-C of RPA70 and DBD-D of RPA32. As such, it is very important for DNA metabolism and has essential roles in recombination and particularly replication (Wold, 1997). The structure of RPA70 (DBD-A and DBD-B) with DNA shows that the ssDNA lies in a channel that extends from one subdomain to the other (Bochkarev et al., 1997). ssDNA binds to the protein via extensive electrostatic and stacking interactions (Figure 4C).

In NER, RPA is required both for the dual incision and the gap-filling synthesis step (Shivji et al., 1995). Actually it binds to the undamaged strand, thereby allowing the accurate positioning of the endonuclease activities of XPG and ERCC1-XPF (de Laat et al., 1998b). After dual incision it remains bound to the undamaged strand and allows the repair synthesis proteins to take over and restore DNA integrity. Finally, it forms a well-defined complex with XPA and together may work cooperatively to verify the damaged strand and specifically position RPA to the nondamaged strand by probing the DNA conformation (Lee et al., 2003; Segers-Nolten et al., 2002). In this way RPA protects the intact strand from inadvertent nuclease attack.

XPA is a 36kDa metalloprotein that functions exclusively in NER. The central part of the protein has been suggested to interact with RPA70 and DNA (Kuraoka et al., 1996; Li et al., 1995). The NMR structure of this domain showed the existence of a positive charged cleft and confirmed that DNA binding occurs in the loop-rich subdomain and that RPA70 interaction occurs in the zinc binding core (Figure 4B) (Ikegami et al., 1998). It has been shown that the protein has high affinity for kinked DNA substrates (Missura et al., 2001), such as three-way junctions, and it has been suggested that in complex with RPA control the proper assembly of the 'pre-incision complex". Indeed, despite the relatively small size, XPA interacts with a striking number of proteins (XPC-HR23B, TFIIH, RPA, ERCC1-XPF) and the NER reaction cannot complete in the absence of this factor. Recent studies have provided molecular details for the ERCC1-XPA interaction (Tripsianes et al., 2007), in line with the notion that XPA is absolutely required to anchor ERCC1-XPF in the pre-incision complex (Riedl et al., 2003; Volker et al., 2001).

IV) Dual incision by XPG and ERCC1-XPF

XPG belongs to the FEN-1 family of structure specific endonucleases (Lieber, 1997) and is multifunctional, as it is reported to participate in BER as well (Klungland et al., 1999). In NER, XPG carries out the incision at the 3′ side of the damage and stabilizes the protein complex on the locally unwound DNA. Indeed, it is able to incise flap or bubble substrates with defined polarity at junctions between double- and single-stranded DNA at the 3′ arm of ssDNA (Evans et al., 1997; Hohl et al., 2003). Its active site has been mapped to two conserved domains that are separated by a spacer region of 600 amino acids (Hohl et al., 2003; Hwang et al., 1998). The spacer region is responsible for the interaction with TFIIH and probably important for the recruitment of XPG to the NER pre-incision complex (Araujo et

al., 2001; Thorel et al., 2004). Apart from performing the 3′ incision in NER, it has a structural role in stabilizing the fully open DNA bubble structure and permitting the 5′ incision by ERCC1-XPF (Constantinou et al., 1999).

The last factor to join the pre-incision NER complex is ERCC1-XPF (Riedl et al., 2003; Volker et al., 2001), another structure specific endonuclease but with the opposite polarity compare to XPG (de Laat et al., 1998a). ERCC1 and XPF form an obligate heterodimer important for the stability of both partners and the catalytic function. The interaction relies mainly at the C-terminal HhH_2 domains of ERCC1 and XPF proteins (de Laat et al., 1998c; Tripsianes et al., 2005; Tsodikov et al., 2005) and is hydrophobic in nature. The active site resides at the XPF protein (nuclease domain) and consists of the characteristic $ERKX_3D$ signature found in archaeal nucleases (Enzlin and Scharer, 2002). The equivalent central domain of ERCC1, though non-catalytic, plays a significant role in the DNA substrate recognition facilitated by specific interactions with XPA (Riedl et al., 2003; Tripsianes et al., 2007; Volker et al., 2001), which is already present in the open bubble substrate. The subject of the present thesis is the ERCC1-XPF heterodimer and the following chapters provide detailed analysis for the mode of dimerization, as well as for the structure and the contribution of each domain to the coordinated function. In short, it seems that the two proteins have divided the tasks with ERCC1 being responsible for the NER substrate recognition and XPF carrying out the cleavage (Enzlin and Scharer, 2002; Sijbers et al., 1996; Tripsianes et al., 2005). While the arrival of ERCC1-XPF does not induce further changes in the NER intermediate, its enzymatic function triggers the transition from the excision to repair synthesis step (Evans et al., 1997; Tapias et al., 2004).

What remains unclear yet, is the temporal order of the dual incision. Uncoupled incision 3′ to the damaged site by XPG has been observed in the absence of ERCC1-XPF (Mu et al., 1996), but its presence is required for the 5′ incision (Constantinou et al., 1999). Independently of the order of cleavage in both sites of the damaged strand, an oligunucleotide of approximately 30 bases is released containing the lesion. The link to repair synthesis and restoration of the genetic information is RPA, as it remains bound to the gapped DNA after the dual incision (Riedl et al., 2003).

V) Repair synthesis

The same proteins that are also involved in replication accomplish the repair synthesis step. These include, apart from RPA, the polymerase processivity factor PCNA (Nichols and Sancar, 1992), the pentameric clamp loader RFC, the PCNA dependent DNA polymerase δ or ε (Podust et al., 1992), and DNA ligase I. Recent evidence has shown that polymerase κ can carry out the repair synthesis step of NER (Ogi and Lehmann, 2006), a role conventionally ascribed to polymerases δ or ε (Shivji et al., 1995; Syvaoja et al., 1990). RPA is involved in the recruitment of RFC and PCNA (Gomes and Burgers, 2001). Both of them facilitate the anchoring of the polymerase to the DNA template. In a processive manner, the polymerase synthesizes the new DNA using as a template the non-damaged strand and finally the ligase seals the nick (Alberts, 2003).

Evolution of XPF function through formation of dimers

The members of the XPF family are characterized by the presence of the ERCC4 nuclease domain. In humans seven members have been identified: XPF, MUS81, ERCC1, EME1, EME2, FANCM and FAAP24 (Ciccia et al., 2007). Of them, only XPF and MUS81 have nuclease activity due to the conservation of the core nuclease motif ($ERKX_3D$) (Enzlin and Scharer, 2002; Heyer et al., 2003; Nishino et al., 2003). The nucleolytic activity however, depends on heterodimer formation with the non-catalytic members. XPF forms obligate hetrodimer with ERCC1 to deliver the 5´ incision during NER. Likewise, MUS81 forms heterodimer either with EME1 or EME2 and is thought to play a role in processing early recombination intermediates in the restart of stalled replication forks and during meiotic recombination (Ciccia et al., 2007; Hollingsworth and Brill, 2004; Osman et al., 2003). FANCM/FAAP24 heterodimer formation was identified recently, and it functions as a complex in the Fanconi Anemia pathway. Interestingly, in this heterodimer both nuclease domains have degenerated and the complex is catalytically inactive (Ciccia et al., 2007). In all proteins the ERCC4 domain is followed by one or two helix-hairpin-helix motifs (Figure 5A).

The exception of the heterodimeric rule is the archaeal members of the family that form homodimers. Crenarchaeota, such as *Sulfolobus solfataricas* and *Aeropyrum pernix*, contain a short XPF homolog, which consists of the catalytic nuclease domain (ERCC4), followed by a DNA-binding domain containing two consecutive helix-hairpin-helix motifs (HhH_2 domain) (Roberts et al., 2003). It also bears a C-terminal PCNA-interacting peptide required for cleavage (Roberts et al., 2003). Euryarchaeota, such as *Pyrococcus furiosus*, posses a long form of XPF containing an extra functional helicase domain at the N-terminus that stimulates the nuclease activity (Figure 5A) (Nishino et al., 2005b). Both archaeal proteins cleave the replication fork or 3´ flap DNA structures in the homodimeric form (Nishino et al., 2005a; Nishino and Morikawa, 2002; Roberts and White, 2005). The homodimer in both cases is held together by two independent interfaces: one between the HhH_2 domains and a second one between the nuclease domains (Figure 5D and 6B) (Newman et al., 2005; Nishino et al., 2003).

I) HhH_2 domains and DNA binding

The HhH motif has been characterized as a DNA binding structural unit that recognizes the phosphate backbone of DNA in a sequence-independent manner (Doherty et al., 1996). It has been detected in a variety of protein families exemplified by DNA polymerases, DNA ligases, and DNA glycosylases (Doherty et al., 1996; Thayer et al., 1995). Structurally, the motif forms into a pair of anti-parallel α-helices connected by a hairpin-like loop. This loop is involved in interactions with DNA (Bruner et al., 2000; Sawaya et al., 1997) and usually contains a consensus glycine-hydrophobic amino acid-glycine sequence pattern (G*h*G), where *h* is a hydrophobic residue, most commonly Ile, Val or Leu. Further sequence and structural analyses unveiled many examples of two consecutive HhH motifs (HhH_2 domain) connected by a linker helix (Shao and Grishin, 2000) that form a five helical entity. Available structures of HhH_2 protein–DNA

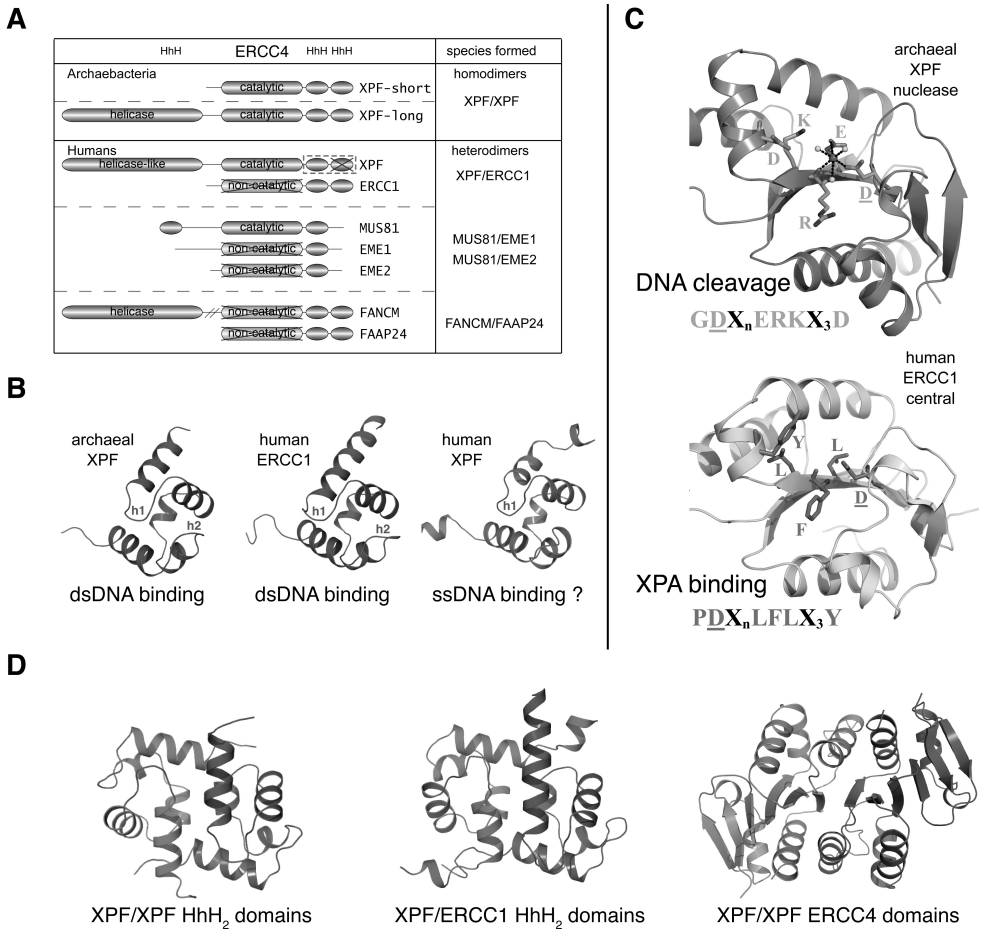


Figure 5. The XPF protein family. (A) Domain organization for the members of the XPF family from archaea and humans, and the dimeric species formed. ERCC4 and HhH domains are indicated at the top. (B) Fold of the HhH2 domains adapted upon dimerization. The canonical double hairpin motif binds to DNA minor groove, whereas the altered fold of the human XPF may be competent for single-stranded DNA binding. Proper hairpins are indicated and colored brown for each domain. (C) The fold of the ERCC4 domain. The conserved catalytic site from the archaeal nuclease with the coordinated divalent cation is shown in stick representation (top) and the corresponding substitutions in the human ERCC1 central domain that mediate XPA binding in stick representation as well (bottom). (D) Dimer interfaces for the HhH2 domains of the XPF homodimers and the human XPF/ERCC1 heterodimer and for the ERCC4 domains of the XPF homodimers. For the homodimers one protomer is shown in light color. (Full color on page 99)

complexes (Ariyoshi et al., 2000; Bruner et al., 2000; Hollis et al., 2000) show that hairpins of both helical pairs recognize both strands of DNA from the minor groove side. In the case of the Holliday junction binding protein RuvA, four subunits form the exact 4-fold symmetric complex, in which the HhH_2 domain in each subunit anchors an arm of the Holliday junction (figure 6A) (Ariyoshi et al., 2000; Yamada et al., 2004). In particular, the two glycine residues of the conserved *GhG* element of the first

HhH motif form hydrogen bonds with the phosphate oxygens of adjacent nucleotides on one strand of the DNA duplex. Similarly, the glycines of the second HhH motif form hydrogen bonds with the phosphate oxygens of the complementary DNA strand of the minor groove. Polar residues after each HhH motif are close enough for electrostatic interactions with the phosphate oxygens of either strand. In other words, the protein-DNA contacts are built on the sugar-phosphate DNA backbone allowing the HhH_2 domain to recognise the DNA in a sequence non-specific manner (figure 6).

Previous structural studies in our lab demonstrated that the C-terminal domain of the bacterial UvrC belongs to the HhH_2 domain (Singh et al., 2002). The structure has shown that is a structurally compact entity with a well-defined hydrophobic core. In addition, chemical shift perturbation experiments indicated that both hairpins are involved in symmetric contacts with the DNA-backbone, in agreement with the RuvA-DNA crystal structure. UvrC is essential for 5´ incision in the prokaryotic nucleotide excision repair process and is functionally homologous to the ERCC1/XPF complex. The studies on UvrC prompt us to explore the structural and functional features of the human heterodimeric counterpart.

The crystal structures of the archaeal XPF homodimers revealed the first case of a dimer of HhH_2 domains (Newman et al., 2005; Nishino et al., 2005a). Each HhH_2 domain of the dimer forms an integral five-helical domain bearing two functional helix-hairpin-helix motifs related by pseudo-two-fold symmetry (Figure 5B). This motif is highly hydrophobic and dimerization ensures integrity for the domain of each protomer (Newman et al., 2005; Nishino et al., 2005a). The structure of the archaeal

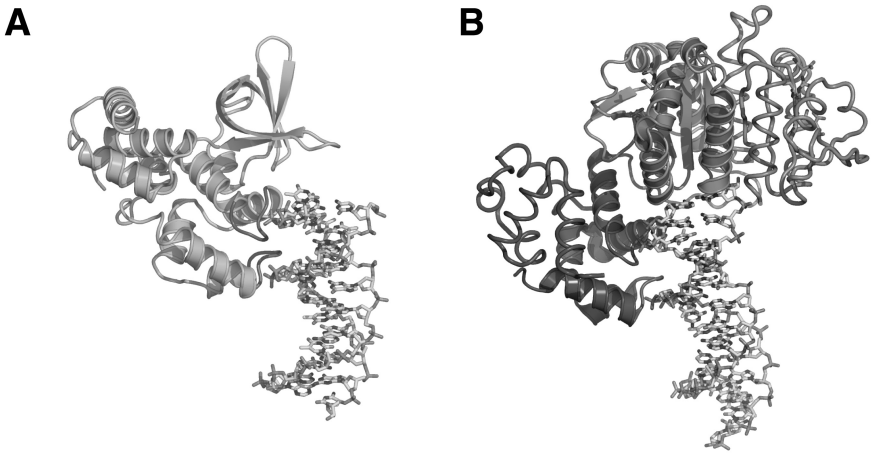


Figure 6. DNA recognition by the HhH_2 domains. (A) A DNA arm of the Holliday junction is recognized on the minor groove side by the two HhH motifs of one RuvA protomer. The interaction involves hydrogen bonding of the main chain amides with the phosphate oxygens of both strands (B) Similarly, one HhH_2 domain of the archaeal XPF homodimer bridges the two phosphodiester backbones across the minor groove of the DNA. In both cases, the protein-DNA interactions are built on the sugar-phosphate backbone of the duplex allowing the HhH_2 domain to recognize dsDNA in a sequence-independent manner. For the XPF homodimer, protomer A is in a cartoon and protomer B in a worm representation. (Full color on page 100)

XPF in complex with dsDNA confirmed their essential role in DNA recognition. dsDNA is mostly bound to the HhH_2 domain of one protomer and the terminal base pair contacts the nuclease domain (Figure 6B) (Newman et al., 2005). The first G-I-G hairpin motif binds two backbone phosphates of one DNA strand and the second G-I-G hairpin binds two phosphates of the opposite strand. Several positively charged residues immediately after the hairpins also contribute to the interaction with the minor groove.

In the human ERCC1/XPF heterodimer the C-terminal domains are critical for complex formation (de Laat et al., 1998c). Choi and co-workers (Choi et al., 2005) observed that the ERCC1 domain adopts structural integrity only if it is refolded in the presence of the XPF domain to yield a heterodimer with 1:1 ratio of the two partners. The refolding scheme enabled them to selectively label either ERCC1 or XPF domain and perform cross-saturation NMR experiments. In this way they mapped the interacting residues in both sides and showed that the heterodimerization is mainly hydrophobic in nature (Choi et al., 2005). Similar results were obtained by our solution NMR studies (Tripsianes et al., 2005). Furthermore, in **Chapter 4** we show that a single ERCC1 mutation at the hydrophobic interface with XPF rearranges the packing between the partners and destabilizes the heterodimer.

Soon after the crystal structure of the same interacting domains was reported (Tsodikov et al., 2005). The structure confirmed that the heterodimer consists of two structurally similar, tandem HhH_2 domains apposed in a pseudo twofold symmetrical arrangement. The two domains interact through an extensive network of hydrophobic interactions. The structural similarity with archaeal XPF homodimer was noted as well. Functional analysis on these domains however, indicated that two molecules of ssDNA are required to saturate binding by one ERCC1/XPF heterodimer (Tsodikov et al., 2005). Therefore the authors proposed that both ERCC1 and XPF HhH_2 domains bind equally the single strands of relevant substrates, such as bubble and splayed arm, that are known to represent the favoured targets for the enzymatic activity of the full-length heterodimer.

Our structural studies on these domains have shown that human ERCC1 and XPF share the HhH_2 motif of the archaeal XPF and complex in the same manner as the archaeal species do (Figure 5D) (Tripsianes et al., 2005; **Chapter 2**). Furthermore the ERCC1 HhH_2 domain can fold only in the presence of the XPF corresponding domain, with the latter acting as a folding scaffold. The dimer interface is predominantly hydrophobic and the major contacts are between two helices and the C-terminal tail of each partner. Interestingly, the human XPF second hairpin is altered to a turn and this domain deviates structurally from the canonical HhH_2 motif observed in the archaeal members and the human ERCC1 (Figure 5B). This suggests that although human XPF retains the structural integrity upon heterodimerization its function may have diverged or even been lost. NMR chemical shift mapping of residues involved in a stem-loop DNA binding has shown that only the ERCC1 HhH_2 domain is involved in DNA binding (Tripsianes et al., 2005; **Chapter 2**). In line with the crystal structures of RuvA and archaeal XPF when bound to DNA, the equivalent residues of

both ERCC1 hairpins were perturbed in the DNA titration, but not the corresponding residues of XPF. This is in agreement with our structural analysis and the preference of the human heterodimer for ds- to ss-junctions that consist of only one DNA duplex. Therefore we suggested that the HhH_2 domain of ERCC1 binds to the minor groove of the DNA before the junction (Figure 5B). The main difference with the crystallographic studies is that our data distinguish functionally the ERCC1 and XPF HhH_2 domains, an observation with implications for the obligate nature of the heterodimerization.

Overall, crystal and solution structures from archaea and humans provide evidence that XPF homodimers and ERCC1/XPF heterodimer dimerize in the same way through the HhH_2 domains (Figure 5D) (Newman et al., 2005; Nishino et al., 2005a; Tripsianes et al., 2005; Tsodikov et al., 2005).

II) ERCC4 domains and the corresponding functions

The crystallographic analysis of the ERCC4 (nuclease) domain with respect to the archaeal nuclease interface revealed that the endonuclease fold comprises a six-stranded β-sheet flanked by several α-helices (Nishino et al., 2003). The two protomers make similar dimer contacts via two helices and an edge β-strand (Figure 5D) (Newman et al., 2005). The tertiary structure of each protomer is similar to that of the type II restriction endonucleases (Nishino et al., 2005a; Nishino and Morikawa, 2002). The signature motif $ERKX_3D$ with an extension of GDX_n at the N-teminus (GDX_nERKX_3D), can be nicely superimposed with the $PDX_n(D/E)XK$ motif of the type II restriction endonucleases (Figure 5C). As in type II restriction enzymes, the conserved acidic residues participate in coordinating the divalent metal ion (Mg^{2+} or Mn^{2+}). Other polar residues of the signature motif participate in hydrogen bonding with the coordinated water molecules. The conserved Lys residue is thought to deprotonate the nucleophilic water molecule for phosphodiester bond breakage. The conserved Arg is unique in the structure-specific endonucleases and directly coordinates the divalent cation with its carbonyl oxygen atom. Moreover, it forms a salt bridge with a Glu side chain, which in turn coordinates the divalent cation through a water molecule. Mutational analyses of these conserved residues in the archaeal and human XPF proteins have firmly established their roles in catalysis upon homo- or hetero-dimerization (Enzlin and Scharer, 2002; Nishino et al., 2003).

The central domain of ERCC1 is the equivalent of the XPF nuclease. This domain is devoid of the conserved catalytic residues and therefore is not active. The crystal structure (Tsodikov et al., 2005) has shown that ERCC1 central domain resembles the fold of type II restriction endonucleases and superimposes very well on the nuclease domain of the archaeal XPFs. DNA binding experiments with structurally diverse DNAs revealed that this domain is a single-stranded DNA binder. In addition, the higher affinity for a 5´ overhang over a 3´ overhang substrate indicates that the central domain of ERCC1 binds to ssDNA with defined polarity. Based on these finding, it was suggested that this domain would bind the uncleaved strand of a bubble-type substrate for NER (Tsodikov et al., 2005).

The solution structure of the ERCC1 central domain is in very good agreement

with the crystal structure (Tripsianes et al., 2007; **Chapter 3**). Together these studies have extended the structural similarity to XPF and ERCC1 proteins (Figure 5C). We have confirmed by EMSA experiments that the central domain binds to ssDNA. Furthermore, we have identified the DNA binding site on the ERCC1 structure by chemical shift perturbation experiments. Our data suggest that the central domain binds three to four bases of the single strand. The DNA binding properties of this domain may account for the altered specificity of the human enzyme compared to the prototype XPF homodimers. In addition, we have shown that the central domain interacts specifically with a small region of the XPA factor. This interaction is very important as it targets the heterodimer to the NER pre-incision complexes and positions the XPF nuclease for DNA incision (Riedl et al., 2003; Volker et al., 2001). Remarkably, the XPA binding site of ERCC1 corresponds structurally to the active site of the archaeal nucleases (Figure 5C). XPA and ssDNA binding sites are independent of each other and the two interactions can happen simultaneously. These data show that the central domain of ERCC1 plays an important role in recognizing the NER intermediates to be cleaved by the XPF nuclease (Tripsianes et al., 2007; **Chapter 3**).

Structural data for the human XPF nuclease are not available. The strong conservation of the catalytic residues argues that the fold is identical to that of the archaeal counterparts and very similar to the ERCC1 central domain. Recent data in our lab indicate that this domain has no structural integrity of its own but it is stabilized by the presence of the ERCC1 central domain (unpublished data K.T.).

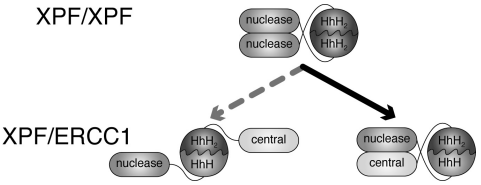


Figure 7. Domain organization of the archaeal homodimers and the human heterodimer with respect to the HhH_2 and ERCC4 domains. (Full color on page 100)

This implies that in ERCC1/XPF exists a second dimer interface analogous to the archaeal nuclease interface. The structure of the archaeal homodimer has revealed that the flexible linker between the HhH_2 domain and the nuclease domain enables the two functional domains to dimerize independently and permits formation of an overall asymmetric conformation. The same may hold true for the human ERCC1/XPF heterodimer (Figure 7).

III) Mode of action of the archaeal XPF homodimer

The crystal structure of the archaeal XPF homodimer in complex with dsDNA provides useful means in understanding the function of both homo- and hetero-dimeric enzymes, despite the clear differences in their substrate specificity. The XPF homodimers cleave 3´ flaps and fork DNA substrates (Roberts and White, 2005) preferentially over splayed duplexes, which is a favoured substrate for the human ERCC1/XPF heterodimer (de Laat et al., 1998a). The archaeal substrates contain two segments of continuous DNA that can be engaged simultaneously by the HhH_2 domains of the dimer. In the crystal structure of XPF-dsDNA only the HhH_2 domain of protomer A is bound to DNA, but the HhH_2 domain of protomer B is also potentially capable of binding DNA. Therefore, the authors have

modelled how the protomer B might bind to the DNA by generating a second putative dsDNA molecule (Newman et al., 2005). In this case, the modelled DNA lies close to the catalytic centre of protomer A and the putative 3´ overhanging strand can enter the active site of protomer A, in agreement with the expected polarity for the cleaved strand and the enzymatic activity. According to this model protomer A engages the downstream duplex and protomer B the upstream duplex of a 3´ flap substrate, the branched DNA is bent by 90 degrees and allows the nuclease of protomer A to be catalytically active (Figure 8, top left).

Complementary studies in archaea are consistent with this model for the homodimeric function (Nishino et al., 2005a). In this work mutations were introduced in HhH_2 or nuclease domains of the homodimer that disrupt the function of each domain (DNA binding or DNA cleavage) but do not impair the dimer interfaces. By using different tags they were able to create heterodimers of all possible combinations, where one or more domains of the dimer have been inactivated and assess what combinations are required for the nuclease activity. The constructs that contained functional nucleases and one HhH_2 domain impaired displayed more than 10-fold reduction in the activity, while two impaired HhH_2 domains in combination with functional nucleases abolished the cleavage activity completely. On the other hand, when one active site was retained with functional HhH_2 domains almost full activity was restored. Interestingly, inactivation of one HhH_2 and one nuclease domain had the same effect like the construct with only one impaired HhH_2 domain, regardless of the mutations were in *cis* or *trans*. These data show that HhH_2 domains are equally important for recognising the branched DNA to be cleaved by one active site. DNA footprinting analysis in the same study, demonstrated that the HhH_2 domains occupy the two arms of the fork DNA, whereas one of the catalytic domains is located in the vicinity of the junction to perform the cleavage (Nishino et al., 2005a). The structural model and the biochemical data indicate an asymmetry in the function of the archaeal XPF homodimers.

IV) Evolutionary aspects of ERCC1/XPF partnership and their implications for function

Structures deposited in the Protein Data Bank over the last years have drawn many parallels for the archaeal XPF homodimers and human ERCC1/XPF heterodimer. Our functional analyses, on the other hand, have highlighted interesting disparities for the human partners that do not exist in the homodimeric case (Tripsianes et al., 2005). The clear similarities between homodimers and heterodimers, with respect to the fold and organization of their domains, suggest that the ERCC1 and XPF genes have derived from duplication of the ancestral XPF gene (Gaillard and Wood, 2001). The duplication of the gene encoding the primitive multi-functional protein (archaeal XPF) yields two independent proteins, which may experience relaxation of functional constraints and increased rate of mutations (Roth et al., 2007). Accordingly the daughter genes can acquire complementary loss-of-function mutations in the independent ancestral subfunctions (DNA binding or DNA cleavage), such that both proteins are required to produce the ancestral XPF function. This process is called subfunctionalization (Force et

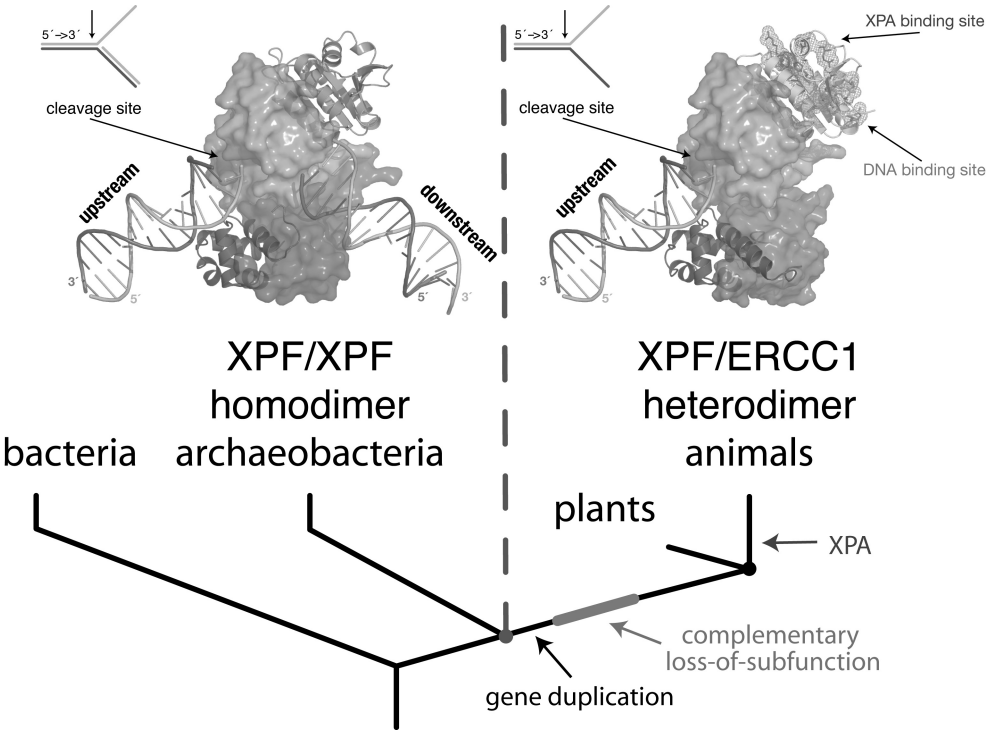


Figure 8. Schematic represenation for the evolution of XPF and ERCC1 genes in the eukaryal lineage. XPF gene duplication is followed by subfunctionalization resulting in obligate heterodimer formation (see text). Models for the function of homodimers and heterodimers reconstructed from the structural and functional data available for both XPF enzymes, in agreement with the basic principles of subfunctionalization (top). For the homodimer one protomer is shown in a cartoon and the other in a surface representation. Accordingly, for the heterodimer ERCC1 is in a cartoon and XPF in a surface representation. (Full color on page 101)

al., 1999; Lynch and Conery, 2003) and explains the preservation of the duplicate genes (XPF and ERCC1) in the eukaryal lineage (Figure 8). The specialization of the complementary functions resulted in the mutual dependence of ERCC1 and XPF proteins (concerning both protein stability and enzymatic activity) and the shift towards obligate heterodimerization. In the heterodimer ERCC1 confers the DNA binding activity (HhH_2 domain) (Tripsianes et al., 2005) and XPF the nuclease activity (Enzlin and Scharer, 2002), both present in the ancestral XPF protein (Newman et al., 2005) (Figure 5 and 8).

Phylogenetic analyses in archaea based on informational pathways (replication, transcription, translation etc.) suggest that euryarchaeota appear more similar to the eukarya than do the crenarchaeota (Gribaldo and Brochier-Armanet, 2006). As has been proposed (White, 2003), the long XPF form of euryarchaeota (*P. furiosus*) seems to be the common ancestor of the human proteins. In the course of eukaryal evolution the DEAH helicase function of the original XPF genomic locus has degenerated, while the sister ERCC1 genomic locus entirely lost this domain. Alternatively, and very likely, the XPF genomic locus was duplicated

partially, omitting the N-terminal part that corresponds to the helicase domain (Figure 5A).

This evolutionary scenario, apart from explaining the transformation from homo- to hetero- association, it allows also understanding the differences in DNA substrate specificity for the homo- and hetero- versions of the enzyme. The homodimers utilize twice the DNA binding activity of their HhH_2 domains and only one nuclease domain for DNA cleavage (figure 8). After subfunctionalization, degeneration of the nuclease activity in the ERCC1 gene does not disturb the enzymatic activity since the second nuclease is functionally redundant *ab initio*. This is not the case for the DNA binding activities mediated by the HhH_2 domains. The degeneration of this domain in the XPF gene alters the binding properties of the heterodimer because it now has only one functional HhH_2 domain compared to two in the homodimeric precursor. As such it does not target anymore 3´ flap substrates consisting of two DNA duplexes but stem-loop or splayed-arms with one duplex that is accommodated by the functional domain of ERCC1 (Figure 8). In addition, the single hairpin of the degenerated XPF domain may have preference for ssDNA explaining further the shift in the DNA substrate recognition (Figure 5B). By partitioning the tasks in this way the heterodimeric function retains the polarity of the homodimer, as both cleave the single-strand emanating in the 3´ direction from the junction (Figure 8).

Finally, in the context of NER, ERCC1/XPF targeting is facilitated by specific interactions of the ERCC1 central domain with the XPA protein. Surprisingly, this interaction is mediated by the degenerated active site of ERCC1. XPA is the last invention of the eukaryal NER (Figure 8) and has an architectural role in complex with RPA. Together they control the correct three-dimensional assembly of the NER intermediates prior to endonucleolytic cleavage (Missura et al., 2001). The otherwise non-functional ERCC1 domain offered the ground for new molecular interactions with XPA, thereby increasing the efficiency of ERCC1/XPF recruitment to legitimate NER substrates and the fidelity of the eukaryotic DNA repair. The adoption of this novel function by ERCC1 would not have been possible without the prior partitioning of the ancestral XPF tasks to the duplicate genes.

Scope of the thesis

The present thesis describes the structural and functional studies of the ERCC1/XPF structure-specific endonuclease. Our goal was to address the properties of the obligate dimerization and decipher the role of the ERCC1 non-catalytic subunit in the coordinated function. **Chapter 1** provides an overview of the current knowledge on the molecular mechanisms of NER, a multi-protein genome caretaker mechanism that employs ERCC1/XPF in the incision step of the reaction. In addition it discusses recent insight on the structures and function of the archaeal XPF homodimers and the human ERCC1/XPF heterodimer. In **Chapter 2** we analyze the molecular details of the ERCC1/XPF interaction, as revealed by the solution structure of their HhH_2 domains. Both proteins share the HhH_2 fold, but XPF exhibits an altered second hairpin. However, the canonical domain of ERCC1 requires the presence of the XPF partner to fold

properly. Functional analysis indicates that only ERCC1 is involved in DNA binding. In **Chapter 3** we have determined the structure of the so-called ERCC1 central domain. This domain has structural homology with the nuclease domain from the archaeal XPF homodimers, but lacks the active site residues and therefore is inactive. We have shown that the degenerated catalytic site of ERCC1 is responsible for interactions with the XPA protein. In addition, we have identified the ssDNA-binding site on the surface of the structure. XPA and ssDNA interactions can take place at the same time through distinct binding sites of the ERCC1 central domain. Finally, in **Chapter 4**, we provide structural explanation for the malfunction of the first reported human ERCC1 deficiency. The mutant ERCC1 HhH_2 domain dimerizes with the corresponding domain of XPF in a way analogous to the wild type structure. The ERCC1 mutation alters the local network of interaction with XPF and is responsible for the lower stability of the mutant dimer. Furthermore, the second hairpin of ERCC1 deviates from the proper geometry indicating additional problems in the DNA recognition and the enzymatic activity of the mutant ERCC1/XPF heterodimer.

Chapter two

The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair

Konstantinos Tripsianes[1], Gert E. Folkers[1], Eiso AB[1], Devashish Das[1], Hanny Odijk[2], Nicolaas G.J. Jaspers[2], Jan H.J. Hoeijmakers[2], Robert Kaptein[1] and Rolf Boelens[1]

[1] Department of NMR spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Department of Cell Biology and Genetics, Erasmus Medical Center, Rotterdam, The Netherlands

Structure, **13**, 1849-1858, (2005)

The human ERCC1/XPF complex is a structure-specific endonuclease with defined polarity that participates in multiple DNA repair pathways. We report the heterodimeric structure of the C-terminal domains of both proteins responsible for the ERCC1/XPF complex formation. Both domains exhibit the double helix-hairpin-helix motif $(HhH)_2$, and are related by a pseudo-2-fold symmetry axis. In the XPF domain, the hairpin of the second motif is replaced by a short turn. The ERCC1 domain folds properly only in the presence of the XPF domain, which implies a role for XPF as a scaffold for the folding of ERCC1. The intersubunit interactions are largely hydrophobic in nature. NMR titration data show that only the ERCC1 domain of the ERCC1/XPF complex is involved in DNA binding. On the basis of these findings we propose a model for the targeting of XPF nuclease via ERCC1-mediated interactions in the context of nucleotide excision repair.

Introduction

Genomes are vulnerable to a plethora of DNA-damaging agents, of both endogenous and environmental origin, which can compromise vital processes such as transcription and replication. To cope with these threats, all organisms have developed a complex defense network, including a variety of repair mechanisms, each suited for specific classes of DNA lesions (Hoeijmakers, 2001). Any defect in this network invariably leads to genomic or chromosomal instability, either spontaneous or after genotoxic challenge. On the cellular level, unrepaired damage is considered the driving force of carcinogenesis as well as organismal ageing.

A remarkable structure-specific DNA nuclease family in eukaryotes is represented by the heterodimeric ERCC1/XPF complex. This enzyme is known to specifically cleave near junctions between single-stranded (ss) and duplex (ds) DNA (de Laat et al., 1998a) where the single strand has a 5′-3′ polarity by recognizing the DNA tertiary structure itself and not the nucleotide sequence. The structure-specific endonuclease ERCC1/XPF performs an essential late step in the nucleotide excision repair (NER) process (Volker et al., 2001), where it nicks the damaged DNA strand at the 5′ side of a helix-distorting lesion. This action of ERCC1/XPF, together with XPG, a complementary structure-specific endonuclease attacking on the other side of the lesion at the ss-ds transition, is required for removal of the damaged segment and the subsequent resynthesis of the segment-spanning gap. A significant contribution of the ERCC1 subunit to NER is interaction with the XPA protein, which binds to the lesion in conjunction with the strand-opening helicase complex TFIIH and the single-strand binding protein RPA (Araujo and Wood, 1999; de Laat et al., 1999; Riedl et al., 2003). As such, it directs its XPF partner to a site of NER action and contributes to correct positioning of the strand incisions (Li et al., 1995).

Inherited defects in the NER process cause the serious prototype repair disorders xeroderma pigmentosum (XP) and Cockayne syndrome (CS), highlighting an extreme risk of UV-induced skin cancer and many features of accelerated ageing, respectively (de Boer and Hoeijmakers, 2000). Mice targeted for knockout (KO)

versions of genes involved in NER usually reflect the human pathology of XP and CS to a considerable extent. In contrast, ERCC1- or XPF-targeted KO mice display a much more severe phenotype with multi-organ involvement, severe runting and death before weaning (Tian et al., 2004; Weeda et al., 1997). This observation points to other functions of this endonuclease outside the context of NER. In line with this finding, XPF mutations associated with XP are relatively rare and always hypomorphic, whereas a first case of human ERCC1 deficiency remains to be reported.

One of these additional roles of ERCC1/XPF is in the recombinational repair of interstrand crosslinks, where it is required for a yet unidentified late step (Niedernhofer et al., 2004). ERCC1- or XPF-deficient hamster mutants are therefore exquisitely hypersensitive to DNA crosslinking agents (Prasher et al., 2005), much more than they are to UV-induced pyrimidine dimers, the classical substrates for NER. This function in crosslink repair may also explain the observations that expression levels of ERCC1/XPF in non-small cell lung cancer may anticorrelate with the response to antitumor therapy with the DNA crosslinker *cis*-platinum (Simon et al., 2005).

In addition, the homologous complexes of *S.cerevisiae* (RAD10/RAD1) and *Drosophila* (ERCC1/MEI9) are known to function in both mitotic and meiotic recombination. Indeed, ERCC1/XPF is also absolutely required for targeted recombination in mouse embryonic stem cells (Niedernhofer et al., 2001). Finally, there is recent evidence that ERCC1/XPF mediates genome integrity by yet another way, since it can prevent uncapped telomeres from creating chromosomal end-to-end fusions or double-minute chromosomes (Zhu et al., 2003).

For the action of ERCC1/XPF, heterodimer formation is essential, where the actual nuclease domain is contributed by the XPF subunit (Enzlin and Scharer, 2002) and the presence of ERCC1 is indispensable for nuclease activity. Details of ERCC1's function in the actual nucleolytic action are still unknown. The XPF homolog of archaebacteria acts as a homodimer (Newman et al., 2005; Nishino et al., 2003), in the absence of an ERCC1 homolog. However, in eukaryotic cells, formation of the heterodimeric complex is required for the stability of both components. The interaction of the subunits depends on the presence of their C-terminal regions (de Laat et al., 1998b), which both carry putative double helix-hairpin-helix $(HhH)_2$ motifs (Figure 4A). These domains are known to mediate DNA binding, and it has been suggested that they are responsible for the correct positioning of the endonuclease at junctions between ds- and ss-DNA where the single strand moves 5′ to 3′ away from the junction (de Laat et al., 1998a). The phenomenon of mutual dependence has hampered overproduction and study of the separate human subunits in bacterial and insect expression systems in the past. Moderate amounts have been generated in baculovirus-infected insect cells, provided both subunits are expressed simultaneously (Enzlin and Scharer, 2002).

Since the affinity of ERCC1/XPF for single-strand to double-strand transitions is the common denominator in all its known functions, we have set out to isolate and characterize the heterodimeric ERCC1/XPF interface containing the $(HhH)_2$ motifs.

In this study, we report the three-

dimensional structure of human ERCC1/XPF consisting of the last 83 residues of XPF and the last 78 residues of ERCC1, which are sufficient for heterodimerization. Analysis of the structure allows us to map in atomic detail the interacting residues and explain the role of the C-terminal region of ERCC1 $(HhH)_2$ domain for the proper complex formation (de Laat et al., 1998b). NMR titrations established the HhH motifs of ERCC1 as the DNA binding unit, an observation that revises our understanding of the tasks each subunit performs.

Results

Purification and biophysical characterization of the C-terminal parts of XPF and ERCC1

Initial attempts to overexpress the C-terminal part of ERCC1 in *E.coli* cells resulted in insoluble and aggregated protein. Attempts to resolubilize the protein domain in various manners failed. This might be explained by our previous findings indicating that ERCC1 can only form a functional fold in complex with XPF. Therefore, the C-terminal parts of XPF and ERCC1 were coexpressed using a bicistronic expression vector encoding the two subunits, with XPF being the first in order and ERCC1 containing a His_6 tag at the C terminus. We obtained soluble proteins with very high expression yield, and the elution profile in the gel filtration resulted in one species with the predicted size of the heterodimeric complex. Furthermore, the protein fingerprint spectrum (^{15}N-^{1}H HSQC) revealed the expected number of residues, confirming complex formation and a compact fold. These data are in accordance with the majority of the biochemical observations that illustrate the significance of the close partnership for

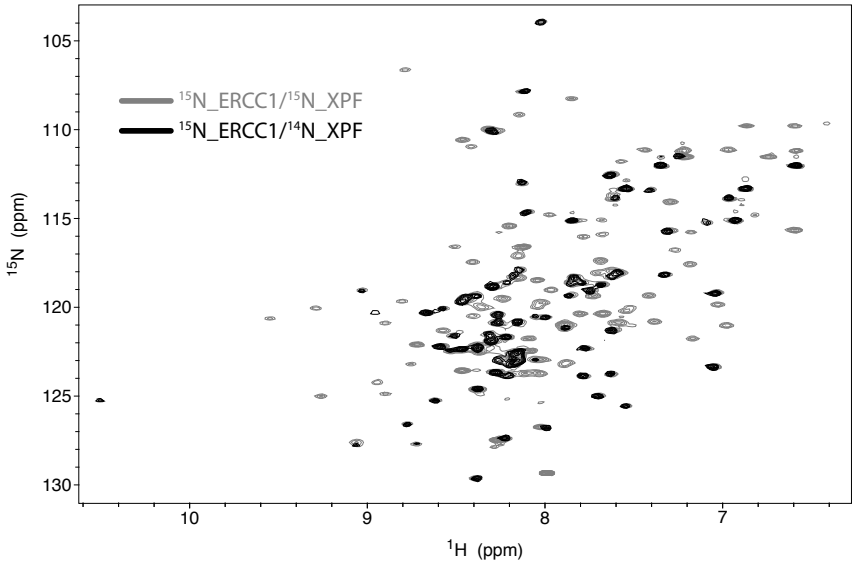


Figure 1. Refolding of the ERCC1/XPF complex. The figure shows the ^{15}N-^{1}H HSQC spectrum of the ERCC1/XPF complex (gray) and the ^{15}N-^{1}H HSQC spectrum of ^{15}N-labeled ERCC1 (black) refolded in the presence of unlabeled XPF. The overlay of the crosspeaks of ERCC1 in the two spectra confirms that the refolded complex forms the native conformation (for more details, see Material and Methods).

the smooth functioning of ERCC1 and XPF proteins (Sijbers et al., 1996a).

A vast amount of free XPF was present in the unbound fraction after Ni^{2+} chromatography. We also purified the free XPF fraction and recorded its ^{15}N-^{1}H HSQC, which showed a large dispersion, indicating a folded conformation (data not shown). Urea unfolding and refolding of the complex of the ERCC1 and XPF domains results in free XPF, as can be concluded from the ^{15}N-^{1}H HSQC spectrum, and a precipitate of the ERCC1 domain. Refolding into a heterodimeric ERCC1/XPF complex was possible when His_6-tagged ERCC1 was immobilized on a Ni^{2+} affinity column and XPF was circulated in excess over this column while gradually decreasing the urea concentration (Figure 1). From that we conclude that the ERCC1 domain can only fold in the presence of the XPF domain under our in vitro conditions.

Structure determination

The solution structure of the 19 kDa complex between the C-terminal domain of ERCC1 (residues 220-297) and the C-terminal domain of XPF (residues 823-905) was solved by heteronuclear double- and triple-resonance NMR spectroscopy by using uniformly ^{15}N- and $^{15}N/^{13}C$-labeled protein. The structure was determined on the basis of 4547 experimental NMR restraints and 197 dihedral angle restraints. A large number of intermolecular distance restraints were collected (442), which permitted the positioning of the two subunits with respect to each other. No distance violations larger than 0.4 Å and dihedral angle violations larger than 5° were found. A summary of the structural and restraints statistics is given in Table 1.

Table 1. Structural statistics of the structure ensemble of the heterodimeric ERCC1/XPF protein

Rmsd (Å) with respect to mean [a] (backbone/heavy)	
ERCC1	0.28 ± 0.06 / 0.80 ± 0.11
XPF	0.30 ± 0.07 / 0.78 ± 0.13
Complex	0.33 ± 0.06 / 0.81 ± 0.10
Number of experimental restraints	
Intraresidue NOEs	876
Sequential NOEs (|i - j| = 1)	1210
Medium range NOEs (1<|i - j|<4)	1252
Long-range NOEs (|i - j|>4)	767
Interprotein	442
Total NOEs	4547
Dihedral angle restrains	197
Hydrogen bonds	53
Restraint violations	
NOE distances with violations >0.4 Å	0.00 ± 0.00
Dihedrals with violations >5°	0.00 ± 0.00
Rmsd for experimental restraints	
All distance restraints (4547) (Å)	0.0186 ± 0.0008
Torsion angles (197) (°)	0.5555 ± 0.049
CNS energies after water refinement	
E_{vdw} (kcal/mol)	-604 ± 15
E_{elec} (kcal/mol)	-6482 ± 101
Rmsd from idealized covalent geometry	
Bonds (Å)	0.01 ± 0.00
Angles (°)	1.29 ± 0.03
Impropers (°)	1.39 ± 0.06
Ramachandran analysis	
Residues in the favored regions (%)	94.37 ± 0.93
Residues in additional allowed regions (%)	5.59 ± 0.97
Residues in generously allowed regions (%)	0.04 ± 0.20
Residues in disallowed regions (%)	0.00 ± 0.00

[a] Residues 227-294 of ERCC1 and residues 831-896 of XPF

Figure 2A shows an overlay of the 20 lowest energy conformers obtained after the structure calculations.

Structure description

The structure of the core of the complex, comprising residues 234 to 294 of ERCC1 and residues 836 to 895 of XPF, exhibits a global pseudosymmetry with a very similar architecture of the associated partners. The fold of both subunits is built up by helical secondary elements. In particular, the fold of ERCC1 is a characteristic example of the double helix-hairpin-helix $(HhH)_2$ motif (Shao and Grishin, 2000), which

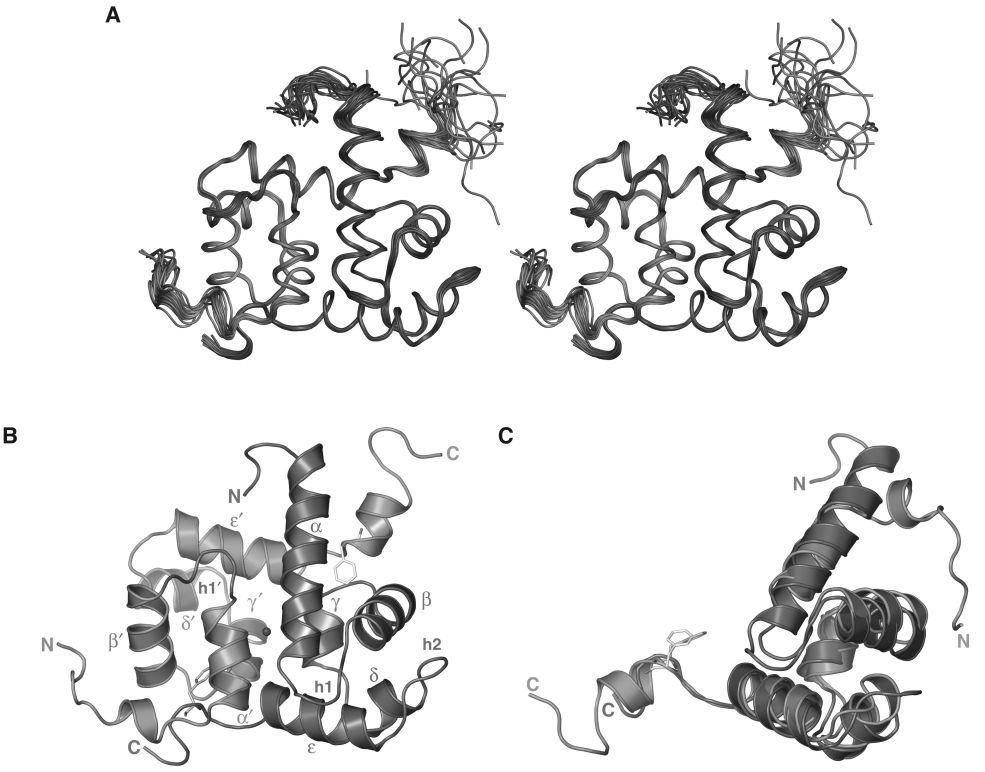


Figure 2. Three dimensional structure of the C-terminal domains of the human ERCC1/XPF complex. (A) Backbone stereoview of the ensemble of final 20 structural conformers. ERCC1 is colored blue, XPF is colored green, and the hairpins in both subunits are brown. (B) Cartoon representation of the lowest-energy model. The purple sphere depicts the centre of the pseudo-2-fold symmetry axis. Helices are denoted with Greek letters, from α to ε for ERCC1 and α′ to ε′ for XPF, while hairpins are indicated as h1, h2, and h1′. Also shown are Phe residues at the C termini of each subunit. Phe293 of ERCC1 is colored orange and Phe894 of XPF is colored yellow. (C) ERCC1 and XPF are superimposed, showing the overall fold similarity. Color conventions as in (B). (Full color on page 102)

is present in a variety of protein families involved in non-sequence-specific DNA binding such as DNA-polymerases, ligases, and nucleases. Each motif forms into a pair of antiparallel α-helices connected by a hairpin-like loop. The ERCC1 helices α and β form the first HhH element, and helices δ and ε the second one connected by a short helix, γ. The angle between the helical vectors on the first and second pair is 68° and 49°, respectively. On the other side, the XPF fold is almost the same, except for a minor, but significant, difference. Although the relative orientation of the XPF helices agrees well with the $(HhH)_2$ fold, only the first hairpin is present. It is formed by helices α′ and β′ with an angle of 51° (helices of XPF are indicated by a prime sign). The hairpin between helices δ′ and ε′ has been replaced by a β turn of only three residues. These helices are inclined at an angle of 55°. Although XPF retains only one complete HhH element, the global fold still resembles the $(HhH)_2$ structure (Figure 2B and 2C). The rmsd of the individual ERCC1 and XPF structures is 2.5 Å (65 Cα atoms) with Z-score of 8.1.

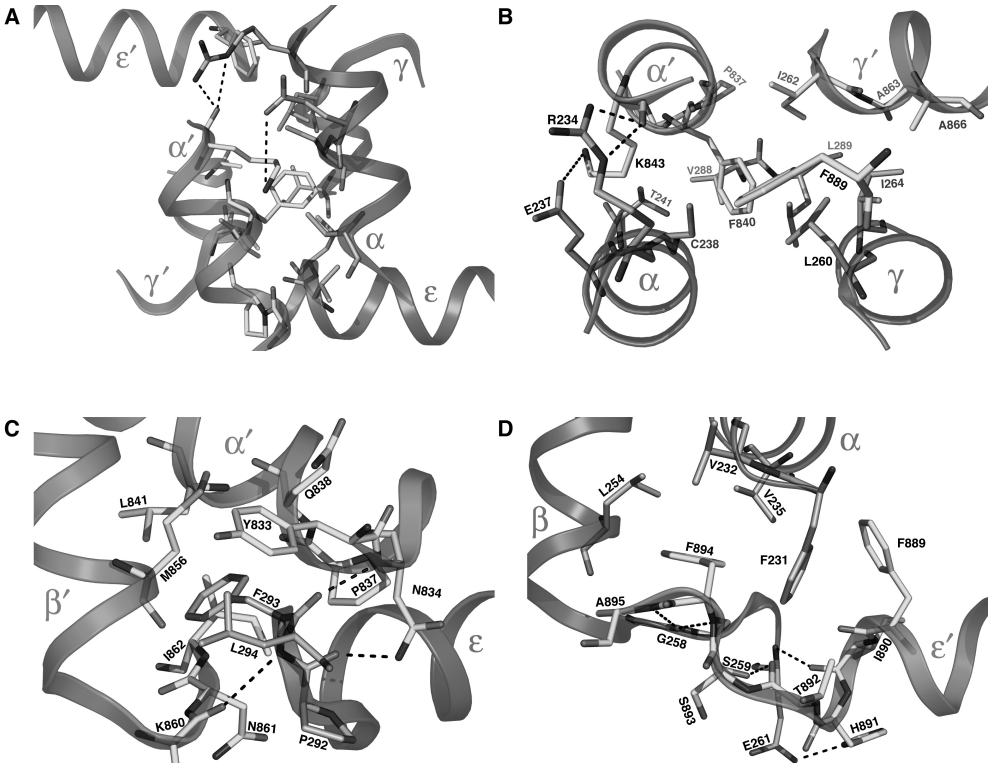


Figure 3. Protein-protein interactions along the C-terminal domains of ERCC1 and XPF. (A) View of the central interacting core. Helices α in the front, helices γ at the back, and the tips of helices ε that run perpendicularly build up the interacting surface. Residues participating in contacts are depicted in stick representation. The backbone of ERCC1 is shown in blue, XPF is shown in green, and interacting amino acids from either side in orange and yellow, respectively. Intermolecular hydrogen bonds are indicated with black dotted lines. (B) Top view of the central core. For the sake of clarity, helices ε have been removed. (C) The tight hydrophobic packing of ERCC1 Phe293 to its XPF cavity and (D) the same for Phe894 of XPF in the hydrophobic pocket of ERCC1. In (B), (C) and (D) amino acids are numbered according to the native sequences. (Full color on page 103)

Protein-protein interactions

The two domains of ERCC1 and XPF have a large interaction surface with an area of 1534 Å^2. The orientation of helices α and γ from both subunits resembles a four-helical bundle, although helices γ are very short. Only the helix γ′ of XPF deviates from the parallel orientation, and it is tilted by 35° with respect to its ERCC1 mate (Figure 3B). Both ERCC1 and XPF helical pairs form relatively flat and extremely hydrophobic surfaces that are shielded in the protein-protein interface. This central hydrophobic core seems to be the main contributor of the association, since it confers more than half of the interaction surface area. A key determinant of the nonbonded interactions is Phe840 of XPF, which protrudes in the middle of the symmetric helical array (Figure 3A and 3B). The aromatic ring of Phe840 makes contacts with Cys238, Thr241, Leu260, and Leu289 of ERCC1, which are located in helices α, γ, and ε. The rest of the intermolecular contacts are restricted to residues that belong to symmetry-related helices. Lys843 in helix

α′ of XPF interacts via its aliphatic side chain with Glu237, Cys238, and Thr241 of ERCC1 helix α. In half of the structures in the ensemble, the carboxyl of Lys843 of XPF forms bifurcated hydrogen bonds to the guanidinium group of ERCC1 Arg234. Moreover, the side chain amide group of the same Lys forms a salt bridge with the carboxyl side chain of Glu237 of ERCC1. Regarding the short γ helices, hydrophobic interactions are present between Ile264 of ERCC1 and Ala863 and Ala866 of XPF. The hydrophobic core further contains Phe889 of XPF and Val288 and Leu289 of ERCC1.

In previous studies it has been noted that the C-terminal residues of each subunit were essential for complex formation (de Laat et al., 1998b). The structure clarifies the important role of the two Phe residues. The ring of ERCC1 Phe293 fits perfectly to a hydrophobic cavity that is formed by the HhH motif of XPF and consists of residues Pro837, Gln838, Leu841, Met856, Asn861, and Ile862. The hydrophobic pocket has a large contact surface of 280 $Å^2$. In this region, the backbone of ERCC1 forms three hydrogen bonds with XPF that anchor the sidechain of Phe293 to the XPF cavity. The sidechains of Tyr833 from XPF and Leu294 from ERCC1 cover the entry of the cavity and lock the pocket, with Phe293 buried in it (Figure 3C). A similar pocket is built up by the first HhH motif of ERCC1, which accommodates, in this case the ring of XPF Phe894. The pocket consists of ERCC1 residues Phe231, Val232, Val235, and Leu254, and the contact surface is smaller than that of XPF and has an area of 220 $Å^2$. It is stabilized by five intermolecular hydrogen bonds from the XPF backbone (Figure 3D). Residues that form the interaction surface are conserved in all mammals, while generally conservative substitutions are found in other eukaryotes, thereby preserving the hydrophobic nature of the complex.

Helix-hairpin-helix motifs and DNA binding

A DALI search in the Protein Data Bank using the individual protomers as queries detected the same set of homologous structures for both ERCC1 and XPF. In all cases, ERCC1 showed better scores, because it exhibits the canonical $(HhH)_2$ motif. The $(HhH)_2$ domain of ERCC1 closely resembles the $(HhH)_2$ domain of the archaeal homodimeric XPF from *Aeropyrum pernix* (PDB code 2BGW; rmsd 1.9 Å) (Newman et al., 2005) and the holiday junction-recognizing $(HhH)_2$ domain of RuvA (PDB code 1C7Y; rmsd 1.9 Å) (Ariyoshi et al., 2000).We also noticed significant similarity to the C-terminal domain of the bacterial repair protein UvrC (PDB code 1KFT; rmsd 2.1 Å) (Singh et al., 2002), which exhibits the same double HhH organization (Figure 4B). We have shown before that the UvrC $(HhH)_2$ domain is required and sufficient for recognition of the single-to-double strand junction. To address whether the ERCC1/XPF HhH domains mediate DNA binding, we performed electrophoretic mobility shift assays.

Weak binding (K_d of 39 ± 13 μM) was detected on a stem-loop DNA substrate that was shown to be the optimal substrate for ERCC1/XPF in nuclease activity assays (de Laat et al., 1998a). We found that binding is specific for probes that contain single-double strand junctions, as complexes were also formed with a bubble and fork substrate, whereas no binding was detected on dsDNA or ssDNA

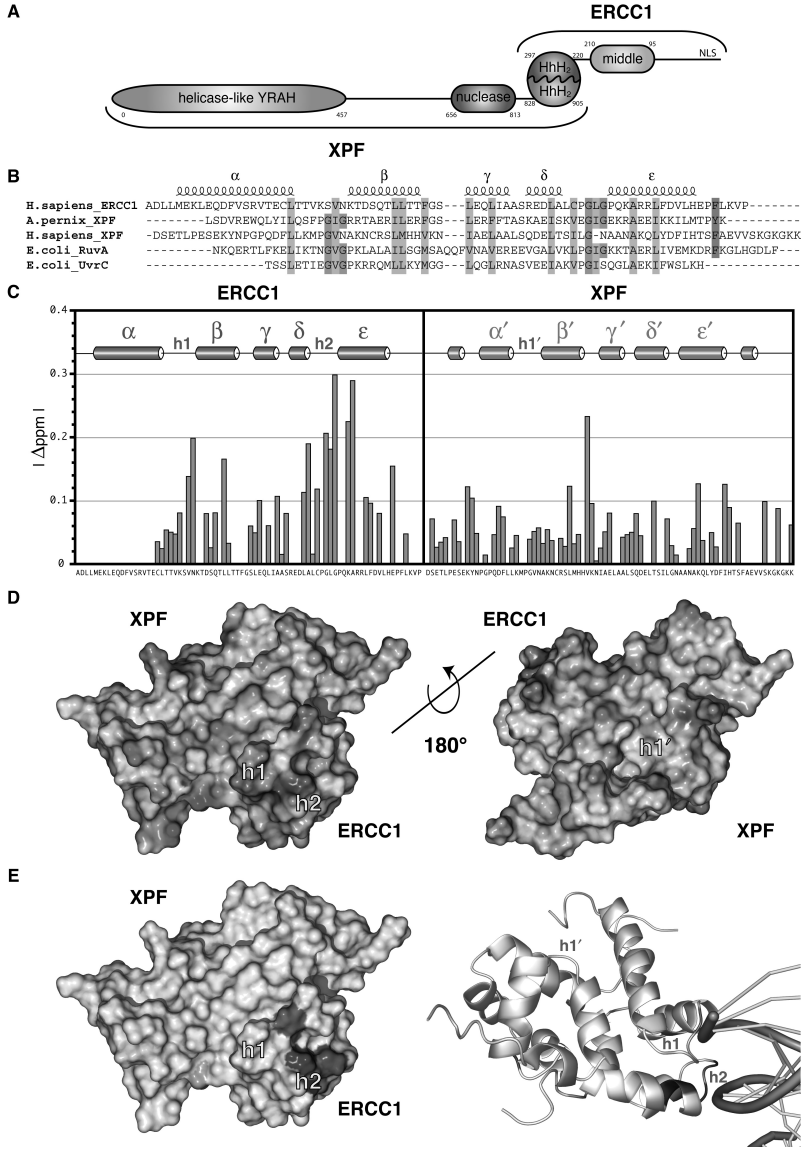


Figure 4. DNA binding by the ERCC1/XPF complex. (A) Domain organization of the ERCC1/XPF complex. (B) Sequence alignment of selected $(HhH)_2$ domains from the three kingdoms. Hydrophobic residues are highlighted in cyan, Gly that belong to hairpin motifs are highlighted in orange and the aromatic residues at the end of the domains are highlighted in green. The observed secondary elements of ERCC1 are indicated above the sequences. (C) Normalized chemical shift changes upon DNA titration versus the ERCC1 and XPF sequences. Missing bars indicate either Pro residues or unresolved chemical shifts due to peak overlap. (D) Two views of surface representation rotated by 180º and colored according to their electrostatic surface potential at ±8kB T/e for positive (blue) or negative (red) charge potential by using the program GRASP (Nicholls, 1993). (E) Observed chemical shift changes on the protein surface and model for the interaction of ERCC1/XPF with DNA by fitting ERCC1 on the archaeal $(HhH)_2$ domain. The blue parts denote shifting residues in both representations, and the intensity of the color corresponds to the absolute shift value (cut-off value of 0.1). (Full color on page 104)

probes (Supplementary Figure 1; page 46). Though weaker, the binding preference of the isolated $(HhH)_2$ interacting domains is similar to the observed in vitro incision preference of full-length recombinant ERCC1/XPF protein (de Laat et al., 1998a), which shows a K_d of 100 nM for the same stem-loop substrate (unpublished data). In this respect, it is likely that other domains of the native complex contribute to DNA binding as well. Furthermore, it has been shown before that recruitment of ERCC1/XPF in the context of NER is stabilized by various other proteins, including XPA-RPA proteins and the presence of TFIIH and XPG in the preincision complex (Riedl et al., 2003; Volker et al., 2001)

To determine the interaction surface for DNA binding, ^{1}H-^{15}N HSQC spectra of the protein were recorded with successive additions of a stem-loop DNA substrate. Although the chemical shift perturbations that we observed upon titration are relatively small, the most prominent effects are clustered in the second hairpin of ERCC1, with a few in its first hairpin (Figure 4C and Supplementary Figure 2; page 46). In particular, the maximum amide chemical shift perturbations were detected in Gly276 and Gly278 along with the intervening Leu277 forming the classical GhG hairpin. In the crystal structures of protein-DNA complexes containing these hairpin motifs, the glycines form hydrogen bonds with adjacent phosphate oxygens of the DNA, and a positively charged residue after the second Gly makes polar contacts with the phosphate backbone. Lys281 of ERCC1 is also shifted significantly, suggesting the involvement of this residue in a protein-DNA contact. The first hairpin of ERCC1 is more distinct in terms of sequence composition and lacks the GhG signature. Unfortunately, due to overlap, we were not able to follow the shifts of three critical residues of the first hairpin, Ser244, Lys247, and Thr248. However, the shifts at Val245 and Asn246 demonstrate that also this hairpin is in contact with DNA. There are also randomly distributed shifts both in ERCC1 and in the XPF side. They probably reflect indirect perturbation effects possibly involving small rearrangements in the hydrophobic packing induced by DNA binding. No shifts, however, are observed in the hairpin of XPF strongly suggesting that this subunit does not directly participate in DNA recognition under the present in vitro conditions. Furthermore, the positive charge mainly in the hairpins of ERCC1 agrees well with the ERCC1 residues found in the proximity of DNA in our experiments (Figure 4D and 4E).

Discussion

ERCC1 forms an obligate heterodimer with XPF

In this study, we show that the driving force of complex formation is mainly hydrophobic interaction, and we show that this yields a compact architecture that enhances the stability of each subunit. Both domains adopt similar structures in the final complex related by a pseudo-2-fold symmetry axis, with ERCC1 displaying the canonical $(HhH)_2$ fold and XPF displaying a very similar fold. The DNA binding domain from ERCC1 resembles the archaeal XPF more closely than the eukaryotic XPF does. The opposite is true for the flanking endonuclease domain, which is inactivated in ERCC1 protein (Figure 4A). In archaebacteria, ERCC1 is absent, while

in all sequenced eukaryotes, ranging from fungi, parasites, plants, and invertebrates to vertebrates, ERCC1 and XPF homologs have been found. Furthermore, the archaeal orthologous repair protein, like the human protein, forms a dimeric structure. The structural resemblances of human ERCC1 and archaeal XPF with regard to the $(HhH)_2$ domains agree with the suggested common origin for the eukaryotic ERCC1 and XPF proteins (Gaillard and Wood, 2001).

The requirement for the formation of a stable ERCC1 through heterodimer association reflects the importance of the biological function of ERCC1. We have demonstrated that the double hairpin motif on the ERCC1 side mediates DNA binding to a stem-loop substrate, an equivalent of the NER bubble, which is the most preferred substrate of eukaryotic ERCC1/XPF. Although in vivo the folding process is probably more complicated and could involve chaperone-mediated posttranslational interactions with XPF, our in vitro biochemical data support the finding that ERCC1 functionality is strictly dependent on the XPF domain as a folding scaffold. While this manuscript was under preparation, a biophysical characterization of the same interacting domains was reported (Choi et al., 2005). Although in this study ERCC1 and XPF C-terminal domains were expressed separately, stabilization of ERCC1 was achieved only in the presence of XPF, in full accordance with our refolding scheme.

This finding, together with the high resolution structure presented here, is related to two studies that have reported sensitivity of the native complex to short C-terminal ERCC1 truncations and highlighted a role for Phe293 in stability and function (de Laat et al., 1998b; Sijbers et al., 1996b). The structure of the complex suggests that locking of this residue to its partner pocket links directly to the formation of the second hairpin motif of ERCC1 and indirectly to DNA recognition capability and function. Evolutionary conservation of Phe293 in all ERCC1 family members underscores its importance. Deletion of the Phe "hook" leads to loss of functionality, presumably because a large interacting area is abolished (Sijbers et al., 1996b).

DNA binding is mediated by the HhH domain of ERCC1

The observed small chemical shift changes in the core of the ERCC1/XPF complex upon addition of DNA (Figure 4) suggest that small, local conformational changes are required to accommodate the DNA substrate for the human ERCC1/XPF heterodimer. In this respect, we note that the $(HhH)_2$ motif of ERCC1 resembles more closely the corresponding domain of the archaeal homodimeric XPF bound to DNA than that of the free archaeal XPF structure (Newman et al., 2005). It is known that HhH domains recognize various substrates, but what really determines the specificity remains largely unclear. The crystal structures of the archaeal XPF homodimer and bacterial RuvA bound to DNA have revealed a similar pattern of contacts between GhG hairpins and the minor groove of DNA. Also, in bacterial UvrC (Singh et al., 2002) similar or identical residues that form the double HhH motif were found to be crucial for DNA binding. We observed the same residue contacts for the second hairpin of ERCC1 in our chemical shift perturbation experiments. Even on the first hairpin, which is less

homologous with the classic GhG hairpins, the perturbed residues agree very well with the crystallographic complexes. Sequence analysis of ERCC1 and XPF members revealed that residues critically involved in DNA binding in our NMR data are strictly conserved in ERCC1 family, but are mostly absent in XPF. Remarkably, most of these ERCC1 residues are present and conserved in archaeal XPF and are shown to be in contact with DNA (Newman et al., 2005).

Our observation that only the $(HhH)_2$ domain of ERCC1 binds to DNA points to an interesting difference in topology between human ERCC1/XPF and the archaeal homodimeric XPF proteins. The crystal structure of the archaeal XPF suggests that $(HhH)_2$ and nuclease domain of the same protomer are involved in DNA contacts. In the human case it has been shown that only XPF carries a nuclease domain preceding its C-terminal domain (Enzlin and Scharer, 2002). Thus for ERCC1/XPF it appears that DNA binding is located in one subunit (ERCC1) and the nuclease activity in the other subunit (XPF).

Functional implications of the intimate association between the HhH domains of ERCC1 and XPF

ERCC1/XPF is the last factor arriving at the DNA damage site, resulting in the mature preincision NER complex. It has been shown that the binding of ERCC1/XPF depends on a multitude of sequential prerequisites, including the TFIIH DNA unwinding step (Mone et al., 2004), the presence of XPG at the opposite cleavage site, and the binding of XPA-RPA (Riedl et al., 2003; Volker et al., 2001). Indeed, the middle domain of ERCC1 can interact specifically with XPA (Bessho et al., 1997; Li et al., 1995). Furthermore, we have shown here that the ERCC1 $(HhH)_2$ domain is sufficient to mediate DNA binding (Figure 4). We propose that ERCC1 combines DNA binding activity and specific interactions with XPA to recruit the endonuclease activity of XPF to the incision complex. This guarantees that XPF cleavage is restricted only to correctly processed DNA templates. This model for ERCC1/XPF is in full accordance with the assembly of NER factors at sites of DNA damage in a sequential and interdependent manner. Our interpretation emphasizes the functional distinction of the tightly associated partners with ERCC1 making all the contacts with DNA and XPA that will target the nuclease of XPF to perform the strand cleavage 5´ upstream of the lesion.

Material and Methods

Protein expression and purification

Earlier experiments revealed that ERCC1 was expressed in inclusion bodies and that resolubilization of ERCC1 with urea and refolding did not result in a properly folded protein. In addition, we have mapped the interaction domains between ERCC1 and XPF to residues 224-297 and 814-905 for ERCC1 and XPF, respectively (de Laat et al., 1998b). To obtain a stable heterodimeric complex with minimal flanking sequences, putative domain boundaries were determined using bioinformatics tools as described previously (Folkers et al., 2004). For both domains, several N-terminal deletion constructs around the previously identified domain (ERCC1: 220, 224,234; XPF: 813, 823, 832) were tested by using in vitro interaction assays (de Laat et al., 1998b). From these interaction experiments, we concluded that the minimal interaction

domain is composed of 220-297 and 823-905 for ERCC1 and XPF, respectively. Therefore, we constructed a bicistronic expression vector in which we first cloned the XPF sequence 823-905 (all oligonucleotide sequences used are available upon request) in the NcoI-BamH1 site of pET28b. Subsequently, in the BamH1-XhoI site of this XPF expression construct, the ERCC1 fragment 219-297, extended with a 27 base pair linker containing an optimal internal ribosome binding site, was cloned, resulting in a bicistronic expression construct where only the ERCC1 domain contains the His_6 tag.

Overexpression and His tag purification of ^{15}N and $^{15}N/^{13}C$ ERCC1/XPF complexes was performed as described before (Folkers et al., 2004), except that elution was performed with 400 mM EDTA. After elution, dialysis against NMR buffer (50 mM $NaPO_4$, 100 mM NaCl [pH 7.0]) was performed using a 3K cutoff dialysis tube (SpectraPor). Subsequently, the complex was loaded on a Superdex 75 column in the same buffer. The ERCC1/XPF complex eluted from this column as a single species at an elution time as expected for a heterodimer of this size. Even when large amounts of the complex were loaded on the column, no monomeric ERCC1 or XPF was detected, and, together with the observed symmetric elution profile for the ERCC1/XPF complex, we conclude that ERCC1 and XPF are in tight association. After concentration of this fraction to 0.5-1.5 mM using ultrafiltration with a 3K filter (amicon; Milipore), protease inhibitor cocktail (Complete EDTA free; Roche) was added.

For on-column refolding experiments, the ERCC1/XPF complex (^{15}N labeled) was first loaded on the nickel-charged metal chelating column of 7.8 ml (Poros MC20; Applied Biosystems) under native conditions (50 mM $NaPO_4$, 150 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol, 0.2 mM PMSF [pH 8.0]) by using a BioCad Vision (Applied Biosystems). Subsequently, the XPF moiety was eluted from the column by using a denaturing buffer (native buffer containing 6 M urea). The 20 ml of eluted XPF fraction (from a nonlabeled elution) was reloaded with a flow rate of 5ml/min by continuous recycling of this XPF- containing fraction on the column while simultaneously mixing the eluate with an increasing amount of native buffer. In this procedure, the urea was gradually decreased from 6 M to a final 200 mM concentration over a total volume of 600 ml. After washing with native buffer, the ERCC1/XPF complex was eluted with 400 mM of EDTA and purified as described above.

DNA binding

Electrophoretic mobility shift assays were performed in a binding buffer containing 50mM $NaPO_4$, 100 mM NaCl, 20 mg/ml BSA and 10% (v/v) glycerol, by using radio-labeled stem-loop, bubble, fork, dsDNA (10bp), and ssDNA (20 mer) probes, as described before (Singh et al., 2002). Unless otherwise indicated, for all experiments, 50 μM ERCC1/XPF was used. For supershift, 0.5 ml of HIS-probeHRP conjugate (Pierce) was added to the reaction mixture prior to addition of probe. MagneHis (Promega) was added to the reaction mixture and cleared by using a magnetic stand, either prior to the addition of DNA or after complex formation, yielding identical results.

For DNA titrations we used 0.125 mM

of ^{15}N-labeled ERCC1/XPF complex in NMR buffer and added increasing amounts of a hairpin sequence in an identical buffer, with 22 unpaired bases (5′-GCCAGCGC TCGGTTTTTTTTTTTTTTTTTTTT TTCCGAGCGCTGGC-3′) until a final concentration of 0.25 mM was reached. ^{15}N-^{1}H HSQC spectra were recorded for the different titration points. Normalized chemical shift changes (δ) for the non-overlapping residues were calculated using the equation: $\delta = [(5\delta_{HN})^2 + (\delta_N)^2]^{0.5}$ (Grzesiek et al., 1996).

NMR Spectroscopy

All NMR spectra were recorded on Bruker AVANCE 700 and AVANCE 900 spectrometers equipped with triple-resonance gradient probes at 295.5 K, using [U-^{13}C, ^{15}N]- and [U-^{15}N]-labeled proteins. Backbone and side-chain ^{1}H, ^{15}N, and ^{13}C resonances for ERCC1/XPF were assigned using the following set of 3D triple-resonance experiments: HNCACB, CBCA(CO)NH, HN(CA)HA, HBHA(CBCACO)NH, HNCO, HN(CA)CO, C(CO)NH-TOCSY, H(CCO)NH-TOCSY, H(C)CH-TOCSY, and (H)CCH-TOCSY spectra acquired at 700 MHz. A full set of 3D heteronuclear-edited NOE spectra were then recorded at 900 MHz for structure determination: 3D NOESY-(^{13}C,^{1}H)-HSQC (τ_{mix}=75ms; tuned once for aliphatic and once for aromatic ^{13}C), 3D NOESY-(^{15}N,^{1}H)-HSQC (τ_{mix}=80ms), 3D (^{13}C)-HMQC-NOESY-(^{13}C,^{1}H)-HSQC (τ_{mix}=75ms), and 3D (^{13}C)-HMQC-NOESY-(^{15}N-^{1}H)-HSQC (τ_{mix}=80ms, tuned for aliphatic ^{13}C). These were complemented by homonuclear 2D NOE spectra recorded without (τ_{mix}=80ms) and with H[^{15}N] suppression in F2 (τ_{mix}=80ms). The triple-resonance and heteronuclear 3D experiments were performed essentially as described by (Cavanagh et al., 1996). All spectra were processed using the NMRPipe software package (Delaglio et al., 1995) and analyzed with Sparky (Goddard and Kneller, 2001).

Structure calculations

Automatic NOE assignment and structure calculations were performed using the CANDID module of the program CYANA (Güntert et al., 1997; Herrmann et al., 2002). The quality of the structures was improved in an iterative procedure in which CANDID runs were followed by manual inspection of the preliminary structures to find additional resonance assignments from the NOE spectra. Hydrogen bond restraints were defined when they were consistent with the secondary shift data and expected NOE contacts. Manual NOE peak assignments were generally not held constant in the CANDID runs, but used to create accurate spectrum-specific chemical shift lists and to check the consistency of subsequent CANDID runs. The final CANDID run was performed with CYANA version 2.0, using Ramachandran and sidechain rotamer restraints for every cycle except for the last one. In the final cycle, fixed stereospecific assignments for prochiral groups were used if available. Finally, the set of NOE based restraints determined by CANDID, together with restraints for 53 H-bonds and 220 ϕ and ψ torsion angle restraints derived from TALOS (Cornilescu et al., 1999) were used in a water refinement run using CNS (Brünger et al., 1998), according to the standard RECOORD protocol (Nederveen et al., 2005). The final structures were validated with WHATIF (Vriend, 1990) and PROCHECK (Laskowski et al., 1993;

Morris et al., 1992). Molecular images were generated with PyMol/NUCCYL (DeLano, 2002).

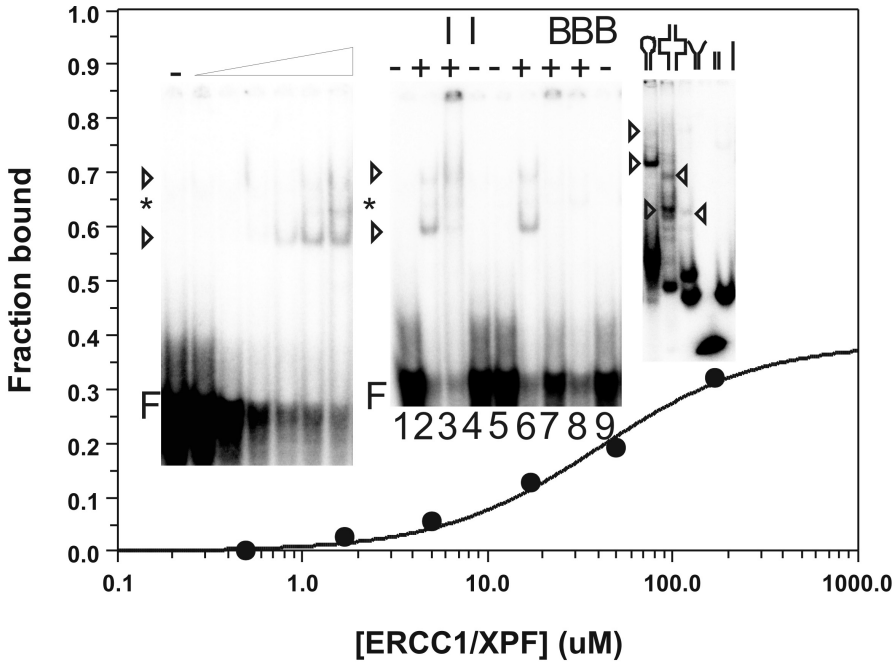
Accession Numbers

Coordinates have been deposited in the RCSB Protein Data Bank with accession code 1Z00, and the chemical shifts have been deposited in the BMRB with accession number 6551.

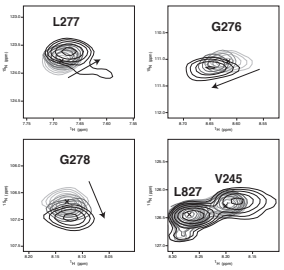
Acknowledgements

The authors are grateful to Dr. Tammo Diercks for expert NMR assistance. This work was financially supported by the Research Council for the Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW) and by the Center for Biomedical Genetics. H.O., N.G.J.J., and J.H.J.H. were supported by Euratom grants nrs FIGH-CT-2002-00207 and LSHG-CT-2005-512113.

Supllementary information



Supplementary Figure 1. DNA binding to single-double strand junction DNA probes by the $(HhH)_2$ domains of ERCC1 and XPF. The electrophoretic mobility shift assay was performed with the indicated amount of ERCC1-XPF C-terminal domains (μM), using a stem-loop substrate. For a representative experiment the fraction bound at indicated concentration is plotted, note that no binding saturation is observed (maximum fractional occupancy ~0.4), most probably caused by dissociation during electrophoresis. The left inset panel shows an autoradiogram of this binding experiment using 0.5, 1.7, 5, 17, 50, 170 μM ERCC1-XPF $(HhH)_2$ domains. The middle inset panel, lanes 1-4, show a supershift on a stem-loop substrate using HIS-probeHRP conjugate (I) in the presence (+) or absence (-) of 50 μM ERCC1/XPF. Lanes 5-9 of the middle panel show that MagnaHIS beads (B) mediated ERCC1-XPF depletion (ERCC1 contains a C-terminal his-tag) prevents complex formation on a stem-loop substrate when cleared either prior to (lane 7) or after complex formation (lane 8) in the presence (+) of 50 μM ERCC1/XPF. The right panel shows complex formation on stem-loop (20), a bubble and a fork both with 10 unpaired bases, dsDNA (10 bp) and ssDNA (20 bases). All probes have been described before (Singh et al., 2002). * Refers to a non-specific complex as it is not supershifted by the HIS-probeHRP conjugate (lane 3) and not depleted by incubation with MagnaHIS beads (lane 7, 8), the triangles refer to specific complexes, free probe is indicated with F.



Supplementary Figure 2. NMR titration study of DNA binding by the ERCC1/XPF complex. The indicated shifts belong to hairpins of ERCC1. From light gray color to black color DNA-protein ratios are 0, 0.325, 0.9 and 2.

Chapter three

Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization

Konstantinos Tripsianes, Gert E. Folkers, Chao Zheng, Devashish Das, Jeffrey S. Grinstead, Robert Kaptein and Rolf Boelens

Department of NMR spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Nucleic Acids Research, **35**, 5789-5798, (2007)

Human ERCC1/XPF is a structure-specific endonuclease involved in multiple DNA repair pathways. We present the solution structure of the non-catalytic ERCC1 central domain. Although this domain shows structural homology with the catalytically active XPF nuclease domain, functional investigation reveals a completely distinct function for the ERCC1 central domain by performing interactions with both XPA and single-stranded DNA. These interactions are non-competitive and can occur simultaneously through distinct interaction surfaces. Interestingly, the XPA binding by ERCC1 and the catalytic function of XPF are dependent on a structurally homologous region of the two proteins. Although these regions are strictly conserved in each protein family, amino acid composition and surface characteristics are distinct. We discuss the possibility that after XPF gene duplication, the redundant ERCC1 central domain acquired novel functions, thereby increasing the fidelity of eukaryotic DNA repair.

Introduction

During DNA replication, recombination and repair, double-stranded DNA inevitably forms three- or four-way junctions, bubbles, flaps or broken ends with single-stranded extensions. These irregular structures must be processed correctly in order to successfully complete DNA metabolism and thereby maintain genome integrity. This task is accomplished by structure-specific endonucleases specialized in pruning downstream of branch, flap or bubble structures by incision at junctions between double- and single-stranded DNA (Nishino and Morikawa, 2002). Inactivation or malfunctioning of these enzymes causes genetic defects or cancer, underlying their importance in genome stability.

A remarkable class of structure-specific endonucleases in humans is XPF. The protein family is characterized by the presence of the ERCC4 domain and consists of seven members (XPF, MUS81, ERCC1, EME1, EME2, FANCM and FAAP24) (Ciccia et al., 2007). Only XPF and MUS81 have nuclease activity, which is mediated by the conserved core nuclease motif (ERKX3D) (Enzlin and Scharer, 2002; Heyer et al., 2003; Nishino et al., 2003). Their catalytic function depends on heterodimer formation with the non-catalytic family members. XPF forms an obligate complex with ERCC1 and functions primarily in nucleotide excision repair (NER), a versatile pathway able to detect and remove a variety of DNA lesions induced by UV light and environmental carcinogens. The ERCC1/XPF heterodimer has additional roles in DNA interstrand cross-link (ICL) repair (Niedernhofer et al., 2004) and telomere maintenance (Zhu et al., 2003). The symptoms of the first patient with inherited ERCC1 deficiency (Jaspers et al., 2007) and of a patient with a novel XPF mutation (Niedernhofer et al., 2006) are distinct from the classical NER phenotype, and underscore the pleiotropic function of ERCC1/XPF.

In contrast to eukaryotes, archaea have a single homolog of the XPF endonuclease that forms homodimers. The archaeal XPF minimally consists of the catalytic nuclease domain followed by a DNA-binding domain containing two consecutive helix-hairpin-helix motifs (HhH_2 domain). Dimerization occurs between both the nuclease and

HhH_2 domains of each subunit (Figure 1A) (Nishino et al., 2003). The strong preference of the archaeal endonucleases for 3′ flap DNA substrates is explained by the function of their individual domains. Structural data suggest a model for DNA binding where one HhH_2 domain of the dimer binds to an upstream and the other to a downstream DNA duplex, sharply bending the DNA substrate and thereby allowing one active site of the dimeric nuclease domain to cleave the 3′ protruding single strand (Newman et al., 2005). Combination of mutations in nuclease and HhH_2 domains that do not disrupt the dimeric interfaces but impair the individual functions (DNA cleavage or DNA binding), provide evidence for this model (Nishino et al., 2005).

The DNA substrate specificity is different for the human ERCC1/XPF heterodimer. The human enzyme is known to specifically incise hairpin, bubble or splayed-arm DNA substrates (de Laat et al., 1998; Tsodikov et al., 2005), which consist of only one duplex (upstream). Previous data indicated the role of the C-terminal ERCC1 'canonical' HhH_2 domain in engaging the upstream DNA duplex of a hairpin substrate (Tripsianes et al., 2005). This function requires heterodimerization with the C-terminal HhH domain of the human XPF partner to form a complex analogous to the archaeal HhH_2 dimer interface (Tripsianes et al., 2005; Tsodikov et al., 2005). Previous structural data (Tsodikov et al., 2005) and the data presented here illustrate the structural similarity between the human ERCC1 central domain and the archaeal XPF nuclease domain. Unlike the archaeal nuclease dimer interface, there is no evidence that the corresponding domains of human ERCC1 and XPF interact with each other in solution (Figure 1A) (Tsodikov et al., 2005).

The catalytic function of the ERCC1/XPF endonuclease is crucial for NER. NER operates through a 'cut and patch' mechanism by excising and removing a short stretch of DNA containing the lesion, and subsequently restoring the genetic information by repair synthesis using the undamaged strand as the template. The incision step involves the sequential and coordinated assembly of the DNA damage sensor XPC-HR23B, the transcription/repair factor TFIIH, the architectural protein XPA, the ubiquitous ssDNA-binding protein RPA, and the two structure-specific endonucleases ERCC1/XPF and XPG, responsible for the incisions 5′ and 3′ to the damaged site, respectively (Gillet and Scharer, 2006). A main role in the progress of the reaction is attributed to XPA, which in complex with RPA probes the DNA helix conformations (Missura et al., 2001) and participates in multiple interactions with the other NER factors (Gillet and Scharer, 2006). XPA is required for the recruitment of ERCC1/XPF in the NER pre-incision complex (Riedl et al., 2003; Volker et al., 2001). Consistent with this regulatory function, *in vitro* studies with recombinant proteins and *in vivo* studies using a yeast two-hybrid system have demonstrated interactions between XPA and ERCC1 (Li et al., 1995).

Here, we report the solution structure of the ERCC1 central domain (cERCC1) and investigate its interactions with XPA and DNA at the molecular level. Using biochemical studies and NMR titration experiments, we have identified two distinct interaction surfaces of cERCC1 that mediate XPA and DNA binding.

Interestingly, the two interactions can take place simultaneously to yield a cERCC1/DNA/XPA ternary complex, which in turn explains the important role of ERCC1 in targeting its catalytic XPF partner to the NER pre-incision complex.

Results and Discussion

Structure analysis

We have determined the solution structure of the ERCC1 central domain (cERCC1) to elucidate the molecular details of its proposed functions. We have collected a large number of distance and dihedral angle restraints that yielded an ensemble of 20 conformers with very good convergence (Figure 1C). The quality of the structure can be judged by the summary of the structural and restraint statistics given in Table 1. All the secondary structure elements are well defined including the short loops that join them. cERCC1 comprises a compact architecture and folds as a six-stranded

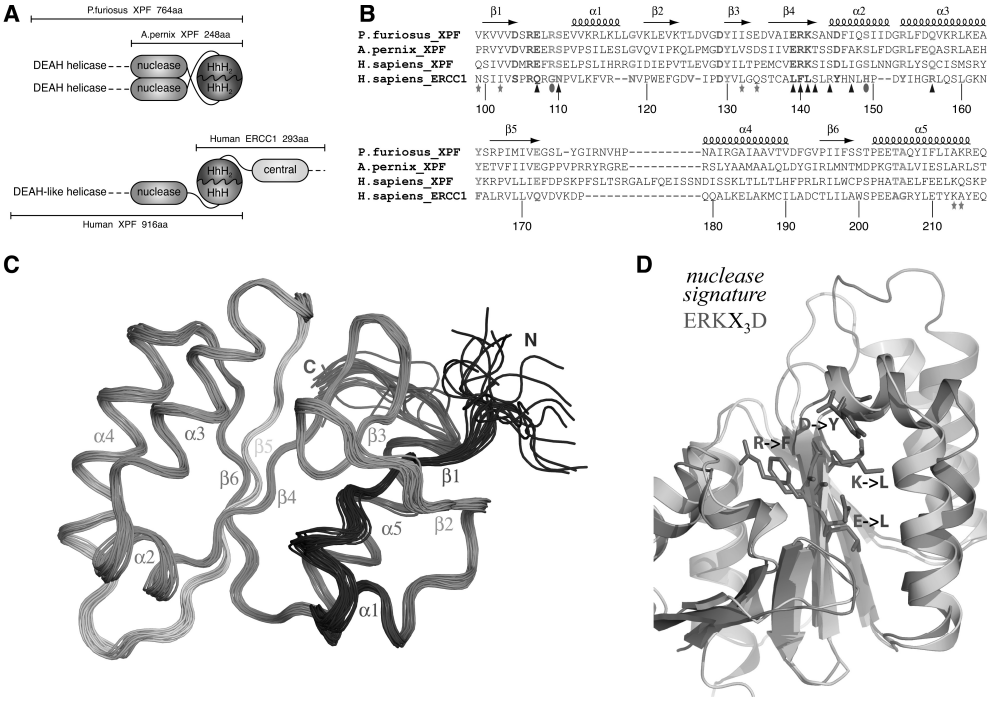


Figure 1. (A) Domain organization of the archaeal XPF homodimeric members and the human ERCC1/XPF heterodimer. (B) Structure-based sequence alignment of the nuclease XPF domains from archaea to human and the corresponding central domain of human ERCC1. Secondary structure elements of the prototype XPF nuclease fold are indicated at the top of the sequences. Catalytic residues in the XPF nucleases are colored red and their corresponding substitutions in ERCC1 blue. Other invariant residues in XPF domains are depicted in orange, while the ERCC1 equivalents are depicted in cyan. Residues of cERCC1 perturbed largely upon XPA titration are indicated by blue triangles and those appear only in the final complex by brown ellipses. Green asterisks indicate cERCC1 residues perturbed by DNA titration. cERCC1 sequence is numbered at the bottom. (C) Ensemble of the final 20 structural conformers of cERCC1 as determined by solution NMR. Secondary structure elements and N- and C-termini are labeled. (D) Superposition of the crystal XPF nuclease structure (2bgw) from *A. pernix* (purple) and the solution NMR structure (2jpd) of human cERCC1 (yellow). Emphasis is given to the nuclease signature and the corresponding substitutions in cERCC1. (Full color on page 105)

β-sheet flanked by 5 α-helices on both sides. As described before (Tsodikov et al., 2005), this fold is reminiscent of the type II restriction endonucleases, to which the catalytic domain of XPF also belongs.

Overall, the solution NMR structure of cERCC1 is in very good agreement with the crystal structure of the same domain (2a1i) (rmsd 1.1 Å for 108 Cα atoms). The only noticeable difference relates to the last few residues of helix α3 and the loop connecting this structure element with β5. The presence of a mercury atom (linked to C137) in the crystal structure may account for the substantial side-chain rearrangements within this loop. When compared to the crystal structures of the archaeal nucleases, helices α1, α2, and strands β1, β2 appear to be shorter in the solution structure of cERCC1. Solution cERCC1 and the crystal structures of archaeal XPF nucleases (2bgw and 1j23) contain the same number of secondary structure elements arranged in a very similar manner (rmsd 1.8 and 2.0 Å for *Aeropyrum pernix* and *Pyrococcus furiosus* nucleases, respectively, for 108 Cα atoms), although in both cases the primary sequence homology is very low (Figure 1B and 1D).

Remarkably, conserved residues of XPF nuclease scattered in the primary sequence superimpose structurally with residues conserved in the ERCC1 sequence family (Figure 1 and 3). Whereas the conservation in XPF is directly related to the catalytic function, ERCC1 has preserved the same fold but lacks the essential residues for catalysis. The fold similarities, coupled with the obligate nature of the heterodimerization (Sijbers et al., 1996), are in full agreement with the common origin of the two proteins (Gaillard and Wood, 2001).

Table 1. Structural statistics of the structure ensemble of the human cERCC1 (residues 96-219)

Rmsd (Å) with respect to mean [a]	
Backbone / Heavy atoms	0.40 ± 0.04 / 0.89 ± 0.08
Number of experimental restraints	
Intraresidue NOEs	632
Sequential NOEs (li - jl = 1)	678
Medium range NOEs (1<li - jl<4)	479
Long-range NOEs (li - jl>4)	880
Total NOEs	2669
Dihedral angle restrains	174
Hydrogen bonds	31
Restraint violations	
NOE distances with violations >0.3 Å	0.00 ± 0.00
Dihedrals with violations >3°	0.00 ± 0.00
Rmsd for experimental restraints	
All distance restraints (2669) (Å)	0.013 ± 0.005
Torsion angles (174) (°)	0.579 ± 0.077
CNS energies after water refinement	
E_{vdw} (kcal/mol)	-614 ± 17
E_{elec} (kcal/mol)	-4914 ± 61
Rmsd from idealized covalent geometry	
Bonds (Å)	0.01 ± 0.00
Angles (°)	1.15 ± 0.03
Impropers (°)	1.34 ± 0.06
Ramachandran analysis	
Residues in the favored regions (%)	90.68 ± 2.04
Residues in additional allowed regions (%)	8.16 ± 1.71
Residues in generously allowed regions (%)	0.81 ± 0.65
Residues in disallowed regions (%)	0.36 ± 0.68

[a] Residues 101-215 of cERCC1

Minimal domains for ERCC1-XPA interaction

Truncation studies of ERCC1 have mapped the XPA interaction site to a region between residues 91 and 118 (Li et al., 1994). From the cERCC1 structure and its compact fold, we predict that such truncations will have a devastating effect on the structural integrity of the central domain (96-219). For XPA, the ERCC1 interaction region seems to be located in a small stretch containing two highly conserved motifs rich in glycine and glutamic acid residues (72-GGGFILEEEEEEE-84, conserved residues are underlined) (Li et al., 1995). We have prepared a HIS-GST-XPA construct

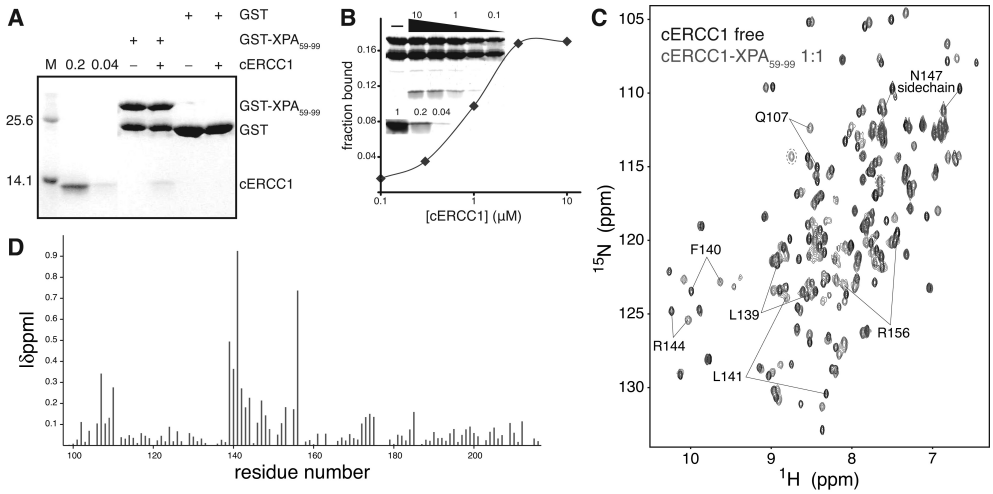


Figure 2. ERCC1-XPA interactions. (A) GST pull-down assay with 3 μg of GST or GST-XPA fusion proteins in the presence or absence of 2 μM cERCC1. Here, 0.2 and 0.04 refer to respectively 20 and 4% of the input cERCC1 protein present in the GST pull-down assay. (B) Semi quantitative GST pull-down assay showing the fraction of cERCC1 bound to 3 μg of GST-XPA at the indicated [cERCC1] (μM). The upper part of inset shows a representative GST pull-down assay, where the various cERCC1 concentrations used are depicted above (10, 3, 1, 0.3 and 0.1 μM). The lower panel shows respectively 100, 20 and 4% of cERCC1 added to the assay. (C) Chemical shift perturbation of the cERCC1 ^{1}H-^{15}N HSQC upon complex formation with XPA. Free cERCC1 spectrum is in black and XPA-bound spectrum in blue, while brown circles in the bound spectrum indicate G109 and H149 resonances. (D) Normalized chemical shift changes between free and XPA-bound forms versus the cERCC1 sequence. The p.p.m. difference for G109 and H149 was calculated by their resonances in the free cERCC1 spectrum at pH 5.5, and are colored brown. (Full color on page 106)

(59-99) containing the conserved stretch, and examined its ability to interact with cERCC1 in a GST pull-down assay. Indeed, the GST XPA fusion protein was able to specifically bind cERCC1 (Figure 2A), confirming that this domain of ERCC1 is sufficient for interactions with XPA. No binding of cERCC1 to GST bound agarose beads or uncharged agarose beads was observed (Figure 2A, data not shown). This XPA peptide is unstructured (Buchko et al., 2001) and appeared very sensitive to proteolysis both during overexpression in *Escherichia coli* and as pure protein after extensive purification (Figure 2A and B). We next performed a semi-quantitative GST pull-down assay (Jonker et al., 2005) and estimated an apparent K_d of 1 μM. Since we reached only 20% binding saturation at the two highest protein concentrations (3 and 10 μM of cERCC1) under these experimental conditions, we could not determine the binding constant more accurately (Figure 2B).

Functional disparity of the XPF nuclease and ERCC1 central domains

The cERCC1-XPA interaction specificity has been explored using NMR titration experiments. Titration with the HIS-XPA peptide causes extensive specific changes in the cERCC1 ^{1}H-^{15}N HSQC NMR spectra. The titrations show slow exchange regime with respect to the NMR chemical shift timescale, in agreement with the determined affinity (Figure 2C). We have been able to assign all the amide resonances of cERCC1 in the bound form (1:1 complex

stoichiometry) by using 3D NOESY-(^{15}N, ^{1}H)-HSQC spectra. The pattern of the backbone amide NOEs does not change substantially compared to the free cERCC1 protein. This indicates that the overall structure is maintained, and led us to conclude that the observed chemical shift perturbations arise from specific contacts with the XPA peptide.

The most significant perturbations map to the surface of the V-shaped groove of the cERCC1 structure and involve many residues that compose the corresponding DNA cleavage site of the archaeal XPF endonucleases (Figures 2D and 3). Interestingly, among the largest chemical shifts of cERCC1 upon XPA binding are those of L139, F140 and L141. These ERCC1 residues correspond to the characteristic strongly conserved ERK catalytic triad of the XPF nuclease domain (Figure 1). The rest of the significantly perturbed residues are polar or positively charged (Q107, N110, S142, R144, N147 and R156), are competent for hydrogen bonding with the XPA peptide and are more conserved than the surrounding regions of the cERCC1 V-shaped groove (Figure 3). Additionally, of special interest are the amide signals of G109 and H149, which are not observed in the free protein spectrum due to fast exchange with the solvent at pH 8 (these amide resonances are observable at pH 5.5). However, both give sharp signals at the end of the titration, an indication that the solvent exchange in the free protein is quenched in the complex. These residues reside at the rim of the V-shaped binding site and may be involved in hydrogen bonding with the XPA peptide (Figure 3).

Although ERCC1 residues forming the XPA-binding site are different from the acidic and basic residues which form the active site in XPF (Enzlin and Scharer, 2002), there are no significant changes in the backbone fold and the spatial side-chain orientations in the structure remain unaffected. However, their different

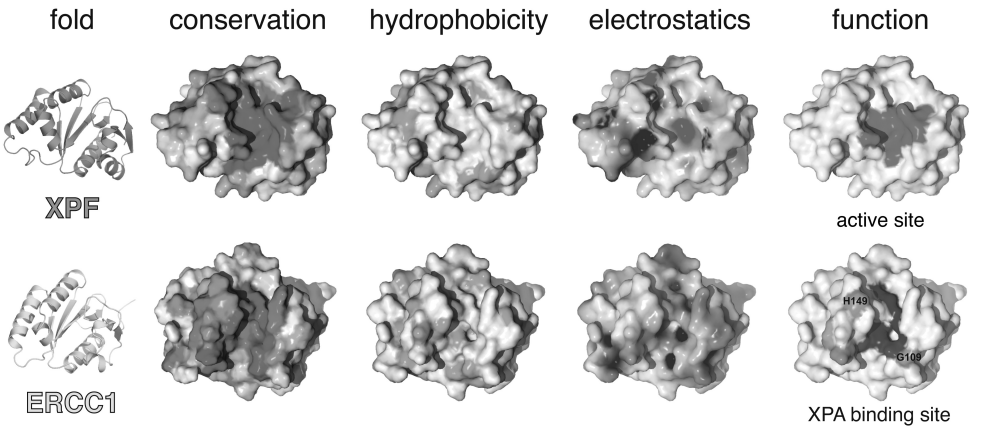


Figure 3. Common fold, different properties and distinct functions for XPF and ERCC1. Sequence conservation either for XPF or ERCC1 proteins is colored from white (non-conserved) to red (highly conserved). The opposite face has no significant conservation for either protein. Hydrophobic side chains are colored orange. Electrostatic surface potentials were calculated using APBS (Baker et al., 2001) and colored blue for positive or red for negative charge potential. Active site residues of XPF were chosen based on previous mutational studies (Enzlin and Scharer, 2002) and the XPA-binding site of ERCC1 was identified by our NMR titrations. (Full color on page 106)

chemical properties result in a distinct charge and hydrophobicity distribution on the structure surface (Figure 3). While XPF displays negative charge at the DNA cleavage site (important for divalent cation coordination), the equivalent XPA binding site in ERCC1 is neutral or slightly positively charged. Therefore, the ERCC1-XPA interaction is a combination of hydrophobic and electrostatic interactions possibly involving the glutamic acid stretch of XPA (Li et al., 1995) and two conserved positively charged residues (R106 and R156) of cERCC1.

DNA binding by the ERCC1 central domain

Given the proposed common origin of ERCC1 and XPF (Gaillard and Wood, 2001), it can be expected that the degenerated catalytic site of ERCC1 could still bind DNA in a similar fashion as the archaeal XPF active site (Tsodikov et al., 2005). Because our ERCC1-XPA interaction data render this unlikely, we explored whether cERCC1 is able to bind DNA as has been suggested previously (Tsodikov et al., 2005).

We performed an electrophoretic mobility shift assay (EMSA) where increasing amounts of cERCC1-HIS protein were added to ssDNA. While we saw a loss of free probe, we failed to see any complex formation under these conditions. If the same protein-DNA complex was loaded on a 3.5% agarose gel, we detected a weak smeary complex at the highest protein concentration (Figure 4A). However, by using a HIS-GST-cERCC1 protein, we observed formation of the complex with ssDNA (Figure 4B). This suggests that the C-terminal his-tag interferes with DNA binding. Therefore, we used the HIS-GST-cERCC1 protein, which we cleaved with thrombin to obtain the untagged protein. DNA binding with the untagged cERCC1 demonstrated a faster migrating complex with various ssDNA substrates (data not shown). HIS-GST-cERCC1/DNA complex formation is prevented by incubation of the reaction mixture with MagneHIS magnetic beads or glutathione agarose beads, which depletes HIS-GST-cERCC1 from the reaction mixture either prior to or after the addition of DNA. If beads were added after complex formation, a small but significant decrease in the amount of free probe was observed. Depletion of the HIS-GST tagged cERCC1 protein from the binding reaction confirms that the protein-DNA complex is indeed formed by the HIS-GST-cERCC1 protein (Figure 4C).

In agreement with earlier studies, under these conditions no binding was observed for dsDNA (Tsodikov et al., 2005), while fairly comparable binding affinities were found for both ssDNA and bubble substrates with 10 or 20 unpaired bases. We calculated an apparent K_d of 2.5 ± 0.7 μM for the HIS-GST-cERCC1 protein bound to a bubble10 (b10) substrate in the presence of 100 mM NaCl (Figure 4B). Corroborating our results, an equilibrium binding titration experiment using fluorescence anisotropy has demonstrated ssDNA binding for the same domain with an apparent K_d of 10 μM (Tsodikov et al., 2005). The relatively small difference in affinity can be well explained by different salt conditions or substrate used.

For the NMR titrations we performed a thrombin cleavage on HIS-GST-cERCC1, to remove the his-tag. The untagged cERCC1 ^{1}H-^{15}N HSQC spectrum is identical to that of cERCC1-HIS, except for the two

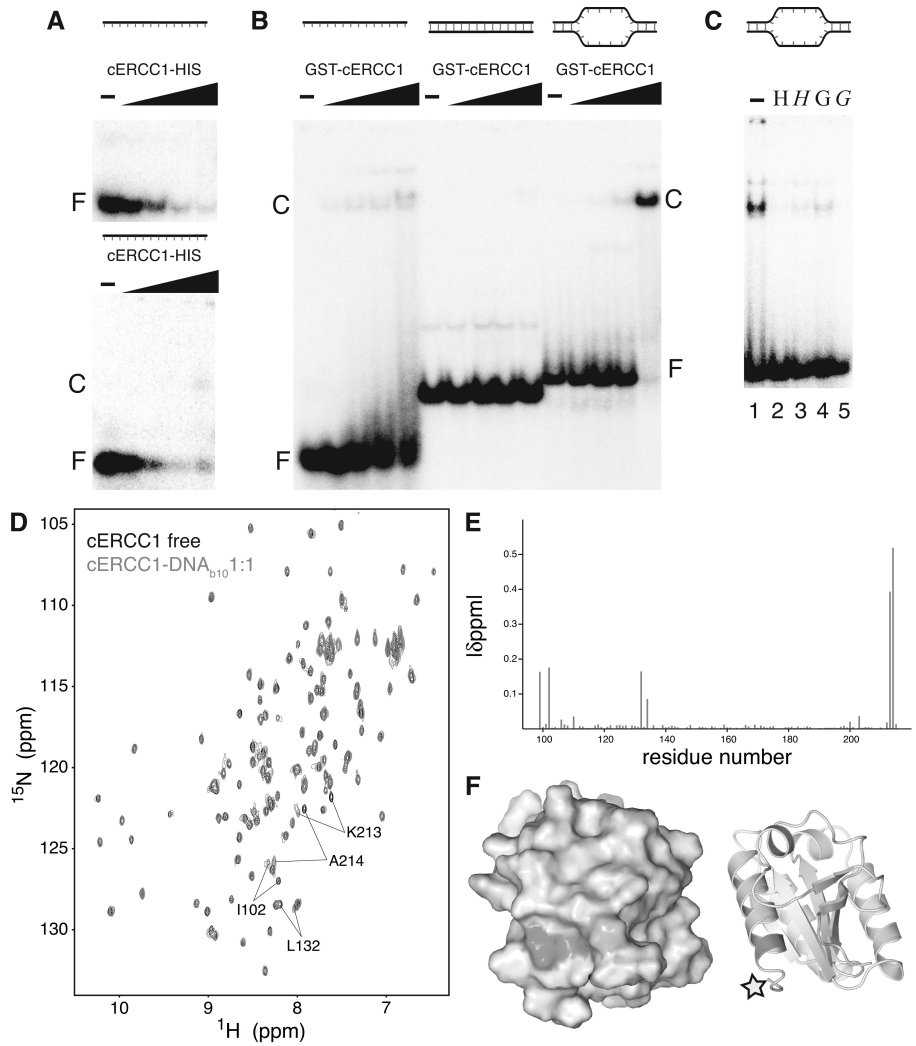


Figure 4. DNA binding by cERCC1. (A) EMSA with a 20-mer ssDNA substrate using increasing concentrations of cERCC1-HIS protein (0, 0.04, 0.2, 1 and 5 µM), loaded on a 7.5% acrylamide gel (upper panel) or a 3.5% agarose gel (lower panel). (B) EMSA using ssDNA (20-mer), dsDNA (30bp), and bubble10, a dsDNA with 10 unpaired bases as substrate, in the presence of 0, 0.125, 0.25, 0.5 and 1 µM HIS-GST-cERCC1. (C) Inhibition of cERCC1-b10 DNA complex formation ([cERCC1] = 0.5 µM) by depletion of HIS-GST-cERCC1 from the EMSA reaction by the addition of MagneHis beads (H) or GST agarose beads (G) prior to or after addition of DNA (italic). F: free DNA, C: cERCC1-DNA complex. (D) Chemical shift perturbation after addition of b10 DNA. The spectrum of the free protein is shown in black and the spectrum after addition of equimolar amount of b10 in green. (E) Normalized chemical shift changes between free and b10-bound forms versus the cERCC1 sequence. (F) Surface and cartoon representations of cERCC1 colored according to normalized chemical shifts. Yellow star indicates the position of the C-terminal his-tag that was interfering with DNA binding. (Full color on page 107)

amides preceding the artificial his-tag (Supplementary Figure 1; page 62). Addition of the b10 substrate to this cERCC1 protein caused a limited number of specific perturbations that were unambiguously assigned. Again, the perturbed resonances

exhibit slow exchange behavior in the NMR titrations and indicate contacts with the DNA. The affected resonances include the backbone amides of N99, I102, L132, K213, A214 and the side-chain amide of Q134 (Figure 4D and E). Compared to the free-protein spectrum we only miss T211, which cannot be identified. The perturbation induced by the b10 DNA was complete at equimolar concentration with cERCC1. Chemical shift mapping reveals that all perturbations are in the vicinity of the C-terminus of the cERCC1 construct, consistent with the presence of a flexible tag interfering with DNA binding. The DNA-binding region resides at the surface of the structure where the N- and C-termini meet (Figure 4F). The perturbed residues belong to well-defined structure elements. The only positively charged residue we identified (K213) is fully conserved in the ERCC1 proteins and may explain the dependence of DNA binding on salt concentration. All the other affected residues are less conserved than the residues involved in XPA binding. The composition of the affected residues together with the slow exchange behavior suggests a large contribution to the binding by hydrophobic interactions with the DNA bases.

Summarizing, both the biochemical and the NMR experiments show that the ERCC1 central domain binds to DNA. The DNA-binding site we identified on cERCC1 by the NMR titrations suggests that cERCC1 contacts only a small part of the ssDNA, probably three to four bases. The situation is different in the full-length heterodimer, where other DNA binding surfaces, such as previously established for the C-terminal ERCC1/XPF domains (Tripsianes et al., 2005; Tsodikov et al., 2005), assist in DNA recognition.

cERCC1 can bind simultaneously to XPA and ssDNA

Chemical shift perturbation experiments revealed distinct binding surfaces for ssDNA and XPA, and therefore, it is probable that both can bind simultaneously to cERCC1. We performed EMSA experiments in the presence of an equimolar or 5-fold excess of the HIS-XPA peptide. Addition of XPA did not influence binding of cERCC1 to ssDNA, while the binding affinity of XPA for the cERCC1 domain under these conditions would suggest that the vast majority of cERCC1 is in complex with XPA. We did not observe formation of a slower migrating HIS-GST-cERCC1/ssDNA/XPA super complex, possibly because the increase in mass by the addition of the HIS-XPA peptide was too small to discern (data not shown).

To independently confirm the formation of a ternary cERCC1/ssDNA/XPA complex, we added the HIS-XPA peptide to the cERCC1-bubble10 protein-DNA complex and followed the chemical shift perturbations in the ^{1}H-^{15}N HSQC spectra. As shown in Figure 5, addition of an equimolar amount of the XPA peptide causes the same amide displacements as in the titration with the free protein. On the contrary, most amides influenced by the DNA binding remain unaffected by the presence of XPA (Figure 5A). I102 and N110 were perturbed by both XPA and b10 DNA when the titrations were performed independently. In the ternary complex, however, I102 remains close to the position as in the DNA bound form. Conversely, the side chain amide of N110, although affected by the DNA, adopts the distinct XPA

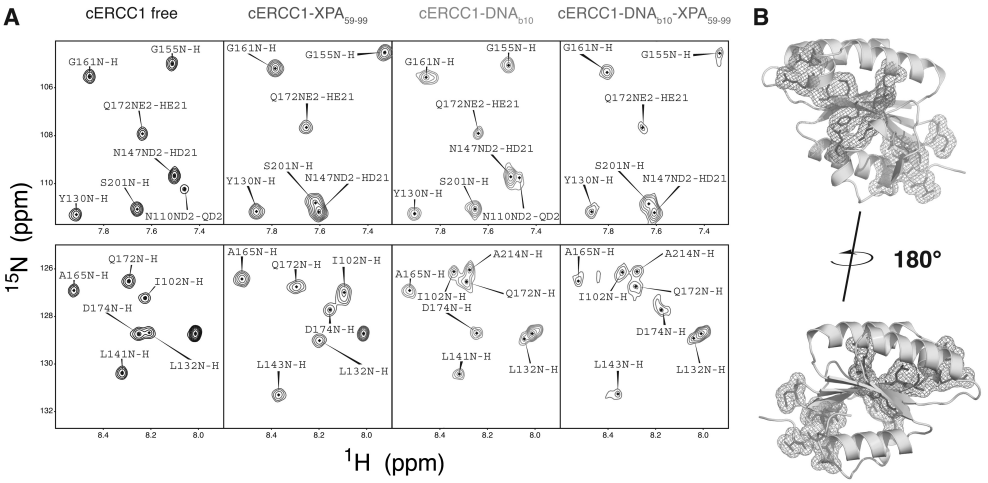


Figure 5. Dual function of cERCC1. (A) Indicative portions of the ^{1}H-^{15}N HSQC spectrum for the free cERCC1, bound to either XPA (1:1) or DNA (1:1) separately, and the ternary complex (1:1:1). (B) Views of the two binding sites on cERCC1 structure. (Full color on page 108)

bound position upon XPA addition. These data strongly indicate the formation of the ternary ERCC1/DNA/XPA complex. DNA and XPA binding sites are distinct, and the two interactions can happen simultaneously (Figure 5B).

Functional role of the ERCC1 central domain in the heterodimer

Since the catalytic activity has been preserved in the human XPF protein through strong conservation of the nuclease signature (Enzlin and Scharer, 2002), the human nuclease fold should be identical to that of the archaeal counterparts. In that sense, human XPF nuclease and human ERCC1 central domains are expected to exhibit the same fold. Moreover, the ERCC1 and XPF HhH_2 domains feature a common architecture and come into tight association exactly in the same way that the HhH_2 domains of both archaeal species do (Newman et al., 2005; Nishino et al., 2005; Tripsianes et al., 2005). Therefore, in structural terms the human ERCC1 and XPF proteins share the essential architectural subunits observed in the short XPF homodimer of *A. pernix*. The structural similarities within the XPF family provide additional evidence for the proposed common origin of the human ERCC1 and XPF proteins (Gaillard and Wood, 2001). Because ERCC1 is absent in archaea, this gene is thought to have been acquired from an ancient XPF gene duplication in the eukaryal lineage. The ERCC1 and XPF genes have subsequently evolved by the process called subfunctionalization (Force et al., 1999; Lynch and Conery, 2003). This model suggests that after gene duplication, both copies may be reciprocally preserved through the fixation of complementary loss-of-subfunction mutations, which results in a partitioning of the tasks of the ancestral gene. From the functions present in the ancestral XPF protein (archaea), human ERCC1 retained the canonical HhH_2 domain that acts as a DNA-binding domain (Tripsianes et al., 2005), while XPF retained the catalytic activity (Enzlin and

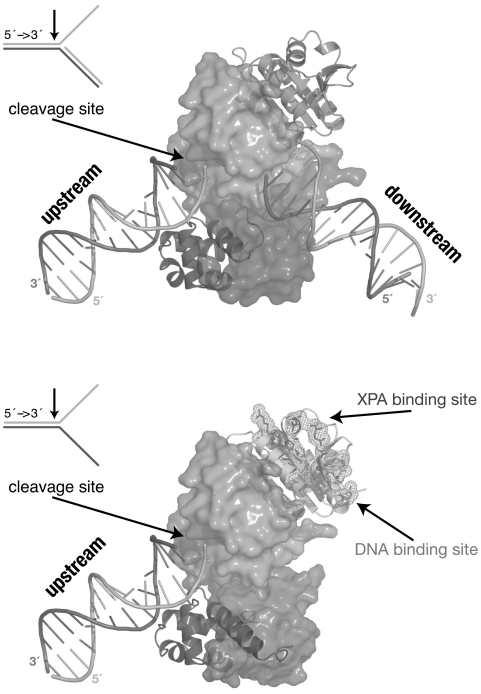


Figure 6. Models for the function of the archaeal homodimeric XPF (top) and human ERCC1/XPF heterodimer (bottom) constructed from the structure of the XPF homodimer bound to dsDNA (2bgw), the free structure of ERCC1/XPF C-terminal interacting domains (1z00) and the current structure of the ERCC1 central domain (2jpd). For the homodimer, one protomer is shown in a cartoon and the other in a surface representation. Accordingly, for the heterodimer ERCC1 is in a cartoon and XPF in a surface representation. The protein domains in both cases are colored as in Figure 1A. (Full color on page 108)

Scharer, 2002). Once the separation of the ancestral subfunctions occurred in ERCC1 and XPF genes, only the heterodimeric protein complex could restore the original function. Additionally, due to genetic alterations, the second helix-hairpin-helix motif of human XPF degenerated, yet the fold remained crucial for stabilizing the corresponding intact domain of human ERCC1 (Tripsianes et al., 2005). Similarly, the central domain of ERCC1 lost the catalytic activity by sequence drift, but despite adoption of a novel dual function (XPA interaction and DNA binding), the 3D fold was preserved.

We suggest a mechanistic model for the heterodimeric function based on the model for the homodimeric archaeal homolog (Figure 6) (Newman et al., 2005; Roberts and White, 2005). Human ERCC1 corresponds to the archaeal protomer that binds with the HhH_2 domain to the upstream duplex. Human XPF corresponds to the archaeal protomer that recognizes the downstream duplex. Since it cannot make contacts with the minor groove (Tripsianes et al., 2005), the specificity of the human heterodimer has shifted to splayed-arms substrates consisting of only one duplex. This does not exclude contribution of the XPF HhH domain to ssDNA interactions, as reported before (Tsodikov et al., 2005). We show here that the central domain of ERCC1 is also involved in ssDNA binding. Additional DNA interactions are likely for the nuclease of XPF, analogous to the archaeal case (Roberts and White, 2005). Furthermore, ssDNA binding by cERCC1 will be stimulated by specific interactions with XPA, already present in the NER pre-incision complex and possibly bound to the DNA via its own DNA-binding site. Most importantly, through these coordinated interactions of the ERCC1 domains the XPF protein will be positioned to cleave the 3´ protruding strand (limon), thereby retaining the polarity present in the archaeal homodimer.

Our model underlines the significant role of ERCC1 in the context of the full-length heterodimer. XPF is the catalytic module but the ERCC1 domains guarantee that the enzymatic activity is targeted

properly. The presence of multiple distinct DNA binding surfaces within the ERCC1/XPF/XPA/RPA repair protein intermediate coordinates cleavage to occur only when the DNA damage is recognized correctly by the NER machinery. The duplication of the ancestral XPF gene within the eukaryal kingdom resulted in an obligate heterodimer through loss of function (with an altered HhH_2 domain of XPF and a degenerated catalytic domain of ERCC1) and adoption of novel functions (ssDNA and XPA binding of cERCC1). This permitted additional quality control mechanisms through a more complicated molecular interaction network, mediated by the novel functional domains, and thereby improving the fidelity of DNA damage repair.

Materials and Methods

Cloning, protein expression and protein purification

ERCC1 central domain (cERCC1, residues 96 to 219) was PCR-amplified from a vector containing the full-length ERCC1 gene and subcloned into a pET28b (Novagen) expression vector between BamHI and XhoI sites. Downstream of the XhoI site a linker together with a his-tag has been engineered. XPA constructs were derived from a full-length XPA vector purchased from RZPD (clone ID IRAUp969B1273D6) and subcloned into a pLICHIS vector using the enzyme free cloning (EFC) strategy (de Jong et al., 2006). The same procedure was followed for the GST fusion proteins, either XPA or cERCC1 (pLICHISGST vector). As a result all the EFC constructs contain an N-terminal his-tag, while the cERCC1 construct bears a C-terminal his-tag. The nucleotide sequences of the cloned DNAs were confirmed by sequencing. The cERCC1-HIS, HIS-XPA, HIS-GST-XPA, and HIS-GST-cERCC1 proteins were expressed in BL21(DE3) Rosetta cells (Novagen) and were subject to two-step purification as has been described (Folkers et al., 2004). When appropriate, HIS-GST-cERCC1 was cleaved with thrombin in the presence of 1 mM Ca^{2+}, 50 mM Tris (pH 8.0) and 100 mM NaCl in order to obtain the untagged protein. Finally, the protein samples were exchanged either to 50 mM sodium phosphate (pH 5.5), 100 mM NaCl for the cERCC1 structure determination, or to 50 mM Tris (pH 8.0), 100 mM NaCl for the binding studies.

NMR spectroscopy

Multidimensional NMR experiments were carried out at 290 K on Bruker AVANCE 600 and 900 NMR spectrometers equipped with TXI triple-resonance probes in 50 mM sodium phosphate (pH 5.5), 100 mM NaCl by using the cERCC1-HIS protein. Spectra were processed using the NMRPipe software package (Delaglio et al., 1995) and analyzed with Sparky (Goddard and Kneller, 2001). The ^{1}H, ^{15}N and ^{13}C resonance assignments are 97% complete and were made using standard triple resonance techniques, 3D NOESY-(^{15}N, ^{1}H)-HSQC spectra, and 3D NOESY-(^{13}C, ^{1}H)-HSQC spectra (both with a mixing time of 60 ms) (Cavanagh et al., 1996). The chemical shifts have been deposited at the BMRB with accession number 15240.

Structure calculations

Automatic NOE assignment and structure calculations were performed using the program CYANA version 2.0 (Herrmann et al., 2002). Hydrogen bond restraints were

defined when they were consistent with the secondary chemical shift data and expected NOE contacts and only for the helical parts of the protein. The set of 2669 NOE restraints determined by CYANA, together with restraints for 31 hydrogen bonds and ϕ and ψ torsion angle restraints (174) derived from TALOS (Cornilescu et al., 1999) were used in a water refinement run according to the standard RECOORD protocol (Nederveen et al., 2005) utilizing CNS (Brünger et al., 1998). Molecular images were generated with PyMol (DeLano, 2002). Coordinates have been deposited at the PDB with accession code 2jpd and the structural restraints at the BMRB with entry number 15240.

EMSAs

All the DNA substrates were 5′ ^{32}P-labeled with T4 polynucleotide kinase and purified on a polyacrylamide gel under native conditions. The nucleotide sequences for ss20, b10 and ds30 have been described before (Singh et al., 2002). Per reaction, 100 fmol of substrate was incubated with the appropriate amount of ERCC1 protein in binding buffer (50 mM Tris, pH 8.0, 100 mM NaCl, 10% glycerol and 1 mg/ml BSA). After incubation for 30 min at 4 °C, samples were loaded onto 7.5% native polyacrylamide gels containing 0.5× TBE and run at 4 °C. Alternatively Electrophoretic mobility shift assay (EMSA) reactions were loaded on a 3.5% agarose gel containing 0.5% TBE, run at 4 °C or 20 °C yielding essentially identical results. Gels were visualized and quantified using a phosphor imager (BioRad) as described before (Singh et al., 2002). For depletion experiments, 20 µl of MagneHis beads (Promega) or GST agarose beads (Sigma), were extensively washed with binding buffer, and after buffer removal the EMSA reaction was added to the beads. After 15 minutes incubation with regular mixing, the beads were removed from the EMSA reaction using a magnet or centrifugation and the reaction mixture was loaded on acrylamide gel or, when appropriate, the substrate was added to the depleted reaction mixture.

GST pull-down assay

GST pull-down assay was performed and quantified as described before (Jonker et al., 2005) using the indicated cERCC1 concentrations and 3-5 µg of GST proteins in 150 µl of 50 mM Tris, 100 mM NaCl, 10% glycerol, 1 mM DTT, 0.2 mM PMSF and 20 mg/ml of BSA (pH 8.0). The semi-quantitative experiments were performed in 50, 150, 500, 1500 and 5000 µl buffer with a constant amount of cERCC1 and GST-XPA.

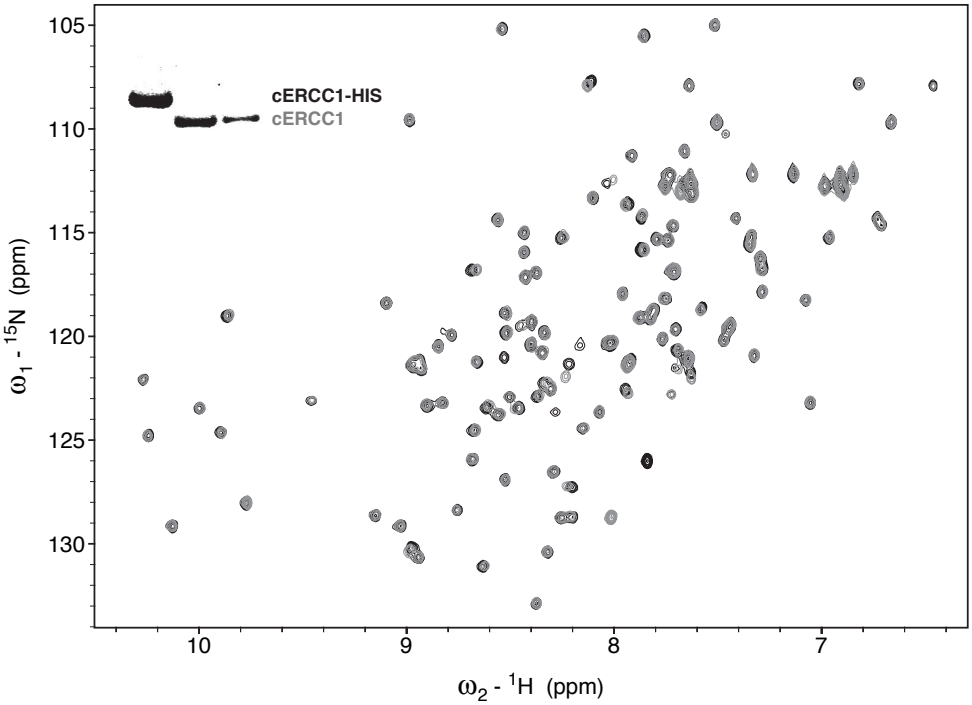
NMR titrations

cERCC1 titrations with XPA and DNA were performed on a Bruker AVANCE 700 NMR spectrometer and were monitored with 2D ^{1}H-^{15}N HSQC experiments. In all cases the concentration of cERCC1 was 0.2 mM in 50 mM Tris (pH 8.0), 100 mM NaCl. XPA and DNA were dissolved in the same buffer and therefore salt and pH did not vary throughout the experiments. DNA binding was feasible only with the cERCC1 construct lacking the C-terminal his-tag, while XPA binding was identical, regardless of the his-tag presence. Normalized chemical shift changes were calculated by using the equation: $\delta = ([\delta_{HN}]^2 + [\delta_N/6]^2)^{0.5}$.

Acknowledgements

We would like to thank Koos Jaspers and Jan Hoeijmakers for providing the full-length vectors of ERCC1 and XPF, and for many fruitful discussions. This work was financially supported by the Netherlands Organization for Scientific Research, Chemistry Council (NWO-CW), by the Center of Biomedical Genetics, the Netherlands, and by the EU project Spine2-Complexes (Contract no 031220). Funding to pay the Open Access publication charges for this article was provided by NWO-CW.

Supllementary information



Supplementary Figure 1. ^{1}H-^{15}N HSQC spectra of the C-terminal his-tag cERCC1 (black) and the corresponding untagged protein (gray). Inset shows the difference in mass for the two proteins.

Chapter four

Structural insight to the first ERCC1 deficiency

Konstantinos Tripsianes[1], Nicolaas G.J. Jaspers[2], Rolf Boelens[1] and Gert E. Folkers[1]

[1] Department of NMR spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Department of Cell Biology and Genetics, Erasmus Medical Center, Rotterdam, The Netherlands

manuscript in preparation

The ERCC1/XPF heterodimer functions as an endonuclease in nucleotide excision repair and other genome maintenance mechanisms. The function requires tight association between the double helix-hairpin-helix motifs (HhH_2 domains) of each protein. The first and hitherto only reported ERCC1 deficiency identified a single mutation at the interaction surface with the XPF partner and was characterized by severe developmental symptoms (Jaspers et al., 2007). To discern the molecular details of the ERCC1/XPF malfunction we determined the structure of the minimal mutant heterodimer. Denaturation studies show that the minimal mutant heterodimer is less stable when compared to the wild-type ERCC1/XPF complex, in agreement with the in vivo observations that instability causes reduced protein amounts for both ERCC1 and XPF. The mutation *per se* does not impair heterodimerization but in addition alters the architecture of the ERCC1 HhH_2 domain. This deformation most likely prevents ERCC1 from DNA binding, since the geometry of the ERCC1 hairpins does not allow anymore for symmetric contacts with the minor groove of the DNA. Our preliminary data suggest that the two effects, reduced stability and ERCC1 hairpin deformation, may account together for the clinical symptoms of the patient.

Introduction

Nucleotide excision repair (NER) is a genome caretaker mechanism responsible for removing helix-distorting DNA lesions, most notably ultraviolet photodimers. It involves about 30 proteins, 8 of which (XP-A to G, and TTDA) are associated with human pathology (Bootsma et al., 2002). According to the mode of damage recognition, NER is divided in two subpathways. In global genome NER (GG-NER), XPC-HR23B and DDB proteins recognize DNA lesions throughout the genome, whereas in transcription-coupled NER (TC-NER), blockage of transcribing RNA-PolII on the damaged DNA template is believed to initiate the repair reaction with the aid of the TCR-specific proteins CSA and CSB (Hoeijmakers, 2001). After damage recognition the DNA is melted, the damaged strand is incised in both sides by specific endonucleases, and gap filling is done by replication polymerases and polymerase κ (Ogi and Lehmann, 2006).

NER defects lead to two distinct phenotypes. GG-NER impairment is related with the syndrome xeroderma pigmentosum (XP), characterized by a 1000-fold elevated risk of sun-induced skin cancer (Bootsma et al., 2002). Defective TC-NER, on the other hand, causes the segmental progerias Cockayne syndrome (CS) and trichothiodystrophy (TTD) without an increase in cancer incidence (Bootsma et al., 2002). In addition, several NER proteins have functions outside NER, leading to complex and pleiomorphic disease phenotypes.

ERCC1/XPF is a structure specific endonuclease required for NER to incise the damaged strand 5′ to the lesion (Sijbers et al., 1996). XPF is the catalytic module (Enzlin and Scharer, 2002) and ERCC1 is essential for DNA binding (Tripsianes et al., 2005; Tripsianes et al., 2007; Tsodikov et al., 2005). The function strictly relies on heterodimerization, which is mediated through the C-terminal tandem helix-

hairpin-helix (HhH_2) domains of both proteins (de Laat et al., 1998; Tripsianes et al., 2005). The obligate ERCC1/XPF heterodimer has additional, NER-independent roles in homologous recombination (Niedernhofer et al., 2001), in the repair of interstrand DNA crosslinks (Niedernhofer et al., 2004), and the protection of telomeres (Munoz et al., 2005; Zhu et al., 2003). Both ERCC1 and XPF knock-out mice have growth retardation and short life span (Tian et al., 2004; Weeda et al., 1997). These symptoms are much more severe than XP, CS or TTD, and indicate the pleiotropic function of ERCC1/XPF.

In humans, most of the XPF mutations are associated with mild xeroderma pigmentosum and considerable levels of residual repair capacity (Matsumura et al., 1998; Sijbers et al., 1998). A unique case was reported recently, where the patient harbors a homozygous non-conservative substitution (R153P) and exhibits profound crosslink sensitivity and dramatic progeroid symptoms (Niedernhofer et al., 2006). Although the point mutation occurs outside the ERCC1 interacting domain, both XPF and ERCC1 protein levels are strongly reduced in the fibroblasts of the patient.

With respect to ERCC1, only a single case of human inherited deficiency has been described so far (Jaspers et al., 2007). The clinical features of the patient were very severe, compatible with a diagnosis of cerebro-oculo-facio-skeletal syndrome (COFS), and led to death in early infancy. Genomic locus sequencing from cells from the patient and his parents revealed two point mutations. The one allele harbors a C->T transition that converts Gln158 to a translational stop signal and is predicted to produce a non-functional truncated ERCC1 polypeptide. The other allele carried a C->G transversion that changes Phe231 to Leu (F231L). This mutation resides in the C-terminal domain of ERCC1 that mediates the interaction with the corresponding C-terminal domain of XPF. The mutant ERCC1/XPF heterodimer complemented partially the resistance to UV and the crosslinking agent mitomycin C. This was attributed to the low levels of the mutant heterodimer compared to the wild-type controls (Jaspers et al., 2007).

Previous structural studies have highlighted the molecular details of the ERCC1/XPF interaction (Tripsianes et al., 2005). Phe231, which is conserved among mammals, is directly involved in hydrophobic contacts with XPF residues. The structure of the mutant heterodimer presented here, shows that the substitution to Leu does not alter the structural properties of the ERCC1/XPF association. Both proteins retain their structure secondary elements that determine the HhH_2 fold. ERCC1, however, rearranges its hydrophobic core and destabilizes the heterodimer due to the local disturbance of the interaction with XPF at the mutation point. These changes have an additional effect in the second hairpin of ERCC1, which deviates significantly from the proper geometry. We suggest that the abnormalities of the patient are both a consequence of the lower mutant stability and the malformed DNA binding domain of ERCC1 that is required to target the XPF nuclease for processing irregular DNA structures.

Results and Discussion

Mutant heterodimer behavior in vitro

The mutant ERCC1/XPF was over-

expressed in *E.coli* host cells essentially as described before (Tripsianes et al., 2005). After purification using a nickel-charged metal chelating column, by taking advantage of the his-tag at the C-terminus of the ERCC1, the ratio between the two subunits was 1:1 confirming dimer formation, in accordance with our previous studies on the wild type domains. During size exclusion chromatography there was no indication of complex dissociation, under low salt concentration (100 mM NaCl). In addition, the fingerprint HSQC spectrum showed dispersion of a folded protein and the expected number of peaks, like the wild type case (Figure 1A). These data indicate that the mutant heterodimer has structural integrity.

We first analyzed the difference in stability between mutant and wild type by monitoring the heat denaturation of the heterodimer by NMR. The HSQC spectra provide good means because we can follow the gradual disappearance of ERCC1 peaks upon complex disruption and the concomitant formation of XPF homodimers that result in a distinct set of peaks (Das et al., 2007). In each case, wild type or mutant, ERCC1 peaks begun to shift at 33 °C. As the temperature was raised, ERCC1 peaks decreased in intensity and completely disappeared at 53 °C for the mutant and at 60 °C for the wild type. At the final temperatures the only species in solution were XPF homodimers (Supplementary Figure 1; page 73). This illustrates that the mutant is less resistant to temperature denaturation than the wild type.

These data are in agreement with the in vivo observations, where the authors

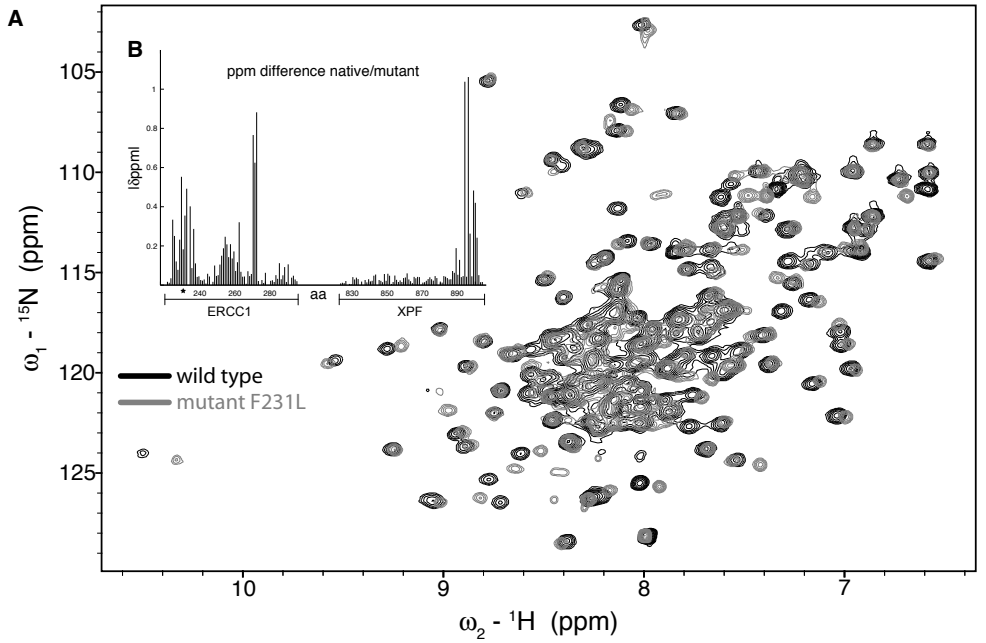


Figure 1. (A) Overlay of 15N-1H HSQC spectra from the wild type heterodimer (black) and the mutant heterodimer (gray). (B) Average chemical shift differences between wild type and mutant conformations. A star indicates the mutation site.

suggested that the reduced activity of the mutant heteroduplex and the low levels of both proteins are due to the instability of the mutant ERCC1/XPF (Jaspers et al., 2007). It might be that the other domains of the full-length heterodimer, which are not present in our protein, influence indirectly the proper ERCC1/XPF complex formation. To understand whether the reduced stability caused by the ERCC1 F231L mutation is due to a different architecture of the complex, we determined the structure of the mutant ERCC1 domain in complex with the XPF interacting partner.

Structure description

We have assigned 90% of all the possible ^{1}H, ^{15}N and ^{13}C resonances of the mutant heterodimer. The structure was determined on the basis of 3228 distance restraints and 201 dihedral angle restraints (Table 1). Compared to the wild type structure we determined before, the NOEs for the mutant represent 71% of those of the wild type, while the dihedral angle restraints are of the same number. The smaller amount of NOEs is partly explained by the fewer NMR experimental data used and/or possibly by the inherent dynamic properties of the mutant ERCC1/XPF complex. In particular, the intermolecular NOEs are less than half of the wild type structure determination, and therefore the present structure should be considered as a starting point towards understanding the mutant malfunction. Despite the lower number of restraints, structure calculations yielded an ensemble of 20 conformers with good convergence (Figure 2A). A summary of the structural and restraint statistics is given in Table 1. The structural analysis has been done in exactly the same way as in the analysis of the wild type structure and therefore the comparison of the two structures is straightforward.

Table 1. Structural statistics of the structure ensemble of the mutant ERCC1/XPF heterodimer

Rmsd (Å) with respect to mean [a] (backbone/heavy)	
ERCC1	0.57 ± 0.11 / 1.19 ± 0.17
XPF	0.47 ± 0.10 / 1.06 ± 0.16
Complex	0.63 ± 0.12 / 1.19 ± 0.15
Number of experimental restraints	
Intraresidue NOEs	680
Sequential NOEs (|i - j| = 1)	919
Medium range NOEs (1<|i - j|<4)	880
Long-range NOEs (|i - j|>4)	558
Interprotein	191
Total NOEs	3228
Dihedral angle restrains	201
Hydrogen bonds	53
Number of consistent violations	
NOE distances with violations >0.3 Å	2
Dihedrals with violations >5°	4
Rmsd for experimental restraints	
All distance restraints (3228) (Å)	0.026 ± 0.003
Torsion angles (201) (°)	1.378 ± 0.096
CNS energies after water refinement	
E_{vdw} (kcal/mol)	-629 ± 23
E_{elec} (kcal/mol)	-6388 ± 67
Rmsd from idealized covalent geometry	
Bonds (Å)	0.01 ± 0.00
Angles (°)	1.32 ± 0.05
Impropers (°)	1.62 ± 0.09
Ramachandran analysis	
Residues in the favored regions (%)	89.41 ± 1.67
Residues in additional allowed regions (%)	10.08 ± 1.55
Residues in generously allowed regions (%)	0.28 ± 0.48
Residues in disallowed regions (%)	0.25 ± 0.41

[a] Residues 227-294 of ERCC1 and residues 831-896 of XPF

The global architecture of the mutant heterodimer does not deviate from that of the wild type. Each partner consists of five helices that give rise to the characteristic double helix-hairpin-helix motif (HhH_2), which complex in a symmetry related manner (Figure 2A and B). The interaction surface between mutant ERCC1 and XPF is hydrophobic and covers an area of 1250 $Å^2$. In the case of the wild type the corresponding area is 1530 $Å^2$. The hydrophobic intermolecular contacts are

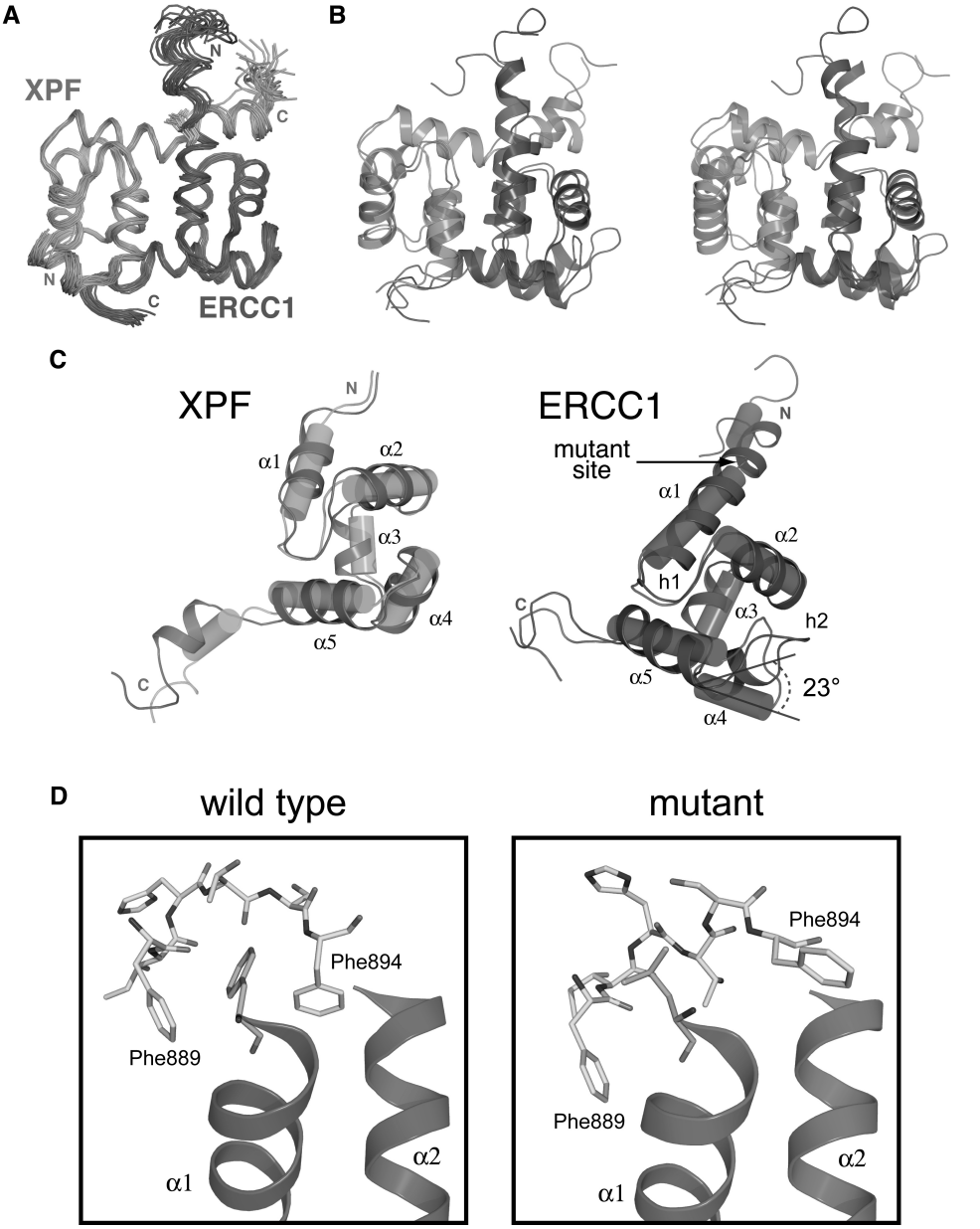


Figure 2. Structural differences induced by the F231L mutation of ERCC1. (A) Backbone superposition of the 20 lowest energy structures for the mutant heterodimer comprising the ensemble. Leu231 of ERCC1 is shown in a stick representation. Both partners are indicated together with their N- and C-termini. (B) Structural overlay of wild-type and mutant hetrodimers superimposed on the XPF (left) and ERCC1 side (right). Wild type conformations are shown in dark transparent color. (C) Structural comparison for each partner in wild type (cartoon) and mutant heterodimeric complexes (cylinders). (D) Local network of interactions for Phe231 (wild type) and Leu231 (mutant). (Full color on page 109)

important for the stability of the complex and the smaller interacting surface in the mutant heterodimer is in agreement with the lower stability. Structural comparison of the individual subunits between the wild type and mutant states shows that the XPF five-helical core remains unaffected (rmsd 1.2 Å, for 60 Cα atoms), whereas the ERCC1 fold exhibits local distortions when compared to its wild type architecture (rmsd 2.0 Å, for 60 Cα atoms) (Figure 2C). The overall rmsd value for the wild type and mutant hetrodimeric structures is 2.2 Å.

Consequences of the F231L mutation in the heterodimeric structure

The first noticeable difference between wild type and mutant structures relates to the mutant site. Whereas Phe231 is part of the first ERCC1 helix, the substitution to Leu induces a kink to the backbone that disrupts the helix continuity (Figure 2B). This is evidenced by the change in the ψ dihedral angle from -52° (Phe231) to -33° (Leu231) that falls outside the typical α-helix range (-60° to -45°). In line with this, the backbone protons of Phe231 of the wild-type complex experience upfield shifts with respect to the CSI random coil values, which are indicative of a helical configuration. On the other hand, the backbone protons of Leu231 in the mutant complex fall in the range of the corresponding CSI random coil values, supporting the local backbone distortion observed in the mutant structure. The F231L substitution results in a different pattern of sidechain interactions with the C-terminal residues of the XPF partner. This can be seen by the chemical shift difference between the wild type and mutant HSQC spectra that involve mainly residues in the vicinity of the ERCC1 mutant site and the C-terminal tail of XPF (Figure 1B).

In the native structure the bulky ring of Phe231 stacks against the aromatic ring of Phe889 from the XPF tail and the following Phe894 is accommodated into the ERCC1 hydrophobic pocket. In addition there are five intermolecular hydrogen bonds that stabilize locally the XPF backbone, including the amide proton of Phe894 (Tripsianes et al., 2005). In the mutant structure the Leu231 sidechain occupies a smaller volume and is not competent for π-stacking interactions. Phe889 of XPF compensates for the ring absence by pointing deeper into the central hydrophobic core formed between ERCC1 and XPF. The backbone conformation of the following XPF residues diverts by facing the Leu sidechain, and the ring of Phe894 cannot reach the ERCC1 cavity. Instead, it flips out with the ring facing the helical distortion at the mutation site of ERCC1. Table 2 shows the chemical shifts for the three ERCC1 residues that accommodate the Phe894 ring in the wild type but not in the mutant heterodimeric structure. Overall the F231L substitution rearranges the local interaction network between ERCC1 and XPF, and alters the backbone and sidechain orientation of the interacting residues to a large extent (Figure 2D). These changes affect the intermolecular hydrogen bonds in the region as well. In the 20 conformers of the mutant ensemble we find consistently only

Table 2. Backbone chemical shifts for the ERCC1 residues that perform hydrophobic contacts with the ring of XPF Phe894 in the wild type heterodimer. Corresponding shifts in the mutant structure where the particular interaction is absent.

ERCC1 residue	wild type $^{15}N/^{1}H$	mutant $^{15}N/^{1}H$
Val232	118.180/8.392	117.550/8.045
Val235	121.000/8.585	121.882/8.973
Leu254	119.588/8.144	118.937/8.316

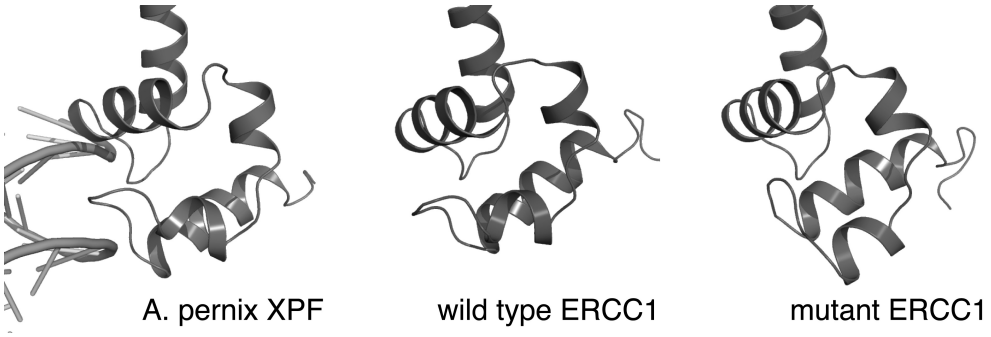


Figure 3. DNA binding mode by the HhH_2 domain of *A. pernix* XPF and the geometry of the ERCC1 hairpin motifs in the wild type and mutant ERCC1/XPF heterodimeric structures. (Full color on page 110)

two hydrogen bonds between ERCC1 and XPF, with no participation of Phe894. This is another structural explanation for the lower stability of the mutant heterodimer.

The HhH_2 integrity and fold is determined by the presence of key hydrophobic residues on each of the five helices (Shao and Grishin, 2000). In the case of ERCC1 and XPF HhH_2 domains the C-terminal aromatic residues of each protein, Phe293 of ERCC1 and Phe894 of XPF, are absolutely required to complete the hydrophobic core of their partner (Tripsianes et al., 2005). This further explains the necessity for ERCC1/XPF dimer formation to mutually support the integrity of each HhH_2 domain. In the mutant heterodimeric structure most of the ERCC1 hydrophobic sidechains reorient to compensate for the loss of contacts with the XPF Phe894 at the expense of the overall heterodimeric stabilty.

The local reorganization however, propagates to the rest of the ERCC1 structure and affects critically the fold at distant regions. The most prominent changes are observed in helix 4, where Leu271, Ala272, and Leu273 exhibit larger chemical shift differences with respect to the wild type conformation, than the ERCC1 residues at the proximity of the mutation (Figure 1B). Compared to the native structure helix 4 is inclined by 23° in the mutant conformation and distorts the second hairpin (Figure 2D). These changes that occur 25 Å away from the mutant site alter significantly the geometry of the second ERCC1 hairpin. The second hairpin folds backwards to fill in the gap created by the tilt of the fourth helix in order to support the new network of hydrophobic interactions.

Double helix-hairpin-helix geometry and the role in protein function

Hairpin motifs are found in many proteins like DNA-polymerases, ligases, glycosylases and nucleases, participating in DNA transactions. Most HhH motifs are arranged in pairs and each pair forms into a five helical HhH_2 domain. These domains are rather unique given that they recognize the DNA in non-sequence specific manner. In accord with the protein-DNA structures available the contacts are not built around the bases but on the sugar phosphate backbone of the minor groove (Ariyoshi et al., 2000; Newman et al., 2005) (Figure 3). The proper geometry of the two consecutive hairpins, which are related by pseudo-2-fold symmetry, allows symmetric contacts with the DNA backbone. Previous

functional analysis has indicated the role of the ERCC1 HhH_2 domain in binding the DNA duplex of substrates targeted by the ERCC1/XPF endonuclease (Tripsianes et al., 2005). The ERCC1 conformation adapted in the mutant structure suggests that ERCC1 is incompetent in engaging the DNA, due to the distortion of the second hairpin from the proper geometry (Figure 3). In the homologous archaeal XPF protein, single mutations in the hairpin residues that are strictly conserved, reduced the nuclease activity ten-fold due to poor DNA binding (Nishino et al., 2005).

Our in vitro data corroborate the in vivo findings that instability accounts for the low levels of the mutant ERCC1/XPF heterodimer (Jaspers et al., 2007). In addition the structural analysis suggest that the mutation poses an extra problem in the DNA recognition by ERCC1. It appears therefore that the low activity observed in the fibroblasts of the patient is a combination of reduced protein amounts and malfunction in the coordinated activity of the mutant ERCC1/XPF. However, we should note that extra experiments are needed to validate the rearrangements observed in the initial mutant structure presented here, and to prove finally, whether or not the mutation affects the DNA binding properties of the heterodimer.

Conclusions

In the cases that XPF mutations are well characterized (Niedernhofer et al., 2006; Sijbers et al., 1998), as well as in the first reported ERCC1 point mutation (Jaspers et al., 2007), the levels of both proteins are strongly reduced in the fibroblasts of each patient. The clinical features, however, vary from mild xeroderma pigmentosum to progeroid syndromes for XPF, whereas the ERCC1 F231L mutation results in severe congenital and developmental abnormalities. These symptoms are partly explained by the low cellular content of the proteins. For the first ERCC1 deficiency, our structural data explain well the reduced in vivo stability. However, we noted that the F231L mutation might affect the coordinated function of ERCC1/XPF heterodimer by reducing the DNA binding activity of the complex. Such an additional defect, which might provide useful means in understanding the unique ERCC1 phenotype, requires clearly support by biochemical experiments that are currently under way in our laboratory.

Materials and Methods

Cloning, protein expression and protein purification

The mutant ERCC1 clone was prepared by using an extended mutagenic 5′ primer during PCR amplification from the wild-type construct (residues 220-297, F231L). Accordingly it was cloned into a pET28b bicistronic expression vector between BamHI and XhoI sites substituting the wild-type coding region, while upstream of the BamHI site is located the XPF cistron (residues 823-905). Downstream of the XhoI site a linker together with a His-6 tag has been engineered. Therefore, the ERCC1 mutated domain contains a C-terminal His-6 tag. The nucleotide sequence of the mutant clone was confirmed by sequencing. Protein expression and purification of the mutant heterodimer were identical with the wild type preparations (Tripsianes et al., 2005) yielding NMR samples of ~0.7 mM.

NMR spectroscopy

Multidimensional NMR experiments were carried out at 290 K on Bruker AVANCE 700 spectrometer in 50mM phosphate (pH=7), 100 mM NaCl buffer. Spectra were processed using the NMRPipe software package (Delaglio et al., 1995) and analyzed with Sparky (Goddard and Kneller, 2001). Triple resonance spectra were recorded essentially as described in (Cavanagh et al., 1996). Backbone and side-chain ^{1}H, ^{15}N and ^{13}C resonances were assigned by using H(CCO)NH-TOCSY and C(CO)NH-TOCSY experiments and the resonances of the wild type complex as a reference guide. These two experiments allowed us to assign all the atoms of the mutant amino acid (Leu231). Extra assignments were obtained by the manual inspection of 3D NOESY-(^{13}C, ^{1}H)-HSQC and 3D NOESY-(^{15}N, ^{1}H)-HSQC spectra (τ_{mix}=75ms) to reach finally a 90% completeness. The 3D NOESY spectra were complemented by homonuclear 2D NOE spectra recorded without and with H[^{15}N] suppression in F2 (τ_{mix}=75ms).

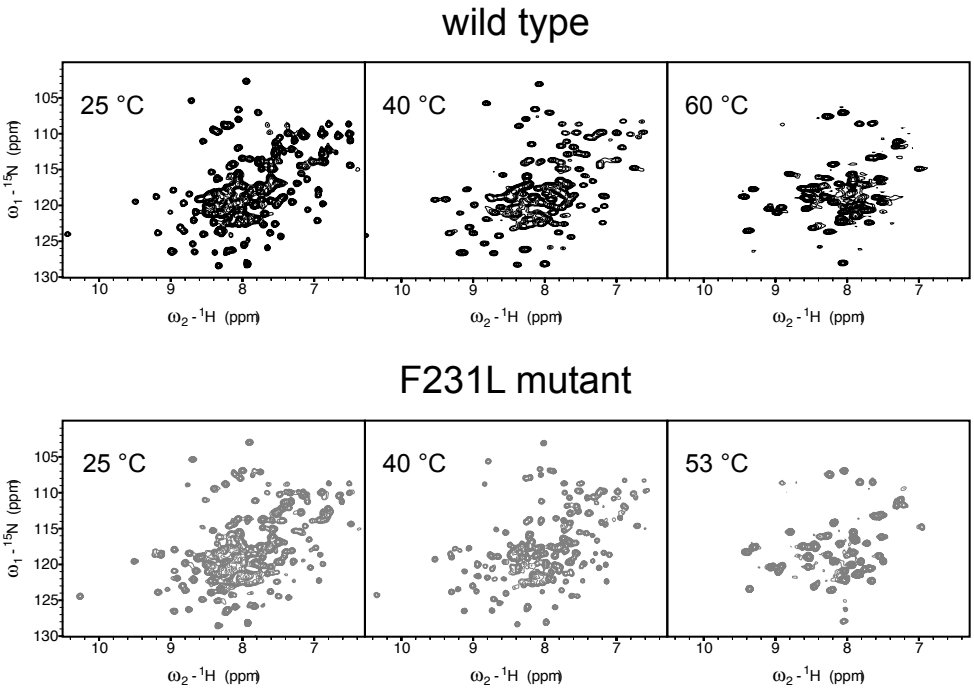
Structure calculations

Automatic NOE assignment and structure calculations were performed using the the program CYANA version2.0 (Güntert et al., 1997; Herrmann et al., 2002). Hydrogen bond restraints were defined when they were consistent with the secondary chemical shift data and expected NOE contacts. The set of NOE based restraints determined by CYANA, together with restraints for 44 H-bonds and 201 ϕ and ψ torsion angle restraints derived from TALOS (Cornilescu et al., 1999) were used in a water refinement run using CNS (Brünger et al., 1998), according to the standard RECOORD protocol (Nederveen et al., 2005). The final structures were validated using WHATIF (Vriend, 1990) and PROCHECK (Laskowski et al., 1993; Morris et al., 1992). Molecular images were generated with PyMol (DeLano, 2002).

Acknowledgements

This work was financially supported by the Netherlands Organization for Scientific Research, Chemistry Council (NWO-CW).

Supllementary information



Supplementary Figure 1. ^{15}N-^{1}H HSQC spectra for wild type and mutant ERCC1/XPF upon increasing temperatures. Notice the formation of identical XPF homodimeric species at different final temperatures.

References

Aboussekhra, A., Biggerstaff, M., Shivji, M. K., Vilpo, J. A., Moncollin, V., Podust, V. N., Protic, M., Hubscher, U., Egly, J. M., and Wood, R. D. (1995). Mammalian DNA nucleotide excision repair reconstituted with purified protein components. Cell 80, 859-868.

Adimoolam, S., and Ford, J. M. (2002). p53 and DNA damage-inducible expression of the xeroderma pigmentosum group C gene. Proc Natl Acad Sci U S A 99, 12985-12990.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 92, 291-294.

Alberts, B. (2003). DNA replication and recombination. Nature 421, 431-435.

Amundson, S. A., Patterson, A., Do, K. T., and Fornace, A. J., Jr. (2002). A nucleotide excision repair master-switch: p53 regulated coordinate induction of global genomic repair genes. Cancer Biol Ther 1, 145-149.

Araki, M., Masutani, C., Maekawa, T., Watanabe, Y., Yamada, A., Kusumoto, R., Sakai, D., Sugasawa, K., Ohkuma, Y., and Hanaoka, F. (2000). Reconstitution of damage DNA excision reaction from SV40 minichromosomes with purified nucleotide excision repair proteins. Mutat Res 459, 147-160.

Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., and Hanaoka, F. (2001). Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. J Biol Chem 276, 18665-18672.

Araujo, S. J., Nigg, E. A., and Wood, R. D. (2001). Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. Mol Cell Biol 21, 2281-2291.

Araujo, S. J., Tirode, F., Coin, F., Pospiech, H., Syvaoja, J. E., Stucki, M., Hubscher, U., Egly, J. M., and Wood, R. D. (2000). Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. Genes Dev 14, 349-359.

Araujo, S. J., and Wood, R. D. (1999). Protein complexes in nucleotide excision repair. Mutat Res 435, 23-33.

Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2000). Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer. Proc Natl Acad Sci U S A 97, 8257-8262.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98, 10037-10041.

Balajee, A. S., and Bohr, V. A. (2000). Genomic heterogeneity of nucleotide excision repair. Gene 250, 15-30.

Bardwell, A. J., Bardwell, L., Tomkinson, A. E., and Friedberg, E. C. (1994). Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. Science 265, 2082-2085.

Batty, D., Rapic'-Otrin, V., Levine, A. S., and Wood, R. D. (2000). Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. J Mol Biol 300, 275-290.

Bessho, T., Sancar, A., Thompson, L. H., and Thelen, M. P. (1997). Reconstitution of human excision nuclease with recombinant XPF-ERCC1 complex. J Biol Chem 272, 3833-3837.

Bochkarev, A., and Bochkareva, E. (2004). From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. Curr Opin Struct Biol 14, 36-42.

Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature 385, 176-181.

Bochkareva, E., Korolev, S., Lees-Miller, S. P., and Bochkarev, A. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. Embo J 21, 1855-1863.

Bootsma, D. (2001). The "Dutch DNA Repair Group", in retrospect. Mutat Res 485, 37-41.

Bootsma, D., Kraemer, K. H., Cleaver, J. E., and Hoeijmakers, J. H. J (2002). The Genetic Basis of Human Cancer. New York: McGraw-Hill Medical Publishing Division.

Bruner, S. D., Norman, D. P., and Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. Nature 403, 859-866.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54 (Pt 5), 905-921.

Buchko, G. W., Isern, N. G., Spicer, L. D., and Kennedy, M. A. (2001). Human nucleotide excision repair protein XPA: NMR spectroscopic studies of an XPA fragment containing the ERCC1-binding region and the minimal DNA-binding domain (M59-F219). Mutat Res 486, 1-10.

Bunick, C. G., Miller, M. R., Fuller, B. E., Fanning, E., and Chazin, W. J. (2006). Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. Biochemistry 45, 14965-14979.

Burns, J. L., Guzder, S. N., Sung, P., Prakash, S., and Prakash, L. (1996). An affinity of human replication protein A for ultraviolet-damaged DNA. J Biol Chem 271, 11607-11610.

Buschta-Hedayat, N., Buterin, T., Hess, M. T., Missura, M., and Naegeli, H. (1999). Recognition of nonhybridizing base pairs during nucleotide excision repair of DNA. Proc Natl Acad Sci U S A 96, 6090-6095.

Cavanagh, J., Fairbrother, J. W., Palmer, G. A. III, and Skelton, J. N. (1996). Protein NMR Spectroscopy. Academic Press, San Diego, CA.

Choi, Y. J., Ryu, K. S., Ko, Y. M., Chae, Y. K., Pelton, J. G., Wemmer, D. E., and Choi, B. S. (2005). Biophysical characterization of the interaction domains and mapping of the contact residues in the XPF-ERCC1 complex. J Biol Chem 280, 28644-28652.

Chu, G., and Chang, E. (1988). Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. Science 242, 564-567.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A. R., Laghmani el, H., Joenje, H., McDonald, N., de Winter, J. P., et al. (2007). Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Mol Cell 25, 331-343.

Cleaver, J. E. (1968). Defective repair replication of DNA in xeroderma pigmentosum. Nature 218, 652-656.

Coin, F., Bergmann, E., Tremeau-Bravard, A., and Egly, J. M. (1999). Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. Embo J 18, 1357-1366.

Constantinou, A., Gunz, D., Evans, E., Lalle, P., Bates, P. A., Wood, R. D., and Clarkson, S. G. (1999). Conserved residues of human XPG protein important for nuclease activity and function in nucleotide excision repair. J Biol Chem 274, 5637-5648.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13, 289-302.

Coverley, D., Kenny, M. K., Munn, M., Rupp, W. D., Lane, D. P., and Wood, R. D. (1991). Requirement for the replication protein SSB in human DNA excision repair. Nature 349, 538-541.

Das, D., Tripsianes, K., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., Boelens, R., and Folkers, G. E. (2007). The HhH domain of the human DNA repair protein XPF forms stable homodimers. Proteins [Epub ahead of print].

de Boer, J., and Hoeijmakers, J. H. (2000). Nucleotide excision repair and human syndromes. Carcinogenesis 21, 453-460.

de Jong, R. N., Daniels, M. A., Kaptein, R., and Folkers, G. E. (2006). Enzyme free cloning for high throughput gene cloning and expression. J Struct Funct Genomics 7, 109-118.

de Laat, W. L., Appeldoorn, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998a). DNA structural elements required for ERCC1-XPF endonuclease activity. J Biol Chem 273, 7835-7842.

de Laat, W. L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998b). DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. Genes Dev 12, 2598-2609.

de Laat, W. L., Jaspers, N. G., and Hoeijmakers, J. H. (1999). Molecular mechanism of nucleotide excision repair. Genes Dev 13, 768-785.

de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G., and Hoeijmakers, J. H. (1998c). Mapping of interaction domains between human repair proteins ERCC1 and XPF. Nucleic Acids Res 26, 4146-4152.

De Weerd-Kastelein, E. A., Keijzer, W., and Bootsma, D. (1972). Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. Nat New Biol 238, 80-83.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6, 277-293.

DeLano, W. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.

Dip, R., Camenisch, U., and Naegeli, H. (2004). Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair. DNA Repair (Amst) 3, 1409-1423.

Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996). The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. Nucleic Acids Res 24, 2488-2497.

Dualan, R., Brody, T., Keeney, S., Nichols, A. F., Admon, A., and Linn, S. (1995). Chromosomal localization and cDNA cloning of the genes (DDB1 and DDB2) for the p127 and p48 subunits of a human damage-specific DNA binding protein. Genomics 29, 62-69.

Dubaele, S., Proietti De Santis, L., Bienstock, R. J., Keriel, A., Stefanini, M., Van Houten, B., and Egly, J. M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. Mol Cell 11, 1635-1646.

Enzlin, J. H., and Scharer, O. D. (2002). The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. Embo J 21, 2045-2053.

Evans, E., Fellows, J., Coffer, A., and Wood, R. D. (1997). Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. Embo J 16, 625-638.

Fitch, M. E., Cross, I. V., and Ford, J. M. (2003a). p53 responsive nucleotide excision repair gene products p48 and XPC, but not p53, localize to sites of UV-irradiation-induced DNA damage, in vivo. Carcinogenesis 24, 843-850.

Fitch, M. E., Nakajima, S., Yasui, A., and Ford, J. M. (2003b). In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. J Biol Chem 278, 46906-46910.

Folkers, G. E., van Buuren, B. N., and Kaptein, R. (2004). Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. J Struct Funct Genomics 5, 119-131.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531-1545.

Freedberg, I. M., Eisen, A. Z., Wolff, K., Austen, K. F., Goldsmith, L. A., and Katz, S. I. (2003). Fitzpatrick's Dermatology In General Medicine. New York: McGraw-Hill Medical Publishing Division.

Friedberg, E. C. (2001). How nucleotide excision repair protects against cancer. Nat Rev Cancer 1, 22-33.

Friedberg, E. C., Walker, G.C. & Siede, W. (1995). DNA Repair and Mutagenesis. American Society of Microbiology Press, Washington DC.

Fujiwara, Y., Masutani, C., Mizukoshi, T., Kondo, J., Hanaoka, F., and Iwai, S. (1999). Characterization of DNA recognition by the human UV-damaged DNA-binding protein. J Biol Chem 274, 20027-20033.

Gaillard, P. H., and Wood, R. D. (2001). Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. Nucleic Acids Res 29, 872-879.

Giglia-Mari, G., Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N. G., Raams, A., Argentini, M., van der Spek, P. J., et al. (2004). A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat Genet 36, 714-719.

Gillet, L. C., and Scharer, O. D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. Chem Rev 106, 253-276.

Goddard, T. D., and Kneller, D. G. (2001). SPARKY 3. University of California, San Francisco.

Gomes, X. V., and Burgers, P. M. (2001). ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. J Biol Chem 276, 34768-34775.

Green, C. M., and Almouzni, G. (2002). When repair meets chromatin. First in series on chromatin dynamics. EMBO Rep 3, 28-33.

Green, C. M., and Almouzni, G. (2003). Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. Embo J 22, 5163-5174.

Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. Philos Trans R Soc Lond B Biol Sci 361, 1007-1022.

Grzesiek, S., Bax, A., Clore, G. M., Gronenborn, A. M., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., and Wingfield, P. T. (1996). The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. Nat Struct Biol 3, 340-345.

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273, 283-298.

Guzder, S. N., Sung, P., Prakash, L., and Prakash, S. (1996). Nucleotide excision repair in yeast is mediated by sequential assembly of repair factors and not by a pre-assembled repairosome. J Biol Chem 271, 8903-8910.

Hara, R., Mo, J., and Sancar, A. (2000). DNA damage in the nucleosome core is refractory to repair by human excision nuclease. Mol Cell Biol 20, 9173-9181.

He, Z., Henricksen, L. A., Wold, M. S., and Ingles, C. J. (1995). RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. Nature 374, 566-569.

He, Z., and Ingles, C. J. (1997). Isolation of human complexes proficient in nucleotide excision repair. Nucleic Acids Res 25, 1136-1141.

Heflich, R. H., and Neft, R. E. (1994). Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. Mutat Res 318, 73-114.

Henning, K. A., Li, L., Iyer, N., McDaniel, L. D., Reagan, M. S., Legerski, R., Schultz, R. A., Stefanini, M., Lehmann, A. R., Mayne, L. V., et al. (1995). The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. Cell 82, 555-564.

Henricksen, L. A., Umbricht, C. B., and Wold, M. S. (1994). Recombinant replication protein A: expression, complex formation, and functional characterization. J Biol Chem 269, 11121-11132.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319, 209-227.

Hess, M. T., Schwitter, U., Petretta, M., Giese, B., and Naegeli, H. (1997). Bipartite substrate discrimination by human nucleotide excision repair. Proc Natl Acad Sci U S A 94, 6664-6669.

Hey, T., Lipps, G., Sugasawa, K., Iwai, S., Hanaoka, F., and Krauss, G. (2002). The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. Biochemistry 41, 6583-6587.

Heyer, W. D., Ehmsen, K. T., and Solinger, J. A. (2003). Holliday junctions in the eukaryotic nucleus: resolution in sight? Trends Biochem Sci 28, 548-557.

Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. Nature 411, 366-374.

Hohl, M., Thorel, F., Clarkson, S. G., and Scharer, O. D. (2003). Structural determinants for substrate binding and catalysis by the structure-specific endonuclease XPG. J Biol Chem 278, 19500-19508.

Hollingsworth, N. M., and Brill, S. J. (2004). The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. Genes Dev 18, 117-125.

Hollis, T., Ichikawa, Y., and Ellenberger, T. (2000). DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA. Embo J 19, 758-766.

Hoogstraten, D., Nigg, A. L., Heath, H., Mullenders, L. H., van Driel, R., Hoeijmakers, J. H., Vermeulen, W., and Houtsmuller, A. B. (2002). Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. Mol Cell 10, 1163-1174.

Houtsmuller, A. B., Rademakers, S., Nigg, A. L., Hoogstraten, D., Hoeijmakers, J. H., and Vermeulen, W. (1999). Action of DNA repair endonuclease ERCC1/XPF in living cells. Science 284, 958-961.

Huang, J. C., and Sancar, A. (1994). Determination of minimum substrate size for human excinuclease. J Biol Chem 269, 19034-19040.

Huang, J. C., Svoboda, D. L., Reardon, J. T., and Sancar, A. (1992). Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. Proc Natl Acad Sci U S A 89, 3664-3668.

Huang, J. C., Zamble, D. B., Reardon, J. T., Lippard, S. J., and Sancar, A. (1994). HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease. Proc Natl Acad Sci U S A 91, 10394-10398.

Hwang, B. J., Ford, J. M., Hanawalt, P. C., and Chu, G. (1999). Expression of the p48 xeroderma pigmentosum gene is p53-dependent and is involved in global genomic repair. Proc Natl Acad Sci U S A 96, 424-428.

Hwang, J. R., Moncollin, V., Vermeulen, W., Seroz, T., van Vuuren, H., Hoeijmakers, J. H., and Egly, J. M. (1996). A 3' --> 5' XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription. J Biol Chem 271, 15898-15904.

Hwang, K. Y., Baek, K., Kim, H. Y., and Cho, Y. (1998). The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. Nat Struct Biol 5, 707-713.

Ikegami, T., Kuraoka, I., Saijo, M., Kodo, N., Kyogoku, Y., Morikawa, K., Tanaka, K., and Shirakawa, M. (1998). Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. Nat Struct Biol 5, 701-706.

Itin, P. H., Sarasin, A., and Pittelkow, M. R. (2001). Trichothiodystrophy: update on the sulfur-deficient brittle hair syndromes. J Am Acad Dermatol 44, 891-920.

Jamieson, E. R., and Lippard, S. J. (1999). Structure, Recognition, and Processing of Cisplatin-DNA Adducts. Chem Rev 99, 2467-2498.

Jaspers, N. G., Raams, A., Silengo, M. C., Wijgers, N., Niedernhofer, L. J., Robinson, A. R., Giglia-Mari, G., Hoogstraten, D., Kleijer, W. J., Hoeijmakers, J. H., et al. (2007). First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. Am J Hum Genet 80, 457-466.

Jones, C. J., and Wood, R. D. (1993). Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA. Biochemistry 32, 12096-12104.

Jonker, H. R., Wechselberger, R. W., Boelens, R., Folkers, G. E., and Kaptein, R. (2005). Structural properties of the promiscuous VP16 activation domain. Biochemistry 44, 827-839.

Kartalou, M., and Essigmann, J. M. (2001). Recognition of cisplatin adducts by cellular proteins. Mutat Res 478, 1-21.

Keeney, S., Chang, G. J., and Linn, S. (1993). Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. J Biol Chem 268, 21293-21300.

Kim, J. K., and Choi, B. S. (1995). The solution structure of DNA duplex-decamer containing the (6-4) photoproduct of thymidylyl(3'-->5')thymidine by NMR and relaxation matrix refinement. Eur J Biochem 228, 849-854.

Klungland, A., Hoss, M., Gunz, D., Constantinou, A., Clarkson, S. G., Doetsch, P. W., Bolton, P. H., Wood, R. D., and Lindahl, T. (1999). Base excision repair of oxidative DNA damage activated by XPG protein. Mol Cell 3, 33-42.

Kong, S. E., and Svejstrup, J. Q. (2002). Incision of a 1,3-intrastrand d(GpTpG)-cisplatin adduct by nucleotide excision repair proteins from yeast. DNA Repair (Amst) 1, 731-741.

Kowalczykowski, S. C. (2000). Some assembly required. Nat Struct Biol 7, 1087-1089.

Kraemer, K. H. (1997). Sunlight and skin cancer: another link revealed. Proc Natl Acad Sci U S A 94, 11-14.

Kuraoka, I., Morita, E. H., Saijo, M., Matsuda, T., Morikawa, K., Shirakawa, M., and Tanaka, K. (1996). Identification of a damaged-DNA binding domain of the XPA protein. Mutat Res 362, 87-95.

Kusumoto, R., Masutani, C., Sugasawa, K., Iwai, S., Araki, M., Uchida, A., Mizukoshi, T., and Hanaoka, F. (2001). Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. Mutat Res 485, 219-227.

Laine, J. P., and Egly, J. M. (2006). When transcription and repair meet: a complex system. Trends Genet 22, 430-436.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystalography 26, 283-291.

Lee, J. H., Park, C. J., Arunkumar, A. I., Chazin, W. J., and Choi, B. S. (2003). NMR study on the interaction between RPA and DNA decamer containing cis-syn cyclobutane pyrimidine dimer in the presence of XPA: implication for damage verification and strand-specific dual incision in nucleotide excision repair. Nucleic Acids Res 31, 4747-4754.

Lee, J. H., Park, C. J., Shin, J. S., Ikegami, T., Akutsu, H., and Choi, B. S. (2004). NMR structure of the DNA decamer duplex containing double T*G mismatches of cis-syn cyclobutane pyrimidine dimer: implications for DNA damage recognition by the XPC-hHR23B complex. Nucleic Acids Res 32, 2474-2481.

Legerski, R., and Peterson, C. (1992). Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. Nature 359, 70-73.

Li, L., Elledge, S. J., Peterson, C. A., Bales, E. S., and Legerski, R. J. (1994). Specific association between the human DNA repair proteins XPA and ERCC1. Proc Natl Acad Sci U S A 91, 5012-5016.

Li, L., Lu, X., Peterson, C. A., and Legerski, R. J. (1995). An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. Mol Cell Biol 15, 5396-5402.

Li, L., Peterson, C. A., Lu, X., and Legerski, R. J. (1995). Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. Mol Cell Biol 15, 1993-1998.

Lieber, M. R. (1997). The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. Bioessays 19, 233-240.

Lindahl, T., and Wood, R. D. (1999). Quality control by DNA repair. Science 286, 1897-1905.

Lisby, M., Mortensen, U. H., and Rothstein, R. (2003). Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. Nat Cell Biol 5, 572-577.

Lynch, M., and Conery, J. S. (2003). The evolutionary demography of duplicate genes. J Struct Funct Genomics 3, 35-44.

MacInnes, M. A., Dickson, J. A., Hernandez, R. R., Learmonth, D., Lin, G. Y., Mudgett, J. S., Park, M. S., Schauer, S., Reynolds, R. J., Strniste, G. F., et al. (1993). Human ERCC5 cDNA-cosmid complementation for excision repair and bipartite amino acid domains conserved with RAD proteins of Saccharomyces cerevisiae and Schizosaccharomyces pombe. Mol Cell Biol 13, 6393-6402.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature 399, 700-704.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P. J., Bootsma, D., et al. (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. Embo J 13, 1831-1843.

Matsumura, Y., Nishigori, C., Yagi, T., Imamura, S., and Takebe, H. (1998). Characterization of molecular defects in xeroderma pigmentosum group F in relation to its clinically mild symptoms. Hum Mol Genet 7, 969-974.

McAteer, K., Jing, Y., Kao, J., Taylor, J. S., and Kennedy, M. A. (1998). Solution-state structure of a DNA dodecamer duplex containing a Cis-syn thymine cyclobutane dimer, the major UV photoproduct of DNA. J Mol Biol 282, 1013-1032.

Min, J. H., and Pavletich, N. P. (2007). Recognition of DNA damage by the Rad4 nucleotide excision repair protein. Nature 449, 570-575.

Missura, M., Buterin, T., Hindges, R., Hubscher, U., Kasparkova, J., Brabec, V., and Naegeli, H. (2001). Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair. Embo J 20, 3554-3564.

Mitchell, D. L. (1988). The relative cytotoxicity of (6-4) photoproducts and cyclobutane dimers in mammalian cells. Photochem Photobiol 48, 51-57.

Moggs, J. G., and Almouzni, G. (1999). Chromatin rearrangements during nucleotide excision repair. Biochimie 81, 45-52.

Mone, M. J., Bernas, T., Dinant, C., Goedvree, F. A., Manders, E. M., Volker, M., Houtsmuller, A. B., Hoeijmakers, J. H., Vermeulen, W., and van Driel, R. (2004). In vivo dynamics of chromatin-associated complex formation in mammalian nucleotide excision repair. Proc Natl Acad Sci U S A 101, 15933-15937.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. Proteins 12, 345-364.

Mu, D., Hsu, D. S., and Sancar, A. (1996). Reaction mechanism of human DNA repair excision nuclease. J Biol Chem 271, 8285-8294.

Mu, D., Park, C. H., Matsunaga, T., Hsu, D. S., Reardon, J. T., and Sancar, A. (1995). Reconstitution of human DNA repair excision nuclease in a highly defined system. J Biol Chem 270, 2415-2418.

Munoz, P., Blanco, R., Flores, J. M., and Blasco, M. A. (2005). XPF nuclease-dependent telomere loss and increased DNA damage in mice overexpressing TRF2 result in premature aging and cancer. Nat Genet 37, 1063-1071.

Nance, M. A., and Berry, S. A. (1992). Cockayne syndrome: review of 140 cases. Am J Med Genet 42, 68-84.

Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., et al. (2005). RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59, 662-672.

Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P. P., White, M. F., and McDonald, N. Q. (2005). Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. Embo J 24, 895-905.

Ng, J. M., Vermeulen, W., van der Horst, G. T., Bergink, S., Sugasawa, K., Vrieling, H., and Hoeijmakers, J. H. (2003). A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. Genes Dev 17, 1630-1645.

Nicholls, A. (1993). GRASP: Graphical Representation and Analysis of Surface Properties. (New York: Columbia University).

Nichols, A. F., and Sancar, A. (1992). Purification of PCNA as a nucleotide excision repair protein. Nucleic Acids Res 20, 2441-2446.

Niedernhofer, L. J., Essers, J., Weeda, G., Beverloo, B., de Wit, J., Muijtjens, M., Odijk, H., Hoeijmakers, J. H., and Kanaar, R. (2001). The structure-specific endonuclease Ercc1-Xpf is required for targeted gene replacement in embryonic stem cells. Embo J 20, 6540-6549.

Niedernhofer, L. J., Garinis, G. A., Raams, A., Lalai, A. S., Robinson, A. R., Appeldoorn, E., Odijk, H., Oostendorp, R., Ahmad, A., van Leeuwen, W., et al. (2006). A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. Nature 444, 1038-1043.

Niedernhofer, L. J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A. F., de Wit, J., Jaspers, N. G., Beverloo, H. B., Hoeijmakers, J. H., et al. (2004). The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. Mol Cell Biol 24, 5776-5787.

Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., and Hanaoka, F. (2005). Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. Mol Cell Biol 25, 5664-5674.

Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2003). X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. Structure (Camb) 11, 445-457.

Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2005a). Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. Structure 13, 1183-1192.

Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., and Morikawa, K. (2005b). Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing. Structure 13, 143-153.

Nishino, T., and Morikawa, K. (2002). Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. Oncogene 21, 9022-9032.

O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C., and Wood, R. D. (1994). XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. Nature 371, 432-435.

O'Donovan, A., and Wood, R. D. (1993). Identical defects in DNA repair in xeroderma pigmentosum group G and rodent ERCC group 5. Nature 363, 185-188.

Ogi, T., and Lehmann, A. R. (2006). The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair. Nat Cell Biol 8, 640-642.

Ortolan, T. G., Tongaonkar, P., Lambertson, D., Chen, L., Schauber, C., and Madura, K. (2000). The DNA repair protein rad23 is a negative regulator of multi-ubiquitin chain assembly. Nat Cell Biol 2, 601-608.

Osman, F., Dixon, J., Doe, C. L., and Whitby, M. C. (2003). Generating crossovers by resolution of nicked Holliday junctions: a role for Mus81-Eme1 in meiosis. Mol Cell 12, 761-774.

Podust, V. N., Georgaki, A., Strack, B., and Hubscher, U. (1992). Calf thymus RF-C as an essential component for DNA polymerase delta and epsilon holoenzymes function. Nucleic Acids Res 20, 4159-4165.

Prasher, J. M., Lalai, A. S., Heijmans-Antonissen, C., Ploemacher, R. E., Hoeijmakers, J. H., Touw, I. P., and Niedernhofer, L. J. (2005). Reduced hematopoietic reserves in DNA interstrand crosslink repair-deficient Ercc1(-/-) mice. Embo J 24, 861-871.

Rademakers, S., Volker, M., Hoogstraten, D., Nigg, A. L., Mone, M. J., Van Zeeland, A. A., Hoeijmakers, J. H., Houtsmuller, A. B., and Vermeulen, W. (2003). Xeroderma pigmentosum group A protein loads as a separate factor onto DNA lesions. Mol Cell Biol 23, 5755-5767.

Reardon, J. T., Nichols, A. F., Keeney, S., Smith, C. A., Taylor, J. S., Linn, S., and Sancar, A. (1993). Comparative analysis of binding of human damaged DNA-binding protein (XPE) and Escherichia coli damage recognition protein (UvrA) to the major ultraviolet photoproducts: T[c,s]T, T[t,s]T, T[6-4]T, and T[Dewar]T. J Biol Chem 268, 21301-21308.

Ridgway, P., and Almouzni, G. (2000). CAF-1 and the inheritance of chromatin states: at the crossroads of DNA replication and repair. J Cell Sci 113 (Pt 15), 2647-2658.

Riedl, T., Hanaoka, F., and Egly, J. M. (2003). The comings and goings of nucleotide excision repair factors on damaged DNA. Embo J 22, 5293-5303.

Roberts, J. A., Bell, S. D., and White, M. F. (2003). An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. Mol Microbiol 48, 361-371.

Roberts, J. A., and White, M. F. (2005). An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. J Biol Chem 280, 5924-5928.

Rodriguez, K., Talamantez, J., Huang, W., Reed, S. H., Wang, Z., Chen, L., Feaver, W. J., Friedberg, E. C., and Tomkinson, A. E. (1998). Affinity purification and partial characterization of a yeast multiprotein complex for nucleotide excision repair using histidine-tagged Rad14 protein. J Biol Chem 273, 34180-34189.

Roth, C., Rastogi, S., Arvestad, L., Dittmar, K., Light, S., Ekman, D., and Liberles, D. A. (2007). Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J Exp Zoolog B Mol Dev Evol 308, 58-73.

Rubbi, C. P., and Milner, J. (2003). p53 is a chromatin accessibility factor for nucleotide excision repair of DNA damage. Embo J 22, 975-986.

Sancar, A. (2003). Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. Chem Rev 103, 2203-2237.

Sarasin, A. (2003). An overview of the mechanisms of mutagenesis and carcinogenesis. Mutat Res 544, 99-106.

Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997). Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry 36, 11205-11215.

Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J. H., and Egly, J. M. (1994). The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. Embo J 13, 2388-2392.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P., and Egly, J. M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science 260, 58-63.

Scherly, D., Nouspikel, T., Corlet, J., Ucla, C., Bairoch, A., and Clarkson, S. G. (1993). Complementation of the DNA repair defect in xeroderma pigmentosum group G cells by a human cDNA related to yeast RAD2. Nature 363, 182-185.

Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J. M. (2000). Molecular structure of human TFIIH. Cell 102, 599-607.

Schweizer, U., Hey, T., Lipps, G., and Krauss, G. (1999). Photocrosslinking locates a binding site for the large subunit of human replication protein A to the damaged strand of cisplatin-modified DNA. Nucleic Acids Res 27, 3183-3189.

Segers-Nolten, G. M., Wyman, C., Wijgers, N., Vermeulen, W., Lenferink, A. T., Hoeijmakers, J. H., Greve, J., and Otto, C. (2002). Scanning confocal fluorescence microscopy for single molecule analysis of nucleotide excision repair complexes. Nucleic Acids Res 30, 4720-4727.

Shao, X., and Grishin, N. V. (2000). Common fold in helix-hairpin-helix proteins. Nucleic Acids Res 28, 2643-2650.

Shivji, K. K., Kenny, M. K., and Wood, R. D. (1992). Proliferating cell nuclear antigen is required for DNA excision repair. Cell 69, 367-374.

Shivji, M. K., Eker, A. P., and Wood, R. D. (1994). DNA repair defect in xeroderma pigmentosum group C and complementing factor from HeLa cells. J Biol Chem 269, 22749-22757.

Shivji, M. K., Podust, V. N., Hubscher, U., and Wood, R. D. (1995). Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. Biochemistry 34, 5011-5017.

Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y. F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., et al. (1996). Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell 86, 811-822.

Sijbers, A. M., van der Spek, P. J., Odijk, H., van den Berg, J., van Duin, M., Westerveld, A., Jaspers, N. G., Bootsma, D., and Hoeijmakers, J. H. (1996b). Mutational analysis of the human nucleotide excision repair gene ERCC1. Nucleic Acids Res 24, 3370-3380.

Sijbers, A. M., van Voorst Vader, P. C., Snoek, J. W., Raams, A., Jaspers, N. G., and Kleijer, W. J. (1998). Homozygous R788W point mutation in the XPF gene of a patient with xeroderma pigmentosum and late-onset neurologic disease. J Invest Dermatol 110, 832-836.

Simon, G. R., Sharma, S., Cantor, A., Smith, P., and Bepler, G. (2005). ERCC1 expression is a predictor of survival in resected patients with non-small cell lung cancer. Chest 127, 978-983.

Singh, S., Folkers, G. E., Bonvin, A. M., Boelens, R., Wechselberger, R., Niztayev, A., and Kaptein, R. (2002). Solution structure and DNA-binding properties of the C-terminal domain of UvrC from E.coli. Embo J 21, 6257-6266.

Smerdon, M. J. (1991). DNA repair and the role of chromatin structure. Curr Opin Cell Biol 3, 422-428.

Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., van der Spek, P. J., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1996). HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro. Mol Cell Biol 16, 4852-4861.

Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell 2, 223-232.

Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., and Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. Genes Dev 15, 507-521.

Sugasawa, K., Shimizu, Y., Iwai, S., and Hanaoka, F. (2002). A molecular mechanism for DNA damage recognition by the xeroderma pigmentosum group C protein complex. DNA Repair (Amst) 1, 95-107.

Svejstrup, J. Q., Wang, Z., Feaver, W. J., Wu, X., Bushnell, D. A., Donahue, T. F., Friedberg, E. C., and Kornberg, R. D. (1995). Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. Cell 80, 21-28.

Szymkowski, D. E., Lawrence, C. W., and Wood, R. D. (1993). Repair by human cell extracts of single (6-4) and cyclobutane thymine-thymine photoproducts in DNA. Proc Natl Acad Sci U S A 90, 9823-9827.

Syvaoja, J., Suomensaari, S., Nishida, C., Goldsmith, J. S., Chui, G. S., Jain, S., and Linn, S. (1990). DNA polymerases alpha, delta, and epsilon: three distinct enzymes from HeLa cells. Proc Natl Acad Sci U S A 87, 6664-6668.

Szymkowski, D. E., Yarema, K., Essigmann, J. M., Lippard, S. J., and Wood, R. D. (1992). An intrastrand d(GpG) platinum crosslink in duplex M13 DNA is refractory to repair by human cell extracts. Proc Natl Acad Sci U S A 89, 10772-10776.

Takao, M., Abramic, M., Moos, M., Jr., Otrin, V. R., Wootton, J. C., McLenigan, M., Levine, A. S., and Protic, M. (1993). A 127 kDa component of a UV-damaged DNA-binding complex, which is defective in some xeroderma pigmentosum group E patients, is homologous to a slime mold protein. Nucleic Acids Res 21, 4111-4118.

Tanaka, K., Miura, N., Satokata, I., Miyamoto, I., Yoshida, M. C., Satoh, Y., Kondo, S., Yasui, A., Okayama, H., and Okada, Y. (1990). Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. Nature 348, 73-76.

Tanaka, K., Satokata, I., Ogita, Z., Uchida, T., and Okada, Y. (1989). Molecular cloning of a mouse DNA repair gene that complements the defect of group-A xeroderma pigmentosum. Proc Natl Acad Sci U S A 86, 5512-5516.

Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Scharer, O. D., Coin, F., Coulombe, B., and Egly, J. M. (2004). Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. J Biol Chem 279, 19074-19083.

Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. Embo J 14, 4108-4120.

Thoma, F. (1999). Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. Embo J 18, 6585-6598.

Thompson, L. H., Brookman, K. W., Weber, C. A., Salazar, E. P., Reardon, J. T., Sancar, A., Deng, Z., and Siciliano, M. J. (1994). Molecular cloning of the human nucleotide-excision-repair gene ERCC4. Proc Natl Acad Sci U S A 91, 6855-6859.

Thorel, F., Constantinou, A., Dunand-Sauthier, I., Nouspikel, T., Lalle, P., Raams, A., Jaspers, N. G., Vermeulen, W., Shivji, M. K., Wood, R. D., et al. (2004). Definition of a short region of XPG necessary for TFIIH interaction and stable recruitment to sites of UV damage. Mol Cell Biol 24, 10670-10680.

Tian, M., Shinkura, R., Shinkura, N., and Alt, F. W. (2004). Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF. Mol Cell Biol 24, 1200-1205.

Tirode, F., Busso, D., Coin, F., and Egly, J. M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol Cell 3, 87-95.

Tripsianes, K., Folkers, G.E., AB, E., Das, D., Odijk, H., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R., and Boelens, R. (2005). The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. Structure 13, 1849-1858.

Tripsianes, K., Folkers, G. E., Zheng, C., Das, D., Grinstead, J. S., Kaptein, R., and Boelens, R. (2007). Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization. Nucleic Acids Res 35, 5789-5798.

Troelstra, C., Odijk, H., de Wit, J., Westerveld, A., Thompson, L. H., Bootsma, D., and Hoeijmakers, J. H. (1990). Molecular cloning of the human DNA excision repair gene ERCC-6. Mol Cell Biol 10, 5806-5813.

Tsodikov, O. V., Enzlin, J. H., Scharer, O. D., and Ellenberger, T. (2005). Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. Proc Natl Acad Sci U S A 102, 11236-11241.

Ura, K., Araki, M., Saeki, H., Masutani, C., Ito, T., Iwai, S., Mizukoshi, T., Kaneda, Y., and Hanaoka, F. (2001). ATP-dependent chromatin remodeling facilitates nucleotide excision repair of UV-induced DNA lesions in synthetic dinucleosomes. Embo J 20, 2004-2014.

Venema, J., Mullenders, L. H. F., Natarajan, A. T., Zeeland, A. A. V., and Mayne, L. V. (1990). The Genetic Defect in Cockayne Syndrome is Associated with a Defect in Repair of UV-Induced DNA Damage in Transcriptionally Active DNA. Proc Natl Acad Sci U S A 87, 4707-4711.

Volker, M., Mone, M. J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, A. A., and Mullenders, L. H. (2001). Sequential assembly of the nucleotide excision repair factors in vivo. Mol Cell 8, 213-224.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J Mol Graph 8, 52-56, 29.

Wakasugi, M., Kawashima, A., Morioka, H., Linn, S., Sancar, A., Mori, T., Nikaido, O., and Matsunaga, T. (2002). DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. J Biol Chem 277, 1637-1640.

Wakasugi, M., and Sancar, A. (1998). Assembly, subunit composition, and footprint of human DNA repair excision nuclease. Proc Natl Acad Sci U S A 95, 6669-6674.

Wakasugi, M., and Sancar, A. (1999). Order of assembly of human DNA repair excision nuclease. J Biol Chem 274, 18759-18768.

Weber, C. A., Salazar, E. P., Stewart, S. A., and Thompson, L. H. (1988). Molecular cloning and biological characterization of a human gene, ERCC2, that corrects the nucleotide excision repair defect in CHO UV5 cells. Mol Cell Biol 8, 1137-1146.

Weeda, G., Donker, I., de Wit, J., Morreau, H., Janssens, R., Vissers, C. J., Nigg, A., van Steeg, H., Bootsma, D., and Hoeijmakers, J. H. (1997). Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. Curr Biol 7, 427-439.

Weeda, G., van Ham, R. C., Masurel, R., Westerveld, A., Odijk, H., de Wit, J., Bootsma, D., van der Eb, A. J., and Hoeijmakers, J. H. (1990). Molecular cloning and biological characterization of the human excision repair gene ERCC-3. Mol Cell Biol 10, 2570-2581.

Westerveld, A., Hoeijmakers, J. H., van Duin, M., de Wit, J., Odijk, H., Pastink, A., Wood, R. D., and Bootsma, D. (1984). Molecular cloning of a human DNA repair gene. Nature 310, 425-429.

White, M. F. (2003). Archaeal DNA repair: paradigms and puzzles. Biochem Soc Trans 31, 690-693.

Winkler, G. S., Araujo, S. J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J. M., Hoeijmakers, J. H., Wood, R. D., Timmers, H. T., and Weeda, G. (2000). TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. J Biol Chem 275, 4258-4266.

Wold, M. S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem 66, 61-92.

Wolffe, A. P. (1997). Chromatin: Structure and Function. Academic Press, New York, NY.

Wood, R. D. (1997). Nucleotide excision repair in mammalian cells. J Biol Chem 272, 23465-23468.

Wood, R. D., Robins, P., and Lindahl, T. (1988). Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. Cell 53, 97-106.

Yamada, K., Ariyoshi, M., and Morikawa, K. (2004). Three-dimensional structural views of branch migration and resolution in DNA homologous recombination. Curr Opin Struct Biol 14, 130-137.

Zamble, D. B., Mu, D., Reardon, J. T., Sancar, A., and Lippard, S. J. (1996). Repair of cisplatin--DNA adducts by the mammalian excision nuclease. Biochemistry 35, 10004-10013.

Zhu, X. D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H., and de Lange, T. (2003). ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Mol Cell 12, 1489-1498.

Zotter, A., Luijsterburg, M. S., Warmerdam, D. O., Ibrahim, S., Nigg, A., van Cappellen, W. A., Hoeijmakers, J. H., van Driel, R., Vermeulen, W., and Houtsmuller, A. B. (2006). Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced dna damage depends on functional TFIIH. Mol Cell Biol 26, 8868-8879.

Summary

The maintenance of DNA integrity is vital for the viability of cells and the health of organisms. The genome is under constant attack from endogenous metabolic byproducts and environmental factors that can alter its chemical structure and corrupt its encoded message. For tolerating and repairing various types of damage cells have an arsenal of ways of responding to such injury. A unique caretaker pathway is Nucleotide Excision Repair (NER) because of its ability to recognize and repair a bewildering array of physically dissimilar lesions that distort the DNA helix and impair base pair stacking. NER is a multiprotein system that operates through a 'cut and patch' mechanism. It is organized in three fundamental steps: (1) DNA damage recognition, (2) DNA damage excision and (3) gap-filling synthesis. **Chapter 1** reviews the research that linked rare genetic diseases to NER deficiency, the subsequent identification of the genes involved in NER, and the advances in modern biology that led in understanding the molecular mechanisms of such a complex repair system that involves over 30 proteins.

An indispensable component of the NER pathway is the ERCC1/XPF structure-specific endonuclease, which is responsible for the cleavage of the damaged DNA strand on the 5′ side of the lesion during the excision step. The aim of the current thesis was to address the structure-function relationship of the obligate ERCC1/XPF heterodimer. In this direction, **Chapter 2** describes the structure of ERCC1/XPF C-terminal interacting domains determined by NMR. ERCC1 and XPF partners share the same fold, the double helix-hairpin-helix motif (HhH_2), and complex in a symmetry related manner. The same dimer interface is observed in the archaeal XPF, which forms homodimers. The archaeal HhH_2 domains bind equally to DNA minor grooves, in agreement with the substrate specificity of the archaeal XPF nuclease. In contrast, human ERCC1/XPF cleaves relevant DNA substrates, which consist of one minor groove only. We have shown that exclusively the HhH_2 domain of ERCC1 mediates minor groove binding. The functional asymmetry in humans, revealed by the structure, is related to the altered second hairpin of XPF. However, the XPF domain acts as scaffold for the folding of the ERCC1 corresponding domain. We concluded that ERCC1 possesses the DNA binding activity that serves to localize the XPF nuclease activity to the NER pre-incision complex and allow for strand incision.

In **Chapter 3** we have determined the structure of the ERCC1 central domain and investigated its contribution to the heterodimeric function. Although this domain shows structural homology with the catalytically active XPF nuclease domain, it has a distinct function by performing interactions with XPA, another NER protein. This function is crucial as it anchors the XPF nuclease in the pre-incision NER complex and subsequently allows cleavage to occur. Remarkably, the XPA binding by ERCC1 and the catalytic function

of XPF are dependent on structurally homologous regions. The disparity in function is shaped by the distinct chemical properties of these regions. Furthermore the ERCC1 central domain can bind ss-DNA. XPA and DNA interactions can happen simultaneously through distinct surfaces and indicate the significant dual role of the ERCC1 central domain in targeting the XPF nuclease to the sites of DNA damage.

Overall, the structural data from human and archaeal species show that the human ERCC1 and XPF proteins share the architectural subunits of the archaeal XPF homodimer. The structural similarities provide evidence for the common origin of the human proteins. Based on our findings in **Chapter 2** and **3**, we proposed that ERCC1 and XPF genes evolved by subfunctionalization, which resulted in the partitioning of the original functions after duplication of the ancestral XPF gene. From the functions present in the archaeal XPF homodimer, human ERCC1 retained the canonical HhH_2 domain that acts as a DNA-binding domain, while human XPF retained the catalytic activity. Once the separation of the ancestral subfunctions occurred in ERCC1 and XPF genes, only the heterodimeric protein complex could restore the original function. This evolutionary scenario is able to explain the preservation of the ERCC1 gene after XPF gene duplication and the concomitant transition from homo- to obligate hetero-association in eukaryotes.

Finally, in **Chapter 4** we have investigated the first reported ERCC1 deficiency in humans. This unique patient carried a single point mutation in the C-terminal domain of ERCC1, which is essential for the interaction with XPF, and exhibited severe clinical symptoms. The structure of the mutant heterodimer retains the global features of the ERCC1/XPF interaction that have been observed in the wild type architecture. However, the ERCC1 mutation disturbs locally the hydrophobic packing between ERCC1 and XPF and reduces the stability of the heterodimer. Due to these changes, ERCC1 rearranges its hydrophobic core and alters the geometry of the second hairpin. According to our findings in the mode of function of the wild type heterodimer, any distortion in the proper geometry of the ERCC1 hairpins will impair the symmetric contacts with the phosphate backbone of the DNA. It is therefore likely that the low enzymatic activity of the mutant ERCC1/XPF accounting for the severe developmental abnormalities of the patient is due to the lower stability of the heterodimer and possibly poor DNA binding by the ERCC1 domain.

Samenvatting

Het behoud van de integriteit van het DNA is van het grootste belang voor levende cellen en dus voor de gezondheid van elk organisme. Het genoom is echter constant onderhevig aan schadelijke invloeden van externe factoren, zoals UV licht en lichaamsvreemde chemische stoffen, maar ook van schadelijke endogene (lichaamseigen) bijproducten van het metabolisme. Deze factoren kunnen de chemische structuur van het DNA veranderen en zo fouten introduceren in de genetische code. Om te voorkomen dat zulke fouten de celfunctie nadelig beinvloeden beschikken cellen over een arsenaal aan mogelijkheden om het DNA te repareren of om om te gaan met beschadigd DNA. Een uniek mechanisme in dit kader is de Nucleotide Excisie Reparatie (Nucleotide Excision Repair; NER). Deze route kan gebruikt worden voor herkenning en reparatie van een breed scala aan beschadigingen die leiden tot vervorming van de karakteristieke DNA-helix structuur en die de stapeling van de basen, de bouwstenen van het DNA, verstoren. Het NER systeem bestaat uit verschillende eiwitten die opereren volgens een 'knip en plak' mechanisme en bestaat uit drie fundamentele stappen: (1) Herkenning van schade aan het DNA, (2) verwijdering van he beschadigde deel van het DNA en (3) synthese van DNA ter opvulling van het onstane gat.In **Hoofdstuk 1** van dit proefschrift wordt een overzicht gegeven van onderzoek waarin een link wordt gelegd tussen NER deficientie en verschillende zeldzame genetische aandoeningen en de hierop volgende identificatie van de genen die betrokken zijn bij NER. Tevens wordt in dit hoofdstuk aandacht besteed aan de ontwikkelingen in de biologie die hebben geleid tot inzicht in de moleculaire mechanismen van een complex reparatiesysteem als NER, waarbij meer dan 30 eiwitten zijn betrokken.

Een onvervangbare component van het NER systeem is ERCC1/XPF, een structuur-specifiek endonuclease (DNA knippend eiwit). Dit eiwit is verantwoordelijk voor het verbreken van de DNA keten aan de zogenaamde 5′ zijde van de plaats van beschadiging gedurende de excisie stap (stap 2). Het doel van het promotieonderzoek was de bepaling van de relatie tussen de structuur van de ERCC1/XPF dimeer en de functie. Als eerste stap wordt hiertoe in **Hoofdstuk 2** de structuur beschreven van de interagerende C-terminale domeinen van ERCC1/XPF, zoals deze zijn bepaald met Nucleaire Magnetische Resonantie (NMR). ERCC1 en XPF hebben beide een karakteristiek helix-haarspeld-helix (HhH_2) motief en vormen een symmetrie gerelateerde dimeer. Een vergelijkbare dimeer wordt gevormd door een XPF nuclease gevonden in archaea, hoewel deze een dimeer vormt met zichzelf: een homodimeer. In de archaeale XPF dimeer binden beide HhH_2 domeinen op gelijke wijze elk aan een zogenaamde kleine groef van het DNA, in overeenstemming met de substraat specificiteit van het archaeaal XPF nuclease. Dit contrasteert echter met de menselijke variant van ERCC1/XPF, die bindt aan een enkele kleine groef van een relevant DNA substraat, waarna dit DNA wordt geknipt. In ons onderzoek tonen wij aan

dat alleen het HhH_2 domein van ERCC1 direct betrokken is bij de binding aan de kleine groef van het DNA. De functionele asymmetrie in menselijk ERCC1/XPF, zoals die blijkt uit de structuur, blijkt met name gerelateerd aan de veranderingen in de tweede haarspeld-structuur van XPF. Hoewel de binding aan het DNA vooral wordt bepaald door het ERCC1 HhH2 domein, speelt het corresponderende XPF domein een belangrijke rol in de vouwing van dit domein. Het algemene beeld dat uit dit onderzoek naar voren komt is dat ERCC1 het vermogen heeft DNA te binden, hetgeen nodig is voor de localisatie van de XPF nuclease activiteit binnen het NER pre-incisie complex en zo de DNA incisie mogelijk maakt.

In **Hoofdstuk 3** wordt de structuur beschreven van het centrale domein van ERCC1 en de bijdrage hiervan aan de functie van de ERCC1/XPF dimeer. Hoewel de structuur overeenkomsten (homologie) vertoont met het katalytisch actieve XPF nuclease domein, is de functie totaal verschillend. Het centrale domein van ERCC1 is niet katalytisch actief, maar ondergaat interactie met een ander eiwit uit het NER systeem, XPA. Deze interactie is cruciaal, aangezien dit het XPF nuclease verankert in het pre-incisie NER complex en zo het knippen van het DNA mogelijk maakt. Opmerkelijk is dat de binding van XPA aan ERCC1 en de katalytische activiteit van XPF nuclease beide afhankelijk zijn van structureel homologe gebieden. Het verschil in functie wordt dus veroorzaakt door de specifieke chemische eigenschappen van deze gebieden. Verder bindt het centrale domein van ERCC1 aan enkel-keten DNA (single-strand DNA; ssDNA). Interacties met XPA en DNA kunnen gelijktijdig plaatsvinden door binding aan verschillende oppervlakken en geven de dubbele rol aan van het centrale domein van ERCC1 voor wat betreft het geleiden van XPF nuclease naar beschadigingen in het DNA.

De structurele overeenkomsten tussen de menselijke ERCC1/XPF dimeer en de archaeale XPF homodimeer wekken de suggestie dat deze eiwitten een gezamenlijke oorsprong hebben. Gebaseerd op de bevindingen in **Hoofdstukken 2** en **3** stellen wij voor dat de genen die coderen voor ERCC1 en XPF zijn geëvolueerd door 'subfunctionalisatie', waarbij de verdubbeling van het oorpronkelijke XPF gen is gevolgd door partitionering van de oorspronkelijke functies. Het gevolg hiervan is een specialisatie van de onderscheiden eiwitten. Van de functies die aanwezig zijn in de XPF homodimeer is in het humaan ERCC1 de DNA bindende activiteit behouden, waarbij het prototypische HhH_2 domein een sleutelrol speelt. In het humaan XPF daarentegen, is de katalytische endonuclease activiteit behouden. Door deze scheiding van de oorspronkelijke, gekoppelde functionaliteit in ERCC1 en XPF is de vorming van een heterodimeer, een complex van twee verschillende eiwitten, noodzakelijk om de totale activiteit (binding en katalyse) te behouden. Dit scenario biedt een verklaring voor het behoud van het gen voor ERCC1 volgend op de verdubbeling van het gen voor XPF, alsmede voor de daarop volgende overgang van homodimeren naar heterodimeren in eukaryoten.

Tot slot wordt in **Hoofdstuk 4** het onderzoek beschreven dat is uitgevoerd naar de eerst gemelde ERCC1 deficiëntie in mensen. Dit betrof een unieke patient met ernstige klinische symptomen. Deze patient bleek een enkele mutatie (single point mutation) te hebben in de C-terminale regio van ERCC1, die essentieel is voor de interactie met XPF. In de structuur van de heterodimeer van de ERCC1 mutant en XPF zijn de globale eigenschappen

behouden, inclusief de interactie tussen de twee eiwitten. Op lokaal niveau zijn echter verschillen waar te nemen in de pakking van de zijketens in de hydrofobe regio tussen ERCC1 en XPF, die leiden tot een verminderde stabiliteit van het complex. Dit heeft tot gevolg dat de hydrofobe kern van ERCC1 herschikt, wat uitmondt in een verandering van de structuur van de tweede lus. Experimentele gegevens wijzen uit dat iedere verandering in de lussen van ERCC1 een negatieve invloed heeft op de symmetrische contacten met de fosfaatgroepen in het DNA. Het is derhalve aannemelijk dat de verminderde stabiliteit van van de ERCC1/XPF mutant en de lagere DNA affiniteit van van ERCC1 de oorzaak zijn van de verminderde enzymatische activiteit, met als gevolg ernstige stoornissen in de ontwikkeling van de patient.

Acknowledgements

First of all, I owe my profound gratitude to my supervisors Professor dr. Rolf Boelens and Professor dr. Robert Kaptein. I am superlatively grateful to Rolf who opened the door by offering me the PhD position and supervising my studies with generous encouragement and outstanding knowledge. The measure of his contribution is that whenever one of his scientific comments particularly galled me, I found on reflection that he was right. Robert has been from the outset both critical and supportive. I thank you for your valuable assistance at difficult times and for your willingness to straighten out many sentence structure of the thesis.

Gert Folkers has been a longtime, indefatigable aide. He did research, especially on biochemical issues, checked everyone else's research, tracked down elusive facts of the project, and allowed nothing to escape his scrutiny.

I am grateful to our collaborators from the Erasmus University, Rotterdam. Jan Hoeijmakers' scientific contribution has been indispensable. I have been enormously fortunate in having discussions with Koos Jaspers. His helpful and insightful suggestions opened up exciting new perspectives for me and shaped many of the ideas presented in this work. Hanny Odijk contributed significantly by providing the precious vector I used throughout the project.

I count myself especially fortunate to have enjoyed the generous assistance of three PhD students: Devashish Das, Julia Rumanuka, and Anding Huang. I wish you great success with the accomplishment of your studies and your future plans.

In the four and a half years I have been working in the laboratory, I have been helped by so many people that I fear I may not remember them all. To everyone who provided advice or support, I am deeply indebted. In particular I would like to thank the following: Hans Wienk, Mark Daniels, Michiel Hilbers, Johan van der Zwan, Aalt-Jan van Dijk, Cyril Dominguez, Leonardus Koharudin, Roberto Kopke-Salinas, Monique Kamphuis, Klaartje Houben, Gloria Fuentes, Jeffrey Grinstead, Eiso AB, Tammo Dierks, Rob de Jong, Hugo van Ingen, Suat Ozdirekcan, Jenny Smit, Tsjerk Wassenaar, Sjoerd de Vries, Marc van Dijk, Mickael Krzeminski, and Alexander Bonvin.

I am most fortunate to be aided by such a towering cast and I can only humbly thank them; any wisdom is theirs; any mistakes my own.

...and a few words in my mother tongue. Θα ήθελα να ευχαριστήσω από τα βάθη της καρδιάς μου, τους γονείς μου Αγγελάκη και Βασιλική Τρυψιάνη, και τον αδερφό μου Χρήστο. Μητέρα, πατέρα σας ευχαριστώ για την ηθική και ψυχολογική υποστήριξη που μου έχετε προσφέρει όλα αυτά τα χρόνια. Αγαπημένε μου αδερφέ σου εύχομαι ένα μελωδικό και δημιουργικό μέλλον.

Last, but first, I must lovingly thank my girlfriend Vassia Arvaniti for tolerating the

brooding presence of ERCC1/XPF in our lives. In this as in everything else, I owe her more than I can put into words. This book is dedicated to her with the author's love.

Curriculum Vitae

Konstantinos Tripsianes was born in Edessa, Greece, on 5 February, 1980. In 1997 he completed the high school studies in his hometown. In December 1998 he attended the Medicine School of Ioannina University, Greece, where he studied Biochemistry. His final year project was on "theoretical prediction and experimental validation of the interactions of the SH3 domain from the muscle protein nebulin" under the supervision of assistant Professor dr. Anastassia Politou. In June 2003 he started his PhD studies at the Bijvoet Center for Biomolecular Research, Utrecht University, on "structural and functional studies of the human ERCC1/XPF DNA repair complex" under the supervision of Professor dr. Rolf Boelens and Professor dr. Robert Kaptein. In December 2007 he joined Technische Universität München, Germany, to contact his post-doctoral studies under the supervision of Professor dr. Michael Sattler.

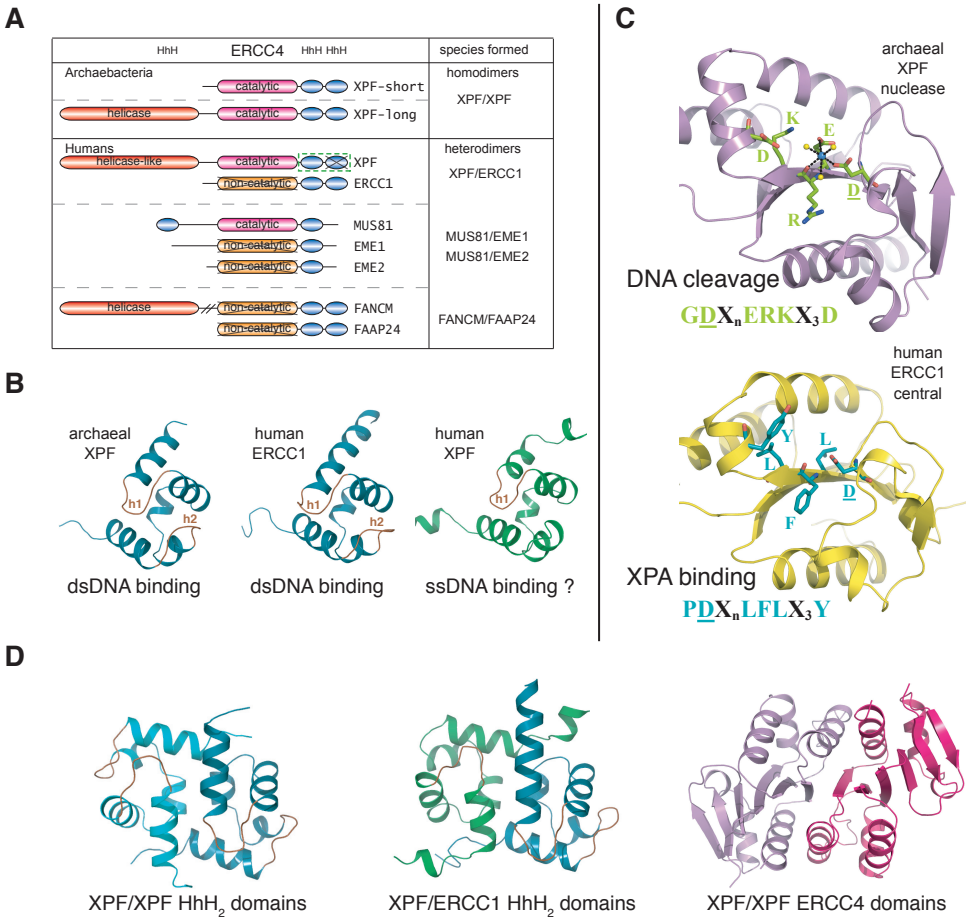
Publications

Das, D., **Tripsianes, K.**, Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., Boelens, R., and Folkers, G. E. (2007). The HhH domain of the human DNA repair protein XPF forms stable homodimers. Proteins [Epub ahead of print].

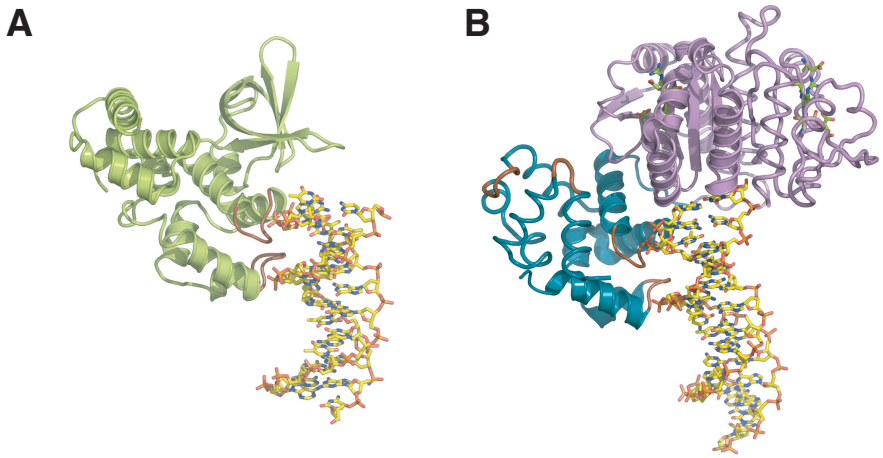
Tripsianes, K., Folkers, G. E., Zheng, C., Das, D., Grinstead, J. S., Kaptein, R., and Boelens, R. (2007). Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization. Nucleic Acids Res 35, 5789-5798.

Tripsianes, K., Folkers, G.E., Ab, E., Das, D., Odijk, H., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R., and Boelens, R. (2005). The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. Structure 13, 1849-1858.

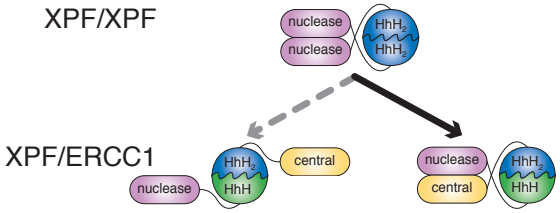
Full color figures



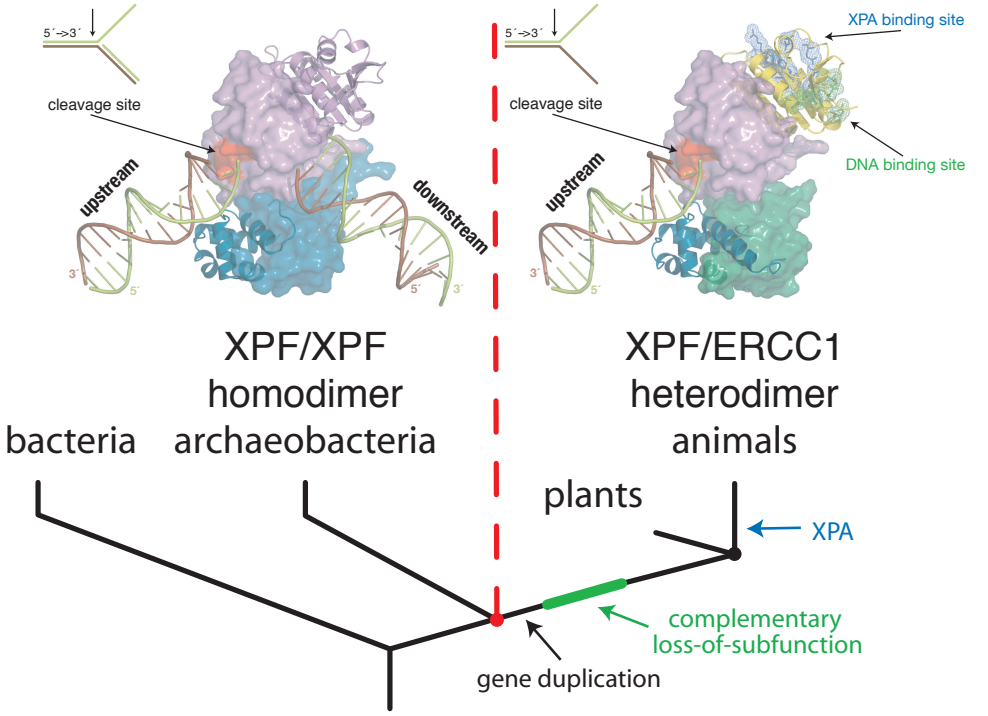
Chapter 1, figure 5. The XPF protein family. (A) Domain organization for the members of the XPF family from archaea and humans, and the dimeric species formed. ERCC4 and HhH domains are indicated at the top. (B) Fold of the HhH2 domains adapted upon dimerization. The canonical double hairpin motif binds to DNA minor groove, whereas the altered fold of the human XPF may be competent for single-stranded DNA binding. Proper hairpins are indicated and colored brown for each domain. (C) The fold of the ERCC4 domain. The conserved catalytic site from the archaeal nuclease with the coordinated divalent cation is shown in stick representation (top) and the corresponding substitutions in the human ERCC1 central domain that mediate XPA binding in stick representation as well (bottom). (D) Dimer interfaces for the HhH2 domains of the XPF homodimers and the human XPF/ERCC1 heterodimer and for the ERCC4 domains of the XPF homodimers. For the homodimers one protomer is shown in light color.



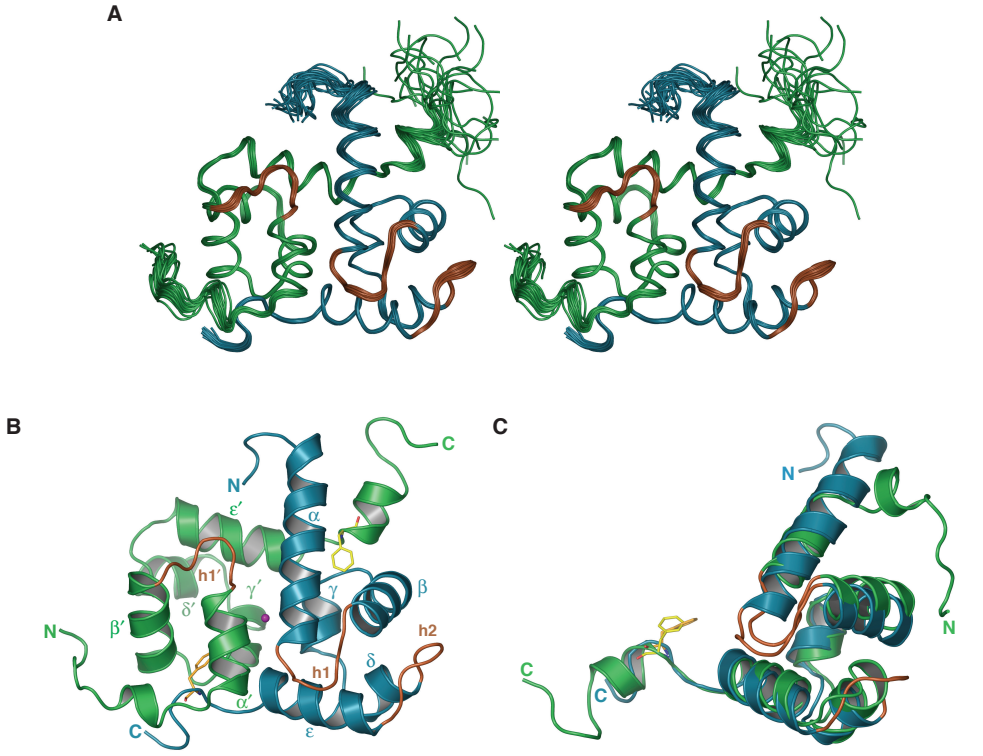
Chapter 1, figure 6. DNA recognition by the HhH_2 domains. (A) A DNA arm of the Holliday junction is recognized on the minor groove side by the two HhH motifs of one RuvA protomer. The interaction involves hydrogen bonding of the main chain amides with the phosphate oxygens of both strands (B) Similarly, one HhH_2 domain of the archaeal XPF homodimer bridges the two phosphodiester backbones across the minor groove of the DNA. In both cases, the protein-DNA interactions are built on the sugar-phosphate backbone of the duplex allowing the HhH_2 domain to recognize dsDNA in a sequence-independent manner. For the XPF homodimer, protomer A is in a cartoon and protomer B in a worm representation.



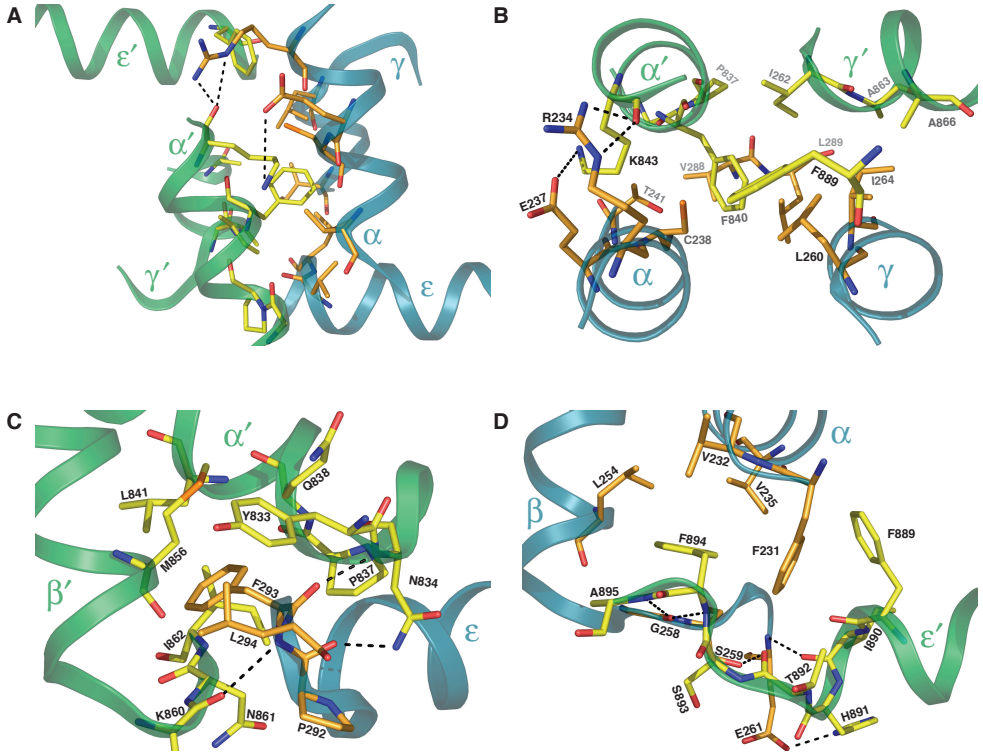
Chapter 1, figure 7. Domain organization of the archaeal homodimers and the human heterodimer with respect to the HhH_2 and ERCC4 domains.



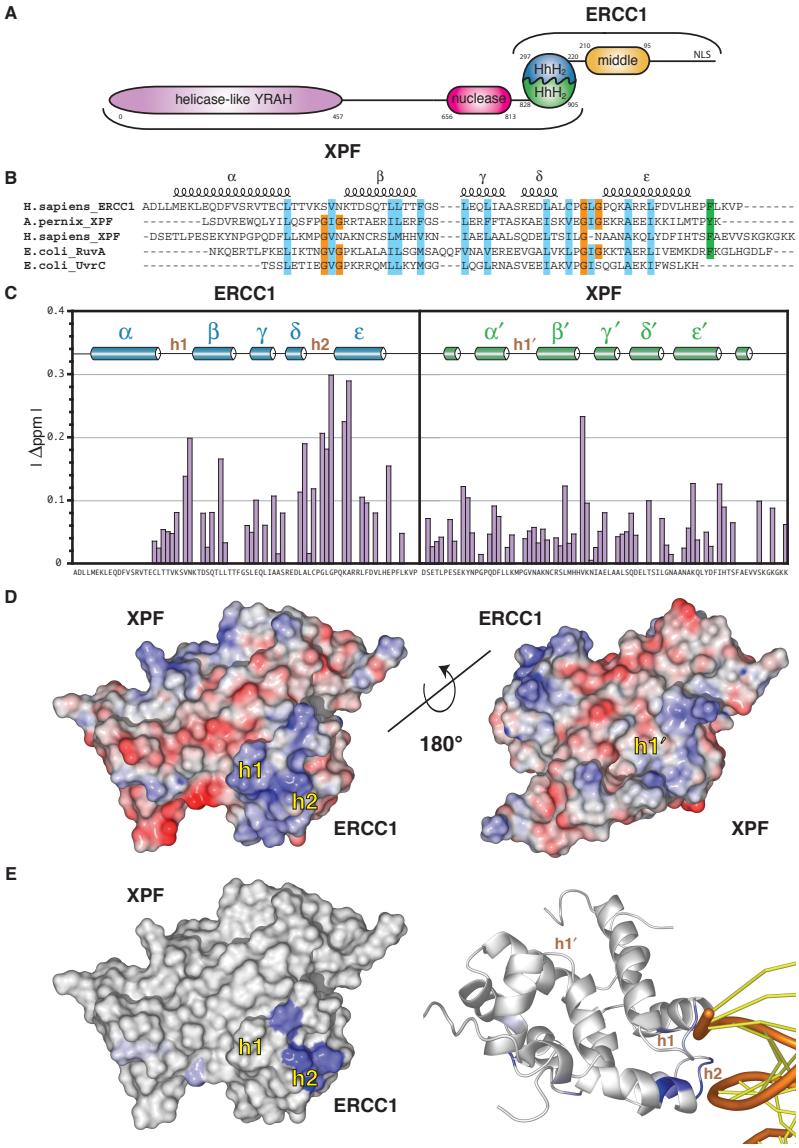
Chapter 1, figure 8. Schematic represenation for the evolution of XPF and ERCC1 genes in the eukaryal lineage. XPF gene duplication is followed by subfunctionalization resulting in obligate heterodimer formation (see text). Models for the function of homodimers and heterodimers reconstructed from the structural and functional data available for both XPF enzymes, in agreement with the basic principles of subfunctionalization (top). For the homodimer one protomer is shown in a cartoon and the other in a surface representation. Accordingly, for the heterodimer ERCC1 is in a cartoon and XPF in a surface representation.



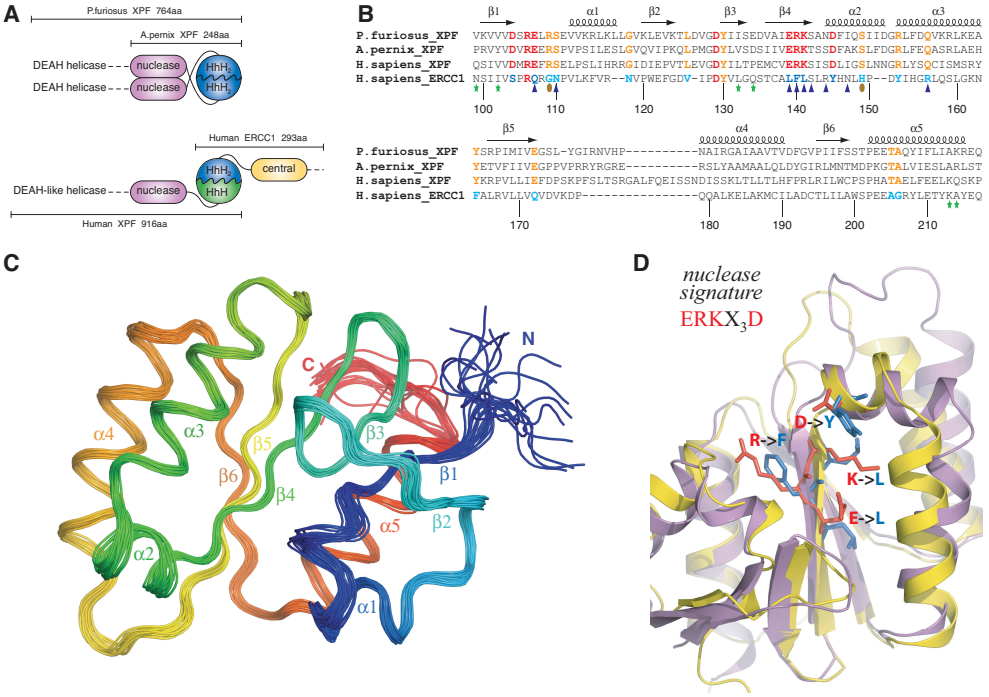
Chapter 2, figure 2. Three dimensional structure of the C-terminal domains of the human ERCC1/XPF complex. (A) Backbone stereoview of the ensemble of final 20 structural conformers. ERCC1 is colored blue, XPF is colored green, and the hairpins in both subunits are brown. (B) Cartoon representation of the lowest- energy model. The purple sphere depicts the centre of the pseudo-2-fold symmetry axis. Helices are denoted with Greek letters, from α to ε for ERCC1 and α′ to ε′ for XPF, while hairpins are indicated as h1, h2, and h1′. Also shown are Phe residues at the C termini of each subunit. Phe293 of ERCC1 is colored orange and Phe894 of XPF is colored yellow. (C) ERCC1 and XPF are superimposed, showing the overall fold similarity. Color conventions as in (B).



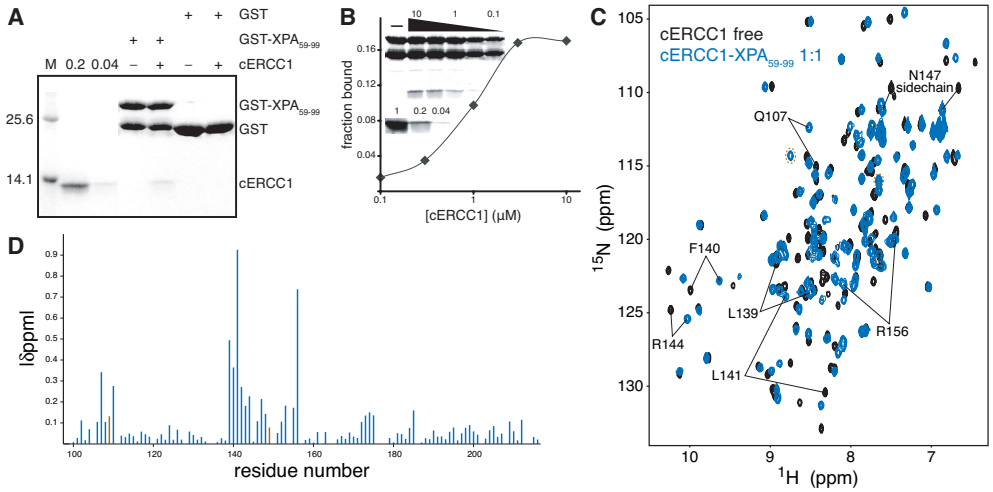
Chapter 2, figure 3. Protein-protein interactions along the C-terminal domains of ERCC1 and XPF. (A) View of the central interacting core. Helices α in the front, helices γ at the back, and the tips of helices ε that run perpendicularly build up the interacting surface. Residues participating in contacts are depicted in stick representation. The backbone of ERCC1 is shown in blue, XPF is shown in green, and interacting amino acids from either side in orange and yellow, respectively. Intermolecular hydrogen bonds are indicated with black dotted lines. (B) Top view of the central core. For the sake of clarity, helices ε have been removed. (C) The tight hydrophobic packing of ERCC1 Phe293 to its XPF cavity and (D) the same for Phe894 of XPF in the hydrophobic pocket of ERCC1. In (B), (C) and (D) amino acids are numbered according to the native sequences.



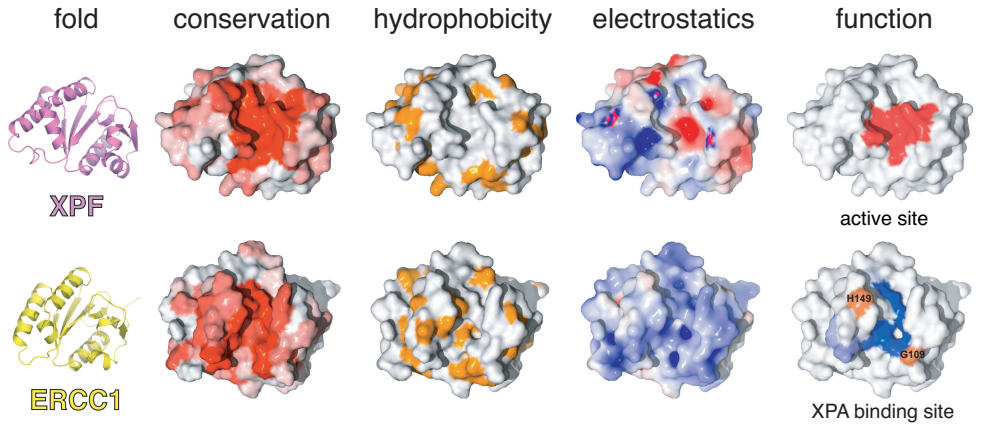
Chapter 2, figure 4. DNA binding by the ERCC1/XPF complex. (A) Domain organization of the ERCC1/XPF complex. (B) Sequence alignment of selected $(HhH)_2$ domains from the three kingdoms. Hydrophobic residues are highlighted in cyan, Gly that belong to hairpin motifs are highlighted in orange and the aromatic residues at the end of the domains are highlighted in green. The observed secondary elements of ERCC1 are indicated above the sequences. (C) Normalized chemical shift changes upon DNA titration versus the ERCC1 and XPF sequences. Missing bars indicate either Pro residues or unresolved chemical shifts due to peak overlap. (D) Two views of surface representation rotated by 180° and colored according to their electrostatic surface potential at ±8kB T/e for positive (blue) or negative (red) charge potential by using the program GRASP (Nicholls, 1993). (E) Observed chemical shift changes on the protein surface and model for the interaction of ERCC1/XPF with DNA by fitting ERCC1 on the archaeal $(HhH)_2$ domain. The blue parts denote shifting residues in both representations, and the intensity of the color corresponds to the absolute shift value (cut-off value of 0.1).



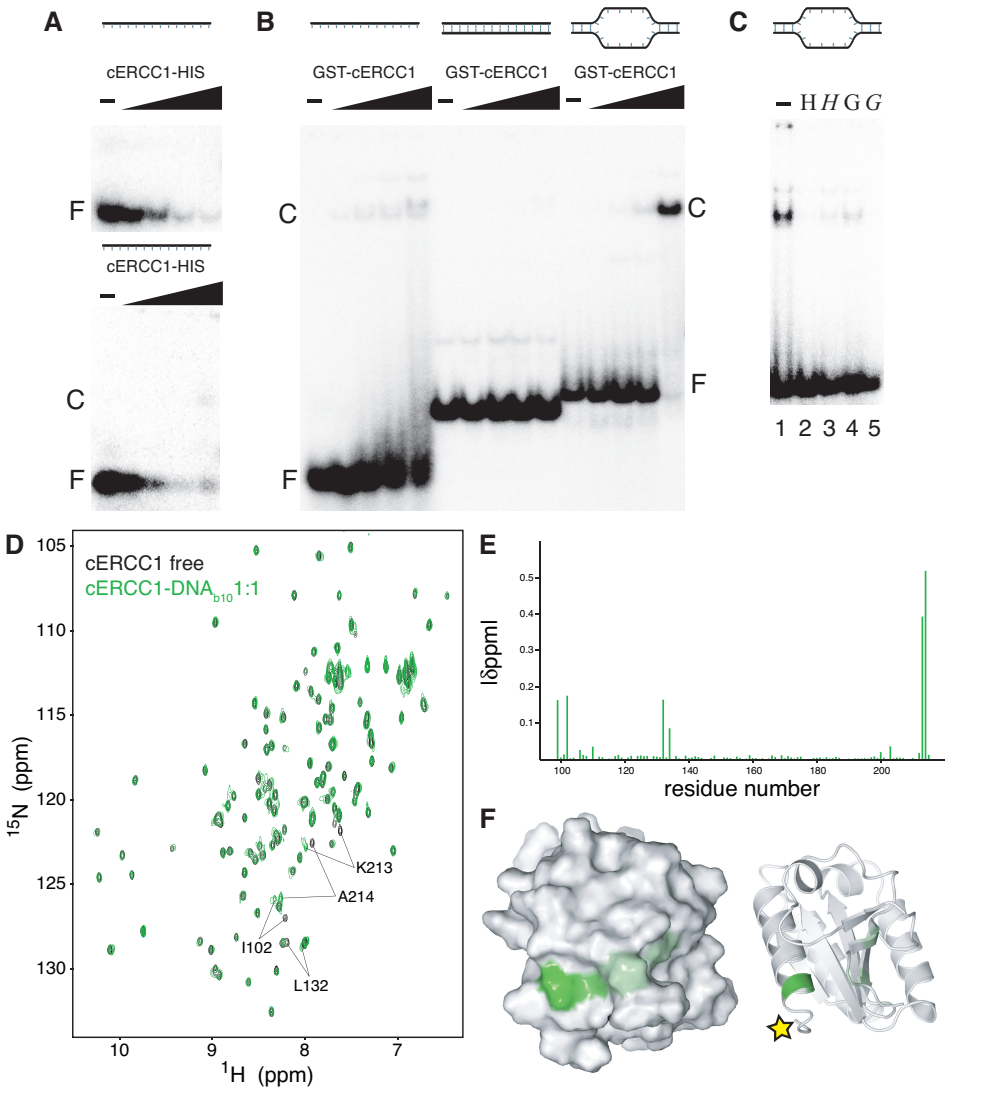
Chapter 3, figure 1. (A) Domain organization of the archaeal XPF homodimeric members and the human ERCC1/XPF heterodimer. (B) Structure-based sequence alignment of the nuclease XPF domains from archaea to human and the corresponding central domain of human ERCC1. Secondary structure elements of the prototype XPF nuclease fold are indicated at the top of the sequences. Catalytic residues in the XPF nucleases are colored red and their corresponding substitutions in ERCC1 blue. Other invariant residues in XPF domains are depicted in orange, while the ERCC1 equivalents are depicted in cyan. Residues of cERCC1 perturbed largely upon XPA titration are indicated by blue triangles and those appear only in the final complex by brown ellipses. Green asterisks indicate cERCC1 residues perturbed by DNA titration. cERCC1 sequence is numbered at the bottom. (C) Ensemble of the final 20 structural conformers of cERCC1 as determined by solution NMR. Secondary structure elements and N- and C-termini are labeled. (D) Superposition of the crystal XPF nuclease structure (2bgw) from *A. pernix* (purple) and the solution NMR structure (2jpd) of human cERCC1 (yellow). Emphasis is given to the nuclease signature and the corresponding substitutions in cERCC1.



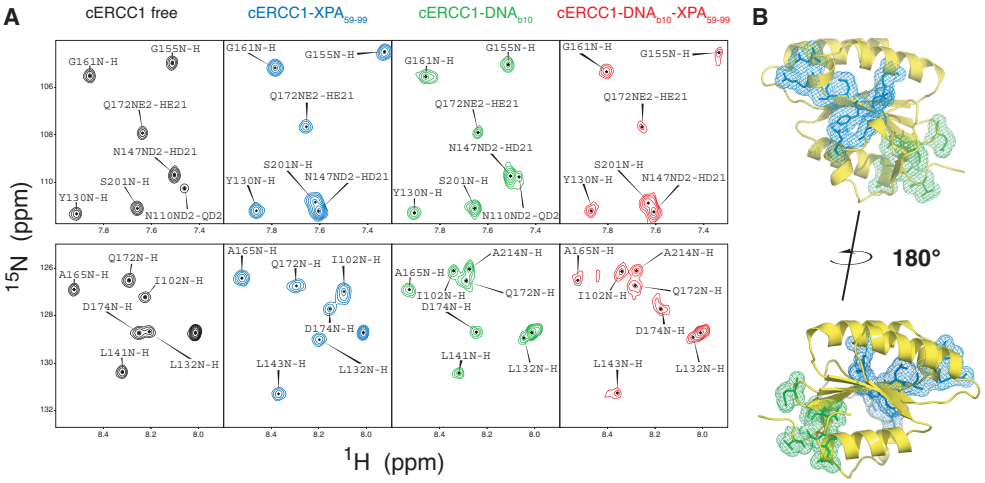
Chapter 3, figure 2. ERCC1-XPA interactions. (A) GST pull-down assay with 3 μg of GST or GST-XPA fusion proteins in the presence or absence of 2 μM cERCC1. Here, 0.2 and 0.04 refer to respectively 20 and 4% of the input cERCC1 protein present in the GST pull-down assay. (B) Semi quantitative GST pull-down assay showing the fraction of cERCC1 bound to 3 μg of GST-XPA at the indicated [cERCC1] (μM). The upper part of inset shows a representative GST pull-down assay, where the various cERCC1 concentrations used are depicted above (10, 3, 1, 0.3 and 0.1 μM). The lower panel shows respectively 100, 20 and 4% of cERCC1 added to the assay. (C) Chemical shift perturbation of the cERCC1 ^{1}H-^{15}N HSQC upon complex formation with XPA. Free cERCC1 spectrum is in black and XPA-bound spectrum in blue, while brown circles in the bound spectrum indicate G109 and H149 resonances. (D) Normalized chemical shift changes between free and XPA-bound forms versus the cERCC1 sequence. The p.p.m. difference for G109 and H149 was calculated by their resonances in the free cERCC1 spectrum at pH 5.5, and are colored brown.



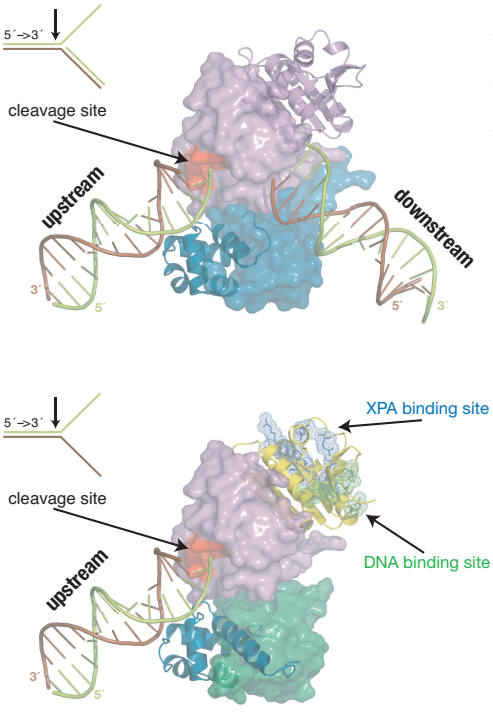
Chapter 3, figure 3. Common fold, different properties and distinct functions for XPF and ERCC1. Sequence conservation either for XPF or ERCC1 proteins is colored from white (non-conserved) to red (highly conserved). The opposite face has no significant conservation for either protein. Hydrophobic side chains are colored orange. Electrostatic surface potentials were calculated using APBS (Baker et al., 2001) and colored blue for positive or red for negative charge potential. Active site residues of XPF were chosen based on previous mutational studies (Enzlin and Scharer, 2002) and the XPA-binding site of ERCC1 was identified by our NMR titrations.



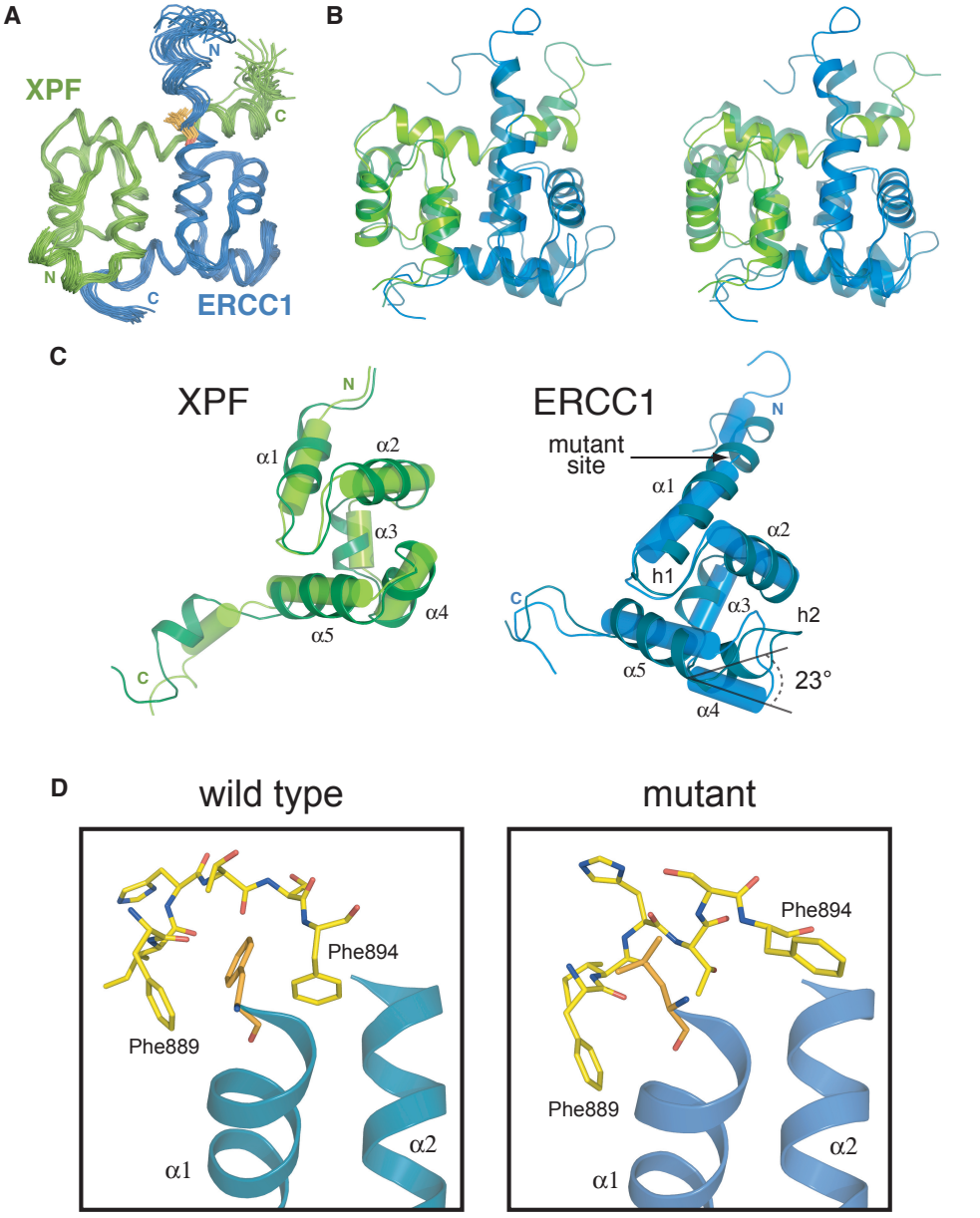
Chapter 3, figure 4. DNA binding by cERCC1. (A) EMSA with a 20-mer ssDNA substrate using increasing concentrations of cERCC1-HIS protein (0, 0.04, 0.2, 1 and 5 µM), loaded on a 7.5% acrylamide gel (upper panel) or a 3.5% agarose gel (lower panel). (B) EMSA using ssDNA (20-mer), dsDNA (30bp), and bubble10, a dsDNA with 10 unpaired bases as substrate, in the presence of 0, 0.125, 0.25, 0.5 and 1 µM HIS-GST-cERCC1. (C) Inhibition of cERCC1-b10 DNA complex formation ([cERCC1] = 0.5 µM) by depletion of HIS-GST-cERCC1 from the EMSA reaction by the addition of MagneHis beads (H) or GST agarose beads (G) prior to or after addition of DNA (italic). F: free DNA, C: cERCC1-DNA complex. (D) Chemical shift perturbation after addition of b10 DNA. The spectrum of the free protein is shown in black and the spectrum after addition of equimolar amount of b10 in green. (E) Normalized chemical shift changes between free and b10-bound forms versus the cERCC1 sequence. (F) Surface and cartoon representations of cERCC1 colored according to normalized chemical shifts. Yellow star indicates the position of the C-terminal his-tag that was interfering with DNA binding.



Chapter 3, figure 5. Dual function of cERCC1. (A) Indicative portions of the ^{1}H-^{15}N HSQC spectrum for the free cERCC1, bound to either XPA (1:1) or DNA (1:1) separately, and the ternary complex (1:1:1). (B) Views of the two binding sites on cERCC1 structure.



Chapter 3, figure 6. Models for the function of the archaeal homodimeric XPF (top) and human ERCC1/XPF heterodimer (bottom) constructed from the structure of the XPF homodimer bound to dsDNA (2bgw), the free structure of ERCC1/XPF C-terminal interacting domains (1z00) and the current structure of the ERCC1 central domain (2jpd). For the homodimer, one protomer is shown in a cartoon and the other in a surface representation. Accordingly, for the heterodimer ERCC1 is in a cartoon and XPF in a surface representation. The protein domains in both cases are colored as in Figure 1A.



Chapter 4, figure 2. Structural differences induced by the F231L mutation of ERCC1. (A) Backbone superposition of the 20 lowest energy structures for the mutant heterodimer comprising the ensemble. Leu231 of ERCC1 is shown in a stick representation. Both partners are indicated together with their N- and C-termini. (B) Structural overlay of wild-type and mutant hetrodimers superimposed on the XPF (left) and ERCC1 side (right). Wild type conformations are shown in transparent color. (C) Structural comparison for each partner in wild type (cartoon) and mutant heterodimeric complexes (cylinders). (D) Local network of interactions for Phe231 (wild type) and Leu231 (mutant).

Chapter 4, figure 3. DNA binding mode by the HhH_2 domain of *A. pernix* XPF and the geometry of the ERCC1 hairpin motifs in the wild type and mutant ERCC1/XPF heterodimeric structures.