Structure and stability of ERCC1-XPF DNA repair complexes

Maryam Faridounnia

2015

ISBN 978-94-6295-175-4
Structure and stability of ERCC1-XPF DNA repair complexes
Maryam Faridounnia
Cover design: Maryam Faridounnia
NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Faculty of Chemistry,
Utrecht University, The Netherlands
April 2015

Printed in The Netherlands by Proefschriftmaken.nl || Uitgeverij BOXPress

Structure and stability of ERCC1-XPF DNA repair complexes

Structuur en stabiliteit van ERCC1-XPF DNA-herstel complexen

(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. G.J. van der Zwaan
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op dinsdag 12 mei 2015 des ochtends te 10.30 uur

door

Maryam Faridounnia
geboren op 12 mei 1980
te Teheran, Iran

Promotor:	Prof. dr. R. Boelens
Copromotor:	Dr. G. E. Folkers

In loving memory of my father

To my mother
With love and eternal appreciation

با درود به روح پر فروغ پدرم که همواره یادش در دلم نور امید به زندگی را می تاباند،

تقدیم به مادر دلسوز و مهربانم
که در لحظات دشواری‌های زندگی همواره یاوری دلسوز و فداکار
و پشتیبانی محکم و مطمئن برایم بوده است.

Table of Contents

List of Abbreviations used

BER	Base Excision Repair
CD	Circular Dichroism
COFS	Cerebro-oculo-facio-skeletal
CS	Cockayne Syndrome
CSP	Chemical Shift Perturbation
DDB	DNA Damage Binding proteins
DSB	Double Strand Break
dsDNA	double strand DNA
ERCC1	Excision Repair Cross Complementation Group 1
FA	Fanconi Anemia
GG-NER	Global Genome Nucleotide Excision Repair
HhH	Helix-hairpin-Helix
HR	Homologous recombination
HSQC	Heteronuclear Single Quantum Coherence spectroscopy
ICL	Interstrand Cross Link
MMEJ	Microhomology mediated End-joining
MMR	Mismatch Repair
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Sequence
NOESY	Nuclear Overhauser Enhancement Spectroscopy
NSCLC	Non-small Cell Lung Cancer
PARP	Poly-ADP-Ribose Polymerase
RIR	Replication-independent Repair
ROS	Reactive Oxygen Species
SPR	Surface Plasmon Resonance
SSA	Single Strand Annealing
SSB	Single Strand Break
ssDNA	single strand DNA
TC-NER	Transcription-coupled Nucleotide Excision Repair
TLS	Translesional Synthesis
XPF	Xeroderma Pigmentosum group F
XRCC1	X-ray Repair Cross Complementation Group 1
$(HhH)_2$	Tandem Helix-hairpin-Helix

Chapter 1

General Introduction

Maryam Faridounnia

Manuscript in preparation

INTRODUCTION

Conservation and stability of genetic information not only relies on the accurate mechanism for copying DNA sequence before cell division, but also on the efforts of the DNA repair machinery in accurate maintenance of the genomic information. Numerous damaging agents endanger the DNA in the genes. A major group of factors responsible for DNA lesions, consists of environmental threats, such as ultraviolet (UV) light, ionizing radiation and genotoxic chemical agents. Another source of DNA lesions are the byproducts of metabolic activity, such as free radicals and reactive oxygen species that, if not removed by an antioxidant defense system, will cause oxidative modifications in DNA. Spontaneous or induced base deamination and faulty nucleotide incorporation also cause lesions to the DNA. These are only a few examples of the many possible modifications that may occur in DNA continuously and which cause thousands of 'spontaneous' DNA lesions each day. If left uncorrected, they would alter the DNA sequence so rapidly that it would be fatal for each living cell. As a result, cells would cease to be viable or become malignant. The first response of the cell is a direct defense against these damaging agents through detoxification mechanisms and oxidative stress response. The biological response to DNA damage occurs through up-regulation of transcription or by inducing cell cycle arrest to enable DNA repair, and if the damage is too severe, apoptosis can be induced (1-4).

Depending on the type of damage, different DNA repair mechanisms have evolved to avoid the accumulation of DNA defects and to preserve the genetic information. The repair instruments range from single protein systems to multistep DNA repair pathways depending on the type of lesion (Figure 1). Repair mechanisms are involved in either the reversal of DNA damage (e.g. photoreactivation,(5,6)) or the excision of damaged components. This excision of damaged bases can be achieved either as free bases in Base Excision Repair (BER), or as nucleotides in Nucleotide Excision Repair (NER). Furthermore, the mispaired bases in DNA are repaired through mismatch repair (MMR) and/or translesional synthesis (TLS). In addition to damaged nucleotides, fractures in DNA threaten the integrity of genome and are treated by DNA repair mechanisms involved in the repair of double strand break (DSB) or single strand break (SSB), such as Homologous Recombination (HR) and NHEJ (Non-Homologous End Joining). A schematic overview of events requiring DNA repair and the different repair pathways is shown in Figure 1. Given the importance of the genome integrity, it is evident that impairments in DNA repair processes can lead to major genomic damage resulting for instance in inherited cancers, neurological disease, and early aging (1,6-9).

The different DNA repair pathways share several steps during the DNA repair process and sometimes even the same factors are found in the different pathways. This is clearly the case for the structure-specific endonuclease ERCC1-XPF (6,10).

The main focus of this thesis is on the heterodimeric complex ERCC1-XPF, more specifically on its stability and its role in DNA damage site recognition. Therefore, the rest of

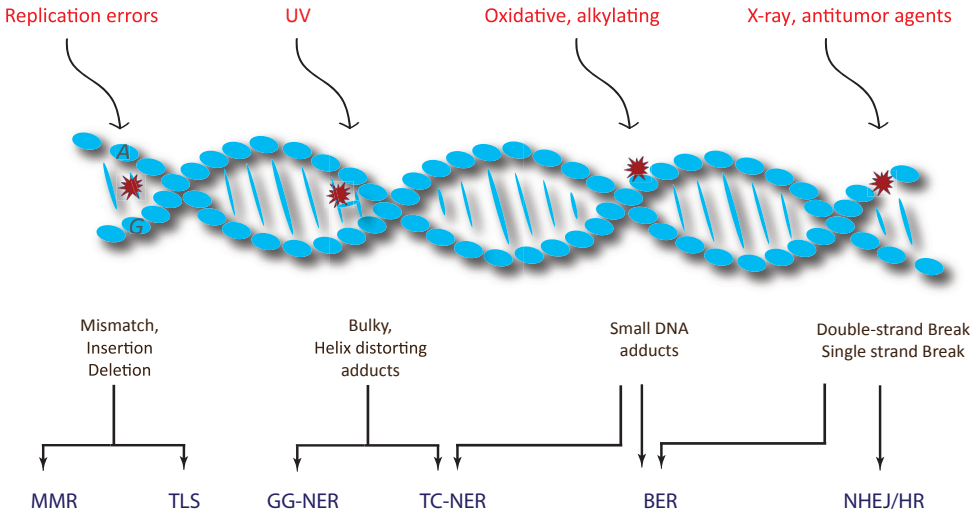


Figure 1. A repertoire of DNA repair mechanisms ensures maintenance of DNA integrity. DNA is continuously exposed to damaging events (top) that cause a range of lesions (middle). To overcome the lesion, the damage triggers a multitude of DNA repair mechanism (bottom). Figure adapted from (3,7,8).

this chapter describes the current knowledge concerning the DNA repair pathways in which ERCC1-XPF is involved, and structural and functional aspects of ERCC1 and XPF.

Function of ERCC1-XPF in DNA repair pathways

The heterodimeric ERCC1-XPF complex is involved in a multitude of mechanisms, and is known for its function among others in nucleotide excision repair (NER). The NER pathway is one of the most extensively studied DNA repair processes. It is characterized by the cyclobutane pyrimidine dimer (CPDs) damages that are formed by exposing the DNA to UV irradiation. The role of the different proteins involved in this pathway, such as ERCC1-XPF, is well known. A multitude of discoveries indicate that ERCC1-XPF further functions in several other DNA repair pathways, to repair for instance SSBs, DSBs, and interstrand cross-links (ICLs), which might be due to its unique catalytic incision properties (6,11).

Nucleotide Excision Repair

In NER, full nucleotides are removed from the damaged genome by a multiprotein repair process. The damages that are recognized by NER, have in common that they cause local bulky distortions in the DNA double helix. The DNA adducts - such as from oxidative DNA damage, thymidine and other cyclobutane pyrimidine dimers (CPDs), intra-strand cross-links, and bulky alkylating adducts induced by chemotherapy (4,12-14) - can be removed by two distinct subpathways: the global genome NER (GG-NER) and the transcription-coupled NER (TC-NER). In GG-NER, the entire genome is examined and the damage is recognized through lesion sensing DNA damage-binding proteins (DDBs), also known as XPE and the XPC-Rad23B complex. On the other hand, in TC-NER RNA Polymerase II that is stalled at a DNA lesion, triggers recognition (Figure 2). When stalled at the lesion, the Cockayne syndrome protein complex CSB-CSA translocates RNA Pol II, so that the DNA becomes available for repair (15-18). However, the mechanism is not fully understood (19,20).

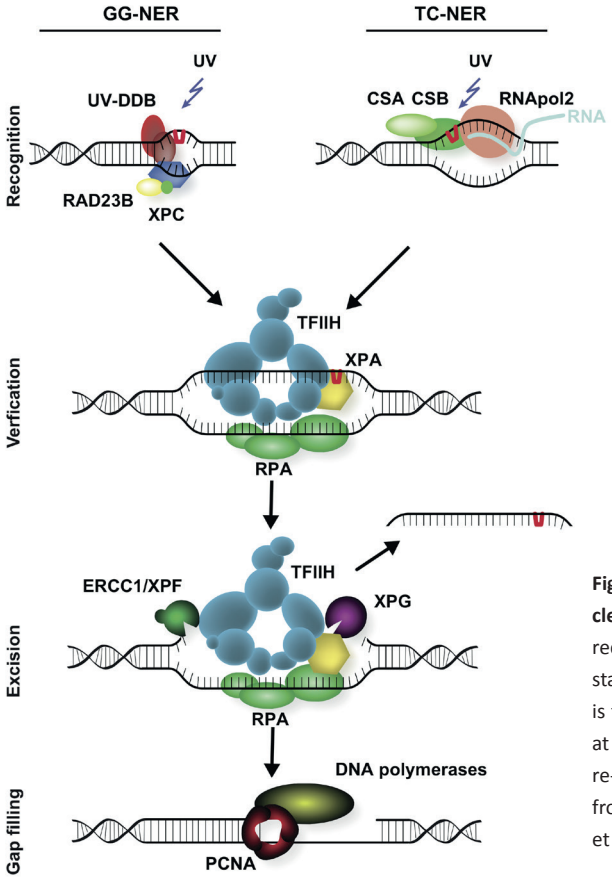


Figure 2. Schematic overview of the Nucleotide Excision Repair pathway. Lesion recognition leads to GG-NER, whereas stalled transcription leads to TC-NER. This is followed by damage verification, incision at 5' and 3' sides of the lesion and DNA re-synthesis. Reprinted with permission from Lans *et al.* (18). Copyright 2012 Lans et al; licensee BioMed Central Ltd.

After damage recognition, the two subpathways go through the same key events: formation of a preincision complex, followed by unwinding and excision of single stranded DNA nearby the lesion, gap filling, and sealing. An assembly of proteins, including transcription factor II H (TFIIH) and XPG, initiates the NER process (7). Subsequently, the XPB subunit of TFIIH binds to the recognized damage site and through its ATPase activity enhances the binding of TFIIH to the damaged DNA (21,22). This is followed by the helicase activity of the TFIIH XPD subunit that unwinds the damaged DNA into a so-called repair bubble, where an open single-stranded DNA structure is formed around the lesion consisting of typically 25-30 unpaired DNA bases containing the lesion. This structure is stabilized by the DNA-binding proteins RPA and XPA. This bubble-like intermediate is the substrate for the dual incision activity of the two NER endonucleases, the ERCC1-XPF complex at the 5´ side and XPG at 3´ side, excising the damage containing 25-30 nucleotide single-stranded fragment. The gap created by the excision of these nucleotides is filled by the replication factors using the intact DNA as template strand. Finally, the remaining nick is ligated by a DNA ligase (3,7,16,18,21,22).

Single-strand break repair

Single-strand breaks (SSB) are one of the most common DNA damages that arise as intermediates of Base Excision Repair (BER), or directly from damage of the DNA deoxyribose by reactive oxygen species (ROS) or by abortive topoisomerase-1 activity (TOP1-SSB). Direct SSB repair of ROS damaged DNA is initiated by recognition and signaling performed by poly-ADP-ribose polymerases (PARP) 1 and 2, while SSBs introduced during BER or by TOP1 activity might require other recognition mechanisms. The PARP proteins, which bind to SSBs and signal their presence, are thought to attract XRCC1. This protein orchestrates the repair of SSBs by forming a scaffold for a number of enzymes responsible for processing and gap filling of DNA breaks (6,23,24).

Since the 3'- or 5'-terminus of many SSBs are distorted, they need end processing to meet the substrate requirements for gap filling and ligation. These distortions can be very diverse, and depending on the type, a repertoire of enzymes is required, including endonucleases that are acquired from different molecular pathways (6,23,24). Related to this, it has been suggested that some of the 3'-termini are removed by ERCC1-XPF (24). Most SSBs gap filling involves insertion of a single nucleotide, followed by XRCC1-dependent ligation. In some cases, when the incision is involved in end processing, gap filling involves insertion of more nucleotides (23,24).

Double-strand break repair

Double-strand breaks (DSBs) can arise exogenously from ionizing radiation or exposure to X-rays, e.g. during cancer treatment, or from a major endogenous source, such as during the replication fork collapse, when the DNA polymerase meets a DNA lesion. In order to avoid fragmentation, translocation and deletion of chromosomal DNA caused by DSBs, cells have de-

veloped several repair pathways (Figure 3). DSBs can have two distinctive structures: one-sided when the replication fork is stalled at a damage site and only one strand is without a partner strand or two-sided, e.g. when two replication forks converge at the damage site (cf. Figure 4 on ICL repair) (25). DSBs are primarily repaired by two main pathways, either by homologous recombination (HR) or by non-homologous end joining (NHEJ). If these two pathways fail to repair the damage, either single-strand annealing or an alternative NHEJ (Alt-NHEJ) pathway known as microhomology (MMEJ) can repair such damage, but with lower efficiency (25-28).

Homologous recombination (HR) repair is the major DSB repair pathway. HR is stimulated by DSBs generated from DNA damages during meiosis and sometimes mitosis. After DSB resection, unidirectional 5'- to 3'-degradation of such a breakage results in single-stranded DNA overhangs. These 3'-overhangs are then coated by replication protein A (RPA) and with the aid of the BRCA2 protein, also known as Fanconi anemia protein FANCD1, multiple RAD51 proteins are loaded on this RPA-coated ssDNA. This process is known as filament formation (25,26,28-30). During this process, ssDNA is prepared to invade a homologous DNA helix that can either be an available sister chromatid or homologous donor DNA sequences in the late S/G2 phase. This strand invasion recruits regions of undamaged DNA to form a template to restore the lost information (25,26,28,31). Then RAD54, assembled with heterochromatin, uses the filament for branch migration on the formed pairing intermediates (heteroduplex Holliday junction or D-loop) (28,32-35). This continues with DNA polymerase extending the missing sequence. Resolving these Holliday junction intermediates can result in either crossovers or non-crossovers products. Other aspects of the repair process may proceed through different subpathways like single-strand annealing (SSA) mechanism (Figure 3)(28,31-34,36). In recent years, the role of ERCC1-XPF in DSB repair has been gradually unraveled. For example, it is now clear that during single-strand annealing, ERCC1-XPF is responsible for the incision of the DNA (37,38).

Non-homologous end joining (NHEJ) is another major mechanism in DSB repair that is activated during G0/G1 and early S-phases of the cell cycle. NHEJ is an end-joining mechanism that directly ligates the two ends in DNA strand breaks. This error-prone process occurs before replication when another copy of DNA is absent. This process does not require any homologous sequence. Therefore it utilizes a machinery that is largely distinct from those used in HR. The Ku70-Ku80 dsDNA-binding heterodimer (KU-DNA) triggers complex formation on the damaged DNA by recruitment of DNA-dependent protein kinase (DNA-PK) to the damaged region of the DNA. Since in most cases the endogenously generated DSBs do not have 3' OH or 5' P overhangs, they lack essential molecular elements for polymerization and ligation. Therefore, the DSB ends have to be processed first, which includes the removal of the inadequacies, followed by polymerization. Only then can ligation be accomplished (6,27,28,39,40). It has been suggested that ERCC1-XPF is involved in NHEJ when DSB ends have extensive 3'-overhangs that need to be trimmed, and that this is regulated by DNA-PKs within the context of the NHEJ complex (41-43).

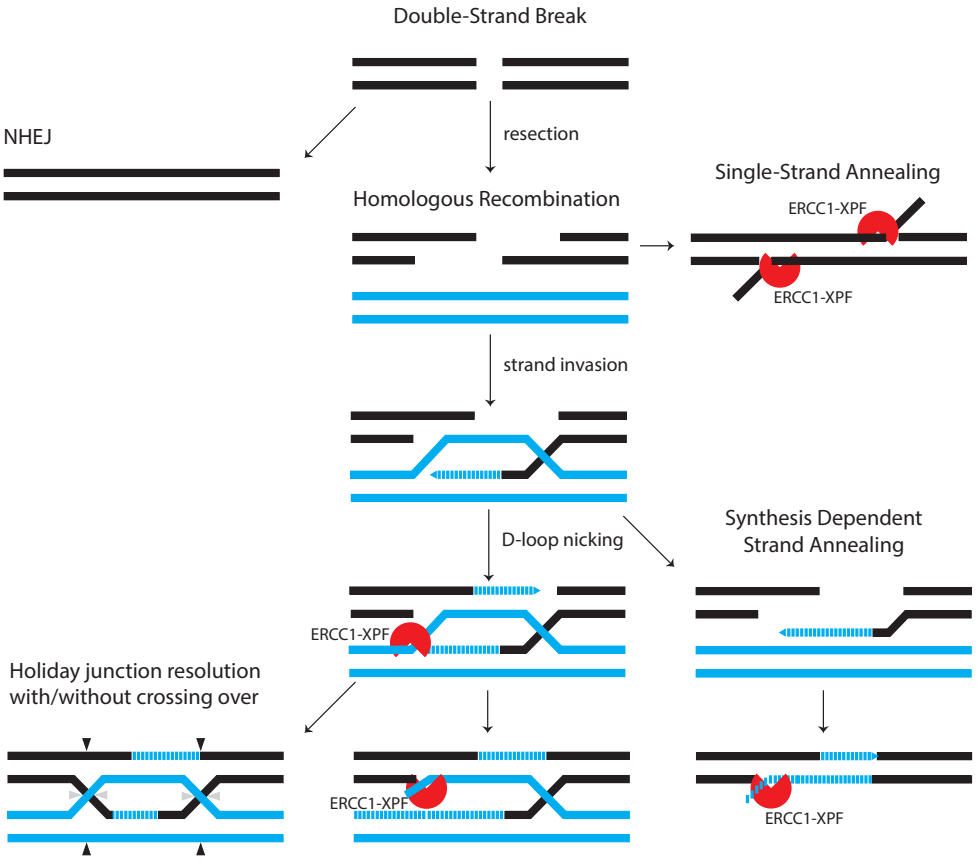


Figure 3 . An overview of eukaryotic DSB repair. DSBs in DNA can be resealed through NHEJ (left) without homology requirement. ERCC1-XPF might trim the ends before resealing. Alternatively, DSBs repair can receive 5' 3 →' single strand resection followed by invasion, which uses another homologous chromatid as template. This is followed either by re-annealing of the invading strands (synthesis-dependent strand annealing) or formation of a heteroduplex Holliday Junction. After a branch migration on this formed junction, DNA polymerase extends the missing sequence (D-loop formation). Also shown is how ERCC1-XPF may be involved at the different mechanisms of DSB repair. DNA copies from two distinct homologous chromatids are represented in blue and black. Adapted from (6,36,38).

Microhomology mediated End-joining (MMEJ) is an alternative NHEJ mechanism that occurs during G1 and S phases. During MMEJ short homologous DNA fragments of 5-20 nucleotides are joined in a rather error-prone manner, often yielding deletions, and leading to genome instability. Many questions remain unresolved about MMEJ. For example, a study by Deng et al. (44) shows that the protein RPA regulates this pathway. However, it is not clear whether it competes with HR components located on the DNA coated with RPA, or if there is an alternative mechanism. One mechanistic model, proposed for MMEJ, starts with end resection, is followed by SSA and DNA end processing, and finalized by ligation (40). Before

ligation, during the end processing, ERCC1-XPF plays a key role in the cleavage of non-complementary 3'-flaps from an annealed intermediate. This excision activity facilitates stable DNA end joining, proper gap filling and ligation (28,40,45).

Interstrand cross-link (ICL) repair

Cytotoxic and antitumor agents can form covalent linkages on DNA in different ways including intra-strand, between two complementary strands, and even between a DNA base and a protein residue. An interstrand cross-link (ICLs), though accountable for only a small amount of the total DNA adducts, is one of the deadliest DNA lesions because this covalent binding prevents DNA from separation and stalls replication and transcription processes, thus causing chromosome rearrangement or break, or cell death (46,47). These lesions arise either exogenously from exposure to toxic mutagens such as cisplatin and carcinogens in cigarette smoke, or endogenously from amino-acid metabolism or lipid peroxidation for instance from a high fat diet. The formed DNA adducts arrest replication, transcription and other activities with unwinding of the DNA as a prerequisite. The interpretation of the response to cross-linking agents is complicated primarily because they do not only produce ICLs but can also cause multiple other DNA lesions that depending on damage type, may trigger a separate repair pathway. Furthermore, a variety of mechanisms (shown in Figure 1) is thought to engage in the repair of ICLs. Therefore, it is difficult to disentangle the contribution of different repair pathways in ICL repair (47,48).

ICL repair can occur during or outside the S-phase of the cell cycle, in the presence or absence of replication proteins. Therefore ICL repair pathways are classified as replication-dependent and replication-independent (Figure 4) (47,49-51). In both ICL repair processes, it is thought that the Fanconi anaemia (FA) network coordinates the major groups of proteins from the NER pathway, the homologous recombination and the DNA translesion synthesis (TLS) pathway. Indeed, mutations in genes regulating the elimination of ICLs can cause FA (46,52). It is increasingly understood that FA proteins play roles in ICL repair through sensing, recognition and processing of ICLs (53).

The *ICL replication/recombination-dependent pathway* is initiated during the late S or G2 phase by a stalled replication fork at the ICL damage site. In general, the presence of an undamaged sister chromatid initiates the assembly of the homologous recombination machinery including RAD54, FA proteins and structure-specific endonucleases such as the ERCC1-XPF complex from NER (Figure 4) (7,46). While the mechanism of ICL repair is still largely unknown, evidence accumulates that ICL damage triggers the FA pathway by stimulating ubiquitylation of FA proteins. This activity is essential for SLX4 (also known as FANCP) loading on this site. SLX4 ensures the participation of the nucleases MUS81-EME1 (both members of the XPF/MUS81 family) and ERCC1-XPF in the incision of DNA flanking the damage site, creating DSBs and unhooking the cross-link at the stalled replication fork. The coexistence of

ERCC1-XPF and SLX4 is essential for unhooking the ICLs. The unhooked fragment is still linked to the template strand through its ICL (54-57). It is believed that FANCD2 ubiquitination assists the translesional polymerases in DNA synthesis at the lesion site, which is followed by the removal of reminiscent DNA flaps by the NER components ERCC1-XPF and XPG. This fully repaired sister chromatid is used for the repair of DSB intermediates through homologous recombination (Figure 4) (47,49,58-60). Impairments in ICL repair can lead to FA disorder, characterized by chromosomal frailty, congenital abnormalities, bone marrow failure, increased cancer rate, and ICL hypersensitivity (47,49-51).

The ICL replication-independent repair (RIR) pathway occurs during the G0/G1 cell cycle phase and is vital for cellular homeostasis especially in post-mitotic cells, such as in neurons or in cells dividing rarely, such as stem cells. RIR probably depends on NER for damage recognition and dual incision utilizing its endonucleases ERCC1-XPF and XPG to unhook the ICL lesion and to excise the DNA ICL lesion. This is followed by repair synthesis of the gap, performed by translesion synthesis (TLS) polymerase activity (Figure 4) (46,57,61).

Roles of ERCC1-XPF beyond DNA repair

It is clear from the above that the ERCC1-XPF heterodimer plays a key role in multiple DNA repair pathways. Nevertheless, not all phenotypes of ERCC1-XPF deficiencies can be interpreted by their primary function in a repair pathway. Many reports associate unexplained phenotypes of ERCC1-XPF, such as developmental disorders and neurodegeneration, with a variety of activities beyond DNA repair. For instance, the ERCC1-XPF endonuclease complex can bind 3' G-rich overhangs of telomeres. Telomere-repeats binding protein TRF2 can bind and inhibit ERCC1-XPF and protect the telomeres (62,63). Another example of ERCC1-XPF function outside DNA repair is its participation in sister chromatid separation during chromosome segregation. Together with MUS81-EME1, the ERCC1-XPF complex can be co-localized by FANCD2 to mitotic chromosomes and it has been shown that ERCC1 down-regulation causes defects in chromosome segregation and mitotic failure (62,64). Finally, a recent study reveals the role of ERCC1-XPF in transcription initiation during cell development. It is proposed that ERCC1-XPF is recruited to gene promoters to fine-tune optimal transactivation of target genes (65,66). This optimization occurs during postnatal development, which can be related to the cerebro-oculo-facio-skeletal (COFS) syndrome in humans with deficient ERCC1 (62,65).

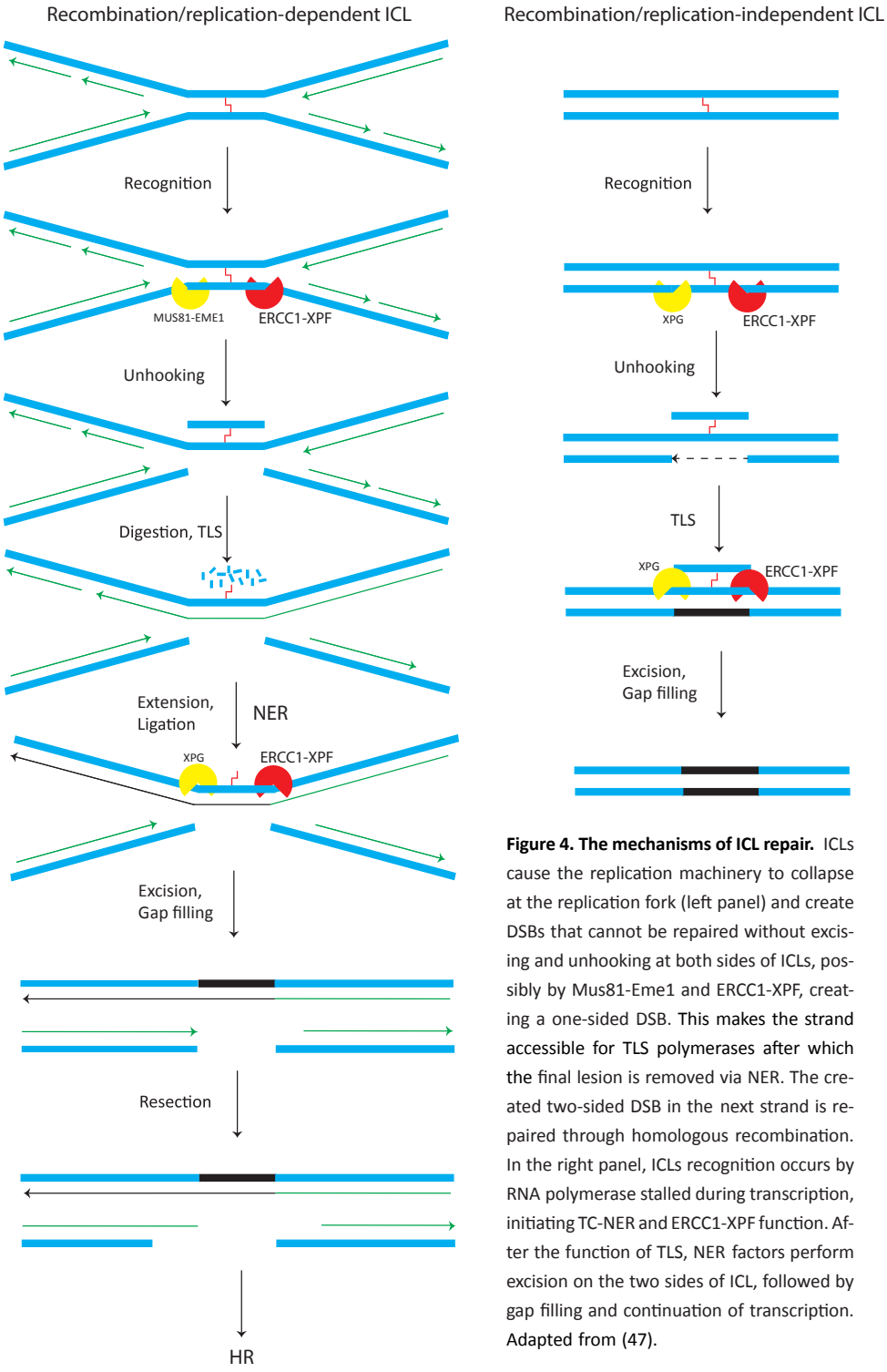


Figure 4. The mechanisms of ICL repair. ICLs cause the replication machinery to collapse at the replication fork (left panel) and create DSBs that cannot be repaired without excising and unhooking at both sides of ICLs, possibly by Mus81-Eme1 and ERCC1-XPF, creating a one-sided DSB. This makes the strand accessible for TLS polymerases after which the final lesion is removed via NER. The created two-sided DSB in the next strand is repaired through homologous recombination. In the right panel, ICLs recognition occurs by RNA polymerase stalled during transcription, initiating TC-NER and ERCC1-XPF function. After the function of TLS, NER factors perform excision on the two sides of ICL, followed by gap filling and continuation of transcription. Adapted from (47).

Physiological consequences of impaired ERCC1-XPF

Several clinically heterogeneous disorders such as susceptibility to cancer, neurological and developmental deficiencies and early onset of aging are associated with inborn defects in the NER proteins ERCC1 and XPF (7,65,67,68). During genome replication, accumulation of lesions due to GG-NER deficiency can cause cancer, whereas mutations in transcription-coupled repair (TC-NER) factors can lead to degenerative phenotypes. This is thought to be because accumulation of lesions in the transcribed genes results in silencing of essential genes or triggers apoptosis (3,7,68,69). Furthermore, in recent years, the role of ERCC1-XPF in other pathways has been understood through syndromes that are beyond symptoms of NER deficiency (51).

Xeroderma pigmentosum (XP), one of the most prominent diseases caused by NER defects, is an inborn disorder conferring dermal damage to sunlight-exposed skin areas. XP has been related to defective protection against UV damage, which results in a cancer-prone skin. XP patients also have an elevated frequency of internal cancers, typically lung and gastro-intestinal cancers, suggestive of NER deficiency in defense against carcinogens from the air and food. In addition, around 20% of XP patients have neurological diseases such as neurodegeneration, dementia, and microcephaly. Since the nervous system is protected from sun, it has been suggested that these patients are deficient in repair of oxidative DNA damage, which causes blockage of transcription. Given that most oxidative lesions (with an exception of cyclopurine-deoxynucleosides) are repaired by base excision repair (BER), the genomic increase in cyclopurine-deoxynucleosides can be the reason for neurological symptoms in XP patients (4,12,19,68,69). Mutations in any of the NER enzyme genes, i.e. *XPA*, *XPB*, *XPC*, *XPD*, *XPE*, *XPF*, *XPG*, plus an *XPV* variant (the genetic disease XPV results from mutations in a gene encoding translesional polymerase η), can cause Xeroderma pigmentosum (3,7,68,69). In contrast to XP patients for whom the onset of cancer typically occurs already before an age of 2 years, XP-F patients have much milder symptoms with a late onset of skin cancer. XP-F patients have considerably reduced levels of both XPF and ERCC1 in fibroblasts, which results in a diminished endonuclease activity during NER and which contributes to a typically mild phenotype in these patients (70). Some rare severe XPF deficiency cases come with neurological deficiencies (71). The most severe case of XPF deficiency was a patient with XFE progeroid syndrome. The patient was completely devoid of TC-NER and GG-NER, but its symptoms were mostly related to its high sensitivity for ICL-inducing agents. Also patients with ERCC1 deficiency, which appears less frequently, suffer from severe symptoms. According to Imoto et al. (72,73), an ERCC1 nonsense mutation K226X combined with a IVS6-G>A splice site mutation on the second allele, has caused Xeroderma pigmentosum with late onset of a progressive and severe neurodegeneration resulting in dementia and cortical atrophy at an age of 15. The patient died at the age of 37 (72,73).

Cockayne syndrome (CS) is a severe disorder that causes accelerated aging. Depending on the severity CS is divided into three types. Depending on the CS type, the life expectancy can be from 3 years old to four decades (74,75). Common symptoms are accelerated neurodegeneration, mental retardation, growth failure, profound cataracts, retinal degeneration and microcephaly, with pneumonia as the frequent cause of death. To date, mutations in *CSA/CSB*, *XPB*, *XPD*, *XPF* and *XPG* genes are known to cause occasional lesions that stall transcription resulting in CS due to the defect in TC-NER repair mechanism (76-78). Even low levels of unrepaired lesions in transcriptional genes, that are not removed by other repair systems, are sufficient to shorten the life span considerably (69,77). A rare variant of CS, XP/CS, show developmental and neurological abnormalities typical of CS, and skin and eyes phenotype typical of XP. The phenotypes of XP and CS overlap because the *XPB*, *XPD* and *XPG* genes belong to a group with similar characteristics, with CS having milder dermatological phenotypes than XP (76,79,80).

Only recently, three cases of CS were identified which involved defects in ERCC1 or XPF. Two of the patients showed the typical CS phenotype, while the third patient had a combined XP, CS and FA phenotype (XPCS/FA). In the XPCS/FA patient the phenotype was caused by the biallelic XPF mutations C236R and R589W (the latter was previously also observed in XP patients) (51,79,81). In one CS patient, the biallelic mutations were C236R in the SF2 helicase domain of XPF and a premature stop codon at Y577 in XPF, while the second CS patient carried a biallelic F231L mutation in ERCC1 which led to death in early life (79,82). This CS patient with F231L mutation that is described by Kashiyama *et al.* (79) is only the second case of F231L mutation in ERCC1, and has the most severe NER deficiency. Before this F231L CS patient, another patient with F231L mutation in ERCC1 was reported by Jaspers *et al.* (82) to have symptoms including a heterogeneous congenital COFS and severe embryonic and postnatal growth failure. Similar to CS, some of the symptoms consisted of reduced birth weight and early microcephaly, followed by brain atrophy, hypotonia and cataracts. COFS was in the past only linked to mutations in CSB, XPD or XPG, though the phenotypes of COFS are more severe (micro-cornea with optic atrophy in COFS vs. pigmentary retinopathy in CS) (77,82,83). The patient described by Jaspers *et al.* presented a new cause for COFS that is caused by a heterozygous biallelic F231L mutation of ERCC1, in which in one allele Q158 was converted to a stop codon resulting in a fully dysfunctional ERCC1, while the other allele carried a C-to-G transversion causing a F231L mutation (82,84). For the Kashiyama *et al.*'s patient, with the biallelic homozygous F231L mutation early onset of CS deficiency without an indication of COFS was observed (79). The findings suggest that residual amounts of deficient ERCC1 could still maintain the TC-NER function at such a level that the severity, as observed for the F231L mutant patient, did not occur.

The *Fanconi anaemia (FA)* disorder, which has a phenotype of progressive bone marrow failure and predisposition to cancer, is associated with proteins that are involved

in ICL repair. The genetic details are not always clear but at least fifteen proteins have been identified to be associated with FA. Recently, the sequencing of all protein-coding genes in an unclassified FA patient revealed mutations in *ERCC4* genes (encoding the XPF protein) (51). The genetic, biochemical and functional analysis of this mutation shows that the function of XPF in ICL repair is drastically reduced without severe NER deficiency. As previously discussed for CS, another patient has also shown a combined phenotype with FA, XP and CS symptoms (79). Summarizing, recent findings underscore a multifunctional role of ERCC1-XPF, existing beyond DNA repair. Though the frequency of diseases due to ERCC1 or XPF deficiencies is not high, it is expected that defects in ERCC1-XPF will be discovered to be associated with several other diseases.

ERCC1 and XPF belong to the XPF nuclease family

The ERCC1 (31KDa) and XPF (103KDa) proteins (Figure 5A) do not function as individual monomers but form a heterodimeric complex. The ERCC1-XPF complex functions as a structure-specific endonuclease and is recruited to perform incision at a ds/ss DNA junction on the 5' side of the damage while XPG incises on 3' side (85-87). Both ERCC1 and XPF belong to the XPF nuclease family, also known as the XPF/MUS81 family. Although the proteins of the XPF family are evolutionarily conserved in eukaryotes, they have no orthologs among eubacteria, despite the similarity of their DNA repair mechanism the UvrABC family of bacterial NER proteins. The N-terminal part of XPF shows similarity to four consecutive motifs in archaeal helicases from superfamily 2. Several homologs of the XPF nuclease family were found in archaeal NER as well (87). The members of this family exist as either heterodimers in eukaryotes or homodimers in archaea (2,6,88-90).

XPF consists of a N-terminal helicase-like domain and a central excision repair cross complementation group 4 (ERCC4) endonuclease domain, followed by a C-terminal helix-hairpin-helix $(HhH)_2$ putative DNA-binding domain. The non-catalytic ERCC1 protein also contains an ERCC4-like domain - known as central domain - that lacks the catalytic motif. Like in XPF, the central domain is followed by a helix-hairpin-helix $(HhH)_2$ (Figure 5A) (2,6). The dimerization takes place through direct interaction between two $(HhH)_2$ domains.

In eukaryotes, several XPF/MUS81 family members have been identified existing as a heterodimeric complex, with one catalytic and one non-catalytic subunit. An exception is FAAP24-FANCM, in which both domains are non-catalytic due to the absence of the catalytic amino-acid residues in the endonuclease domain (91,92). In humans, in addition to ERCC1 and XPF, five other homologs have been identified, namely MUS81, EME1, EME2, FANCM and FAAP24. Of these, only XPF and MUS81 have endonuclease activity (2,93). These proteins have been found in four heterodimeric complexes: XPF-ERCC1, MUS81-EME1, MUS81-EME2 and FANCM-FAAP24; also for example the Rad1-Rad10 (*S. cerevisiae*), Rad16-Swi10 (*S. pombe*) and HIM9-XPF (*C. elegans*) belong to this family (94).

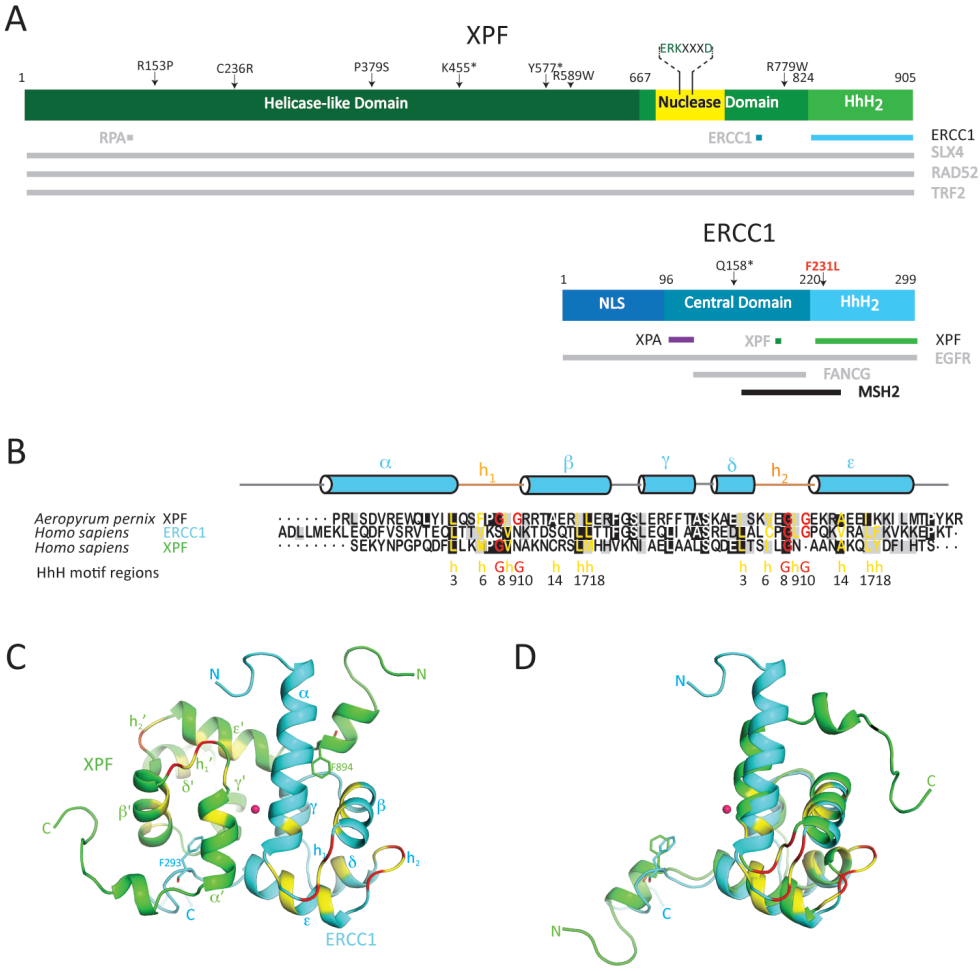


Figure 5. Schematic representation of human ERCC1-XPF. *A,* domain architecture, mutations and interactions. The expansion of key residues is represented above the active site (in yellow) within the XPF nuclease domain. Protein names in grey denote unconfirmed or undefined protein–protein interactions and confirmed interactions are written black. NLS is the abbreviation for nuclear localization sequence. Numbers denote residue number, arrows show mutations, asterisks are used to denote stop codons, and the horizontal bars identify the proteins. *B*, $(HhH)_2$ domains of *A. pernix* XPF and human ERCC1 and XPF. The domains are compared by structure-based alignment using the DaliLite server (99) with the sequence alignment colored in boxshade [grey for similar and black for identical (100)]. The signature residues based on multiple sequence alignment of HhH motifs by Doherty et al. (101) are manually colored (yellow for hydrophobic and red for Glycine) and denoted below the sequence (G indicates conserved glycine residues; h indicates a hydrophobic side chain). The secondary structure elements are shown based on the structure of human ERCC1 $(HhH)_2$ domain. *C,* NMR structure of the heterodimeric complex of the human ERCC1-XFP $(HhH)_2$ domains with the two-fold pseudo-symmetry axis (central red dot). *D*, Overlay of the ERCC1 and XPF $(HhH)_2$ structures using DaliLite (99), showing the overall fold similarity and close overlap of conserved F293 and F894 anchor residues. ERCC1 is depicted in blue, XPF in green, and the HhH motif signature resides are in yellow and red; GhG denotes signature residues for the hairpins. (A) and (B) are adapted from (102) and (92), respectively.

The three heterodimers ERCC1-XPF, MUS81-EME1 and MUS81-EME2 maintain their catalytic activity due to the conserved endonuclease domain in XPF and MUS81 (2,91,95). Of these, of ERCC1-XPF is involved in DNA repair pathways, as discussed above, whereas the other heterodimers have different functions. MUS81-EME1 is involved in the resolution of recombination intermediates containing Holliday junctions during the G2/M phase of the cell cycle (96,97), and the MUS81-EME2 heterodimer is required for processing and restart of stalled replication forks and is restricted to the S-phase of the cell cycle (97). Finally, FANCM-FAAP24 functions as a complex in FA, protecting cells from interstrand crosslink (ICL) and UV damage (2,98).

In contrast to the eukaryotic members of the XPF nuclease family, several archaeal members are functional as homodimers. *P. furiosus* from the archaeal subgroup euryarchaeota forms homodimers of the XPF homolog Hef (helicase-associated endonuclease for forked structure). A shorter form of the XPF homologs found in *S. solfataricus* and *A. pernix* (ApXPF), both from crenarchaeota, lacks the helicase domain, but combines a ERCC4 nuclease domain, a $(HhH)_2$ domain and a C-terminal PCNA-interaction motif (103,104). Human XPF homodimers have also been observed *in vitro* but not *in vivo* (2,105).

The superfamily 2 helicase-like domain

The helicase domains in the human XPF family members belong to the superfamily 2 (SF2) of RNA-helicases. They show a high similarity to the N-terminal part of XPF in seven conserved archaeal helicase families. In human XPF, the degenerate SF2 helicase-like domain is devoid of key residues essential for ATP binding and hydrolysis activity resulting in a non-functional helicase domain. Its function remains elusive. However, deletion of the helicase domain from ERCC1-XPF causes a strong decrease in *in vitro* DNA cleavage activity (106). The disrupted regions in the human SF2 helicase-like domain are localized in two motifs. The first is the MgATP/MgADP interacting GKT sequence, a consensus nucleotide binding of RNA helicases. The second motif contains acidic residues within the DEAD/DEAH box within the α-β domain, which is essential for chelating Mg^{2+} in RNA helicases. In FANCM, the DEAH helicase domain is functional and it harbors an ATP- dependent DNA translocase activity important for mono-ubiquitination. In the archaeal homodimer, Hef, the two characteristic helicase motifs are separated by an α-helical domain that are crucial for ATP hydrolysis and DNA translocation, and contribute to branched DNA recognition (2).

The ERCC4 endonuclease domain

The middle domain of the XPF family proteins is a metal-dependent endonuclease domain, designated as the excision repair cross complementation group 4 (ERCC4) domain. The architecture of Hef, the archaeal homolog of this domain is remarkably similar to that of the nuclease domain of type II restriction endonucleases. These contain a small endonucle-

ase subunit (ca. 18 kDa) with a conserved structural fold common in all restriction enzymes. This subunit is composed of five- or six-stranded β-sheets, flanked on both sides by α-helices. Despite the structural homology between Hef and type II restriction enzymes, sequence homology is mostly limited to the active site residues of Hef. The homologous regions include the $ERKX_3D$ signature motif with an extension GDX_n (GDX_nERKX_3D). This corresponds to the $PDX_n(E/D)XK$ motif of type II restriction enzymes (104). Mutational studies indicated metal binding and/or catalytic function for these residues, while no role was found in DNA binding (93). Similarly as found for type II restriction endonucleases, the conserved acidic residues within the ERCC4 domain motif participate in Mg^{2+}- and Mn^{2+}-ion binding. The polar residues in this motif including the lysine form hydrogen bonds with coordinated water molecules, and activate the water for nucleophilic attack in the endonuclease reaction. The hydrolysis results in breakage of the bond between a phosphate group and the 3'-oxygen of DNA (93,104,107). During this reaction, the divalent metal ion stabilizes the departing groups (108).

The ERCC1 central domain of ERCC1

Although the overall fold of the ERCC1 central domain (cERCC1) is similar to that of the nuclease domains of archaeal species, it lacks the characteristic catalytic residues (106). The non-catalytic domain has a completely distinct function in DNA substrate recognition. Previous NMR studies have shown that cERCC1 can interact to two highly conserved motifs in XPA (rich in glycine and glutamate) (109). The binding site was mapped to residues 92–119 of cERCC1, the region that corresponds to the catalytic 'signature' of XPF nuclease domain. However, in cERCC1, the catalytic residues "ERK" are replaced by "LFL" in human ERCC1 or "LF/YL" in other eukaryotic ERCC1 proteins. It was also noted that cERCC1 can bind ssDNA directly. Chemical shift perturbations mapped the ssDNA-binding site to residues N99, I102, L132, K213, A214 and Q134, which is located at a hydrophobic cleft on an opposite site of cERCC1 as compared to the XPA-binding site (106,109,110). It has been suggested that cERCC1 central domain plays an important role in the positioning of ERCC1-XPF heterodimer near the cleavage site (109) .

In the archaeal homolog of ERCC1-XPF, the nuclease domains are directly interacting. However, stable interactions between the nuclease domain of human XPF and the central domain of ERCC1 have not been observed (109).

The tandem helix-hairpin-helix domain $(HhH)_2$

The C-terminal part of the proteins of the XPF family contains a conserved domain engaged in binding DNA junctions. The domain consists of a tandem Helix-hairpin-Helix $(HhH)_2$ motif. The HhH motif is a characteristic DNA recognition motif, which in general binds nonspecifically to the phosphate backbone of DNA. It can be found in proteins such as DNA and RNA polymerases, DNA ligases, DNA glycosylases and in the NER repair nuclease UvrC (90,111). Two copies of this HhH motif are connected by an α-helix (helix γ) and together

this tandem $(HhH)_2$ motif forms a globular fold. The key residues in HhH motifs are found at positions 3, 6, 8, 9, 10, 14, 17 and 18 (101). Residues 3, 17, and 18 are responsible for hydrophobic packing of the motif, positions 6 and 9 define a type II β-turn. On positions 8 and 10 two glycine residues are generally found, that are important for the β-hairpin conformation (Figure 5B). Finally, position 14 of the HhH motif is generally occupied by a small hydrophobic residue such as alanine, to prevent a steric clash with the first helix of the HhH motif (101,111). This GXG sequence in the β-hairpins h_1 and h_2 is essential for interaction with the minor groove of duplex DNA and this type of sequence binds the phosphates backbone of both DNA strands separated by three base pairs. The hydrophobic residue X within the hairpin is most commonly a hydrophobic Ile, Val or Leu residue.

As mentioned before, proteins of XPF family can form heterodimers. For this, the $(HhH)_2$ domains of each protomer associate together into a stable heterodimeric endonuclease. Thus the tandem Helix-hairpin-Helix domain of XPF family of proteins are required both for DNA binding and dimerization. Dimerization was first observed in the archaeal XPF homolog, which formed a symmetric homodimeric complex. The eukaryotic heterodimers are not symmetric, and it was found for ERCC1-XPF that each of the two monomers has distinct functional roles (112). While the ERCC1 monomer contains a canonical $(HhH)_2$ motif, it was noted that in XPF the second hairpin of the $(HhH)_2$ motif is evolved into a short turn containing only three residues. The tandem heterodimeric $(HhH)_2$ domain is composed of a hydrophobic core and two pseudo-symmetric phenylalanine anchor residues that insert into a complementary cavity of their binding partner (Figure 5C and 5D)(95,106). The structure of the heterodimeric complex of the ERCC1-XFPF $(HhH)_2$ domains has been determined previously by NMR (95) (Figure 5C), and protein-crystallography (106). For complex formation two conserved phenylalanines, namely F293 of ERCC1 and F894 of XPF, are important (113). Taking advantage of the crucial role of these phenylalanines for the stability of the ERCC1-XPF complex, recently the cavity containing the F293 sidechain has been targeted by small molecules, which led to the loss of interaction between the two proteins (114). This induced loss-of-interaction could potentially be used to inhibit the mitigation of the damages that are stimulated during chemotherapy by alkylating agents.

The dimerization of the two $(HhH)_2$ domains implies that there exist two surfaces for potential DNA binding in ERCC1-XPF. It has been observed that for the ERCC1 and XPF archaeal homologs of this family the $(HhH)_2$ domain engages in both upstream and downstream duplex regions of their branched DNA substrates. DNA-binding properties of the heterodimeric ERCC1-XPF complex have been studied before. Tripsianes et al. (95,106) showed that the ERCC1 $(HhH)_2$ domain can interact with DNA through both canonical h_1 and h_2 β-hairpin motifs (95). Later, Das *et al* (112) showed that the XPF $(HhH)_2$ domain, which *in vitro* can form stable homodimers, has a larger interaction interface compared to the heterodimer and lacks the second canonical β-hairpin motif that is essential for dsDNA binding. The DNA-bind-

ing studies revealed that XPF homodimer can bind single-strand DNA (ssDNA). Together, this led to a model for binding of ERCC1-XPF at a junction between dsDNA and ssDNA, as present near the DNA damage site (95,112).

Additional interactions of ERCC1 and XPF

ERCC1 and XPF do not only interact with each other, but also with other proteins (Figure 5A). Some of these interactions, such as with the proteins XPA (in NER) and SLX4 (in ICL repair) were already described. Here, the findings on the interactions of ERCC1 and XPF with these and several other proteins are summarized.

As described above, the central domain of ERCC1 can interact to the ssDNA binding NER protein XPA (109). This can contribute to the DNA binding of ERCC1-XPF to the DNA damage site. It has also been demonstrated that ERCC1-XPF binds to DNA more efficiently in the presence of the replication protein A (RPA). RPA binds undamaged ssDNA and protects it from nucleases. The RPA binding site has been mapped to the N-terminus of XPF, and upon mutation of P85 to serine, XPF is unable to interact with RPA (102,115,116).

During ICL repair SLX4 functions as a molecular scaffold for the assembly of proteins including ERCC1-XPF. This warrants the processing of branched DNA substrates. Biochemical studies demonstrated the interaction between residues in the N-terminal part of SLX4 and ERCC1-XPF, but the precise binding regions have not been identified (102,117).

FANCG-deficient cells are not able to incise DNA at ICLs. It has been found that FANCG can also interact with the central domain in ERCC1 (50) . However, whereas the XPA binding involves residues 92-119 of cERCC1, the FANCG binding was shown to involve residues 96-119 of cERCC1 (50)(Figure 5A). Mutagenesis studies map the interaction to the tetratricopeptide repeats TPR-1, TRP3, TRP-5, and TRP6 in FANCG. It is believed that this interaction plays a crucial role for recruiting ERCC1-XPF to ICLs, analogous to the recruitment of ERCC1-XPF to sites of damage in NER by XPA. Further studies will be necessary to confirm these findings and map the interacting regions in more detail (50,58).

The MutS homolog protein MSH2, a mismatch repair (MMR) protein, is involved in the survival of cisplatin treated cells, probably due to its preferential binding to cisplatin ICLs containing a mismatch (118-120). Since ERCC1-XPF is also involved in cisplatin resistance it was suggested there might exist an interaction between these proteins. Lan et al. reported that during an XPA-independent repair mechanism the ERCC1 C-terminal residues 184–260, partly overlapping with the XPF interaction domain, might be involved in an interaction with MSH2 (118,121).

Furthermore, ERCC1-XPF can be stably associated with hRad52 via interactions between XPF and the N-terminal DNA-binding region of Rad52, in a DNA-independent manner. This interaction can promote the cleavage of 3'-overhangs and the processing of recombination intermediates that have arisen during the repair of double-strand DNA breaks (122).

ERCC1-XPF can also be involved in telomere maintenance and functions through direct or indirect interaction with the telomere repeats binding protein TRF2 independent of TRF2's DNA binding activity. In telomeres DNA loop conformations are formed in long G-rich 3'-overhangs at the ends of chromosomes. These long ssDNAs circles are stabilized by TRF2. ERCC1-XPF associates with TRF2 at these sites and its activity on 3'-overhangs depends on the presence of TRF2 (63). So far, no direct interaction has been reported between TRF2 and ERCC1-XPF. Since also TRF2 interacts with SLX4, it is possible that the effects attributed to ERCC1-XPF are mediated by SLX4 in a ternary complex of XPF, SLX4 and TRF2 (63). The XPF function in TRF2-mediated telomere shortening may not occur through its nuclease activity, since mutations in the conserved nuclease domain have no influence on telomere processing (123).

Recently, Liccardi et al. (124) investigated the interaction of ERCC1-XPF with the epidermal growth factor receptor (EGFR), since its existence had previously been suggested on the basis of a correlation between ERCC1 expression and occurrence of EGFR mutations in non-small cell lung cancer (NSCLC). It was demonstrated that interaction of ERCC1 and EGFR can be a key modulator of repair of DNA strand breaks, independent of NHEJ, and ICLs induced during cancer treatment (124,125).

SCOPE OF THE THESIS

Understanding DNA repair pathways such as NER, DSB and ICL repair is of basic interest for understanding fundamental cellular processes. It also forms the basis for understanding molecular details of diseases when defects occur in these pathways. Additional insight will help us to define relations to other cellular processes, and to find relations between the different DNA repair pathways, ageing and cancer. Detailed knowledge of these interactions can unravel the function of ERCC1 and XPF in the different DNA repair pathways. It forms a foundation for understanding the molecular basis for several related diseases and could lead to therapeutic strategies to overcome those. Moreover, the involvement of ERCC1-XPF in DNA repair, especially the removal of damages induced during platinum-chemotherapy in various cancer types, results in tumor resistance to these treatments. This reflects another horizon in medical importance of ERCC1-XPF, and has increased the attention to possible regulation of ERCC1-XPF expression and activity (14,124). In fact, the correlation between ERCC1 expression and the resistance to chemotherapeutic treatments makes ERCC1 a potential biomarker for the prediction of chemoresistance and patient survival (126). It has been proposed that ERCC1-XPF may be a possible target to overcome the chemo-resistance in cancers treated with cisplatin (114). By increasing our knowledge of the DNA repair pathways, it is my expectation that one day we can treat the related defects and overcome the associated diseases.

In the research described in this thesis the heterodimeric complex of the $(HhH)_2$

domains of ERCC1-XPF was studied, as well as its stability and role in DNA recognition. In **Chapter 2**, the effects of the disease-related F231L ERCC1 mutation (F231L) are described. F231 is part of the cavity of ERCC1, which contains the F894 anchor of XPF. It is shown that the F231L mutation causes a small but significant destabilization of the interaction between ERCC1 and XPF. As described above, loss of function of ERCC1-XPF endonuclease in diverse DNA repair pathways especially ICL repair can in some cases lead to the severe COFS syndrome. In principle, this can explain the observed phenotype of the patient for which the F231L mutation was established.

In **Chapter 3**, a detailed look into the interaction between ERCC1 and XPF in the ERCC1-XPF complex is given. The stability of the ERCC1-XPF complex was analyzed from a thermodynamic and a kinetic perspective and it is discussed which factors contribute to the stability of the heterodimeric complex. Furthermore, a model is presented that explains the preferred formation of ERCC1-XPF heterodimers.

In **Chapter 4**, the model for binding of ERCC1-XPF at a ds/ssDNA junction is examined. This model was initially based on interactions between XPF $(HhH)_2$ homodimers and short ssDNA nucleotides. The original model is validated with interaction studies of heterodimeric ERCC1-XPF and different DNA substrates, including ds/ssDNA forks. The biochemical and biophysical data show that the ERCC1 $(HhH)_2$ domain binds primarily to dsDNA and that at the same time the XPF $(HhH)_2$ domain can bind to ssDNA. In this way, the heterodimeric ERCC1-XPF complex can optimally bind at a ds/ssDNA junction as present during NER DNA repair.

REFERENCES

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) Molecular Biology of the Cell Garland Science, New York
2. Ciccia, A., McDonald, N., and West, S. C. (2008) Structural and functional relationships of the XPF/MUS81 family of proteins. Annual Review of Biochemistry 77, 259-287
3. Hoeijmakers, J. H. J. (2001) Genome maintenance mechanisms for preventing cancer. Nature 411, 366-374
4. Goode, E. L., Ulrich, C. M., and Potter, J. D. (2002) Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiology Biomarkers & Prevention 11, 1513-1530
5. Cook, J. S. (1970) Photoreactivation in animal cells. Photophysiology 5, 191-233
6. Friedberg, E. C., Walker, G. C., and Siede, W. (2006) DNA repair and mutagenesis, American Society for Microbiology, Washington DC
7. Hoeijmakers, J. H. J. (2009) DNA damage, aging, and cancer. New England Journal of Medicine 361, 1475-1485
8. Lord, C. J., and Ashworth, A. (2012) The DNA damage response and cancer therapy. Nature 481, 287-294
9. Cleaver, J. E. (1968) Defective repair replication of DNA in xeroderma pigmentosum. Nature 218, 652-656
10. Nguyen, T. A., Chul-Hwan, L., and Yeon-Soo, S. (2011) Lagging Strands Synthesis and Genomic Stability. in DNA Repair - On the Pathways to Fixing DNA Damage and Errors, InTech. pp 1-25
11. Zhang, N., Liu, X., Li, L., and Legerski, R. (2007) Double-strand breaks induce homologous recombinational repair of interstrand cross-links via cooperation of MSH2, ERCC1-XPF, REV3, and the Fanconi anemia pathway. DNA Repair 6, 1670-1678
12. He, J., Xu, Y., Qiu, L.-X., Li, J., Zhou, X.-Y., Sun, M.-H., Wang, J.-C., Yang, Y.-J., Jin, L., Wei, Q.-Y., and Wang, Y. (2012) Polymorphisms in ERCC1 and XPF genes and risk of gastric cancer in an eastern Chinese population. PloS ONE 7, e49308
13. Schärer, O. D. (2012) Alkyltransferase-like Proteins: Brokers Dealing with Alkylated DNA Bases. Molecular Cell 47, 3-4
14. Kirschner, K., and Melton, D. W. (2010) Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs. Anticancer Research 30, 3223-3232
15. Ellenberger, T. (2012) An Arresting Development in Transcription. Molecular Cell 46, 3-4
16. Marteijn, J. A., Lans, H., Vermeulen, W., and Hoeijmakers, J. H. J. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. Nature Reviews Molecular Cell Biology 15, 465-481
17. Hanawalt, P. C. (2002) Subpathways of nucleotide excision repair and their regulation. Oncogene 21, 8949-8956
18. Lans, H., Marteijn, J. A., and Vermeulen, W. (2012) ATP-dependent chromatin remodeling in the DNA-damage response. Epigenetics and Chromatin 5, 1-14
19. Brooks, P. J. (2007) The case for 8, 5'-cyclopurine-2'-deoxynucleosides as endogenous DNA lesions that cause neurodegeneration in xeroderma pigmentosum. Neuroscience 145, 1407-1417
20. Brooks, P. J. (2008) The 8, 5'-cyclopurine-2'-deoxynucleosides: candidate neurodegenerative DNA lesions in xeroderma pigmentosum, and unique probes of transcription and nucleotide excision repair. DNA Repair 7, 1168-1179
21. Hanawalt, P. C., and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises. Nature Reviews Molecular Cell Biology 9, 958-970
22. Egly, J.-M., and Coin, F. (2011) A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. DNA Repair 10, 714-721
23. Caldecott, K. W. (2007) Mammalian single-strand break repair: mechanisms and links with chromatin. DNA Repair 6, 443-453
24. Caldecott, K. W. (2008) Single-strand break repair and genetic disease. Nature Reviews Genetics 9, 619-631
25. Chapman, J. R., Taylor, M. R., and Boulton, S. J. (2012) Playing the end game: DNA double-strand break repair

pathway choice. Molecular cell 47, 497-510
26. McHugh, P. J., Spanswick, V. J., and Hartley, J. A. (2001) Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. The Lancet Oncology 2, 483-490
27. Kanaar, R., Hoeijmakers, J. H. J., and van Gent, D. C. (1998) Molecular mechanisms of DNA double-strand break repair. Trends in Cell Biology 8, 483-489
28. Goodarzi, A. A., and Jeggo, P. A. (2012) The repair and signaling responses to DNA double-strand breaks. Advances in Genetics 82, 1-45
29. Roy, R., Chun, J., and Powell, S. N. (2012) BRCA1 and BRCA2: Different roles in a common pathway of genome protection. Nature Reviews Cancer 12, 68-78
30. Liu, J., Doty, T., Gibson, B., and Heyer, W. D. (2010) Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nature Structural and Molecular Biology 17, 1260-1262
31. Sinha, M., Watanabe, S., Johnson, A., Moazed, D., and Peterson, C. L. (2009) Recombinational repair within heterochromatin requires ATP-dependent chromatin remodeling. Cell 138, 1109-1121
32. Jaskelioff, M., Van Komen, S., Krebs, J. E., Sung, P., and Peterson, C. L. (2003) Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin. Journal of Biological Chemistry 278, 9212-9218
33. Baudat, F., and de Massy, B. (2007) Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Research 15, 565-577
34. Serrentino, M.-E., and Borde, V. (2012) The spatial regulation of meiotic recombination hotspots: Are all DSB hotspots crossover hotspots? Experimental Cell Research 318, 1347-1352
35. Heyer, W. D., Li, X., Rolfsmeier, M., and Zhang, X. P. (2006) Rad54: The Swiss Army knife of homologous recombination? Nucleic acids research 34, 4115-4125
36. Bugreev, D. V., Pezza, R. J., Mazina, O. M., Voloshin, O. N., Camerini-Otero, R. D., and Mazin, A. V. (2010) The resistance of DMC1 D-loops to dissociation may account for the DMC1 requirement in meiosis. Nature Structural & Molecular Biology 18, 56-60
37. Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H. J., and Niedernhofer, L. J. (2008) ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Molecular and Cellular Biology 28, 5082-5092
38. Al-Minawi, A. Z., Saleh-Gohari, N., and Helleday, T. (2008) The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. Nucleic Acids Research 36, 1-9
39. Moore, J. K., and Haber, J. E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and Cellular Biology 16, 2164-2173
40. McVey, M., and Lee, S. E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends in Genetics 24, 529-538
41. Celli, G. B., and de Lange, T. (2005) DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. Nature Cell Biology 7, 712-718
42. Ahmad, S. I., and Hanaoka, F. (2008) Molecular mechanisms of xeroderma pigmentosum, Springer Science+Business Media, Texas, USA
43. Wu, H., and Roks, A. J. M. (2014) Genomic instability and vascular aging: A focus on nucleotide excision repair. Trends in Cardiovascular Medicine 24, 61-68
44. Deng, S. K., Gibb, B., De Almeida, M. J., Greene, E. C., and Symington, L. S. (2014) RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nature Structural and Molecular Biology 21, 405-412
45. McVey, M. (2014) RPA puts the brakes on MMEJ. Nature Structural and Molecular Biology 21, 348-349
46. Williams, H. L., Gottesman, M. E., and Gautier, J. (2013) The differences between ICL repair during and outside of S phase. Trends in Biochemical Sciences 38, 386-393
47. Huang, Y., and Li, L. (2013) DNA crosslinking damage and cancer-a tale of friend and foe. Translational Cancer

Research 2, 144-154
48. Muniandy, P. A., Liu, J., Majumdar, A., Liu, S.-t., and Seidman, M. M. (2010) DNA interstrand crosslink repair in mammalian cells: step by step. Critical reviews in biochemistry and molecular biology 45, 23-49
49. Huang, M., Kim, J. M., Shiotani, B., Yang, K., Zou, L., and D'Andrea, A. D. (2010) The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. Molecular Cell 39, 259-268
50. Wang, C., and Lambert, M. W. (2010) The Fanconi anemia protein, FANCG, binds to the ERCC1-XPF endonuclease via its tetratricopeptide repeats and the central domain of ERCC1. Biochemistry 49, 5560-5569
51. Bogliolo, M., Schuster, B., Stoepker, C., Derkunt, B., Su, Y., Raams, A., Trujillo, J. P., Minguillón, J., Ramírez, M. J., Pujol, R., Casado, J. A., Baños, R., Rio, P., Knies, K., Zúñiga, S., Benítez, J., Bueren, J. A., Jaspers, N. G. J., Schärer, O. D., De Winter, J. P., Schindler, D., and Surrallés, J. (2013) Mutations in ERCC4, Encoding the DNA-Repair Endonuclease XPF, Cause Fanconi Anemia. The American Journal of Human Genetics 92, 800-806
52. Moldovan, G.-L., and D'Andrea, A. D. (2009) How the fanconi anemia pathway guards the genome. Annual Review of Genetics 43, 223-249
53. Walden, H., and Deans, A. J. (2014) The fanconi anemia DNA repair pathway: Structural and functional insights into a complex disorder. in Annual Review of Biophysics
54. Adair, G. M., Rolig, R. L., Moore-Faver, D., Zabelshansky, M., Wilson, J. H., and Nairn, R. S. (2000) Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination. EMBO Journal 19, 5552-5561
55. Bhagwat, N., Olsen, A. L., Wang, A. T., Hanada, K., Stuckert, P., Kanaar, R., D'Andrea, A., Niedernhofer, L. J., and McHugh, P. J. (2009) XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. Molecular and Cellular Biology 29, 6427-6437
56. Kottemann, M. C., and Smogorzewska, A. (2013) Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. nature 493, 356-363
57. Klein Douwel, D., Boonen, R. A., Long, D. T., Szypowska, A. A., Räschle, M., Walter, J. C., and Knipscheer, P. (2014) XPF-ERCC1 Acts in Unhooking DNA Interstrand Crosslinks in Cooperation with FANCD2 and FANCP/SLX4. Molecular Cell 54, 460-471
58. Deans, A. J., and West, S. C. (2011) DNA interstrand crosslink repair and cancer. Nature Reviews Cancer 11, 467-480Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J. D., Ellenberger, T. E., Schärer, O. D., and Walter, J. C. (2008) Mechanism of replication-coupled DNA interstrand crosslink repair. Cell 134, 969-980
59. Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T. V., Schärer, O. D., Elledge, S. J., and Walter, J. C. (2009) The fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science 326, 1698-1701
60. Williams, H. L., Gottesman, M. E., and Gautier, J. (2012) Replication-independent repair of DNA interstrand crosslinks. Molecular Cell 47, 140-147
61. Besse, B., Olaussen, K. A., and Soria, J.-C. (2013) ERCC1 and RRM1: ready for prime time? Journal of Clinical Oncology 31, 1050-1060
62. Zhu, X.-D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H. J., and de Lange, T. (2003) ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Molecular Cell 12, 1489-1498
63. Naim, V., Wilhelm, T., Debatisse, M., and Rosselli, F. (2013) ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. Nature Cell Biology 15, 1008-1015
64. Kamileri, I., Karakasilioti, I., and Garinis, G. A. (2012) Nucleotide excision repair: new tricks with old bricks. Trends in Genetics
65. Le May, N., Mota-Fernandes, D., Vélez-Cruz, R., Iltis, I., Biard, D., and Egly, J. M. (2010) NER Factors Are Recruited to Active Promoters and Facilitate Chromatin Modification for Transcription in the Absence of Exogenous

Genotoxic Attack. Molecular cell 38, 54-66
66. de Boer, J., and Hoeijmakers, J. H. J. (2000) Nucleotide excision repair and human syndromes. Carcinogenesis 21, 453-460
67. Nouspikel, T. (2009) DNA repair in mammalian cells. Cellular and Molecular Life Sciences 66, 994-1009
68. Rass, U., Ahel, I., and West, S. C. (2007) Defective DNA repair and neurodegenerative disease. Cell 130, 991-1004
69. 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. Human Molecular Genetics 7, 969-974
70. Sijbers, A. M., van Voorst Vader, P. C., Snoek, J. W., Raams, A., Jaspers, N. G. J., and Kleijer, W. J. (1998) Homozygous R788W point mutation in the XPF gene of a patient with xeroderma pigmentosum and late-onset neurologic disease. Journal of Investigative Dermatology 110, 832-836
71. Imoto, K., Boyle, J., Oh, K., Khan, S., Ueda, T., Nadem, C., Slor, H., Orgal, S., Gadoth, N., Busch, D., Jaspers, N. G. J., Tamura, D., DiGiovanna, J. J., and Kraemer, K. H. (2007) Patients with defects in the interacting nucleotide excision repair proteins ERCC1 or XPF show xeroderma pigmentosum with late onset severe neurological degeneration. in Journal of Investigative Dermatology, 127, S92-S92
72. Gregg, S. Q., Robinson, A. R., and Niedernhofer, L. J. (2011) Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. DNA Repair 10, 781-791
73. Spivak, G. (2004) The many faces of Cockayne syndrome. Proceedings of the National Academy of Sciences of the United States of America 101, 15273-15274
74. Brace, L. E., Vose, S. C., Vargas, D. F., Zhao, S., Wang, X. P., and Mitchell, J. R. (2013) Lifespan extension by dietary intervention in a mouse model of Cockayne Syndrome uncouples early postnatal development from segmental progeria. Aging Cell 12, 1144-1147
75. Lehmann, A. R. (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. Biochimie 85, 1101-1111
76. Nouspikel, T. (2009) Nucleotide excision repair: variations on versatility Cellular and Molecular Life Sciences 66, 994 – 1009
77. De Waard, H., De Wit, J., Andressoo, J. O., Van Oostrom, C. T. M., Riis, B., Weimann, A., Poulsen, H. E., Van Steeg, H., Hoeijmakers, J. H. J., and Van Der Horst, G. T. J. (2004) Different effects of CSA and CSB deficiency on sensitivity to oxidative DNA damage. Molecular and Cellular Biology 24, 7941-7948
78. Kashiyama, K., Nakazawa, Y., Pilz, D. T., Guo, C., Shimada, M., Sasaki, K., Fawcett, H., Wing, J. F., Lewin, S. O., Carr, L., Li, T.-S., Yoshiura, K.-I., Utani, A., Hirano, A., Yamashita, S., Greenblatt, D., Nardo, T., Stefanini, M., McGibbon, D., Sarkany, R., Fassihi, H., Takahashi, Y., Nagayama, Y., Mitsutake, N., Lehmann, A. R., and Ogi, T. (2013) Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. The American Journal of Human Genetics 92, 807-819
79. Kraemer, K. H., Patronas, N. J., Schiffmann, R., Brooks, B. P., Tamura, D., and DiGiovanna, J. J. (2007) Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. Neuroscience 145, 1388-1396
80. 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., Theil, A. F., Vermeulen, W., Van Der Horst, G. T. J., Meinecke, P., Kleijer, W. J., Vijg, J., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (2006) A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. Nature 444, 1038-1043
81. Jaspers, N. G. J., Raams, A., Silengo, M. C., Wijgers, N., Niedernhofer, L. J., Robinson, A. R., Giglia-Mari, G., Hoogstraten, D., Kleijer, W. J., and Hoeijmakers, J. H. J. (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. The American Journal of Human Genetics 80, 457-466
82. Andressoo, J. O., and Hoeijmakers, J. H. J. (2005) Transcription-coupled repair and premature ageing. Muta-

tion Research/Fundamental and Molecular Mechanisms of Mutagenesis 577, 179-194
83. Ahmad, A., Enzlin, J. H., Bhagwat, N. R., Wijgers, N., Raams, A., Appledoorn, E., Theil, A. F., Hoeijmakers, J. H. J., Vermeulen, W., and Jaspers, N. G. J. (2010) Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. PLoS Genetics 6, e1000871
84. Bessho, T., Sancar, A., Thompson, L. H., and Thelen, M. P. (1997) Reconstitution of human excision nuclease with recombinant XPF-ERCC1 complex. Journal of Biological Chemistry 272, 3833-3837
85. Park, C.-H., and Sancar, A. (1994) Formation of a ternary complex by human XPA, ERCC1, and ERCC4 (XPF) excision repair proteins. Proceedings of the National Academy of Sciences 91, 5017-5021
86. Moggs, J. G., Yarema, K. J., Essigmann, J. M., and Wood, R. D. (1996) Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1, 3-intrastrand d (GpTpG)-cis-platin adduct. Journal of Biological Chemistry 271, 7177-7186
87. Eisen, J. A., and Hanawalt, P. C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes. Mutation Research - DNA Repair 435, 171-213
88. Sgouros, J., Gaillard, P.-H. L., and Wood, R. D. (1999) A relationship between a DNA-repair/recombination nuclease family and archaeal helicases. Trends in Biochemical Sciences 24, 95-97
89. Singh, S., Folkers, G. E., Bonvin, A. M. J. J., 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 Journal 21, 6257-6266
90. Wang, Y., Han, X., Wu, F., Leung, J. W., Lowery, M. G., Do, H., Chen, J., Shi, C., Tian, C., and Li, L. (2013) Structure analysis of FAAP24 reveals single-stranded DNA-binding activity and domain functions in DNA damage response. Cell research 23, 1215-1228
91. Wienk, H., Slootweg, J. C., Speerstra, S., Kaptein, R., Boelens, R., and Folkers, G. E. (2013) The Fanconi anemia associated protein FAAP24 uses two substrate specific binding surfaces for DNA recognition. Nucleic Acids Research 41, 6739-6749
92. Enzlin, J. H., and Schärer, O. D. (2002) The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. The EMBO Journal 21, 2045-2053
93. Youds, J. L., O'Neil, N. J., and Rose, A. M. (2006) Homologous recombination is required for genome stability in the absence of DOG-1 in Caenorhabditis elegans. Genetics 173, 697-708
94. Tripsianes, K., Folkers, G., Das, D., Odijk, H., Jaspers, N. G. J., Hoeijmakers, J. H. J., 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
95. Sarbajna, S., Davies, D., and West, S. C. (2014) Roles of SLX1-SLX4, MUS81-EME1, and GEN1 in avoiding genome instability and mitotic catastrophe. Genes and Development 28, 1124-1136
96. Pepe, A., and West, S. C. (2014) MUS81-EME2 Promotes Replication Fork Restart. Cell Reports 7, 1048-1055
97. Yang, H., Zhang, T., Tao, Y., Wang, F., Tong, L., and Ding, J. (2013) Structural insights into the functions of the FANCM-FAAP24 complex in DNA repair. Nucleic Acids Research, 10573-10583
98. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16, 566-567
99. Falquet, L., Bordoli, L., Ioannidis, V., Pagni, M., and Jongeneel, C. V. (2003) Swiss EMBnet node web server. Nucleic acids research 31, 3782-3783
100. 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 research 24, 2488-2497
101. McNeil, E. M., and Melton, D. W. (2012) DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. Nucleic Acids Research 40, 9990-10004
102. Roberts, J., Bell, S., and White, M. (2003) An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. Molecular Microbiology 48, 361-371
103. 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

11, 445-457
104. Arcas, A., Fernández-Capetillo, Ó., Cases, I., and Rojas, A. M. (2014) Emergence and evolutionary analysis of the human DDR network: implications in comparative genomics and downstream analyses. Molecular Biology and Evolution 31, 940-961
105. Tsodikov, O. V., Enzlin, J. H., Schärer, O. D., and Ellenberger, T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. Proceedings of the National Academy of Sciences of the United States of America 102, 11236-11241
106. McDaniel, L. D., and Schultz, R. A. (2009) XPF/ERCC4 and ERCC1: their products and biological roles. in Molecular Mechanisms of Xeroderma Pigmentosum, Springer. pp 65-82
107. Pingoud, A., and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. Nucleic Acids Research 29, 3705-3727
108. 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 Research 35, 5789-5798
109. 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. Proceedings of the National Academy of Sciences of the United States of America 91, 5012-5016
110. Shao, X., and Grishin, N. V. (2000) Common fold in helix-hairpin-helix proteins. Nucleic Acids Research 28, 2643-2650
111. Das, D., Folkers, G. E., van Dijk, M., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., and Boelens, R. (2012) The structure of the XPF-ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix-hairpin-helix domains in ss/ds DNA recognition. Structure 20, 667-675
112. de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1998) Mapping of interaction domains between human repair proteins ERCC1 and XPF. Nucleic Acids Research 26, 4146-4152
113. Jordheim, L. P., Barakat, K. H., Heinrich-Balard, L., Matera, E.-L., Cros-Perrial, E., Bouledrak, K., El Sabeh, R., Perez-Pineiro, R., Wishart, D. S., and Cohen, R. (2013) Small molecule inhibitors of ERCC1-XPF protein-protein interaction synergize alkylating agents in cancer cells. Molecular pharmacology 84, 12-24
114. Matsunaga, T., Park, C.-H., Bessho, T., Mu, D., and Sancar, A. (1996) Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. Journal of Biological Chemistry 271, 11047-11050
115. Fisher, L. A., Bessho, M., Wakasugi, M., Matsunaga, T., and Bessho, T. (2011) Role of interaction of XPF with RPA in nucleotide excision repair. Journal of molecular biology 413, 337-346
116. Muñoz, I. M., Hain, K., Déclais, A.-C., Gardiner, M., Toh, G. W., Sanchez-Pulido, L., Heuckmann, J. M., Toth, R., Macartney, T., Eppink, B., Kanaar, R., Ponting, C. P., Lilley, D. M. J., and Rouse, J. (2009) Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. Molecular Cell 35, 116-127
117. Lan, L., Hayashi, T., Rabeya, R. M., Nakajima, S., Kanno, S.-i., Takao, M., Matsunaga, T., Yoshino, M., Ichikawa, M., Te Riele, H., Tsuchiya, S., Tanaka, K., and Yasui, A. (2004) Functional and physical interactions between ERCC1 and MSH2 complexes for resistance to cis-diamminedichloroplatinum (II) in mammalian cells. DNA Repair 3, 135-143
118. Claij, N., and Te Riele, H. (2002) Methylation tolerance in mismatch repair proficient cells with low MSH2 protein level. Oncogene 21, 2873-2879
119. Kothandapani, A., Sawant, A., Dangeti, V. S. M. N., Sobol, R. W., and Patrick, S. M. (2013) Epistatic role of base excision repair and mismatch repair pathways in mediating cisplatin cytotoxicity. Nucleic acids research 41, 7332-7343
120. Bertrand, P., Tishkoff, D. X., Filosi, N., Dasgupta, R., and Kolodner, R. D. (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair. Proceedings of the National Academy of

Sciences of the United States of America 95, 14278-14283
121. Bennardo, N., Cheng, A., Huang, N., and Stark, J. M. (2008) Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genetics 4, e1000110
122. Wu, Y., Zacal, N. J., Rainbow, A. J., and Zhu, X.-D. (2007) XPF with mutations in its conserved nuclease domain is defective in DNA repair but functions in TRF2-mediated telomere shortening. DNA Repair 6, 157-166
123. Liccardi, G., Hartley, J. A., and Hochhauser, D. (2014) Importance of EGFR/ERCC1 interaction following radiation-induced DNA damage. Clinical Cancer Research 20, 3496-3506
124. Gandara, D. R., Grimminger, P., Mack, P. C., Lara Jr, P. N., Li, T., Danenberg, P. V., and Danenberg, K. D. (2010) Association of epidermal growth factor receptor activating mutations with low ERCC1 gene expression in non-small cell lung cancer. Journal of Thoracic Oncology 5, 1933-1938
125. 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

Chapter 2

The ERCC1 mutation F231L causes dissociation of the Nucleotide Excision Repair complex ERCC1-XPF

Maryam Faridounnia, Hans Wienk, Lidija Kovačič, Gert E. Folkers, Nicolaas G.J. Jaspers, Robert Kaptein, Jan H.J. Hoeijmakers and Rolf Boelens

Manuscript under review in JBC

ABSTRACT

The ERCC1-XPF heterodimer, a structure-specific DNA endonuclease, is best known for its function in the Nucleotide Excision Repair pathway. The ERCC1 point mutation F231L, located at the hydrophobic interaction interface of ERCC1 and XPF, leads to severe NER pathway deficiencies. Here, we analyze biophysical properties and report the NMR structure of the complex of the C-terminal $(HhH)_2$ domains of ERCC1-XPF that contains this mutation. The structures of wildtype and the F231L mutant are very similar. The F231L mutation results in only a small disturbance of the ERCC1-XPF interface where, in contrast to F231, L231 lacks interactions stabilizing the ERCC1-XPF complex. This results in a more dynamic complex causing reduced stability and an increased dissociation rate of the mutant complex as compared to wildtype. These data provide a biophysical explanation for the severe NER deficiencies caused by this mutation.

INTRODUCTION

UV light, ionizing radiation, mutagenic compounds, as well as natural cellular metabolism can damage DNA and lead to mutations causing cancer or cell death and senescence contributing to ageing-related disorders (1). As a countermeasure the cell has developed a set of complementary repair mechanisms, one of which is the nucleotide excision repair (NER)3 pathway (2,3). This multi-step 'cut-and-patch' repair system deals with the broad class of helix-distorting lesions, including UV-induced DNA lesions and numerous bulky chemical adducts and intrastrand crosslinks. Two proteins involved in this repair process, ERCC1 (excision repair cross complementation group 1) and XPF (xeroderma pigmentosum complementation group F) form a heterodimeric complex which functions as a structure-specific DNA endonuclease responsible for the 5' incision, when the lesion is excised from the DNA as a 22-30 oligonucleotide in a late step of the NER reaction (4,5). There is considerable evidence for additional NER-independent roles of the ERCC1-XPF heterodimer in homologous recombination (6-9), repair of interstrand DNA crosslinks (ICL) (9,10), telomere length maintenance (11-13), double-strand break (DSB) repair (14) and gene conversion (9,15).

The proper functioning of the ERCC1-XPF complex as endonuclease in NER critically depends on the formation of a heterodimer *via* their C-terminal tandem helix-hairpin-helix $(HhH)_2$ domains (16-19). This contact region of the dimer is not only important for the interaction between the monomers, but also plays a role in DNA binding: the C-terminal domain of ERCC1 is essential for double-strand (ds) DNA binding (18,20,21), whereas the C-terminal $(HhH)_2$ domain of XPF can bind single-strand (ss) DNA (22). We proposed that this would position the ERCC1-XPF complex at a ss/ds-DNA junction, and places the catalytic module of XPF, which is adjacent to its $(HhH)_2$ domain, to the 5' site at some distance from the DNA lesion (22).

In classic cases of XPF deficiency, the DNA repair function is still partly retained and UV sensitivity is only moderately increased (23-25). In recent years, however, several more severe cases of XPF-related deficiencies have been described (10,26-29). Mutational studies on a mouse homolog of human ERCC1 have shown the N-terminus of ERCC1, despite being dispensable for DNA repair, to be important for heterodimer expression (7,30). Inherited deficiencies in ERCC1 are rare in human(28,31,32), probably reflecting the lethality of many genetic mutations in this protein, which became apparent also from studies in mice perinatal mortality (33). The similarity of these symptoms to those resulting from XPF defects in mice combined with the biochemical and cellular parallels support the notion that ERCC1 and XPF function as an obligate complex *in vivo*. Only, two cases involving ERCC1 defects have been reported in human, having in common a F231L point mutation that causes one of the severest cases of NER deficiency known until now. The first case (*165TOR*) is a biallelic heterozygous mutation, one of which is F231L, which leads to severe degeneration of

the brain and spinal cord and death at early infancy (31,34). The second case of ERCC1 deficiency is a biallelic homozygous F231L mutation, resulting in an early onset of Cockayne Syndrome (CS) and patient's death at early infancy. Both cases link the F231L mutation to NER deficiency, where the ERCC1-XPF complex is still present, but in reduced quantities, resulting into slower repair rates (28).

F231 is conserved in ERCC1 of invertebrates and mammals. The residue is located in the C-terminal $(HhH)_2$ domain of ERCC1, and in the wildtype protein structure it has extensive interactions with the $(HhH)_2$ domain of XPF (16,18,20,21). Therefore, in principle the loss of ERCC1-XPF functionality in the F231L mutant could be due to ERCC1 instability, and/or the reduced stability of the entire ERCC1-XPF complex. Furthermore, since the phenylalanine is in close proximity to the hairpin regions of the ERCC1 $(HhH)_2$ domain that are important for its DNA-binding role, the mutation could also affect the DNA-binding properties of ERCC1-XPF complex. In order to analyse these possibilities, we produced and purified the ERCC1-XPF $(HhH)_2$ complex containing the F231L mutation and compared several structural and functional properties to those of wildtype ERCC1-XPF. We show that while the mutant heterodimer preserves the overall fold of the ERCC1-XPF $(HhH)_2$ complex as well as its DNA-binding properties, the mutation causes a side chain reorientation for residue 231 that disrupts interactions with nearby amino acids of XPF. We show that this small, local disturbance in ERCC1-XPF heterodimer leads to a lower stability of the complex due to an increased dissociation rate.

EXPERIMENTAL PROCEDURES

Cloning, Protein Expression and Purification - The construct for the F231L ERCC1 $(HhH)_2$ mutant was prepared by using an extended mutagenic 5´ primer during PCR amplification with the wildtype construct as a template (residues 220-301). Subsequently, it was cloned into the BamHI-XhoI site of the pET28b bicistronic expression vector (18), substituting the wild-type ERCC1 coding region, construction was verified by DNA sequencing. The plasmid was transformed into the Escherichia coli rosetta (DE3) strain, overexpressed, and purified using a Ni^{2+}-NTA-agarose column followed by gel filtration as described for wildtype ERCC1-XPF (18). During NMR analysis it was found out that the expressed protein was a 1:1 mixture of a longer than expected protein, probably due to the presence of a rare alternate start codon (35,36) next to the constructed one, which resulted in seven additional amino acids at the N-terminus of ERCC1 (RIRRRYN). This alternative initiation codon was found for both wildtype and mutant, and did not interfere with DNA binding or stability as judged by NMR.

Surface Plasmon Resonance - DNA binding to the F231L ERCC1-XPF $(HhH)_2$ heterodimer was measured by SPR as described before (37). All solutions for SPR were filtered through a 0.22 µm filter and degassed. The experiments were performed at 12 °C

in 10 mM Hepes, pH 7.5, 50 mM NaCl and 0.005% (w/v) Tween20 (SPR buffer) at 10 µl/min using a Biacore® X system (Biacore AB, Sweden). The mutant and wildtype ERCC1-XPF $(HhH)_2$ complexes were exchanged to the SPR buffer using Zeba Spin Desalting Columns (Thermo Scientific, USA). Protein concentrations were determined before and after each experiment using SDS-PAGE. Low-binding tubes and tips were used to prevent the loss of sample during handling. Mutant or wildtype ERCC1-XPF $(HhH)_2$ complex (0.05 mM), in absence or presence of double strand "bubble10" (i.e. annealed DNA $GGGCGGCGGG(T)_{10}GGCGGGGCGG$ and $CCCGCCGCCC(T)_{10}CCGCCCCGCC$) with concentrations ranging from 0.01 to 100 mM using a dilution factor of 2, was incubated in the SPR buffer for 20 minutes on ice, injected over the NTA CM5 sensor chip. The association was monitored for 60 seconds and the dissociation for 120 seconds. Before each SPR experiment, Ni^{2+} was immobilised on the second flow cell of the sensor chip (Biacore AB, Sweden) while the first flow cell was used as a reference surface. Between consecutive injections, the chip was regenerated with 10 µl of 0.25 M EDTA in 3.5 M guanidine pH 8.0. Data was processed using the Biaevaluation 3.2 software. To calculate the apparent dissociation constant (K_D^{app}) of the DNA binding to the ERCC1-XPF $(HhH)_2$ heterodimer, the relative amount of DNA-free protein was divided by the total DNA concentration and fitted according to a simple 1-to-1 binding model using the software Prism (GraphPad Software Inc, USA). Since the DNA prevented His-tag binding of the ERCC1-XPF $(HhH)_2$ domains on the Ni-NTA surface, the relative amount of DNA-free protein was determined as response values at the end of the loading time of 60 seconds (R_{60}) divided by the R_{60} value of the ERCC1-XPF $(HhH)_2$ heterodimer in absence of DNA (fraction $=R_{60}[ERCC1\text{-}XPF]_{free}/R_{60}[ERCC1\text{-}XPF]_{total}$).

The real-time monitoring of the interaction of ERCC1 with XPF was performed in a flow cell maintaining the laminar flow (BIAcore AB), and with a Sensor Chip Ni^{2+}-NTA (Biacore AB, Sweden) (37). All experiments were performed at a flow rate of 20 µl/minutes using similar conditions as with the SPR for DNA-binding. A set of experiments was conducted at three concentrations of wildtype and mutant ERCC1-XPF $(HhH)_2$ heterodimer in duplicates (0.2, 0.1 and 0.05 µM) and three temperatures (12 ˚C, 20 ˚C and 25 ˚C). The association of the ERCC1-XPF complex on the chip and the dissociation from the chip (expressed as reciprocal of time) was monitored for 60 and 120 seconds, respectively. Identical amounts of proteins were loaded on the chip without preloaded Ni^{2+}, in order to subtract the baseline from each obtained sensogram before analysis. SPR curves were fitted as a single-phase decay model over 60 to 180 seconds using the software GraphPad Prism and the equation:

$$R=(R_U - P)\, e^{(k_{off}^{app}\, t)} + P$$

Here R is the maximum response in arbitrary response units (RU), R0 the response after association (60 seconds), P the plateau at infinite times, and k_{off}^{app} is the sum of the off-rate of the XPF dissociation from His-tagged ERCC1 and of the His-tagged protein from the chip. The SPR chip performance was optimized using different protein concentrations combined with

different Ni^{2+} concentrations and different flow rates, based on the findings of Nieba et al. (37). The protein complex was loaded on the chip with a 12-25 °C temperature-gradient and the dissociation was followed for 1000 seconds. No reliable data could be obtained above 25 °C due to elevated dissociation of His-tagged protein from the chip. The baseline for the buffer was subtracted.

Circular Dichroism - CD spectra were measured using a JASCO J-810 spectropolarimeter as described before (38,39). The instrument was calibrated with d-10-camphor sulfonate, assuming $[\theta]_{291}$ = 7820 deg cm^2 $dmol^{-1}$, and with non-hygroscopic ammonia d-10-camphor sulfonate, assuming $[\theta]_{290.5}$ = 7820 deg cm^2 $dmol^{-1}$. The thermal unfolding of wildtype and mutant ERCC1-XPF-His_6 was measured in the far-UV wavelength range (190-250 nm) under identical conditions (CD buffer: 50 mM $NaPO_4$, 100 mM NaCl, pH 7.0, protein concentrations about 25 µM, path-length cell 1 mm). Ten 10-second scans were recorded and the temperature was increased from 15-85 ˚C in increments of 1 ˚C/minute. The signal intensities at 222 nm were collected for each temperature and analysed as described above (38).

Thermal Shift Assay - Thermal shift experiments were performed in 96-well thin-wall PCR plates sealed with Optical-Quality Sealing Tape. The fluorescence changes were followed on an iCycler iQ PCR Detection System (BioRad) from 15-95 ˚C in increments of 0.2 ˚C/minute by monitoring Sypro Orange fluorescence and analysed (40). The wavelengths used for excitation and emission were 490 and 575 nm, respectively.

NMR spectroscopy - NMR spectra were recorded at 290 K, unless indicated differently, using a Bruker Avance DRX 600 MHz system with a TCI cryoprobe, an Avance II 750 MHz machine with a TXI probe and an Avance III 900 MHz spectrometer with a TCI cryoprobe. Triple resonance experiments (including HMCMCGCBCACONH) were recorded using 0.4 mM $^{13}C/^{15}N$-labeled ERCC1-XPF-His_6 in NMR buffer (50 mM NaPi, 100 mM NaCl, 8% ^2H_2O, pH7.0) (41,42). Spectra were processed with Topspin version 2.1/3.0, or NMRPipe (43) and analysed in Sparky (44) to assign backbone and side-chain ^1H, ^{15}N and ^{13}C resonances. A 2D NOE spectrum (τmix = 80 ms), a 3D NOESY-[^1H,^{15}N]-HSQC spectrum (τ_{mix} = 100ms), and a 3D NOESY-[^1H,^{13}C]-HSQC spectrum (τ_{mix} = 80ms) were recorded. An additional 3D NOESY-[^1H,^{13}C]-HSQC spectrum (τmix = 80ms) was obtained with ERCC1-XPF in D_2O.

For NMR temperature series 2D [^1H,^{15}N]-HSQC spectra were measured at temperatures ranging from 280 to 333 K using 400 µM of ^{15}N-labeled wildtype and mutant ERCC1-XPF in NMR buffer. In total, fifteen 2D [^1H,^{15}N]-HSQC spectra were recorded at 750 MHz and chemical shift changes were analysed with Sparky (45,46).

Proton-deuterium exchange of ^{15}N-labeled mutant and wildtype ERCC1-XPF $(HhH)_2$ heterodimer was measured at 750 MHz in 100 mM NaCl, 50 mM NaPi, pH 6.8. Prior to monitoring H/D exchange, the acquisition parameters were adjusted with wildtype ERCC1-XPF prepared in NMR buffer (47). Dissolving the lyophilized sample in ^2H_2O, transferring to the NMR tube, transfer to the NMR instrument and temperature equilibration resulted in a

total preparation time of 8 minutes. Thereafter the acquisition of series of 2D [^{1}H,^{15}N]-HSQC spectra for following the H/D exchange was started. For each residue the decay of the signal intensities as a function of time was evaluated as described (48).

Phase-modulated CLEAN chemical exchange (CLEANEX-PM) experiments were recorded at 600 MHz as explained by Huang et al. (49). Signal intensities were obtained from 2D (CLEANEX-PM)-FHSQC spectra recorded with a mixing time of 100 ms and for reference compared to a control FHSQC spectrum as described in (41).

NMR structures were calculated using CYANA3.0 (50,51). Distance restraints were derived from NOESY peak intensities. Torsion angle restraints ϕ and ψ were derived from HN, Hα, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts using TALOS+ (52) with an extra error of ±10 degrees. In each CYANA cycle 100 structures were generated and energy-minimized using 10,000 steps of simulated annealing. After initial CYANA runs additional NOE assignments were obtained by manual inspection of the NOESY spectra. Before water refinement, CYANA-derived NOE and H-bond restraints, in addition to the torsion angle restraints derived from TALOS+ (53) were used to calculate 100 structures with CNS (54,55). Water refinement was done on each of these structures using restrained molecular dynamics in a hydrated environment. The twenty lowest energy conformations were validated using WHATIF, WHATCHECK and iCing (56-58). Molecular images were generated using Pymol (59,60).

For the wildtype ERCC1-XPF $(HhH)_2$ heterodimer domain, we observed inconsistencies in the precise length of helix α. While the helix length for the F231L ERCC1-XPF mutant was consistent with that of the crystal structure of wildtype ERCC1-XPF (2A1J), this helix was shorter in the NMR structure of wildtype ERCC1-XPF (1Z00). Therefore the NMR structure of wildtype ERCC1-XPF was recalculated following the same protocol as for the mutant. For this, the previously determined NOE based restraints and restraints for H-bonds (18) were combined with ϕ and ψ torsion angle restraints derived from TALOS+ (53) for the calculation of an ensemble of 100 conformers with CNS (54,55). Water refinement, validation and generation of molecular images were as for F231L ERCC1-XPF.

RESULTS

Similar to the heterodimeric complex of the C-terminal $(HhH)_2$ domains of wildtype ERCC1-XPF (18), the system containing the F231L mutation could only be overexpressed and purified as an obligate complex using a bicistronic expression vector. During size exclusion chromatography there was no indication of dissociation arguing that F231L ERCC1-XPF can form a stable heterodimeric $(HhH)_2$ complex, similar to the wildtype situation.

DNA-binding - Based on the models presented before, in wildtype ERCC1 the helices of the first HhH motif make extensive contacts with the dsDNA minor groove near the ss/ds-DNA junction (22). Therefore, the F231L mutation, located in this first helix, could well interfere with DNA binding of ERCC1-XPF. It is known that bubble-shaped DNA substrates

containing ss/ds-DNA junctions efficiently bind wildtype ERCC1-XPF heterodimer (18,21). One of the highest-affinity DNA probes appears to be “bubble10”, where two double-stranded GC-rich segments flank a region of 20 unpaired thymine bases. We have used the Electrophoretic Mobility Shift Assay and found that both wildtype and F231L mutant ERCC1-XPF bind to DNA with comparable affinity (data not shown). Recently however, we found that the relatively weak DNA binding of ERCC1-XPF can be more accurately analysed using Surface Plasmon Resonance (SPR). We have compared the ability of F231L and wildtype ERCC1-XPF to bind bubble10 DNA. For this, Ni^{2+}-nitrilotriacetic acid (Ni^{2+}-NTA) chips were loaded with wildtype or mutant heterodimer and immobilized on the chips via the C-terminal Histidine-tag of ERCC1. The recorded SPR responses (Supplementary Figure S1) of wildtype and mutant complexes upon DNA binding were used to calculate the dissociation constants (Kd). Both were determined to be about 0.2 mM, i.e. within error the same (Figure 1A). These DNA-binding studies show that it is not likely that impaired DNA binding would be the explanation for the malfunctioning of nucleotide excision repair as seen for the patients with the ERCC1-XPF F231L mutation.

Biophysical characterization - Next, the effect of the F231L mutation on the stability of the ERCC1-XPF $(HhH)_2$ complex was analysed. Therefore, we compared the temperature-dependent behavior of F231L and wildtype ERCC1-XPF $(HhH)_2$ heterodimers using a thermal shift assay (ThermoFluor) (40), which measures the increase in fluorescence upon exposure of hydrophobic patches during protein unfolding. The assay is commonly used to follow the unfolding of a protein with increasing temperatures. For ERCC1-XPF two transitions upon temperature increase were found (Figure 1B), similarly as observed previously using other approaches (38). The first transition at ~50 °C is interpreted as the dissociation of the ERCC1-XPF heterodimer followed by ERCC1 denaturation, and the second transition at ~80 °C is explained as the unfolding of subsequently formed XPF homodimers. We noted a significant difference in the first transition temperatures between the wildtype and F231L complexes, with onsets of dissociation at 52.0 ± 0.5 °C and 47.5 ± 1.0 °C, respectively. Although the difference in transition temperatures for the heterodimers can be interpreted as directly caused by a difference in unfolding of mutant versus wildtype ERCC1-XPF, it could also be the result from structural distortions in ERCC1, in its binding partner XPF, and/or by enhanced dissociation of ERCC1-XPF, which could all lead to increased exposure of hydrophobic surfaces.

Far-UV Circular Dichroism (CD) directly probes changes in secondary structure formation, and ellipticity at 222 nm is indicative for the helical content of proteins (61). For both mutant and wildtype ERCC1-XPF again a biphasic behaviour was observed upon increasing temperature, similarly as previously observed for the wildtype heterodimer (Figure 1C) (18). Also here, the first transition (at a Tm of 51.0 ± 1.0 °C) reflects heterodimer dissociation followed by ERCC1 $(HhH)_2$ unfolding, while the XPF $(HhH)_2$ domain remains

folded and forms a homodimer (38). The high-temperature transition (with a T_m of 76.0 ± 1.0 °C) reflects XPF homodimer dissociation and subsequent unfolding. Comparing the change in ellipticity as a function of temperature for the F231L and wildtype ERCC1-XPF complex, we observe no evident differences in the first transition between the two proteins. The second transition is also very similar for both complexes, which can be expected since in both cases it probably involves the unfolding of the same XPF homodimer. Whereas the CD spectra show that the ERCC1 and XPF secondary structure elements of both wildtype and mutant ERCC1-XPF start to unfold at similar temperatures, the ThermoFluor analysis shows that the mutant heterodimer is influenced already at lower temperature than wildtype. Since CD monitors the unfolding of helices, while the ThermoFluor analysis can also signal dissociation of the heterodimer that precedes the unfolding of the ERCC1 monomer, we propose that the observed differences are indicative for the increased dissociation of F231 ERCC1-XPF $(HhH)_2$ heterodimers due to the F231L mutation.

Association and dissociation of proteins forming complexes can be studied using SPR (62). To study the dissociation of the wildtype and mutant heterodimeric ERCC1-XPF complexes directly, they were immobilized via the C-terminal Histidine-tag of ERCC1 on Ni^{2+}-NTA chips. The increase in SPR response upon loading the ERCC1-XPF complex shows that the complex can be successfully loaded within 50 seconds (Figure 1D). Upon buffer elution a drop of the SPR response is observed till it reaches a plateau at about 50% of the initial response. Given the similar molecular weights of XPF and ERCC1, this reduction can be explained by the dissociation of XPF from ERCC1. Upon continued elution for 1000 s, we observe the return of the SPR response to its starting level due to dissociation of His-tagged ERCC1 from the Ni^{2+}-NTA chip (not shown).

The dissociation of wildtype and mutant ERCC1-XPF was followed by SPR at different heterodimer concentrations, temperatures and salt concentrations (Figure 1D, Supplementary Figure S2). Negative response at temperatures above 25 °C suggests that the SPR results are biased by ERCC1 precipitation. At 25 °C and lower temperatures, at lower ionic strength the dissociation rates k_{off}^{app} for wildtype and mutant seem only marginally different (Table 1, Supplementary Figure S2). For example, at 12 °C and low ionic strength, the k_{off}^{app} for the F231L mutant was 1.9 ± 0.3 × 10^{-2} s^{-1}, while for wildtype ERCC1-XPF k_{off}^{app} was 1.3 ± 0.2 × 10^{-2} s^{-1}. At higher ionic strength the k_{off}^{app} for the wildtype ERCC1-XPF is similar to the k_{off}^{app} at 50 mM NaCl, with a value of 1.43 ± 0.03 × 10^{-2} s^{-1}, whereas for the F231L mutant the dissociation rates k_{off}^{app} significantly increases to 3.0 ± 0.3 × 10^{-2} s^{-1} (Figure 1D).

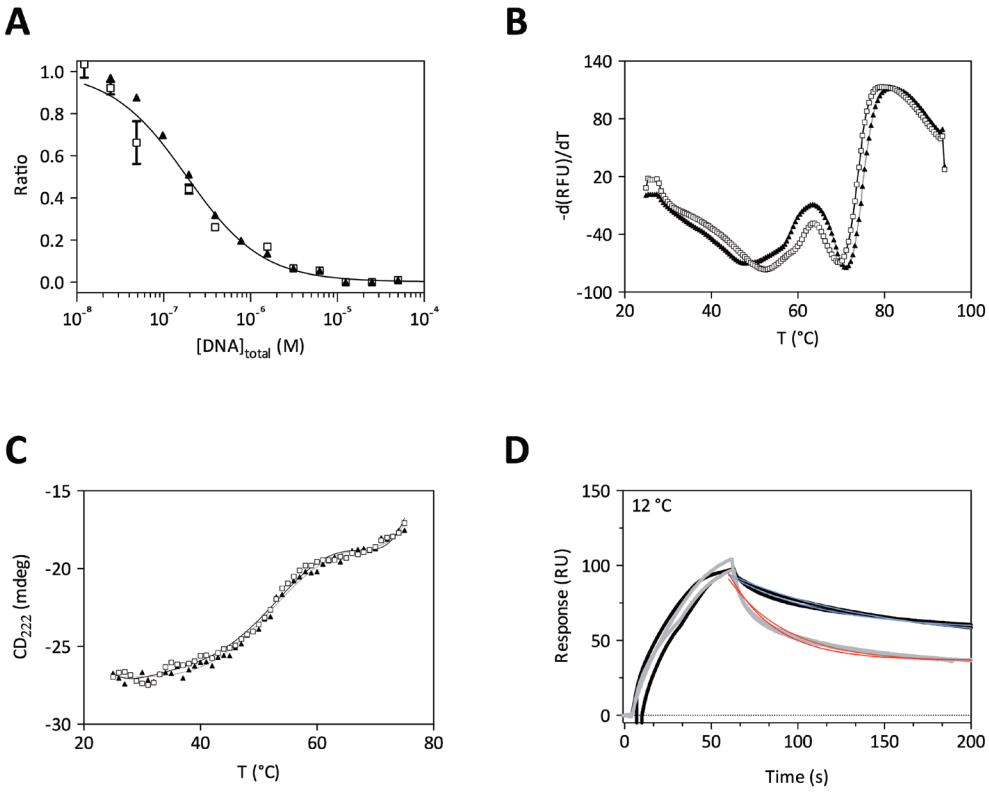


FIGURE 1. **DNA binding and stability of the wildtype ¬and F231L ERCC1-XPF $(HhH)_2$ heterodimers.** *A*, DNA binding by wildtype and mutant ERCC1-XPF measured using SPR. The figure displays ratio of R_{60} values of ERCC1-XPF $(HhH)_2$ in the presence and absence of bubble10 DNA (R_{60} $[ERCC1\text{-}XPF]_{free}$ /R_{60} $[ERCC1\text{-}XPF]_{total}$) as a function of the concentration of DNA. *B*, thermal shift assay. The stability of ERCC1-XPF $(HhH)_2$ is measured using the Sypro Orange Thermofluor assay (40). The empty squares and black line indicate the data for wildtype, the black triangles and grey line the data for the mutant. *C*, thermal stability measured by CD spectroscopy. Color scheme as in (B). *D*, dissociation of ERCC1-XPF measured by SPR. The graphs show the dissociation of XPF from immobilized wildtype ERCC1 (shown in grey and fitted in red) and immobilized F231L ERCC1 (shown in black and fitted in blue at 150 mM NaCl and 12°C).

TABLE 1

Dissociation rates for wildtype and F231L mutant ERCC1-XPF $(HhH)_2$ heterodimers measured by SPR

ERCC1-XPF	k_{off}^{app} $[10^{-2}\ s^{-1}]$		
50 mM NaCl	12°C	20°C	25°C
wildtype	1.3±0.2	1.5 ±0.1	2.3±0.3
F231L	1.9±0.3	2.0±0.5	3.0±0.1
150 mM NaCl	12°C		
wildtype	1.43±0.03		
F231L	3.0±0.3		

NMR spectral differences between wildtype and F231L ERCC1-XPF - Since NMR assignments for the wildtype complex were already available (18), we report here those for the F231L mutant ERCC1-XPF $(HhH)_2$ complex. The 2D [^{1}H,^{15}N]-HSQC spectra of the wildtype and F231L mutant ERCC1-XPF $(HhH)_2$ complexes are similar, but do show differences, both in chemical shift (Figure 2 and Supplementary Figure S3) and in intensity (Figure 2). Comparing the chemical shifts (Figure 2A, 2C and Supplementary Figure S3) it is noted that differences are mostly close to the F231L mutation site (i.e. for ERCC1 helix α L227, Q229, D230, V232, S233 and V235), but significant chemical shift differences are also found for residues further away, both in ERCC1 and in XPF $(HhH)_2$. Among these distant residues are M224, E237, T252-L260, L263, R268 and G278 in ERCC1 helices α and γ, and F894-S899 in the XPF C-terminal helix. Whereas the local chemical shift differences could be merely the effect of the different chemical nature of F231 and L231, the distant differences may reflect structural differences, which could influence the association and dissociation behavior of the ERCC1-XPF complex. Figure 2C shows that the observed chemical shift differences are most pronounced for residues that locate in the wildtype structure at the ERCC1-XPF interface, suggesting that the monomeric folds of ERCC1 and XPF are very similar in both complexes and that the mutation mainly affects contacts between ERCC1 and XPF.

Intensity differences between the 2D [^{1}H,^{15}N]-HSQC spectra of wildtype and F231L mutant ERCC1-XPF were noted as well (Figure 2B). The graph shows a plateau that coincides with regions that in the wildtype ERCC1-XPF $(HhH)_2$ complex have stable secondary structure. F231L ERCC1-XPF shows reduced intensity for residues that are directly contacting the mutation site in the wildtype structure, i.e. for residues in helices α, β and γ of ERCC1 (Q229, D230, V232, S233, T252, L253, T255, F257, G258, S259) and for residues in the C-terminal helix of XPF (T892, S893, F894). In fact, lack of signal intensity might be the reason that the amide proton of L231 could not be unambiguously assigned. The differences between the wildtype and F231L ERCC1-XPF HSQC signal intensity likely reflect increased local dynamics in F231L ERCC1-XPF at the mutation site.

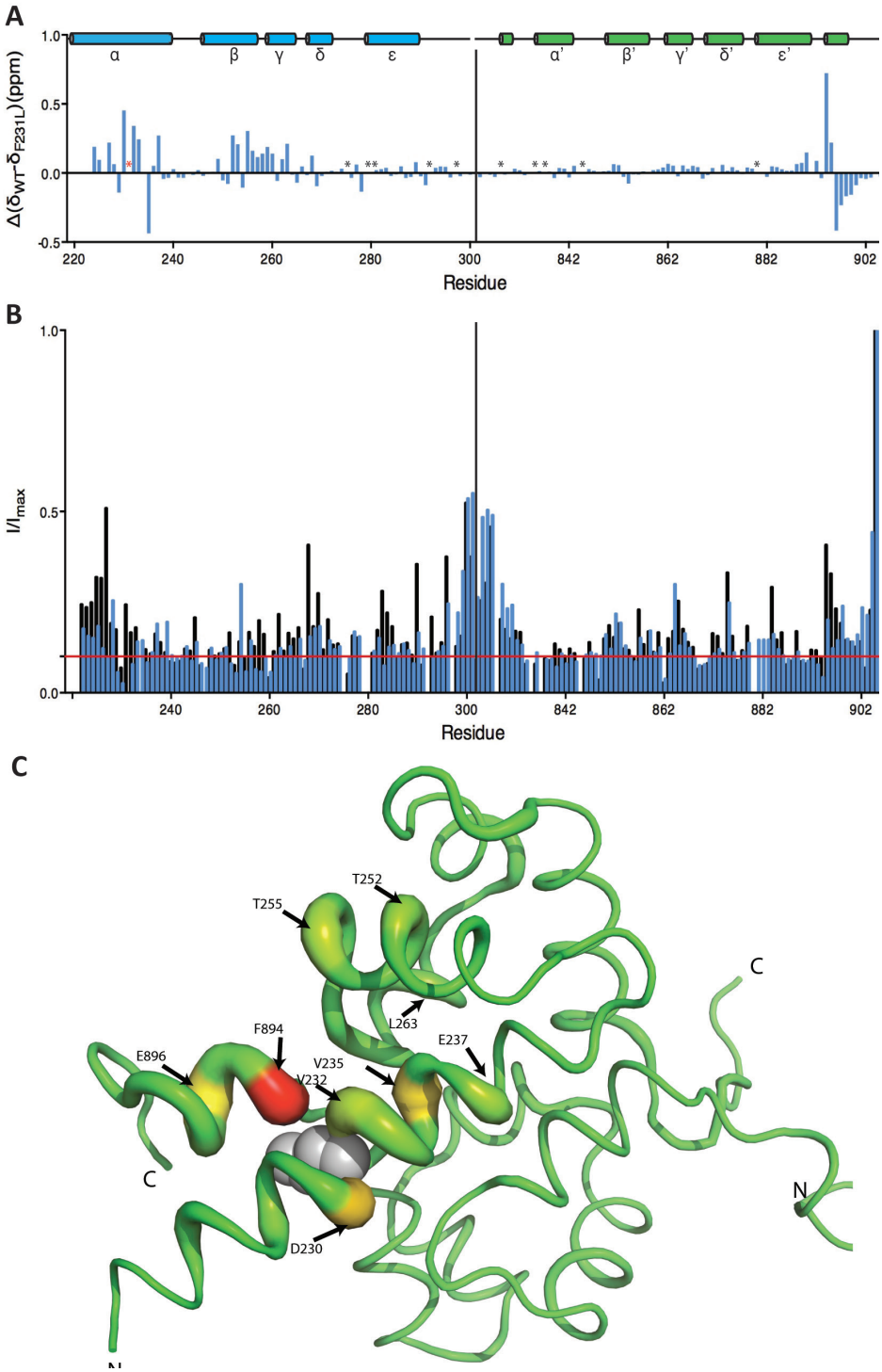
Amide proton exchange - The interface of the complex of the $(HhH)_2$ domains of wildtype ERCC1-XPF is formed by hydrophobic interactions involving several aromatic phenylalanine residues (of which two important ones are the pseudo-symmetric anchors F293 (in ERCC1) and F894 (in XPF)), and several intermolecular hydrogen bonds (S259 HG…O T892, E261 HN…O I890, F293 HN…O K860, N834 HN…O F293, N834 HD21…O P292, H891 HD1…OE2 E261 and A895 HN…O G258) (18,21). The importance of hydrogen bonding for the stability of the ERCC1-XPF complex, and the effect of the F231L mutation on this stability was studied by amide proton/deuterium (H/D) exchange. For this, we acquired a series of 2D [^{1}H,^{15}N]-HSQC spectra of wildtype and mutant ERCC1-XPF $(HhH)_2$ heterodimer freshly dissolved in ^2H_2O and compared the H/D exchange for all residues that were still visible after the dead-time of the experiment (i.e. eight minutes) and were non-overlapping in the

spectra of the wildtype ERCC1-XPF heterodimer as well as the mutant F231L. A comparison of the H/D exchange could be made for residues C238, T240-V242, L253, T255, S259-E261, L263, A265, S267, L271, A272, A282, L285, H290, L294 of ERCC1 and the residues D839, L842, K847, M856, V859, N861-A863, L873, T874, I876, L877, A883, Q885-T892 of XPF. Overall, we note similar H/D exchange rates for wildtype and mutant ERCC1-XPF $(HhH)_2$ heterodimer, with slow exchanging amide protons in the core of the complex and quick exchange for residues that are more surface exposed. Only the ERCC1 residue L253 of the F231L mutant shows significant faster exchange than in the wildtype, which suggests an increased solvent accessibility for the helix of ERCC1 (Supplementary Figure S4).

Whereas H/D exchange measures slow (minutes to hours) amide proton exchange (63), CLEANEX-PM experiments can probe fast (sub-second) solvent exchange (49). Also with CLEANEX-PM no significant exchange differences were observed between the wildtype and mutant ERCC1-XPF. The only exception for which solvent exchange was reduced in the mutant is the amide proton of V235 in helix α of ERCC1, which is four residues apart from the F231L mutation site and hydrogen bonded to residue 231 in the wildtype structure. If any, this suggests even a slightly increased stability of helix α in F231L ERCC1 as compared to the wildtype system.

In summary, between the wildtype and mutant ERCC1-XPF $(HhH)_2$ heterodimers differences in amide proton chemical shifts and in NMR intensities are seen for regions in direct contact with the mutation site and for residues localized near the F894 anchor of the interface of ERCC1-XPF. The solvent exchange experiments indicate that the integrity of the ERCC1 and XPF monomers is hardly affected by the F231L mutation. In line with the thermal stability analysis this also indicates that the reduced stability of ERCC1-XPF that we observed, is not caused by differences in secondary structure stability, but is rather the result of different association or dissociation properties of the $(HhH)_2$ complex.

FIGURE 2. NMR chemical shift and [^{1}H,^{15}N]-HSQC signal intensity analysis of wildtype and F231L ERCC1-XPF $(HhH)_2$ heterodimers. A, chemical shift differences between the mutant and wildtype. The figure shows the difference for amide proton chemical shifts $\Delta HN(\delta_{WT}-\delta_{F231L})$ (in ppm). Unassigned residues are marked with an asterisk, for the mutation site in red. B, variation in [^{1}H,^{15}N]-HSQC cross-peak intensities. The intensities for the wildtype are depicted in black and for the F231L mutant in blue. Signal intensities are normalized to the highest signal intensity (K905). The horizontal line is a measure for the signal intensity for residues in secondary structure elements. Regions with HSQC intensities below average, with significant lower intensity for the mutant compared to the wildtype ERCC1-XPF $(HhH)_2$ heterodimer are indicated by an asterisk.


A
α
β
γ
δ
ε
α'
β'
γ'
δ'
ε'
Δ(δWT-δF231L)(ppm)
Residue
B
I/Imax
Residue
C
T252
T255
L263
E237
F894
E896
V235
V232
D230
C
N
C
N

Temperature-dependent chemical shift changes - Temperature-dependent chemical shift changes for the amide protons of wildtype and mutant ERCC1-XPF are usually linear as a function of temperature, and the $\Delta\delta/\Delta T$ gradients can be related to hydrogen bond strengths. (45,46). Slow amide proton exchange combined with a temperature coefficient more positive than −4.5 ppb/K indicates the presence of a hydrogen bond, while fast exchange combined with a temperature coefficient more negative than −4.5 ppb/K indicates the absence of a hydrogen bond. The residues with significant differences between wildtype and mutant (cf Table 2, Supplementary Figure S5) are mostly located in direct vicinity of the mutation site or around the anchor in its neighborhood, F894, indicating that the mutation causes mainly a local structural effect. However, overall, with only a few exceptions, the temperature coefficients are more negative for the F231L mutant than wildtype ERCC1-XPF, suggesting a reduced overall stability of F231L ERCC1-XPF.

TABLE 2

Residues with significantly different proton temperature coefficients ($\Delta\delta/\Delta T$) for wildtype and F231L mutant ERCC1-XPF $(HhH)_2$ heterodimers measured by NMR

	Wildtype	F231L mutant
	$\Delta\delta/\Delta T$ (ppb/K)	
V235	-2.932	-9.358
C238	2.733	1.095
L239	-2.684	-5.859
L242	-2.932	-4.450
Q251	-6.294	-2.527
L253	-4.080	-2.476
T255	-4.972	-0.051
S259	-5.494	1.977
L263	-3.992	-2.01
A265	-0.062	-2.125
I890	-1.831	-1.350
F894	-2.565	2.415

NMR structure - To see whether the increased dissociation of the F231L mutant ERCC1-XPF complex is caused by structural differences between the wildtype and mutant, we decided to determine the NMR structure of the F231L ERCC1-XPF $(HhH)_2$ heterodimer and compared it to existing structures of wildtype ERCC1-XPF. To allow direct comparison we recalculated the earlier reported wildtype ERCC1-XPF $(HhH)_2$ heterodimer NMR structure with exactly the same parameters as the F231L mutant.

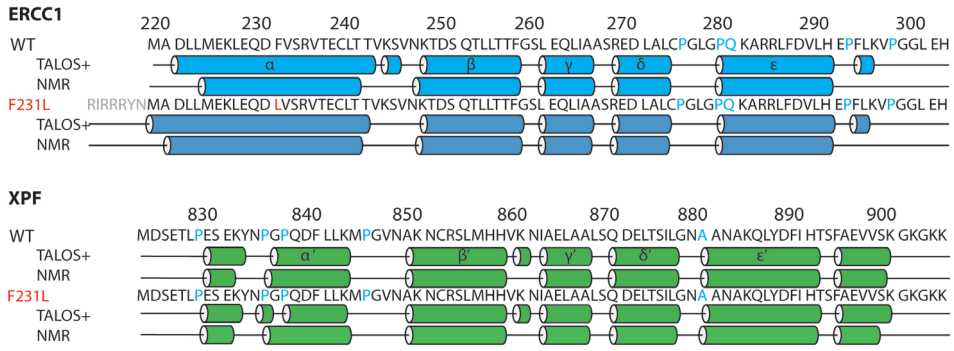


FIGURE 3. **Secondary structure elements of wildtype and F231L ERCC1-XPF $(HhH)_2$ heterodimers.** Top: secondary structure elements for wildtype (WT) and mutant (F231L) ERCC1 predicted by TALOS+, and detected in the lowest energy structure after (re)calculation of the NMR ensembles. Top: ERCC1; Bottom: XPF. Unassigned amide protons are shown in blue in the sequence. The N-terminal extension assigned in mutant ERCC1 is shown in gray.

The NMR structures of the wildtype and F231L ERCC1-XPF $(HhH)_2$ heterodimeric complexes were determined using NOE-based distance restraints and chemical-shift based torsion angle restraints (Figure 4A, Table 3). The total number of NOEs used to calculate the structure of the F231L mutant ERCC1-XPF is lower than that for wildtype. Still, the overall quality of the wildtype and F231L mutant ERCC1-XPF structures is similar. The secondary structure elements in the wildtype and mutant structure ensembles are very similar. The helices α-ε of wildtype ERCC1 (i.e. M224-T240, K247-F257, L260-A265, R268-A272, P279-H290, resp.) and the helices α'-ε' of XPF (i.e. E829-E831, P837-K843, A849-H858, I862-A866, Q870-L877, A880-H891, F894-K900, resp.) are almost invariably present in all conformers of the ensemble of mutant ERCC1-XPF. However, some differences are evident from the NMR structures. For the wildtype heterodimer, assignments are lacking for the N-terminus of ERCC1, which obviously leads to a lower number of NOEs for this part of the protein. Therefore, as pointed out above, helix α of wildtype ERCC1 starts at M224, whereas that of F231L ERCC1 starts at D221 (Figure 4A).

Even though the two NMR ensembles of wildtype and mutant ERCC1-XPF are well-defined and the overall fold is the same, the structures show small local differences. The closest-to-mean RMSD values for all heavy atoms are 1.35 Å for ERCC1 residues 227-294, 1.15 Å for the XPF core residues 831-896 and 1.29 Å for both regions together. This indicates small but genuine differences between the two structures, indicating that the XPF part is more similar than the ERCC1 monomer. After superimposing the backbones, interesting differences are seen for the side chain orientations of the two complexes. In wildtype ERCC1-XPF, F231 is embedded deeply in the core of the $(HhH)_2$ heterodimer structure (Figure 4A). Its aromatic side chain and the backbone show short distances (<4Å) to ERCC1 residues L227-V235 of the same helix, but the residue is also close to S259 and L260 of ERCC1 and

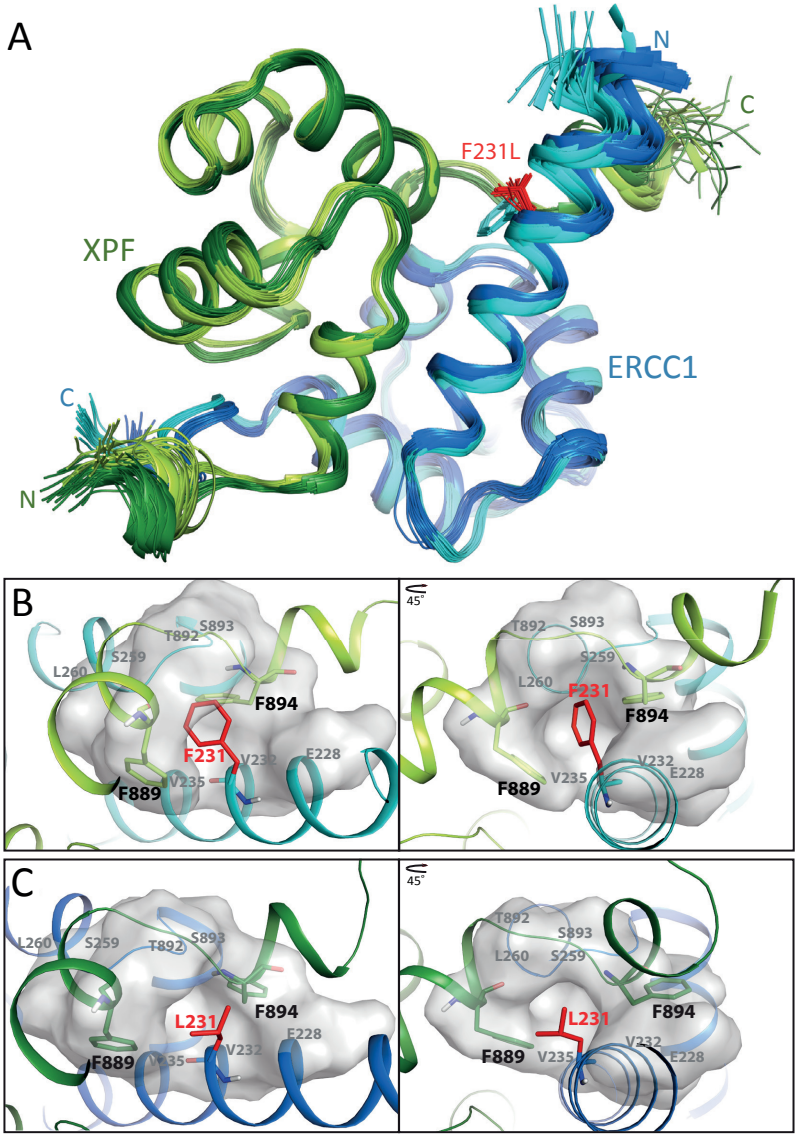


FIGURE 4. Solution structures of wildtype and F231L ERCC1-XPF $(HhH)_2$ heterodimers. *A*, superimposed NMR ensembles of F231L and recalculated wildtype ERCC1-XPF $(HhH)_2$. The backbone of ERCC1 is colored blue (wildtype: cyan; F231L: blue); the backbone of XPF is colored green (wildtype: limon; F231L: dark green). The mutated residue L231 is shown in red. *B*, surface depiction and side chain organization for phenylalanines F231, F889 and F894 of wildtype ERCC1-XPF $(HhH)_2$ and *C*, for residues L231, F889 and F894 of F231L ERCC1-XPF $(HhH)_2$. Backbones, F889 and F894 are colored as in Figure 4*A*; the side chain of F231/L231 is colored in red, the labels in grey depict residues surrounding F231/L231. *C*, the chemical shift differences are mapped on the wildtype ERCC1-XPF $(HhH)_2$ heterodimer structure, the ribbon width is a measure for the amide chemical shift difference between wildtype and mutant ERCC1-XPF $(HhH)_2$ complex. The most pronounced chemical shifts differences are indicated by arrows. In addition, maximum differences are colored red, medium differences as yellow and small or no differences are colored green. The F231L mutation site is represented as grey spheres.

TABLE 3

Statistics of the structure calculations for wildtype and mutant ERCC1-XPF $(HhH)_2$ heterodimer

	Wildtype	F231L mutant
RMSD (Å)[a]		
ERCC1 (bb / heavy atom)	0.21 / 0.53	0.24 / 0.71
XPF (bb / heavy atom)	0.22 / 0.49	0.30 / 0.68
Complex (bb / heavy atom)	0.24 / 0.53	0.30 / 0.72
Number of peaks		
Assigned	11077	8686
Total	12947	11365
Number of experimental restraints		
Intra	895	561
Sequential	1218	738
Medium	1246	599
Long intra	777	303
Long inter	478	223
Total NOE	4614	2424
Dihedrals	197	267
H-bonds	56	63
Consistent violations		
NOE > 0.5	0	0
Dihedral > 5	0	1 (F889 ϕ[b])
WhatCheck[a,c]		
Ramachandran plot appearance	-2.044 +/- 0.209	-2.703 +/- 0.171
2nd generation packing quality	4.312 +/- 1.403	3.156 +/- 1.304
χ 1 – χ2 rotamer normality	-4.879 +/- 0.167	-5.085 +/- 0.322
Backbone conformation	-0.412 +/- 0.279	-4.954 +/- 0.209
Ramachandran plot[a,c]		
Favourable	94.4	96.8
Additionally allowed	5.6	3.2
Generously	0.0	0.0
Disallowed	0.0	0.0

[a] core residues 227-294 (ERCC1) and 831-896 (XPF)

[b] the violation is introduced by the water refinement

[c] Results from iCING

to residues F889, I890, T892, S893, and F894 of XPF. The aromatic side chain of F231 in wildtype ERCC1-XPF is packed between the aromatic rings of F894 and F889, making T-shaped arrangement with respect to each of these neighboring phenylalanines. The F231L mutation does not influence the integrity of helix α, despite a very different side chain orientation of residue L231 as compared to F231 in wildtype (Figure 4A-C, Figure 5). In the mutant, L231 maintains similar distances and angles toward residues L227-V235 of ERCC1 and residues F889 and S893 of XPF. However, the mutation allows the side chain of F894 to reorient. Consequently, at the ERCC1 region flanking the XPF C-terminus, contacts are lost between residues F894, G258, S259 and L260 and the interactions between the segments I890-H891 and L260-E261 are weakened. The reduction in aromatic ring interactions likely explains part of the differences in chemical shifts for segments L227-E237 and D252-L263 of ERCC1 and the segment T892-S899 of XPF (Figure 2).

Due to the F231L mutation, the aromatic side chains of F889 and F894 cannot maintain their T-shaped arrangement with respect to residue 231 (Figure 4B-C). In the wildtype complex, F889 contacts residues D230, F231, R234, V235, C238 and L260 of ERCC1, and M844, P845, the N-terminally located helix ε' (Q885-D888) and segment I890-T892 of XPF. In the mutant, F889 modestly changes its contacts with L260 of ERCC1, while most

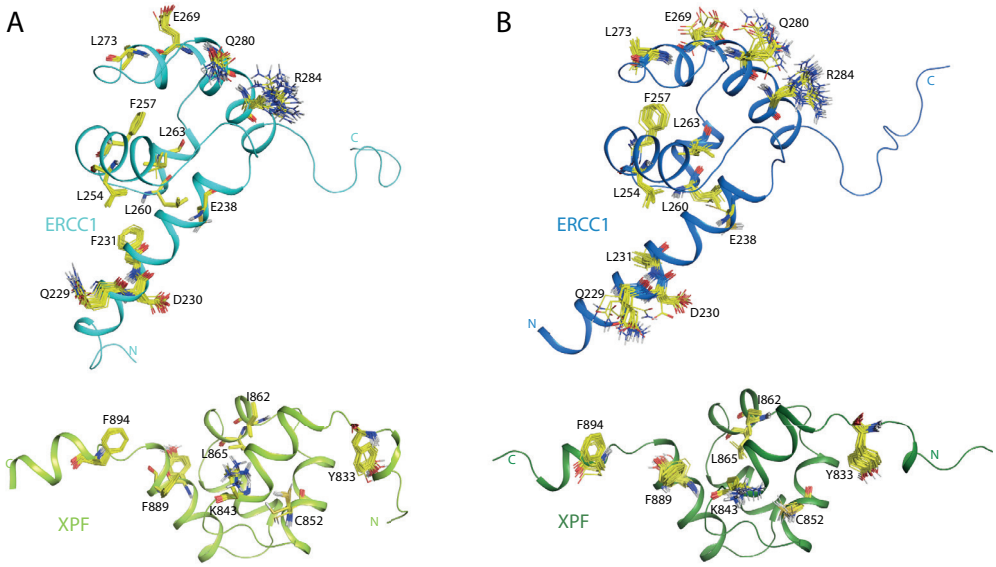


FIGURE 5. Side chain orientations in wildtype and F231L ERCC1-XPF $(HhH)_2$ heterodimer. Side chains of residues showing the largest differences between the NMR ensembles are depicted for (*A*) the (re)calculated complexes of wildtype ERCC1 $(HhH)_2$ (top) and XPF (bottom), (*B*) the calculated complexes of F231L mutant ERCC1 $(HhH)_2$ (top) and XPF (bottom). Backbone coloring is as in figure 4A. Amino acid side chains are shown in yellow, with polar hydrogen in grey, oxygen in red and nitrogen in blue. The monomers of ERCC1 and XPF are separated for better view.

interactions within XPF remain. The situation for the F894 anchor, however, is very different. Where most contacts with the neighboring residues in XPF's F894 interaction pocket (T892-S899) are still present in the mutant, the intermolecular contacts with ERCC1 are strongly altered due to side chain reorientations for residues Q229, L231 and V235 of helix α, residues L254 and F257 of helix β and residue E261 and L263 of helix γ. It was found that in addition to these differences in phenylalanine contacts several other residues of the complex show side chain reorientations. Most of these residues (E269, L273, N280, R284, Y833, C852) are at the surface of the complex (Figure 5), and the observed differences can be due to lack of NOEs. However, for the interface residues K843, I862, L865 and Y887 the story is different: whereas the backbones of the twenty structures are very well superimposed, the differences between the wildtype and mutant consistently show considerable changes in side chain orientation (Figure 5). Finally, the F231L mutation not only leads to a significant reduction in the intermolecular hydrophobic packing involving the F894 anchor, but also a slight increase (0.3±0.01 Å) of the distance between the backbone amide of XPF F894 and the backbone carboxyl of ERCC1 G258 causing the H-bond between the backbone amide of A895 and the backbone carboxyl of G258 to vanish completely. In addition, as a consequence of the reorientation of E261, the imidazolium-carboxyl H-bond between H891 and E261 disappears. Also, the interactions between I264 and XPF residue A863, between L289 and F840, and those between the side chain of K843 and ERCC1 residues E237, C238 and T241 of ERCC1 appear diminished. Overall, the small structural changes initiated by the F231L mutation lead to local, modest conformational effects in the direct vicinity of the mutation, which accumulate in a number of differences affecting the interface between ERCC1 and XPF.

DISCUSSION

At the moment, the most severe cases of ERCC1-XPF deficiency are described for patients with F231L mutation in ERCC1 (28,31). Genetic and functional studies have revealed that the first patient carries a null allele and a F231L allele, and the second patient carries a biallelic F231L mutation. This F231L ERCC1 mutant in complex with XPF retains enzymatic activity. Analogous to patients with XPF mutations, this phenotype is a consequence of low intracellular levels of ERCC1-XPF (28,31). Earlier structural studies of the wildtype ERCC1-XPF complex have shown that the interface of ERCC1-XPF is composed of two pseudo-symmetric hydrophobic cavities tightly embracing anchor residues and a third, large hydrophobic contact region (Figure 6A and 6B) (18,21). The first anchor is formed by the aromatic ring of ERCC1 F293, which precisely fills the cavity composed of residues L294 of ERCC1 and Y833, N834, P837, Q838, L841, M856, N861 and I862 (helices α' and β') of XPF. The second anchor is formed by the XPF residue F894 aromatic ring, which fits into the cavity comprising ERCC1 residues F231, V232, V235, L254, G258, S259 (helices α and β) and XPF residues

F889, T892, S893, A895. The interactions in both anchor regions are stabilized by multiple intermolecular and intramolecular hydrogen bonds. The third contact region lies in between both anchor regions and involves the pseudo-symmetrically positioned ERCC1 helices α, γ and XPF helices α', γ'and ε'. In this region, the aromatic side chain of F840 plays a major role for intermonomer contacts. ERCC1 residues C238, L239, T241, L260, L263, I264, L285, V288 and L289 surround F840 and their contacts to residues G836, P837, Q838, D839, L841, L842, K843, M844 and I862 in the pseudo-symmetrically positioned region of XPF support this interaction. Furthermore, the short γ, ε, γ' and ε' helices show contacts between the side chains of I264, L289 in ERCC1 and A863 and A866 in XPF, and between L260 in ERCC1 and F889 in XPF. In addition, the backbone carbonyl of K843 contacts the side chain guanidinium group of R234, while the NH_3^+-group of the K843 side chain contacts the side chain carboxyl of E237 (Figure 5).

In wildtype ERCC1, F231 is located in helix α. It is part of the cavity containing anchor F894 and completely embedded in the heterodimer core. The side chain of F231 is located between the aromatic rings of F894 and F889 with stabilizing T-shaped arrangements (64). In the F231L mutant, the aromatic side chain of F889 maintains its orientation, whereas residues L231 and F894 reorient their side chains towards the surface of the protein (Figure 6B). This means that not only the interactions of the benzyl group of F231 are missing in the F894 hydrophobic cavity, but also that the benzyl group of F894 can no longer function as an anchor stabilizing the complex. One step further, since the side chain of F894 becomes exposed, a smaller hydrophobic interaction surface is present in the mutant (especially concerning residues in helix β), and the lack of aromatic ring interactions destabilizes ERCC1-XPF by increasing the dissociation rate of the complex.

The reduced hydrophobic character for the interactions within the F231L ERCC1-XPF complex interface implies that the electrostatic contributions to the interactions should relatively increase. Indeed, when increasing the ionic strength, the dissociation rate for wildtype ERCC1-XPF did not change much, whereas it increased for mutant ERCC1-XPF. The higher dissociation rate for the mutant indicates that the interactions at the interface of mutant ERCC1-XPF are more electrostatic in nature than that of wildtype ERCC1-XPF (65,66). A comparison of wildtype and mutant ERCC1-XPF $(HhH)_2$ complexes shows that the majority of structural differences is limited to regions in direct proximity of L231. These include the α and γ helices of ERCC1 and the C-terminal helix of XPF. The amides surrounding L231 also show a significant decrease in [^{1}H,^{15}N]-HSQC signal intensity as compared to those surrounding F231 in the wildtype, and the signal of the amide of L231 was even completely lacking. This loss of signal intensity can be explained by the occurrence of slow internal dynamics caused by millisecond motional effects of the F894 aromatic ring. The observed structural differences, mainly at the interface, could lead to instability of the ERCC1-XPF complex. Our stability analysis of the ERCC1-XPF $(HhH)_2$ domains using far-UV CD spectroscopy and ThermoFluor

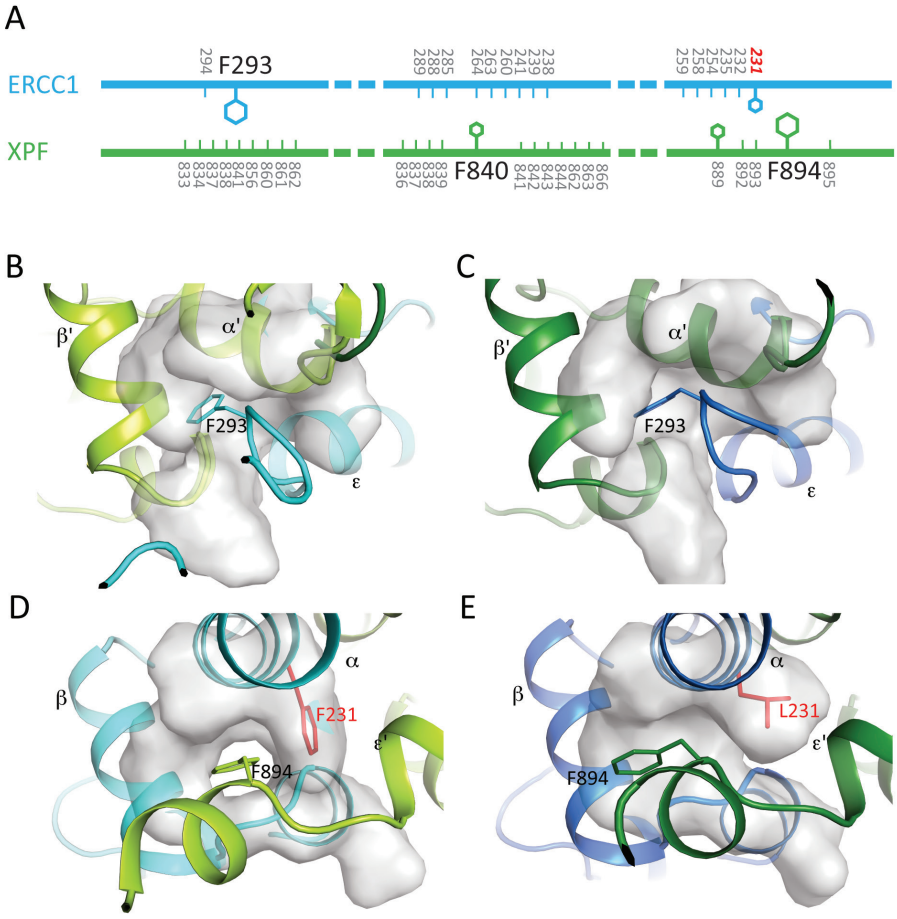


FIGURE 6. Interaction interface of ERCC1 and XPF and the comparison of F293 and F894 cavities at the interface of wildtype and F231L mutant ERCC1-XPF. A, the figure schematically represents the three major interaction regions: Anchor F293/Pocket 1, AnchorF894/Pocket 2 and the Central Domain involving F840. ERCC1 is shown in blue, XPF in green; side chains of important phenylalanines are shown; numbers of residues involved in intermolecular interactions are indicated; residue 231 is labelled in red. The F293-cavity in the wildtype (B) and mutant (C) ERCC1-XPF $(HhH)_2$ complex. The F894-cavity in the wildtype (D) and mutant (E) ERCC1-XPF $(HhH)_2$ complex. Side chains of the anchors F293 and F894 are shown as sticks. The side chain of residue 231(F or L) is shown in red.

analysis shows that the lower temperature stability of the F231L ERCC1-XPF complex is not due to a reduced intrinsic ERCC1 and XPF stability but rather to a faster dissociation of the heterodimer. Indeed, using SPR, we could directly demonstrate that this dissociation is significantly faster for the mutant than for the wildtype complex.

To conclude, the replacement of F231 by leucine appears to have only little effect on the ERCC1 or XPF monomer structures and in vitro stabilities, and to have no direct

functional effect on the DNA binding affinity of ERCC1-XPF at an ss/dsDNA junction. The $(HhH)_2$ heterodimer is still formed, which agrees with the fact that ERCC1-XPF based DNA repair is still partially functional in patients (28). The mutation influences the hydrophobic interactions between ERCC1 and XPF in the heterodimer, causes faster dissociation of the ERCC1-XPF heterodimer and in this way reduces the ERCC1-XPF complex stability. This increased dissociation of the mutant F231L heterodimer can directly result in a diminished amount of ERCC1-XPF complex, which would probably imply that the equilibrium is more shifted to the dissociated state. In the cellular context the separate subunits are unstable because of increased vulnerability to proteolysis. One reason for the intrinsic instability of the individual ERCC1 and XPF molecules might be that uncontrolled DNA binding and endonuclease activities of the separate subunits are a threat to regular DNA metabolism. The lower cellular concentration of the ERCC1-XPF repair complex will reduce the DNA repair capacity and may also lead to the build-up of repair intermediates that wait for the rate-limiting ERCC1-XPF step. This may also disrupt DNA metabolism leading to enhanced cell death and cellular dysfunction, contributing to the premature aging and cancer predisposition of mutants in the ERCC1-XPF complex both in human and mice. This is most dramatically illustrated by the very severe Cockayne syndrome-like features, which are exhibited by the ERCC1-deficient patients.

ACKNOWLEDGEMENTS

Ing. Mark Daniëls and drs. Hans Ippel, Tobias Madl, Klaartje Houben, Hugo van Ingen and Konstantinos Tripsianes are gratefully acknowledged for their contributions to this work.

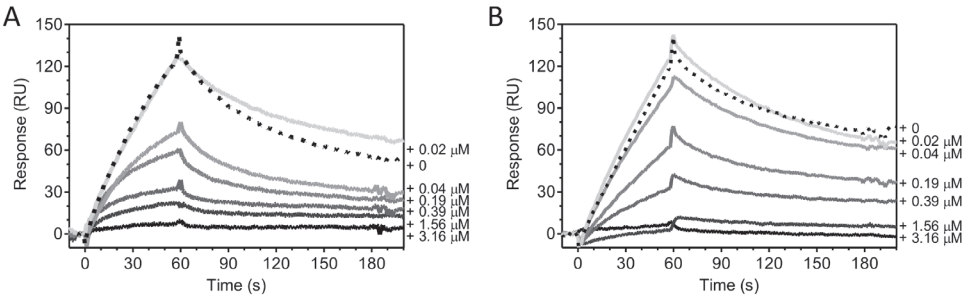
1. Hoeijmakers, J. H. J. (2001) Genome maintenance mechanisms for preventing cancer. Nature 411, 366-374
2. Sugasawa, K., Ng, J. M. Y., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P. M., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. J. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Molecular Cell 2, 223-232
3. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) DNA Repair and Mutagenesis, American Society of Microbiology Press, Washington DC
4. Friedberg, E. C., Aguilera, A., Gellert, M., Hanawalt, P. C., Hays, J. B., Lehmann, A. R., Lindahl, T., Lowndes, N., Sarasin, A., and Wood, R. D. (2006) DNA repair: from molecular mechanism to human disease. DNA Repair 5, 986-996
5. 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., Rademakers, S., De Rooij, J., Jaspers, N. G. J., Hoeijmakers, J. H. J., and Wood, R. D. (1996) Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell 86, 811-822
6. Niedernhofer, L. J., Essers, J., Weeda, G., Beverloo, B., de Wit, J., Muijtjens, M., Odijk, H., Hoeijmakers, J. H. J., and Kanaar, R. (2001) The structure-specific endonuclease Ercc1-Xpf is required for targeted gene replacement in embryonic stem cells. EMBO Journal 20, 6540-6549
7. Sijbers, A. M., van der Spek, P. J., Odijk, H., van den Berg, J., van Duin, M., Westerveld, A., Jaspers, N. G. J., Bootsma, D., and Hoeijmakers, J. H. J. (1996) Mutational analysis of the human nucleotide excision repair gene ERCC1. Nucleic Acids Research 24, 3370-3380
8. Sargent, R. G., Meservy, J. L., Perkins, B. D., Kilburn, A. E., Intody, Z., Adair, G. M., Nairn, R. S., and Wilson, J. H. (2000) Role of the nucleotide excision repair gene ERCC1 in formation of recombination-dependent rearrangements in mammalian cells. Nucleic Acids Research 28, 3771-3778
9. Al-Minawi, A. Z., Lee, Y.-F., Håkansson, D., Johansson, F., Lundin, C., Saleh-Gohari, N., Schultz, N., Jenssen, D., Bryant, H. E., Meuth, M., Hinz, J. M., and Helleday, T. (2009) The ERCC1/XPF endonuclease is required for completion of homologous recombination at DNA replication forks stalled by inter-strand cross-links. Nucleic Acids Research 37, 6400-6413
10. Niedernhofer, L. J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A. F., de Wit, J., Jaspers, N. G. J., Beverloo, H. B., Hoeijmakers, J. H. J., and Kanaar, R. (2004) The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. Molecular and Cellular Biology 24, 5776-5787
11. Wu, Y., Mitchell, T. R. H., and Zhu, X.-D. (2008) Human XPF controls TRF2 and telomere length maintenance through distinctive mechanisms. Mechanisms of Ageing and Development 129, 602-610
12. Muñoz, 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. Nature Genetics 37, 1063-1071
13. Zhu, X.-D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H. J., and de Lange, T. (2003) ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Molecular Cell 12, 1489-1498
14. Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H. J., and Niedernhofer, L. J. (2008) ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Molecular and Cellular Biology 28, 5082-5092
15. Al-Minawi, A. Z., Saleh-Gohari, N., and Helleday, T. (2008) The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. Nucleic Acids Research 36, 1-9
16. de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1998) Mapping of interaction domains between human repair proteins ERCC1 and XPF. Nucleic Acids Research 26, 4146-4152
17. de Laat, W. L., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1999) Molecular mechanism of nucleotide excision

repair. Genes & Development 13, 768-785
18. Tripsianes, K., Folkers, G., Das, D., Odijk, H., Jaspers, N. G. J., Hoeijmakers, J. H. J., 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
19. Ciccia, A., McDonald, N., and West, S. C. (2008) Structural and functional relationships of the XPF/MUS81 family of proteins. Annual Review of Biochemistry 77, 259-287
20. 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 Research 35, 5789-5798
21. Tsodikov, O. V., Enzlin, J. H., Schärer, O. D., and Ellenberger, T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. Proceedings of the National Academy of Sciences of the United States of America 102, 11236-11241
22. Das, D., Folkers, G. E., van Dijk, M., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., and Boelens, R. (2012) The structure of the XPF-ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix-hairpin-helix domains in ss/ds DNA recognition. Structure 20, 667-675
23. Kondo, S., Mamada, A., Miyamoto, C., Keong, C.-H., Satoh, Y., and Fujiwara, Y. (1989) Late onset of skin cancers in 2 xeroderma pigmentosum group F siblings and a review of 30 Japanese xeroderma pigmentosum patients in groups D, E and F. Photo-dermatology 6, 89-95
24. Sijbers, A. M., van Voorst Vader, P. C., Snoek, J. W., Raams, A., Jaspers, N. G. J., and Kleijer, W. J. (1998) Homozygous R788W point mutation in the XPF gene of a patient with xeroderma pigmentosum and late-onset neurologic disease. Journal of Investigative Dermatology 110, 832-836
25. Vermeulen, W. P., Rademakers, S., Jaspers, N. G. J., Appeldoorn, E., Raams, A., Klein, B., Kleijer, W. J., Kjærsgård Hansen, L., and Hoeijmakers, J. H. J. (2001) A temperature-sensitive disorder in basal transcription and DNA repair in humans. Nature Genetics 27, 299-303
26. 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. Molecular and Cellular Biology 24, 1200-1205
27. 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., Theil, A. F., Vermeulen, W., Van Der Horst, G. T. J., Meinecke, P., Kleijer, W. J., Vijg, J., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (2006) A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. Nature 444, 1038-1043
28. Kashiyama, K., Nakazawa, Y., Pilz, D. T., Guo, C., Shimada, M., Sasaki, K., Fawcett, H., Wing, J. F., Lewin, S. O., Carr, L., Li, T.-S., Yoshiura, K.-I., Utani, A., Hirano, A., Yamashita, S., Greenblatt, D., Nardo, T., Stefanini, M., McGibbon, D., Sarkany, R., Fassihi, H., Takahashi, Y., Nagayama, Y., Mitsutake, N., Lehmann, A. R., and Ogi, T. (2013) Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. The American Journal of Human Genetics 92, 807-819
29. Bogliolo, M., Schuster, B., Stoepker, C., Derkunt, B., Su, Y., Raams, A., Trujillo, J. P., Minguillón, J., Ramírez, M. J., Pujol, R., Casado, J. A., Baños, R., Rio, P., Knies, K., Zúñiga, S., Benítez, J., Bueren, J. A., Jaspers, N. G. J., Schärer, O. D., De Winter, J. P., Schindler, D., and Surrallés, J. (2013) Mutations in ERCC4, Encoding the DNA-Repair Endonuclease XPF, Cause Fanconi Anemia. The American Journal of Human Genetics 92, 800-806
30. van Duin, M., Van den Tol, J., Warmerdam, P., Odijk, H., Meijer, D., Westerveld, A., Bootsma, D., and Hoeijmakers, J. H. J. (1988) Evolution and mutagenesis of the mammalian excision repair gene ERCC-1. Nucleic acids research 16, 5305-5322
31. Jaspers, N. G. J., Raams, A., Silengo, M. C., Wijgers, N., Niedernhofer, L. J., Robinson, A. R., Giglia-Mari, G., Hoogstraten, D., Kleijer, W. J., and Hoeijmakers, J. H. J. (2007) First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and

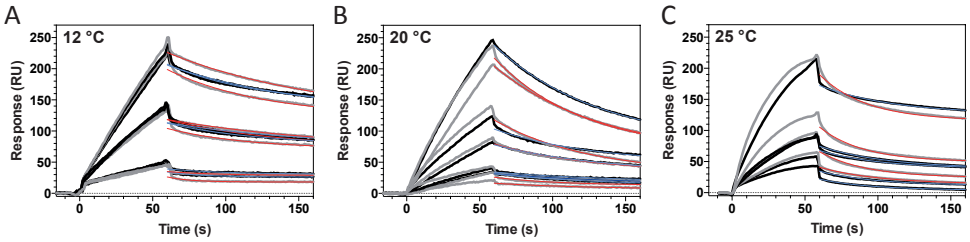
severe developmental failure. The American Journal of Human Genetics 80, 457-466
32. Gregg, S. Q., Robinson, A. R., and Niedernhofer, L. J. (2011) Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. DNA Repair 10, 781-791
33. McWhir, J., Selfridge, J., Harrison, D. J., Squires, S., and Melton, D. W. (1993) Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. Nature genetics 5, 217-224
34. Imoto, K., Boyle, J., Oh, K., Khan, S., Ueda, T., Nadem, C., Slor, H., Orgal, S., Gadoth, N., Busch, D., Jaspers, N. G. J., Tamura, D., DiGiovanna, J. J., and Kraemer, K. H. (2007) Patients with defects in the interacting nucleotide excision repair proteins ERCC1 or XPF show xeroderma pigmentosum with late onset severe neurological degeneration. in Journal of Investigative Dermatology, 127, S92-S92
35. Farabaugh, P. J. (1978) Sequence of the lacI gene. Nature 274, 765-769
36. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12. Science 277, 1453-1462
37. Nieba, L., Nieba-Axmann, S. E., Persson, A., Hämäläinen, M., Edebratt, F., Hansson, A., Lidholm, J., Magnusson, K., Karlsson, Å. F., and Plückthun, A. (1997) BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. Analytical Biochemistry 252, 217-228
38. Das, D., Tripsianes, K., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., Boelens, R., and Folkers, G. E. (2008) The HhH domain of the human DNA repair protein XPF forms stable homodimers. Proteins: Structure, Function, and Bioinformatics 70, 1551-1563
39. Politou, A. S., Thomas, D. J., and Pastore, A. (1995) The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. Biophysical Journal 69, 2601-2610
40. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. Analytical Biochemistry 357, 289-298
41. Cavanagh J., F. J. W., Palmer G. A. , Skelton J. N. (1996) Protein NMR Spectroscopy, Academic Press, San Diego , CA
42. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progress in Nuclear Magnetic Resonance Spectroscopy 34, 93-158
43. 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. Journal of Biomolecular NMR 6, 277-293
44. Goddard, T. D., and Kneller, D. G. SPARKY 3. University of California, San Francisco
45. Baxter, N. J., and Williamson, M. P. (1997) Temperature dependence of 1H chemical shifts in proteins. Journal of Biomolecular NMR 9, 359-369
46. Cierpicki, T., and Otlewski, J. (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. Journal of Biomolecular NMR 21, 249-261
47. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. Quarterly reviews of biophysics 16, 521-655
48. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. Proteins: Structure, Function, and Bioinformatics 17, 75-86
49. Hwang, T.-L., van Zijl, P. C. M., and Mori, S. (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. Journal of Biomolecular NMR 11, 221-226
50. Güntert, P. (2009) Automated structure determination from NMR spectra. European Biophysics Journal 38, 129-143
51. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation

with the new program DYANA. Journal of Molecular Biology 273, 283-298
52. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. Journal of Biomolecular NMR 44, 213-223
53. Cornilescu, G., Hu, J.-S., and Bax, A. (1999) Identification of the hydrogen bonding network in a protein by scalar couplings. Journal of the American Chemical Society 121, 2949-2950
54. Brunger, 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica Section D Biological Crystallography 54, 905-921
55. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Güntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R., and Bonvin, A. M. J. J. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins: Structure, Function, and Bioinformatics 59, 662-672
56. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. Journal of Molecular Graphics 8, 52-56
57. Hooft, R., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures. Nature 381, 272-272
58. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates. Proteins: Structure, Function, and Bioinformatics 12, 345-364
59. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA.
60. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. Journal of Molecular Graphics 14, 51-55
61. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and Œ≤ form of proteins in aqueous solution by circular dichroism. Biochemistry 13, 3350-3359
62. McDonnell, J. M. (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. Current Opinion in Chemical Biology 5, 572-577
63. Colowick, S. P., Kaplan, N. O., and Purich, D. L. (1955) Methods in enzymology, Academic press, New York
64. Chelli, R., Gervasio, F. L., Procacci, P., and Schettino, V. (2002) Stacking and T-shape competition in aromatic-aromatic amino acid interactions. Journal of the American Chemical Society 124, 6133-6143
65. Serdyuk, I. N., Zaccai, N. R., and Zaccai, J. (2007) Methods in molecular biophysics: Structure, Dynamics, Function, Cambridge University Press, Cambridge, UK
66. Finkelstein, A. V., and Ptitsyn, O. B. (2002) Protein Physics, Academic Press, California

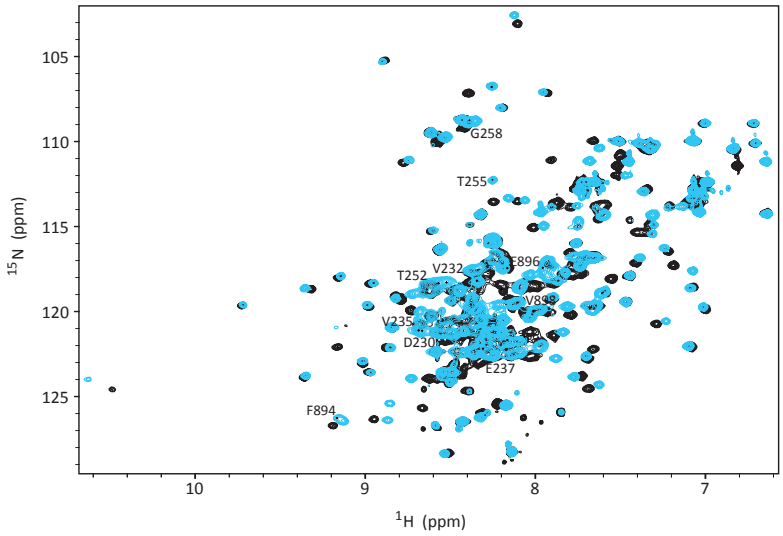
SUPPLEMENTARY FIGURES



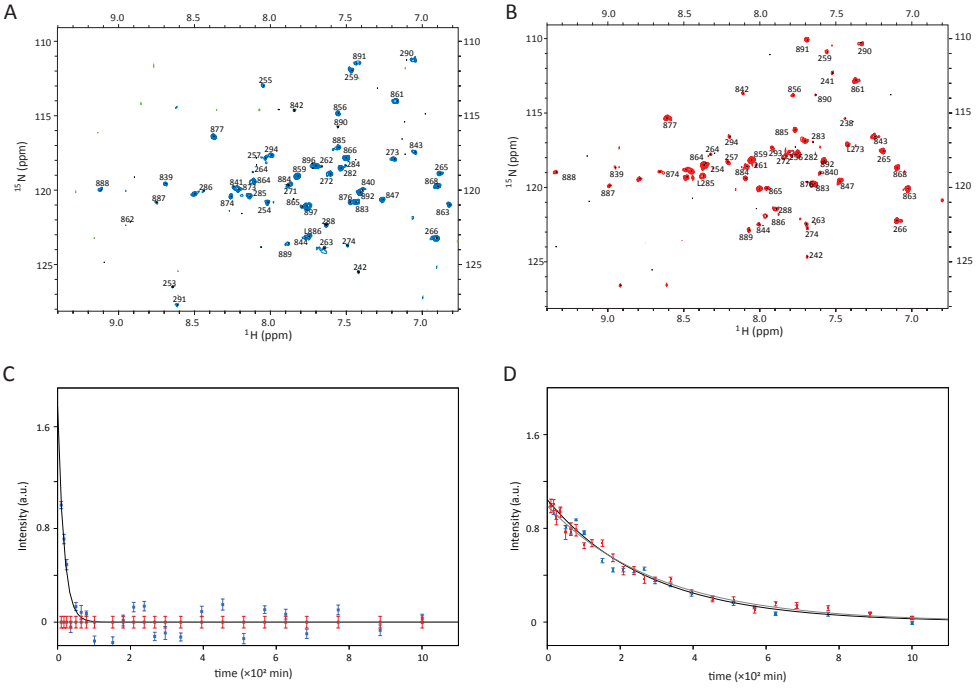
Supplementary Figure S1. DNA binding by wildtype and F231L mutant ERCC1-XPF $(HhH)_2$ heterodimer. SPR with different concentrations of bubble10 DNA binding to the wildtype ERCC1-XPF $(HhH)_2$ heterodimer (A) and the mutant F231L ERCC1-XPF $(HhH)_2$ complex (B).



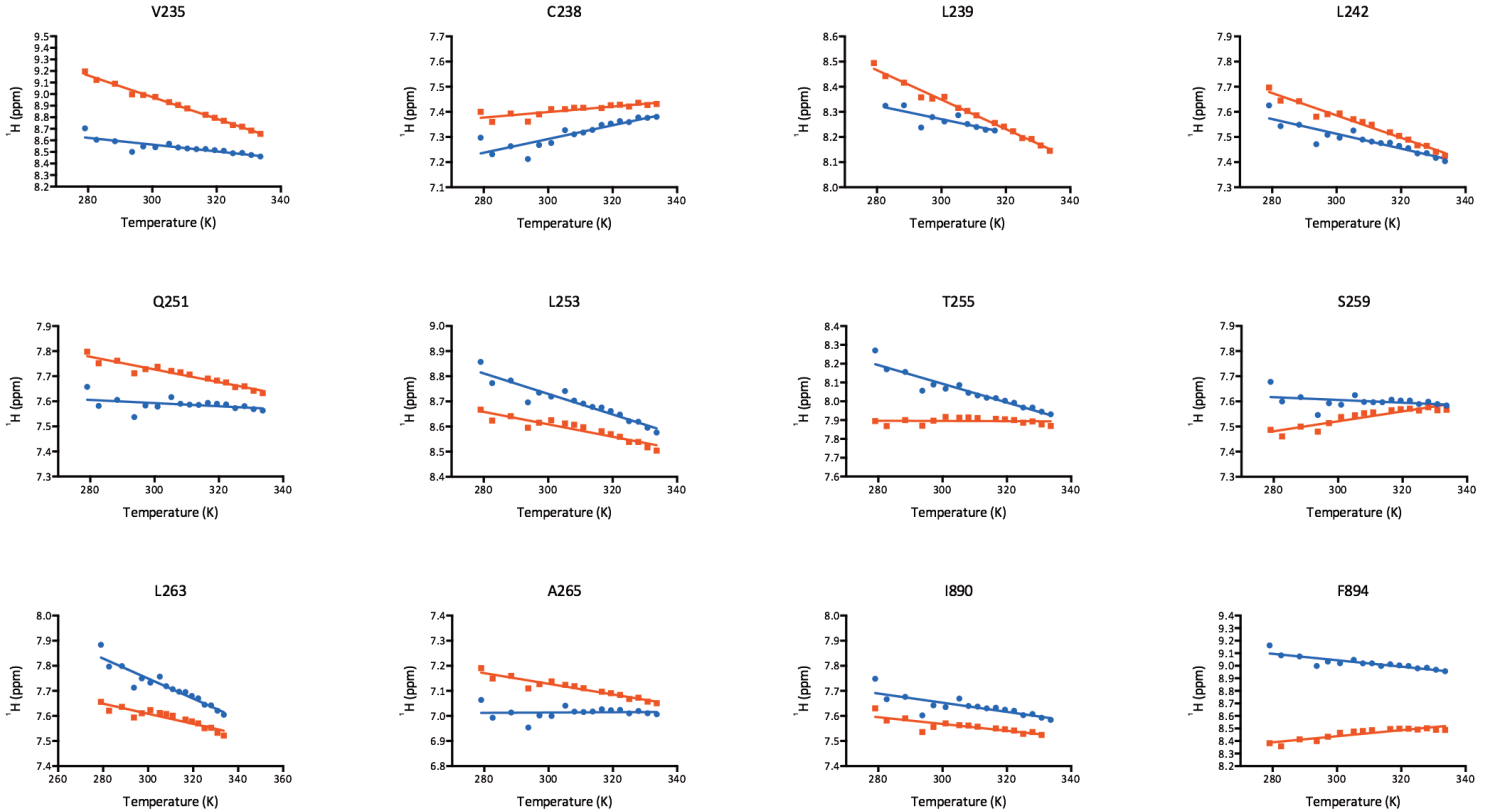
Supplementary Figure S2. Wildtype and F231L ERCC1-XPF $(HhH)_2$ heterodimer dissociation measured by SPR under different conditions. A, 12 °C and 50 mM NaCl. B, 20 °C and 50 mM NaCl. C, 25 °C and 50 mM NaCl. Mutant and wildtype ERCC1-XPF $(HhH)_2$ heterodimers are shown in grey and black, respectively. All the experiments were conducted at three different protein concentrations (0.2, 0.1 and 0.05 µM) and in duplicates. The fits to obtain kinetics values are colored in red (mutant) and blue (wildtype).



Supplementary Figure S3. Differences in proton chemical shift. Overlay of [^{1}H,^{15}N]- HSQC spectra of wildtype (blue) and F231L mutant (black) ERCC1-XPF $(HhH)_2$ heterodimers.



Supplementary Figure S4. Proton-deuterium exchange. A, initial [^{1}H,^{15}N]-HSQCs after the deadtime of the experiment for wildtype (left) and mutant (right) ERCC1-XPF $(HhH)_2$ heterodimers. Time course of the H/D exchange behavior of L253 (B), and V589 (C). Signal intensities for the wildtype system are shown in blue, for the mutant in red.



Supplementary Figure S5. Amide proton chemical shift as a function of temperature.
Wildtype (blue) and mutant (red) ERCC1-XPF Δδ/ΔT-changes.

Chapter 3

Relationship between structure and stability of the ERCC1-XPF complex

Maryam Faridounnia, Lidija Kovačič, Hans Wienk, Rolf Boelens and Gert E. Folkers

Manuscript in preparation

ABSTRACT

XPF forms a heterodimeric complex with ERCC1 that is functional in nucleotide excision repair (NER). *In vitro* XPF can also form homodimers. The homodimeric state is more stable than heterodimeric state at elevated temperatures and in denaturants, however, this conformation is not observed *in vivo*. We now find that the ERCC1-XPF heterodimer can rapidly dissociate. Whereas association of ERCC1-XPF is preferred *in vivo* and *in vitro*, we show that despite this rapid dissociation XPF preferentially re-associates with ERCC1, and that XPF only forms homodimers in the absence of ERCC1. Since the structures of XPF in the homodimer and in the heterodimer differ, our data suggest that XPF dissociated from ERCC1 has the proper conformation to re-associate with ERCC1 but that a structural transition is required for homodimerization.

INTRODUCTION

The ERCC1-XPF complex is best understood for its role in nucleotide excision repair (NER) but plays a key role in several other DNA repair pathways as well (1-10). ERCC1-XPF functions as a structure-specific endonuclease that binds close to the DNA damage site using the C-terminal $(HhH)_2$ domains of ERCC1 and XPF, and the adjacent endonuclease domain in XPF is responsible for DNA cleavage in the repair process.

The *ERCC1* gene was the first cloned gene of the human NER pathway (11) and its role in NER has been thoroughly described (1). Still our knowledge on the structural and kinetic properties of ERCC1 in complex with other NER proteins is very limited. Firstly, in 1993 Biggerstaff *et al.* noted that ERCC1, in complex with XPF, forms a functional endonuclease (12). They also observed reduced amounts of ERCC1 in cells lacking XPF, indicating that ERCC1 is unstable in the absence of its partner (12). Next in 1995, Park *et al.* reported the purification of the full 38 kDa ERCC1 from HeLa cells by using XPA affinity columns, taking advantage of its XPA-binding property (13). This purified ERCC1 was in complex with a 112 kDa endonuclease, concluded to be XPF on the basis of the ERCC1-XPF sequence homology with *S. cerevisiae* Rad10-Rad1 (13,14). Subsequent mutational studies by Sijbers *et al*. (14) on the *ERCC1* gene revealed that its N-terminal 91 aminoacids are poorly conserved and dispensible for its function in DNA repair. However, they also showed that the C-terminal part of ERCC1, which is crucial for its function, contained a strongly conserved tandem Helix-hairpin-Helix [$(HhH)_2$] DNA-binding motif with considerable homology to the NER proteins UvrC of *E. coli* and Swi10 of *S. pombe* (14). Subsequently, de Laat *et al.* (15), using immunoprecipitation analysis, established that ERCC1 and XPF can indeed efficiently form a heterodimeric complex *in vitro* (15). Via deletion analysis they mapped the protein interaction regions to residues 224-297 in ERCC1 and 814–905 in XPF in their C-terminal $(HhH)_2$ domains (15).

A first structure model for the human ERCC1-XPF complex came from studies on the homologous XPF from *Archaea* by Newman *et al.* which showed that the XPF C-terminal $(HhH)_2$ domains can form a stable homodimeric complex (16). Proof for this model came from NMR cross-saturation experiments by Choi *et al*. who showed that *in vitro* the C-terminal part of XPF functions as a scaffold for the folding of the C-terminal part of ERCC1 and that the ERCC1 $(HhH)_2$ domain is only stable as 1:1 complex with XPF (17). First experimental data for the structure of the complex of the $(HhH)_2$ domains of ERCC1 and XPF came from NMR studies by Tripsianes *et al*. (18) and from crystallographic studies by Tsodikov *et al*. (19). Both studies showed that the ERCC1-XPF $(HhH)_2$ domains dimerize in the same manner as the archaeal XPF homodimer in a pseudo-symmetric fashion. It was shown that the stability of the interface of ERCC1 and XPF relies heavily on hydrophobic contacts between the helices α and γ of ERCC1 and α' and γ' of XPF and on the interactions of the C-terminal anchors

F293 of ERCC1 and F894 of XPF in cavities formed by the complementary XPF and ERCC1 monomers, respectively (18,19).

A number of chemotherapeutic cancer treatments, such as with cisplatin, is based on damaging DNA (20). In several instances it has been noted that these damages can be repaired by ERCC1-XPF dependent DNA repair pathways (1,21). The importance of a stable ERCC1-XPF complex for the catalytic activity in DNA repair suggests that interference with ERCC1-XPF complex stability could increase the efficiency of chemotherapy by simultaneously suppressing DNA repair (22). For this, it would be important to better understand which factors contribute to the stability of the ERCC1-XPF complex. Therefore, we analyzed the structure and stability of the ERCC1-XPF complex.

Using Surface Plasmon Resonance (SPR), hydrogen deuterium exchange (HDX) and other NMR spectroscopy techniques we gathered insight into the considerable dynamics of the ERCC1-XPF complex. The ERCC1-XPF heterodimers can rapidly dissociate into monomers. However, no XPF homodimers, reported by several groups to be highly stable (17,19), are formed and ERCC1 re-associates preferentially with XPF. The slow association of XPF homodimers can be explained by structural differences between heterodimeric and homodimeric XPF. Here, the preference of forming ERCC1-XPF heterodimers over XPF homodimers is discussed after comparing the H/D exchange and structural properties of these two protein complexes. Our data suggest that XPF dissociated from ERCC1 has the proper conformation to re-associate with ERCC1 but that a structural transition is required for homodimerization.

EXPERIMENTAL PROCEDURES

Sample Preparation - The heterodimeric $(HhH)_2$ complex of ERCC1-XPF (residues 220-301 of ERCC1 and residues 823-905 of XPF) containing a C-terminal Histidine-tag on ERCC1 was cloned, expressed in *Escherichia coli* BL21 (DE3), and purified using metal-affinity chromatography followed by gel filtration, as described by Tripsianes *et al.* (18).

Surface Plasmon Resonance - Real-time association and dissociation of ERCC1-XPF was monitored using a Biacore® X system in flow cells with laminar flow (Biacore AB), essentially as described (23). The ERCC1-XPF $(HhH)_2$ complex was prepared as described in Chapter 2, and immobilized on a Ni^{2+}-NTA Sensor Chip (Biacore AB, Sweden) via the C-terminal His-tag of ERCC1. The conditions for maximal ERCC1-XPF/Ni-NTA binding were optimized by varying protein and $NiCl_2$ concentrations and by adjusting the loading speed based on the findings of Nieba *et al*. (24). The highest concentration of ERCC1-XPF, at which dissociation of His-tagged ERCC1 could be avoided, was 0.2 μM. At higher concentrations, residual dissociation of His-tagged ERCC1 was observed. Experiments were repeated at three different concentrations (0.2, 0.1 and 0.05 μM) in 10 mM Hepes pH 7.5, 50 mM NaCl, 0.005% Tween20, and each measured at temperatures of 12, 20, 25 and 37 °C using a flow

rate of 20 µl/min. The association and dissociation of the ERCC1-XPF complex from the chip was monitored over two intervals of 60 and 120 seconds duration, respectively. The dissociation rates (k_{off}^{app}) were determined using Prism (GraphPad Software, Inc.) by fitting the SPR curves of the dissociation to a single-phase decay model with the equation:

$$R=(R_0 - R_P)\, e^{(k_{off}^{app}\, t)} + R_P$$

where R is the observed response (in arbitrary response units RU) at time t, R_0 is the initial response before dissociation, R_p is the final response plateau after dissociation, and k_{off}^{app} is the dissociation rate constant for dissociation of XPF from His-tagged ERCC1.

NMR spectroscopy – For comparison, 2D [^{1}H,^{15}N]-HSQC spectra were recorded on a Bruker Avance 750 MHz NMR spectrometer for the ERCC1-XPF $(HhH)_2$ heterodimer and XPF $(HhH)_2$ homodimer at exactly the same pH and temperature. The assignments of heterodimeric XPF-ERCC1 and homodimeric $(XPF)_2$ were obtained from literature (18,25).

Hydrogen Deuterium Exchange – HDX was monitored by NMR, recording a series of 2D [^{1}H,^{15}N]-HSQC spectra on a Bruker Avance 750 MHz NMR spectrometer. A freshly prepared aqueous solution of 400 µM ^{15}N-labeled ERCC1-XPF $(HhH)_2$ heterodimer in 100 mM NaCl, 50 mM NaPi pH 6.8 was lyophilized and redissolved in D_2O. The acquisition parameters for the experiment were optimized with another aqueous ERCC1-XPF sample with the same volume and prepared in the same buffer (26). The deadtime due to preparation of the sample and temperature stabilization of the sample in the NMR spectrometer was 8 minutes. The recorded 2D HSQC spectra were processed using NMRPipe (27) and the HSQC cross-peak intensities were obtained using Sparky(28). For each amide signal the exponential decay was fitted using the Prism software (GraphPad Software, Inc.) using the equation:

$$I(t) = (I_0 - I_\infty)\ exp\ (-k_{ex}\, t) + I_\infty$$

where $I(t)$ is the signal intensity at time t, k_{ex} is the H/D exchange rate constant, I_0 is the initial amide signal intensity and I_∞ the final intensity after full exchange. The intrinsic exchange rates (k_{intr}) for unprotected amino acids were derived from (29,30). Protection factors P were calculated as the ratio:

$$P = k_{intr} / k_{ex}$$

Temperature dependence of the ^{1}H chemical shift - To investigate whether the amide protons are present in hydrogen bonds or not, 2D [^{1}H,^{15}N]-HSQC spectra were recorded at 600 MHz at fifteen different temperatures ranging from 279-333 K using 400 µM of ^{15}N-labeled ERCC1-XPF $(HhH)_2$ heterodimer. The changes in amide proton chemical shifts as a function of temperature were analyzed as described in (31,32) using the software Sparky (28). The temperature coefficients, Δδ/ΔT (ppb/K), were determined by fitting the change in amide proton chemical shift with temperature using a linear least square fit.

Structural analysis – Structural analyses were based on the NMR structures of the ERCC1-XPF heterodimer (1Z00) and the XPF homodimer (2AQ0) (18,25). The dimer interfaces were analyzed using the web services COCOMAP and CMA (33). Distances were

obtained between the protons and nitrogens of all backbone amides of one monomer to all atoms in the other monomer, with a distance cutoff of 8 Å. The obtained distances were used to calculate the neighbor monomer proximity $\sum(1/d)^2$ *(*in $Å^{-2}$*)* for each amide nitrogen and proton, where *d* is the distance between the amide proton or nitrogen in XPF and the atom in the other monomer.

The fold similarity and structure element positions were compared by superimposition of the ERCC1-XPF heterodimer and XPF homodimer structures. The backbone Ca-atoms of heterodimeric and homodimeric XPF were superimposed either for the full-length domain or for the secondary structure elements using the McLachlan algorithm as implemented in the program ProFit (Martin, A.C.R., unpublished, http://www.bioinf.org.uk/software/profit) (15) and RMSD values were calculated for each residue. The atom depths of the amide protons of ERCC1-XPF, i.e. the distance between these atoms and the bulk solvent surrounding the molecule were calculated for the heterodimer (1Z00)(18), and the separate ERCC1 and XPF monomers using the program Depth (34).

RESULTS AND DISCUSSION

Dissociation of the ERCC1-XPF $(HhH)_2$ heterodimer

The dissociation of the heterodimeric ERCC1-XPF complex was monitored directly by SPR. The ERCC1-XPF complex was immobilized via the C-terminal Histidine-tag of ERCC1 on Ni^{2+}-NTA chips and the dissociation of XPF from immobilized ERCC1 was followed. Figure 1 shows the SPR response upon loading ERCC1-XPF onto the chip, followed by a subsequent dissociation event. In principle this dissociation could be either the dissociation of the full ERCC1-XPF complex from the chip, or the dissociation of XPF, while ERCC1 remains attached to the chip. The dissociation reaches a plateau within 2 minutes at approximately 50% of the SPR response upon loading the ERCC1-XPF complex. Only after washing with buffer containing EDTA, the SPR response returns to its original level. Since XPF and ERCC1 have a similar mass, our observations indicate that under these conditions XPF rapidly dissociates from ERCC1 and that there is no significant dissociation of His-tagged ERCC1-XPF from the chip. The dissociation of XPF from the ERCC1-XPF complex was studied at different heterodimer concentrations and at different temperatures. At 25 °C dissociation of ERCC1-XPF occurs with a rate k_{off} of 2×10^{-2} s^{-1}. This rate does not vary considerably with different concentrations, probably because the re-association has only a minor contribution under these conditions.

Previously it has been found that ERCC1-XPF can form highly stable complexes, that allowed protein crystallography and NMR studies (18,19). Thermofluor analysis showed that ERCC1 and XPF are in stable complexes at concentrations far below 1 μM. Also, the complex can be purified as intact heterodimer by gel filtration chromatography (data not shown), even at micromolar concentrations. Based on this, we estimated the dissociation constant K_d

for the ERCC1-XPF complex to be below 0.1 µM. Since the SPR experiments show that ERCC1-XPF can dissociate within 100 seconds (corresponding to a dissociation constant $k_{off} \approx 10^{-2}\ s^{-1}$), it can be reasoned that the association rate (k_{on}) for the formation of the complex of XPF and ERCC1 should be above $10^5\ M^{-1}\ s^{-1}$, which is close to diffusion-limited association (35).

Previously, it was reported that XPF can form stable homodimers (25). In fact, these homodimers were found to be much more temperature and denaturant stable than the ERCC1-XPF heterodimers. This indicates a dissociation rate lower than that of the heterodimers, i.e. a $k_{off} < 10^{-2}\ s^{-1}$. Therefore, assuming that the association of XPF homodimers occurs with a similar rate as the heterodimer, the dissociation constant for the homodimers would be even smaller than that of the heterodimer. Given the high stability of the homodimers combined with the fast dissociation of the heterodimers, dissociation of ERCC1-XPF would lead within several minutes to the formation of homodimers. However, this is not experimentally observed, since it was found that the ERCC1-XPF heterodimers are stable over many weeks at room temperature, for instance in the NMR instrument. Supplementary Figure S1 shows that the heterodimers are in fact stable up to about 50 °C and only at temperatures above that a conversion to homodimers is observed. Under these conditions the concentration

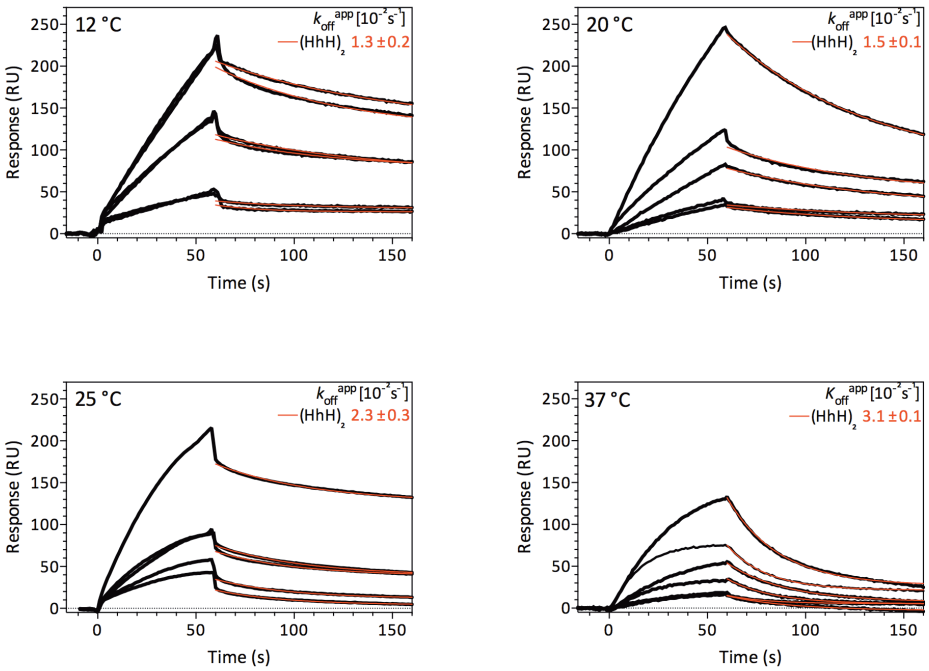


Figure 1. SPR sensogram for the dissociation of the ERCC1-XPF $(HhH)_2$ heterodimer at different temperatures. The ERCC1-XPF $(HhH)_2$ heterodimer complex is immobilized via its ERCC1 his-tag to the Ni^{2+}-NTA surface. The dissociation rate k_{off} is calculated from three different concentrations and two replicates for each concentration. The SPR response is shown in black and the fit in red.

of dissociated ERCC1 decreases due to precipitation (visible as white flakes). One has to conclude that the association of XPF into homodimers is much slower than its re-association with ERCC1 and only when ERCC1 is absent, XPF will associate with itself.

A possible explanation for this counterintuitive effect is that XPF occurs in two different conformations: one as observed in the ERCC1-XPF complex, and another one that is required for the formation of XPF homodimers. When the conversion between these XPF conformations is slow or when the homodimeric XPF conformation is only a small fraction of XPF in the monomeric state, the association of XPF homodimers can be less frequent than the association of XPF with ERCC1. In this manner, ERCC1-XPF could, despite its rapid dissociation, still be more stable than the homodimeric XPF complex. To analyze whether XPF has different conformations when present in a heterodimeric complex with ERCC1 or in a homodimeric form, we focused on [^{1}H,^{15}N]-HSQC spectra, hydrogen deuterium exchange behavior, and previously reported NMR structures.

Comparison of the [^{1}H,^{15}N]-HSQC spectra of XPF in heterodimeric and homodimeric form

The NMR chemical shift is a very sensitive probe of the immediate environment of an atom. Differences in chemical shift provide useful information about differences in molecular structure, for instance due to interaction with other molecules, changes in the local environment or altered H-bond geometries. Differences between backbone amide chemical shifts of residues of XPF in the heterodimeric and homodimeric state could be the result of local changes in environment due to the proximity of ERCC1 (in the heterodimer) or XPF (in the homodimer), due to differences in H-bond strengths and/or because of structural differences in XPF.

In Figures 2A and 2C, the differences in backbone amide proton and nitrogen chemical shift between the $(HhH)_2$ domain of heterodimeric XPF and that of homodimeric XPF are shown. Overall for many residues, significant upfield amide proton chemical shift differences are observed for homodimeric XPF compared to heterodimeric XPF, especially in helices α', γ', and δ' and the region from helix ε' and the C-terminal helix. However, a number of residues including residues K832 and Y833 at the N-terminal, Q838 in helix α', L855 and M856 in helix β', and A863 and A866 in helix γ' show downfield proton shifts for the homodimer with an exceptionally large chemical shift difference for Q838. Strikingly, residues Y833, Q838, L855 and M856 all belong to the cavity anchoring F894 in the XPF homodimer structure, corresponding to the cavity surrounding F293 in ERCC1-XPF. Similarly, in the C-terminal part of XPF, near F894, also chemical shift differences are observed. Apparently, there are substantial differences in the chemical environment of F894 when surrounded by the cavity residues of ERCC1 or by the cavity residues of XPF.

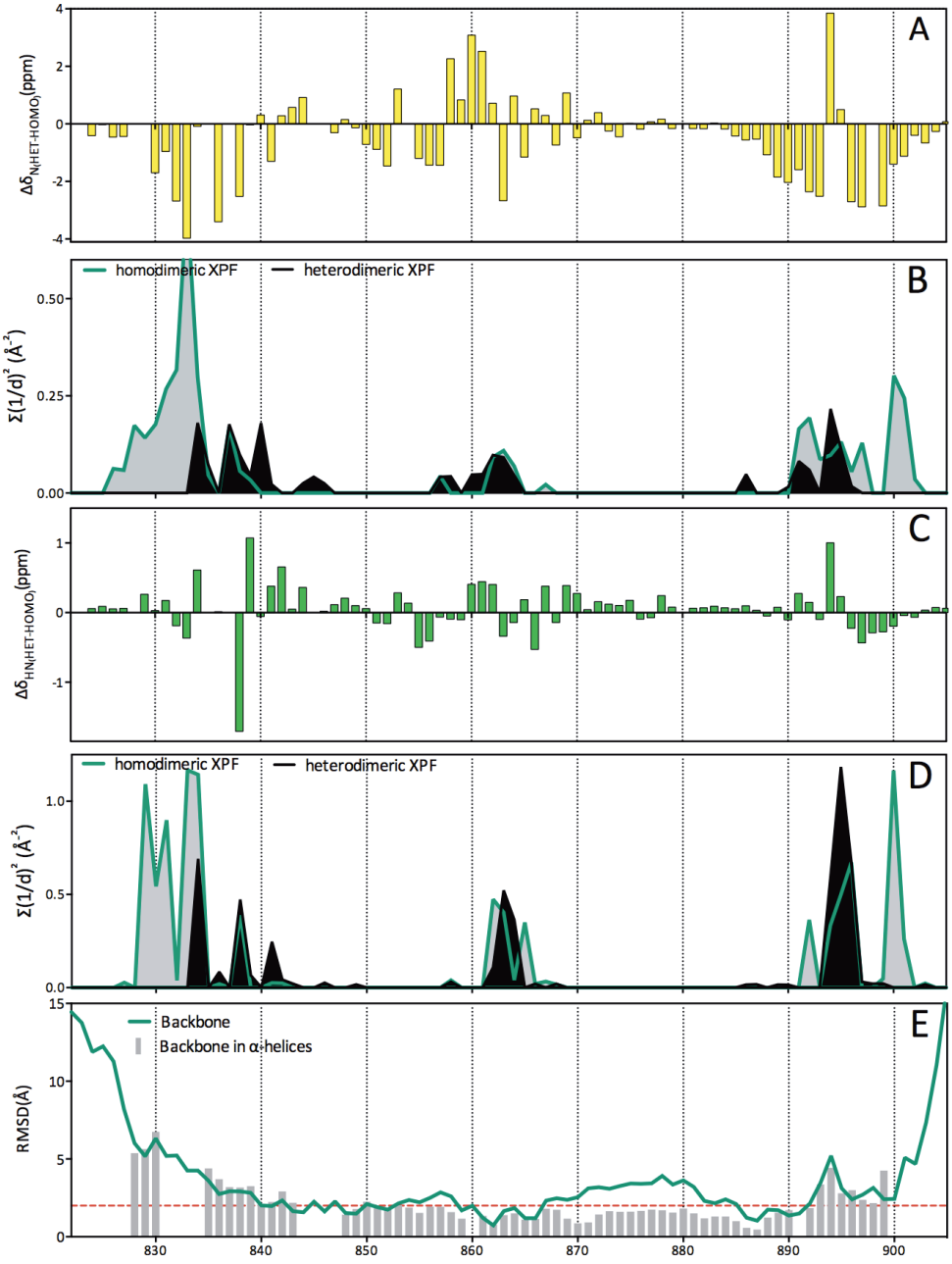


Figure 2. Conformational differences between heterodimeric and homodimeric XPF. *A,* difference in amide nitrogen chemical shift ($\Delta\delta_{N(HET-HOMO)}$). *B,* amide nitrogen surface proximity for heterodimeric XPF (black) and homodimeric XPF (grey). *C,* difference in amide proton chemical shift ($\Delta\delta_{HN(HET-HOMO)}$), *D,* amide proton surface proximity for heterodimeric XPF (black) and homodimeric XPF (grey). *E*, Residue RMSDs by superimposing the backbone Cα-atoms of homodimeric XPF (2AQ0) on heterodimeric XPF (1Z00).

Analysis of the distances of the amide atoms of an XPF monomer with respect to either ERCC1 or XPF shows that most of the observed chemical shift differences can be explained by the difference in local environment (Figure 2B, 2D). It should be noted that the interaction surfaces of XPF in ERCC1-XPF and $(XPF)_2$ are different. In the homodimer the interacting surfaces include the C-terminal and N-terminal parts of XPF, whereas in ERCC1-XPF these are not involved in intermolecular contacts. However, when we compare the chemical shift differences for the nitrogens (Figure 2A) and the protons (Figure 2C) to the neighbor proximities (Figure 2B and Figure 2D, respectively), we also observe that for several amino acids these are not overlapping with any of the interacting regions. These include residues 844, 847, 848 in hairpin h_1', residues 849, 851-856 along helix β', residues 872-878 along helix δ', and residues 869-870 located in hairpin h_2' (Figure 3A, 3B). Only for 844, 855 and 856 other atoms of these residues are still in close proximity, but most of the residues are far from the interface. The fact that their amides show substantial chemical shift differences indicates either substantial structural differences in these regions and/or differences in H-bond strengths between XPF in homodimeric or the heterodimeric complex.

Proton-deuterium exchange and hydrogen bond stability of ERCC1 and XPF in the heterodimer

We measured the hydrogen deuterium exchange rates for the amide protons of ERCC1 and XPF to study monomer stability within the ERCC1-XPF heterodimer. The exchange rates were correlated with their global surface exposure in the complex. Since the HDX could occur either within the intact heterodimer or from ERCC1 and XPF monomers after transient dissociation of the complex, we studied the surfaces of ERCC1 and XPF as monomers as well as heterodimers.

The observed hydrogen deuterium exchange rates of ERCC1-XPF and the derived protection factors are listed in the Supplementary Table S1. The protection factors are also shown in Figure 4, and in Figure 5 the HDX rates are mapped onto the previously determined NMR structure of ERCC1-XPF (18). The backbone amides of many residues in the helices a and several in helix b of ERCC1 show immediate exchange, indicating weak or absence of hydrogen bonds and direct surface exposure. In XPF, we observe immediate exchange for the amides in the short N-terminal helix, the majority of helix β' and the C-terminal part of helix ε'. However, several residues display slow HDX, and for these, even no exchange was observed during 14 days. For those residues for which the exchange rate could be measured, we noted that protection factors for ERCC1 are on average about 5-fold smaller than for XPF indicating that ERCC1 is less stable than XPF in the heterodimer (Supplementary Table S2, Figure 4).

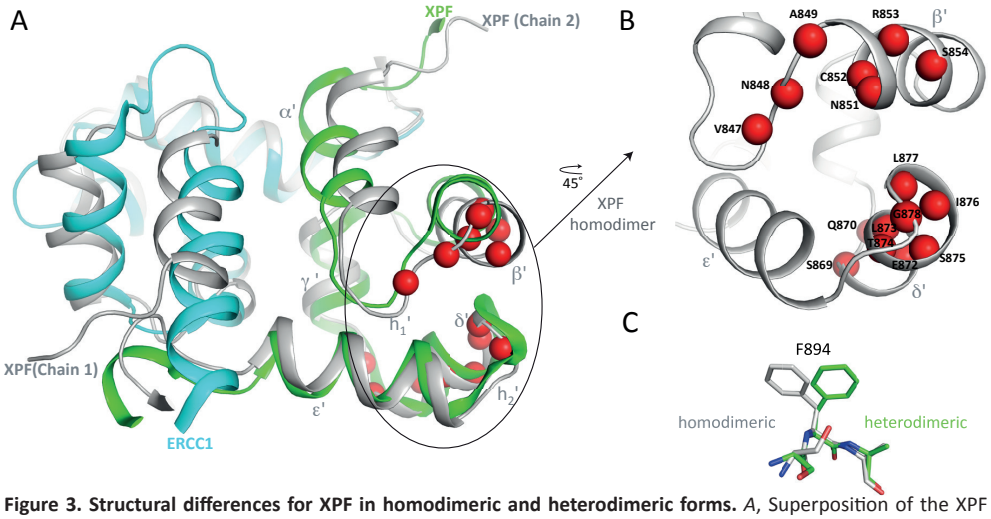


Figure 3. Structural differences for XPF in homodimeric and heterodimeric forms. *A*, Superposition of the XPF $(HhH)_2$ domains in the homodimer and in the ERCC1-XPF heterodimer. The amide nitrogens with large chemical shift differences outside the interaction interface are depicted as red spheres. *B*, Expansion of (A). C, Orientation of the sidechains of F894 in homodimeric and heterodimeric XPF overlaid at the surrounding backbone residues. ERCC1 and XPF in heterodimeric ERCC1-XPF are depicted in cyan and green, respectively. The XPF homodimer is shown in grey.

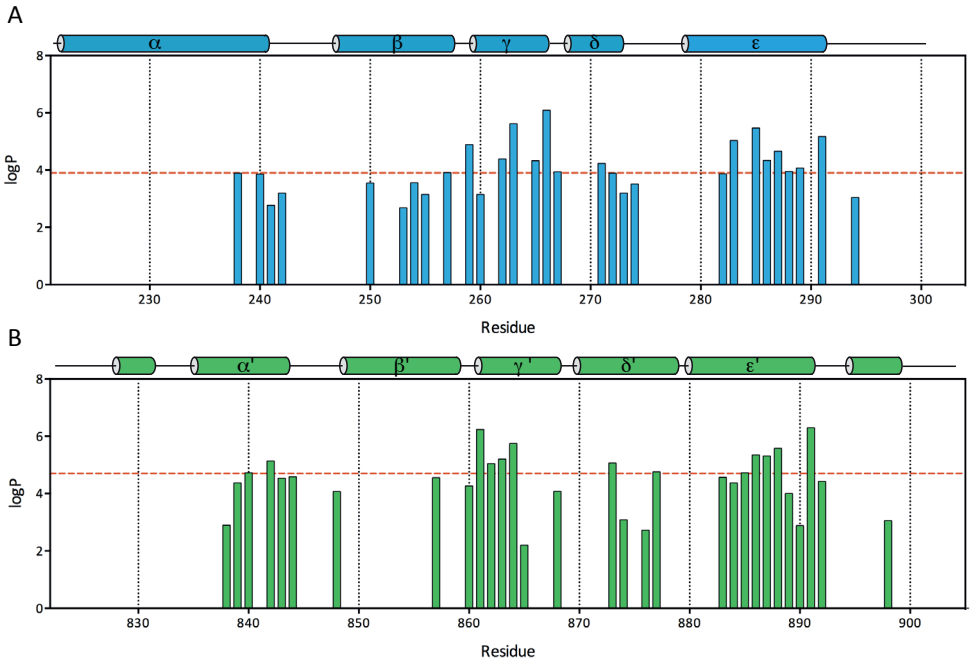


Figure 4. Protection factors for the backbone amide protons of the ERCC1-XPF $(HhH)_2$ heterodimer. The Figure shows the logarithm of protection factors (logP) calculated from H/D exchange rates (k_{ex}) for amide protons of ERCC1 (A) and of XPF as part of the ERCC1-XPF heterodimer (B). Red lines indicate global average protection factors for ERCC1 or XPF.

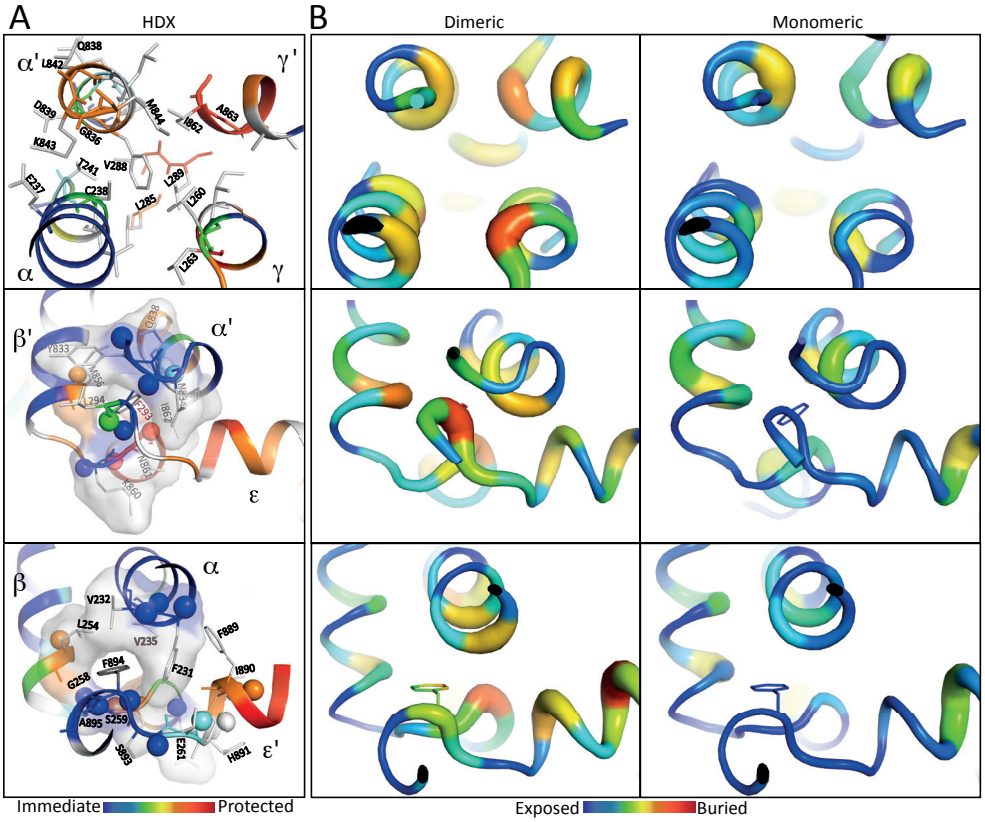


Figure 5. Hydrogen deuterium exchange and amide surface exposure of monomeric and heterodimeric XPF and ERCC1 mapped on the NMR structure. *A*, Amide proton H/D exchange rates, color-mapped onto the core of ERCC1-XPF (top), of the F293 anchor region (middle) and of the F894 anchor region (bottom). The grey for backbone and sidechains denotes residues for which no H/D exchange information could be obtained. The cavities at the anchor regions are shown in a surface representation. The spheres are representing the amide protons of each residue. *B*, Amide proton atom depths for the ERCC1-XPF contact regions assuming heterodimeric (left panel) and monomeric XPF and ERCC1 (right panel).

In addition to HDX, H-bond stability can be analyzed using amide proton temperature coefficients (31). Supplementary Table S1 shows these values for the heterodimeric ERCC1-XPF complex. The temperature coefficients of the amide protons in the loop regions and N-terminal of helices of ERCC1-XPF are below -4.5 ppb/K indicating the absence of stable H-bonds with carboxyl groups (31). For most of the slowly exchanging backbone amide protons the temperature coefficients are above -4.5 ppb/K indicating that these are H-bonded. These H-bonds are found within helices γ, δ and ε of ERCC1, and within helices α', γ', δ' and ε' of XPF. The combined analysis of the HDX rates and temperature coefficients demonstrates that some amide protons in the helices are involved in weak H-bonds, which

is seen for instance for residues 223, 236 and 240 in helix α and residue 249 in helix β of ERCC1 and helix β' and the N-terminal and C-terminal helices of XPF.

Fast hydrogen deuterium exchange can correlate with amide surface exposure, and therefore we estimated the surface exposure of the amide protons of ERCC1 and XPF by calculating their "atom depths". This atom depth is defined as the distance of the bulk solvent to the amide proton and this parameter can be related to HDX rates with some precautions (36-38). In particular for residues near the interface of the heterodimer large differences in atom depths exist for amides in the heterodimeric and monomeric states. In the ERCC1-XPF heterodimer, the amides of the interfacial ERCC1 residues in helices α (231, 234, 235, 238, 239), β (253), γ (260, 263, 264), ε (285, 288, 289) and the anchor residue F293 are deeply buried, and therefore we expect a very slow H/D exchange for those residues when the HDX would occur when ERCC1 is in the heterodimer. This would also apply to the amides in XPF helices α' (837, 838, 840, 841), β' (855, 856), γ' (862, 865), ε' (886, 887, 888 and 890) and the anchor residue F894. When ERCC1 and XPF are in monomeric form, only few amide protons are predicted to exchange slowly (Figure 5B).

As the protection factors in Figure 4 indicate, many interfacial residues of ERCC1-XPF do not fit with HDX when ERCC1 is in a dimeric state. For instance, in Figures 4A and 5A large parts of helices α and γ show little protection, which agrees best with exchange via a monomer (Figure 5B). In contrast, for XPF no evidence is found for exchange via monomers. This indicates that ERCC1 and XPF each follow a different mechanism to explain the observed H/D exchange results. This finding is only possible when ERCC1 is less stable in the monomeric than in the dimeric state. However, when ERCC1 would exist transiently as monomers, XPF monomers should exist simultaneously. Such XPF monomers could react rapidly into homodimers, which is not observed experimentally, at least not at temperatures below 50 °C (even during the lengthy recordings of the 2D [1H,15N]-HSQC spectra during HDX, we never observed the appearance of signals characteristic of XPF homodimer). This rather suggests that fast dissociation of ERCC1 and XPF is followed by fast re-association of the ERCC1 monomers with XPF monomers. In this model the association of XPF to form homodimers does not occur, perhaps because the XPF monomers have no conformation that is suitable for immediate XPF homodimer formation. To test this hypothesis, we checked if XPF has a different structure in the homodimeric and heterodimeric form.

Structure and stability differences of heterodimeric and homodimeric XPF

Das *et al*. already described substantial differences between the homodimeric and heterodimeric conformations of XPF (25), which agrees well with the results from the NMR cross-relaxation experiments by Choi *et al*. (17). This holds not only for the interface region but can also be seen in various other regions. For example, the distance between helix α' and helix γ' is increased to facilitate the accommodation of the bulky side chains

of F840, L841 and F889. The reorientation of helix α' in XPF also accounts for the larger interaction interface in the homodimers as compared to the ERCC1-XPF heterodimer (1760 Å^2 *vs.* 1534 Å^2, respectively) (25). The high stability of the XPF homodimer was additionally attributed to π–stacking of the two F840 sidechains and the T-shaped arrangement of both F889 sidechains at the interface, and several additional intermonomer hydrogen bonds, such as the N861 side chain NH_2 to the F889 backbone CO, the A863 backbone NH to the I890 backbone CO and the K860 sidechain NH to the H891 backbone CO. The backbone-Cα RMSD between XPF in the heterodimeric and the homodimeric forms are substantial for most residues, even when only the helices are fitted (Figure 2E and Figure 3C), and not only towards the C-terminal and N-terminal ends. The overall RMSD of ~3 Å is largely due to both tilting and displacement of helix α' allowing the accommodation of the bulky sidechains of F840, L841 and F889. The different XPF folds include also a different shape for the cavities embracing the phenylalanine anchors: in the XPF homodimers the cavities are more shallow than in the heterodimers with F894 aromatic ring tilting 66˚ towards helix γ and far from helix α of its partner XPF monomer (21). This structural variability implies that XPF has to adapt its conformation to allow the formation of stable homodimers.

Not only the structure, but also the stability of XPF in the homodimeric and heterodimeric forms is different. The observed HDX protection factors of heterodimeric XPF are overall over 10-fold higher than reported for homodimeric XPF by Das *et al.* (25) (cf data for ERCC1-XPF in Figure 2A and for $(XPF)_2$ in Supplementary Figure S2). This suggests that the packing of XPF in ERCC1-XPF is better than in $(XPF)_2$. This may be a counterintuitive finding given the higher temperature stability of the homodimers. One explanation is that upon temperature increase the monomer-to-heterodimer equilibrium is shifted to the monomeric state. While free ERCC1 unfolds and aggregates to form a visible precipitate, the thermal energy allows XPF to adopt a different conformation, which may be less stable than the conformation of XPF in the heterodimer, but which is essential to form stable homodimers.

CONCLUDING REMARKS

The heterodimer of the ERCC1-XPF Helix-hairpin-Helix domains forms a complex that can rapidly dissociate into its ERCC1 and XPF monomeric counterparts. These monomers rapidly re-associate into a heterodimer, so that no ERCC1 or XPF homodimers are formed. In this heterodimeric complex, hydrogen deuterium exchange shows that ERCC1 is less stable than XPF. Similarly, ERCC1 unfolds at lower temperature than XPF. When ERCC1 self-aggregates and precipitates at elevated temperature when unfolded, the remaining XPF monomers become available for homodimer formation. Although these XPF homodimers are extremely stable, their association from XPF monomers that would exist after ERCC1-XPF dissociation must be slower than the association of XPF with free ERCC1 monomers. An explanation for this slow association of XPF monomers can be found in the different fold

of homodimeric XPF *versus* heterodimeric XPF. When the interconversion between these different folds is slow or when the homodimeric XPF fold corresponds to a hardly populated state, the XPF monomer association can be slow, especially at moderate temperatures. However, we postulate that when no ERCC1 is present the homodimer association still occurs and highly stable homodimers can be formed. Support for a model that XPF can have an alternate fold in a sparsely populated state comes from the hydrogen deuterium exchange, which showed a lower intrinsic stability of XPF in the homodimers than of XPF in ERCC1-XPF. These results demonstrate the essence of a correct fold of XPF for association with ERCC1 leading to a functional ERCC1-XPF complex.

Based on work of de Laat *et al*. (15), it can be concluded that the C-terminal HhH domain of both ERCC1 and XPF is required and sufficient for ERCC1-XPF dimer formation. Also, *in vivo* the ERCC1 and XPF proteins are not found in isolation (13,14,39). In ERCC1-defective cells only small amounts of XPF are found (40,41). This indicates that the ERCC1-XPF complex is very stable. Current study shows however, that this heterodimeric complex, which is critical for its function, has a relatively short half-life. The dissociation of ERCC1-XPF would lead to monomers, and, assuming that this behavior is also present in the full-length proteins, this indicates that monomeric species might be formed *in vivo*. However, since these are not found, probably the monomeric proteins are unstable and prone to active degradation (13,14,39). We propose that the structural rearrangement somehow forms a barrier for XPF homodimer formation and that the formation of the functional ERCC1-XPF heterodimeric DNA repair complex is favored. The observed fast association and dissociation of ERCC1 and XPF might also enable continuous replenishment to maintain a functional ERCC1-XPF complex, presenting the direct possibility for regulation of DNA repair regulation via the ERCC1 and XPF cellular levels.

ACKNOWLEDGEMENTS

The authors thank Drs. Devashish Das, Hans Ippel and Anna Vangone for help and useful discussions. This work was financially supported by the Netherlands Foundation for Chemical Research (NWO-CW), the Slovenian Infrastructural Centre for Analysis of Molecular Interaction and the Slovenian Research Agency (Z1-4071).

REFERENCES

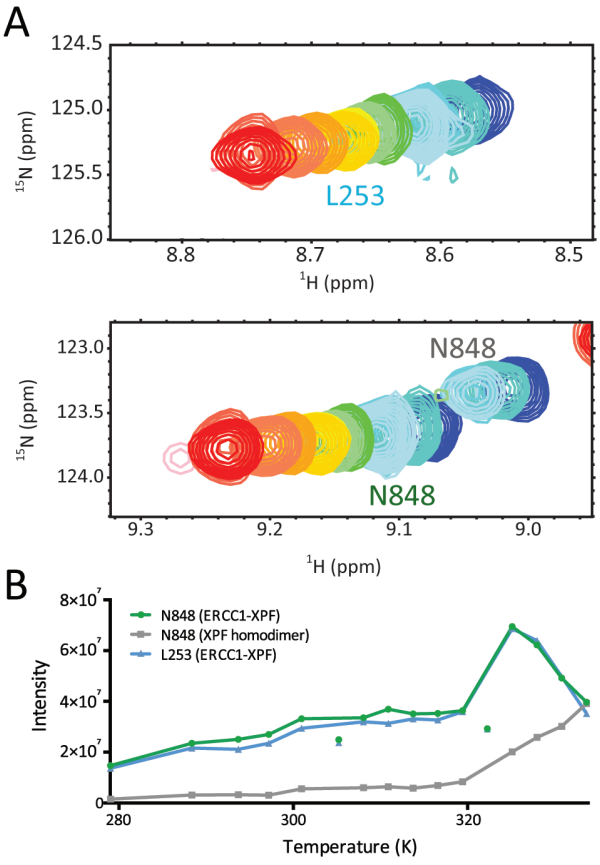
1. Friedberg, E. C., Walker, G. C., and Siede, W. (2006) DNA repair and mutagenesis, American Society for Microbiology, Washington DC
2. Caldecott, K. W. (2008) Single-strand break repair and genetic disease. Nature Reviews Genetics 9, 619-631
3. Hanawalt, P. C. (2002) Subpathways of nucleotide excision repair and their regulation. Oncogene 21, 8949-8956
4. Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H. J., and Niedernhofer, L. J. (2008) ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Molecular and Cellular Biology 28, 5082-5092
5. Goodarzi, A. A., and Jeggo, P. A. (2012) The repair and signaling responses to DNA double-strand breaks. Advances in Genetics 82, 1-45
6. Adair, G. M., Rolig, R. L., Moore-Faver, D., Zabelshansky, M., Wilson, J. H., and Nairn, R. S. (2000) Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination. EMBO Journal 19, 5552-5561
7. Bhagwat, N., Olsen, A. L., Wang, A. T., Hanada, K., Stuckert, P., Kanaar, R., D'Andrea, A., Niedernhofer, L. J., and McHugh, P. J. (2009) XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. Molecular and Cellular Biology 29, 6427-6437
8. Huang, Y., and Li, L. (2013) DNA crosslinking damage and cancer-a tale of friend and foe. Translational Cancer Research 2, 144
9. Zhu, X.-D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H. J., and de Lange, T. (2003) ERCC1/XPF removes the 3′ overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Molecular Cell 12, 1489-1498
10. Besse, B., Olaussen, K. A., and Soria, J.-C. (2013) ERCC1 and RRM1: ready for prime time? Journal of Clinical Oncology 31, 1050-1060
11. Westerveld, A., Hoeijmakers, J. H. J., 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
12. Biggerstaff, M., Szymkowski, D. E., and Wood, R. D. (1993) Co-correction of the ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects in vitro. EMBO Journal 12, 3685-3692
13. Park, C. H., Bessho, T., Matsunaga, T., and Sancar, A. (1995) Purification and characterization of the XPF-ERCC1 complex of human DNA repair excision nuclease. Journal of Biological Chemistry 270, 22657-22660
14. Sijbers, A. M., van der Spek, P. J., Odijk, H., van den Berg, J., van Duin, M., Westerveld, A., Jaspers, N. G. J., Bootsma, D., and Hoeijmakers, J. H. J. (1996) Mutational analysis of the human nucleotide excision repair gene ERCC1. Nucleic Acids Research 24, 3370-3380
15. de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1998) Mapping of interaction domains between human repair proteins ERCC1 and XPF. Nucleic Acids Research 26, 4146-4152
16. 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 Journal 24, 895-905
17. 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. Journal of Biological Chemistry 280, 28644-28652
18. Tripsianes, K., Folkers, G., Das, D., Odijk, H., Jaspers, N. G. J., Hoeijmakers, J. H. J., 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
19. Tsodikov, O. V., Enzlin, J. H., Schärer, O. D., and Ellenberger, T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. Proceedings of the

National Academy of Sciences of the United States of America 102, 11236-11241
20. Zamble, D. B., and Lippard, S. J. (1995) Cisplatin and DNA repair in cancer chemotherapy. Trends in Biochemical Sciences 20, 435-439
21. Zhang, N., Liu, X., Li, L., and Legerski, R. (2007) Double-strand breaks induce homologous recombinational repair of interstrand cross-links via cooperation of MSH2, ERCC1-XPF, REV3, and the Fanconi anemia pathway. DNA Repair 6, 1670-1678
22. Jordheim, L. P., Barakat, K. H., Heinrich-Balard, L., Matera, E.-L., Cros-Perrial, E., Bouledrak, K., El Sabeh, R., Perez-Pineiro, R., Wishart, D. S., and Cohen, R. (2013) Small molecule inhibitors of ERCC1-XPF protein-protein interaction synergize alkylating agents in cancer cells. Molecular pharmacology 84, 12-24
23. McDonnell, J. M. (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. Current Opinion in Chemical Biology 5, 572-577
24. Nieba, L., Nieba-Axmann, S. E., Persson, A., Hämäläinen, M., Edebratt, F., Hansson, A., Lidholm, J., Magnusson, K., Karlsson, Å. F., and Plückthun, A. (1997) BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. Analytical Biochemistry 252, 217-228
25. Das, D., Tripsianes, K., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., Boelens, R., and Folkers, G. E. (2008) The HhH domain of the human DNA repair protein XPF forms stable homodimers. Proteins: Structure, Function, and Bioinformatics 70, 1551-1563
26. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. Quarterly reviews of biophysics 16, 521-655
27. 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. Journal of Biomolecular NMR 6, 277-293
28. Goddard, T. D., and Kneller, D. G. SPARKY 3. University of California, San Francisco
29. Connelly, G. P., Bai, Y., Jeng, M. F., and Englander, S. W. (1993) Isotope effects in peptide group hydrogen exchange. Proteins: Structure, Function and Genetics 17, 87-92
30. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. Proteins: Structure, Function and Genetics 17, 75-86
31. Baxter, N. J., and Williamson, M. P. (1997) Temperature dependence of 1H chemical shifts in proteins. Journal of Biomolecular NMR 9, 359-369
32. Cierpicki, T., and Otlewski, J. (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. Journal of Biomolecular NMR 21, 249-261
33. Sobolev, V., Eyal, E., Gerzon, S., Potapov, V., Babor, M., Prilusky, J., and Edelman, M. (2005) SPACE: A suite of tools for protein structure prediction and analysis based on complementarity and environment. Nucleic acids research 33, W39-W43
34. Tan, K. P., Nguyen, T. B., Patel, S., Varadarajan, R., and Madhusudhan, M. S. (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic acids research 41, W314-321
35. Schreiber, G., Haran, G., and Zhou, H. X. (2009) Fundamental aspects of protein - Protein association kinetics. Chemical Reviews 109, 839-860
36. Bennett, B. C., Gardberg, A. S., Blair, M. D., and Dealwis, C. G. (2008) On the determinants of amide backbone exchange in proteins: A neutron crystallographic comparative study. Acta Crystallographica Section D: Biological Crystallography 64, 764-783
37. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. Quarterly Reviews of Biophysics 16, 521-655
38. Maity, H., Lim, W. K., Rumbley, J. N., and Englander, S. W. (2003) Protein hydrogen exchange mechanism: Local fluctuations. Protein Science 12, 153-160
39. 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., Rademakers, S., De Rooij, J., Jaspers, N. G. J., Hoeijmakers, J. H. J., and Wood, R. D. (1996)

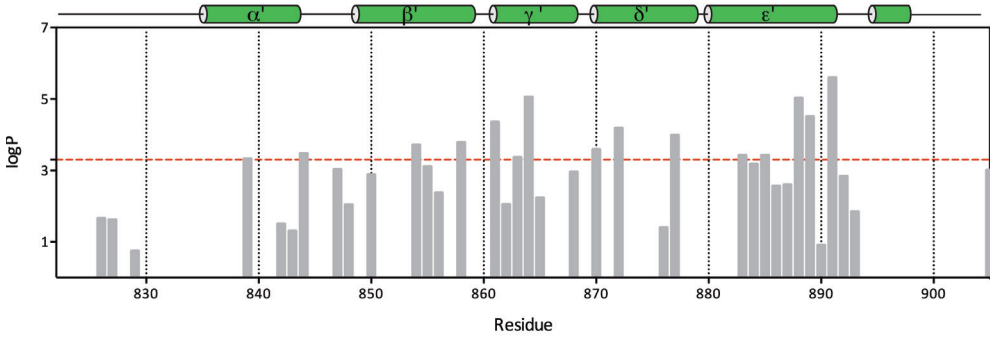
Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell 86, 811-822

40. Gaillard, P.-H. L., and Wood, R. D. (2001) Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. Nucleic acids research 29, 872-879
41. Hayashi, T., Takao, M., Tanaka, K., and Yasui, A. (1998) ERCC1 mutations in UV-sensitive Chinese hamster ovary (CHO) cell lines. Mutation Research - DNA Repair 407, 269-276

SUPPLEMENTAL INFORMATION



Supplementary Figure S1. Conversion of ERCC1-XPF heterodimer into XPF homodimer upon temperature increase. A, Overlays of [^{1}H,^{15}N]-HSQC regions for L253 (blue) and N848 (green) in ERCC1-XPF and of N848 in $(XPF)_2$ (grey) for temperatures ranging from 279-333 K. B, [^{1}H,^{15}N]-HSQC cross-peak intensities as a function of temperature for residues L253 and N848 in heterodimeric ERCC1-XPF and for residue N848 in homodimer XPF.



Supplementary Figure S2. Protection Factors of the backbone amide protons of the XPF homodimer. The protection factors logP are calculated from the ratio of the intrinsic H/D exchange rate of a free amino acid and the observed H/D exchange rate (k_{ex}). The red line indicates a global average protection factor for XPF.

Supplementary Table 1. Amide proton temperature coefficients (ppb/K) and Hydrogen/Deuterium Exchange rates of ERCC1-XPF.

residue#	ppb/K	k_{ex} (in s^{-1}) [A]	logP
222	---	Immediate	0
223	-4.961	Immediate	0
224	-4.101	Immediate	0
228	-2.515	Immediate	0
230	---	Immediate	0
231	---	Immediate	0
232	---	Immediate	0
233	---	Immediate	0
234	---	Immediate	0
235	-2.932	---	---
236	-4.689	Immediate	0
237	---	Immediate	0
238	2.733	4.77E-04	3.89
239	-2.684	---	---
240	-4.724	8.99E-05	3.86
241	-1.179	2.88E-03	2.77
242	-2.932	2.48E-04	3.19
243	-7.802	Immediate	0
244	---	---	---
245	-5.394	Immediate	0
246	---	Immediate	0
248	---	Immediate	0
249	-6.143	Immediate	0
250	-2.102	5.43E-04	3.55
251	-0.6294	Immediate	0
252	---	Immediate	0
253	-4.08	1.08E-03	2.68
254	-2.497	5.50E-05	3.56
255	-4.972	4.58E-04	3.15
257	-3.689	1.41E-04	3.91
258	---	Immediate	0
259	-0.5494	5.50E-05	4.89
260	-6.632	4.58E-04	3.15
261	-1.822	Immediate	0

residue#	ppb/K	k_{ex} (in s^{-1}) [A]	logP
824	---	Immediate	0
825	-6.757	Immediate	0
826	-6.692	Immediate	0
827	-5.924	Immediate	0
829	-6.18	Immediate	0
830	-5.593	Immediate	0
832	-3.563	Immediate	0
833	-6.633	Immediate	0
834	-3.081	Immediate	0
836	-5.785	Immediate	0
838	1.259	1.02E-03	2.9
839	-8.436	1.05E-04	4.37
840	-2.251	9.03E-06	4.73
842	-3.632	1.46E-06	5.14
843	-0.8336	~2.01E-05	4.54
844	-3.447	4.07E-05	4.59
846	-3.944	Immediate	0
847	-2.366	~0.0000305	0
848	-3.943	Immediate	4.07
849	---	Immediate	0
850	-3.477	Immediate	0
853	-4.98	Immediate	0
854	-9.414	Immediate	0
855	-1.21	Immediate	0
856	-2.727	2.07E-05	4.56
857	-4.048	Immediate	0
858	-1.951	---	---
859	-1.692	6.01E-05	---
860	-2.37	Immediate	4.27
861	-1.294	1.40E-07	7.55
862	-3.292	~0.00000013	6.55
863	1.065	5.20E-08	7.14
864	-4.135	1.70E-06	5.75
865	---	1.50E-03	2.2

Supplementary Table 1. (Continued) Amide proton temperature coefficients (ppb/K) and Hydrogen/Deuterium Exchange rates of ERCC1-XPF.

residue#	ppb/K	k_{ex} (in s^{-1}) [A]	logP
262	-4.52	4.20E-05	4.39
263	-3.992	1.30E-06	5.62
265	0.062	3.41E-05	4.33
266	-0.62	~0.000001	6.09
267	-4.964	3.29E-04	3.94
271	-4.3	1.28E-05	4.23
272	-3.195	9.90E-05	3.89
273	-1.16	2.10E-04	3.19
274	-2.466	9.70E-04	3.51
276	---	Immediate	0
277	-2.342	Immediate	0
278	-6.494	Immediate	0
281	-9.072	Immediate	0
282	-0.891	2.19E-04	3.87
283	---	1.36E-05	5.03
285	-5.577	1.86E-06	5.47
286	-3.989	2.00E-05	4.34
287	-2.949	3.93E-05	4.66
288	-1.999	1.80E-05	3.95
289	-0.7346	2.00E-05	4.07
291	-7.421	2.98E-05	5.17
293	---	Immediate	0
294	---	3.30E-04	3.05
295	---	Immediate	0
296	---	Immediate	0
297	---	Immediate	0
298	-8.371	Immediate	0
299	---	Immediate	0
300	---	Immediate	0
301	-4.91	Immediate	0

residue#	ppb/K	k_{ex} (in s^{-1}) [A]	logP
867	0.5484	Immediate	0
868	-1.406	2.69E-05	4.08
869	-8.425	Immediate	0
870	-4.39	Immediate	0
871	---	Immediate	0
872	---	---	0
873	---	2.13E-10	9.04
874	-3.508	5.28E-04	3.09
876	-2.563	8.78E-04	2.72
877	-3.671	2.36E-10	8.91
878	-1.791	Immediate	0
881	---	Immediate	0
882	---	Immediate	0
883	-2.128	6.97E-05	4.57
884	-1.55	4.78E-05	4.37
885	-2.555	3.48E-05	4.73
886	---	9.16E-11	9.75
887	-3.353	2.00E-06	5.31
888	-2.694	9.62E-11	10.26
889	---	4.64E-05	4.01
890	-1.831	3.52E-04	2.88
891	-1.268	1.50E-06	6.3
892	-1.87	1.83E-04	4.43
893	-5.069	Immediate	0
894	-2.565	Immediate	0
895	---	Immediate	0
897	-5.458	Immediate	0
898	---	1.56E-04	3.06
899	-7.714	Immediate	0
900	---	Immediate	0
901	---	Immediate	0
902	---	Immediate	0
903	---	Immediate	0
904	---	Immediate	0
905	-6.92	Immediate	0

[A] Residues for which the signal intensity is already gone before the start of the HSQC recordings are designated as "immediate".

Chapter 4

Single-strand DNA binding by the helix-hairpin-helix domain of XPF contributes to substrate specificity of ERCC1-XPF

Devashish Das, Maryam Faridounnia, Lidija Kovacic, Robert Kaptein, Rolf Boelens, and Gert E. Folkers

Manuscript in preparation

ABSTRACT

The ERCC1-XPF complex, part of the nucleotide excision repair pathway, is required for incision of DNA upstream of the DNA damage. Structural studies of a homodimeric XPF–ssDNA complex allowed us to propose a model where the positioning of ERCC1-XPF to the ss/dsDNA junction is dependent on the joint binding to the DNA binding domain of ERCC1 and XPF. We now use biochemical and structural methods to experimentally validate this model. We show that the homodimeric XPF HhH domain is able to bind to Holliday junction substrates and various ssDNA sequences with a preference for guanine containing sequences. NMR titration experiments and in vitro DNA-binding assays show that, in contrast to the corresponding domain of ERCC1, XPF binds specifically to ssDNA. The HhH domain of ERCC1 preferentially binds dsDNA through the hairpin region. The two separate non-overlapping DNA binding domains in the ERCC1-XPF heterodimer jointly bind to an ss/dsDNA substrate and thereby, at least partially, dictate the incision position during damage removal. Based on structural models, NMR titrations, DNA-binding studies, site-directed mutagenesis, charge distribution and sequence conservation, we propose that the HhH domain of ERCC1 binds to dsDNA upstream of the damage and XPF binds to the non-damaged strand of a repair bubble.

INTRODUCTION

In order to survive, cells must repair a plethora of UV light induced DNA lesions. Therefore, the cells activate the Nucleotide Excision Repair (NER) pathway, a conserved DNA repair mechanism present in all living organism, which can remove a variety of DNA lesions (1,2). Within a mammalian cell nearly 25-30 proteins are known to participate in two NER pathways: global genome and transcription coupled repair (3-5). Mutations in NER genes lead to impaired DNA repair. Presently, a dozen mutations in distinct NER genes have been identified in patients with eight overlapping phenotypes (6,7). Most patients carrying a mutation in NER genes develop two distinct symptoms, sunlight induced skin cancer and segmental progerias without cancer (8,9).

ERCC1 and XPF form a stable heterodimeric complex that is essential for NER, that functions as a structure specific DNA endonuclease and is able to nick at the 5' end of damaged DNA (10-12). Mutations in ERCC1 and XPF genes can be linked to sunlight induced skin abnormalities, late onset of skin cancers, neurodegeneration and premature aging in both human patients and transgenic mice (7-9,13). In the absence of ERCC1 only a marginal amount of XPF is present in fibroblasts and CHO cells (10,12,14-17). This suggests that *in vivo* stability of the full length ERCC1-XPF depends on a tight association between the two proteins (10,12,14). Consistent with this finding, XPF and ERCC1 knockout mice exhibit similar phenotypes (18-20). Furthermore, postnatal phenotypes of XPF and ERCC1 knockout mice suggest additional function for ERCC1-XPF in homologous recombination (21), single-strand annealing (22), inter strand crosslink repair (23,24), telomere maintenance (25,26), and gene-targeting events (27). All these genome regulatory processes require binding of ERCC1-XPF at distinct DNA sequences, involving various protein complexes (28-30).

Biochemical and structural studies revealed that the helix-hairpin-helix (HhH) domain present in the C-terminal part of both proteins is essential for both ERCC1-XPF complex formation (12) and DNA binding (31-33) . Structural studies by us and others showed that the HhH domain of the XPF protein serves as a scaffold for ERCC1 folding, permitting formation of a stable heterodimer (33,34). Using NMR spectroscopy we found that ERCC1 specifically recognizes DNA (33). Furthermore, the C-terminal ERCC1-XPF complex binds more tightly to ss-dsDNA junctions, such as bubble and fork substrates than to either dsDNA or ssDNA alone. This led us to suggest that XPF might also contain an independent DNA-binding domain. To test this hypothesis we took advantage of an earlier observation, that demonstrated that the isolated HhH domain of XPF is able to form a highly stable homodimer (35). Although XPF lacks one residue in the second hairpin motif, this domain adopts a canonical HhH domain structure (35). Recently, using NMR, we showed that the homodimeric XPF HhH domain binds ssDNA. We subsequently determined the solution structure of XPF bound to ssDNA (36). We could show that besides non-specific phosphate backbone contacts involving the second helix of the first HhH motif, a cavity is

formed between the two motifs of the HhH domain where a guanine base is bound. These observations led us to propose that the ERCC1-XPF heterodimer recognizes DNA substrates involving the two individual DNA binding surfaces present in ERCC1 and XPF that bind respectively dsDNA and ssDNA.

Here we confirm and extend this model using *in vitro* DNA binding assays and NMR titration experiments revealing the substrate preference of XPF and the ERCC1-XPF heterodimer for various DNA sequences. We show that XPF binds preferentially to ssDNA, without defined sequence specificity, but with a clear preference for guanine sequences. Chemical shift perturbations (CSPs) in NMR titrations, and *in vitro* DNA binding experiments indicate that ssDNA binding to XPF also occurs in the ERCC1-XPF heterodimeric complex. On the basis of the XPF-ssDNA structure, the chemical shift perturbations for the ERCC1-XPF heterodimer upon addition of various DNA sequences and binding studies using mutated ERCC1 and XPF proteins, we propose a model for the binding of the HhH domains of ERCC1-XPF heterodimers. In this model the concerted binding of the HhH domains of ERCC1 and XPF to dsDNA and ssDNA, respectively, is essential for the correct positioning on the ss/dsDNA junction.

RESULTS

De Laat et al (12) have shown that the C-terminal HhH domains of XPF and ERCC1 are indispensable for heterodimer formation and function. Like the full-length ERCC1-XPF heterodimers, these HhH domains can together form stable complexes with various ss/ds junction containing DNA substrates, like bubble, hairpin and fork (31-33). These findings suggest that structure-specific DNA binding by the ERCC1-XPF heterodimer is dependent on the HhH domain regions of both proteins. The ability of XPF to bind to ssDNA further supports this model (36). To elucidate the contribution of the HhH domain of XPF in ERCC1-XPF substrate preference, we test the binding of homodimeric XPF HhH domain to a variety of DNA substrates, as shown in Supplementary Figure S1.

We first tested the binding of XPF HhH domain homodimers to bubble10 (B10), because the ERCC1-XPF heterodimer can form a stable complex with this DNA sequence as shown previously (33). Surprisingly, we noticed that XPF homodimer binds to this substrate with even higher affinity (Figure 1A and B) than the ERCC1-XPF heterodimer (Figure 1A) (33). Quantification revealed an apparent K_d of 0.5 ± 0.1 μM (Figure 1B) being more than one order of magnitude tighter than reported for the ERCC1-XPF heterodimer (33). It should be mentioned however, that the ERCC1-XPF complex dissociates during electrophoresis. If binding affinity would be determined based on the disappearance of free DNA, both complexes would bind with similar affinities.

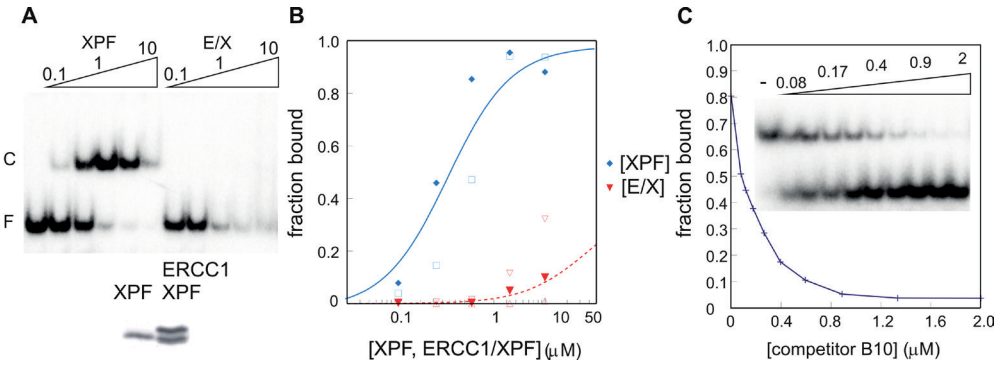


Figure 1. Binding of the HhH domain of XPF to B10 substrate. *A*, Electrophoretic mobility shift assay showing the binding of 0.1, 0.3, 1.0, 3.3, 10 μM of XPF (left panel) or ERCC1-XPF (right panel) to 0.02 μM radio-labeled B10 substrate. Free protein (F) and protein-DNA complex (C) are indicated. *B*, Quantification of a few representative DNA-binding experiments (each experiment with different symbol type) of XPF and ERCC1-XPF bound to the B10 substrate; the fraction bound is plotted as a function of the protein concentrations. Data were fitted as described before (33) and a simulated binding curve with the dissociation constants obtained were plotted. *C,* Fraction of complex formed on B10 bound by 1.25 μM XPF in the presence of 0.08, 0.12, 0.18, 0.26, 0.40, 0.59, 0.89, 1.33 and 2.0 μM non-labeled B10 oligonucleotide in comparison with the binding in the absence (0) of competitor. The inset shows the autoradiogram of the corresponding competition experiment.

The homodimeric HhH domain of XPF binds synergistically to ssDNA

The binding preferences of XPF for various probes were evaluated using competition experiments, where an excess of the non-radiolabeled B10 oligonucleotide is added to the reaction mixture. As shown in Figure 1C, binding of XPF to radiolabeled B10 is effectively competed by a non-labeled oligonucleotide, showing the expected exponential decay curve. The affinity of XPF for various DNA substrates can be determined by comparing the ability of various probes to compete for the binding of XPF to the B10 substrate.

Using these competition assays, we find that XPF is unable to bind to dsDNA or short ssDNA probes, while probes containing single-double strand junctions or longer ssDNA fragments (39nt) are found to be good substrates for the XPF homodimer (Figure 2). Interestingly, for the bubble and hairpin substrates the length of the ssDNA stretch influences the DNA-binding affinity. Figure 2 describes that both hairpin 20 (H20) and B10 are better substrates than probes containing either longer or shorter ssDNA stretches (Figure 2). Also a splayed arm with two ssDNA sequences is a better XPF substrate than any DNA sequence containing one ssDNA strand (data not shown). Taken together, these data show that the XPF HhH homodimer binds to ssDNA. We argue that stable complex formation involves both DNA binding surfaces of the symmetric XPF dimer that can bind simultaneously to either one long ssDNA fragment or to a conformationally restricted DNA containing two ssDNA stretches. In contrast, weaker binding is observed for short ssDNA sequences, hairpin

or bubble substrates. These short sequences may occupy only one binding site of the XPF homodimer. This all supports that the synergistic binding of a ssDNA fragment to the two DNA binding surfaces is required for high-affinity DNA binding by XPF homodimers.

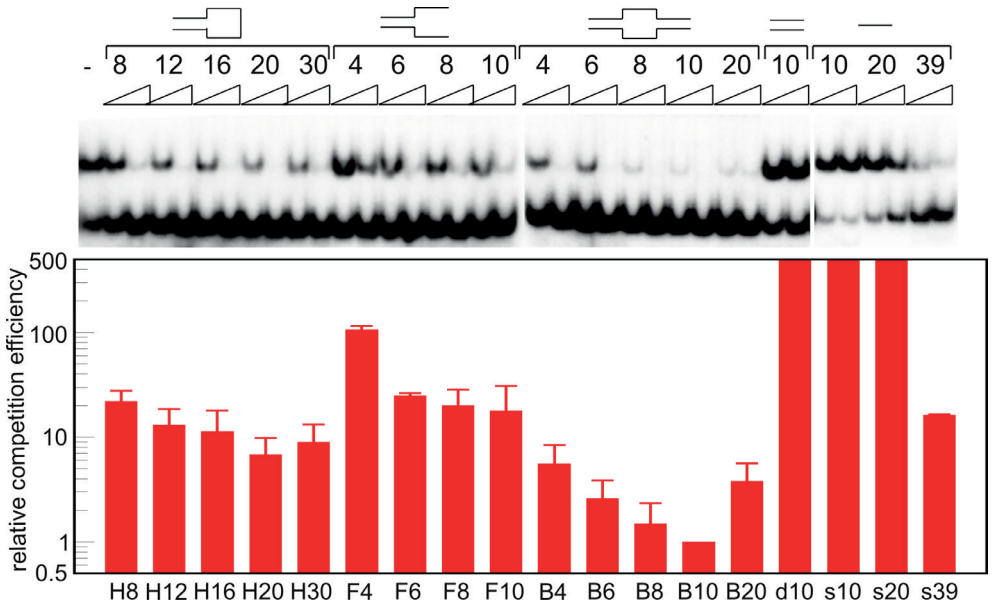


Figure 2. Binding of the HhH domain of XPF to single/double-strand DNA junctions. Representative binding experiment (upper panel) and quantification of the competition experiments (lower panel). Approximately 1µM XPF in the absence (-) or presence of 0.2 or 2 µM of the indicated non-labeled probes is bound to 0.02 µM B10 substrate. The relative competition efficiency is determined by quantification of the fraction bound in the presence of competitor. The ability to compete for B10 binding is compared with the competition obtained with non-labeled B10 substrate as shown in Figure 1C. Using non-linear regression methods this curve is fitted and the competition obtained in the presence of the amount of heterologous probe is compared with the amount of B10 probe required to obtain the same inhibition of binding. E.g. a 10 fold relative competition efficiency means that 10 times more probe is required to obtain the same inhibition of binding by a given concentration of B10 competitor. If no competition is obtained at the highest amount of competitor relative competition efficiency is shown to be higher than 500 fold. Average and standard deviation of 4 independent experiments is presented.

Given the role XPF homologs fulfill in homologues recombination (37), we tested whether the XPF homodimers can bind to a Holliday junction substrate (Supplementary Figure S2). These competition experiments show an 8-fold stronger binding than to B10 DNA. However, since this substrate is produced by annealing 4 individual oligonucleotides, the presence of ssDNA can in principle not be excluded. In addition, binding of XPF to this substrate may induce branch migration, and newly formed ssDNA sequences may serve as a substrate for XPF. To overcome the limitations of these competition experiments we also performed a binding experiment with radio-labeled, gel-purified, Holliday-junction

substrate. As shown in Supplementary Figure S2, the affinity for this probe is similar as for long ssDNA or the B10 substrate. Since no faster migrating complexes were formed we can exclude the occurrence of XPF-dependent strand separation. Taken together, these experiments show that the HhH domain of XPF binds an intact Holliday Junction but it is unable to perform branch migration and/or resolving activity.

Preference of the homodimeric HhH domain of XPF for G-rich ssDNA

To determine whether XPF possesses sequence preference in ssDNA-binding experiments, competition experiments were performed using four different 39 nt sequences. Although all fragments were able to compete for binding to a similar extent, small differences were found (Figure 3A). Since this indicated possible substrate preference, we next performed competition of XPF binding to B10 probe with poly-dG, -dT, -dC sequences of various lengths. Competition with poly dA was not possible as this probe forms a heteroduplex with the poly-dT stretch of the B10 probe used as radio-labeled probe. Irrespective of the probe length, poly-dG fragments bind stronger to XPF homodimers than poly-dT or poly-dC (Figure 3B). Quantification reveals a 4-fold higher affinity for dG_{20} than for dT_{20} or dC_{20} (data not shown). To confirm the binding preference of the homodimeric HhH domain of XPF, we tested the competition of 20 nt homopolymeric ssDNA substrates with a 39 nt ssDNA probe for binding to XPF. XPF shows strongest binding for poly-dG substrates, competing as effective as the unlabeled 39 nt ssDNA substrate (Figure 3C), while poly-dT and poly-dC compete 10-fold less effective. It is remarkable that, the purine poly-dG binds well to XPF, whereas the poly-dA binds at least two orders of magnitude less efficient. Using SPR we find similar binding preferences for the ERCC1-XPF heterodimeric complex (Figure 3E). Addition of DNA prevented the binding of the His-tagged ERCC1-XPF to Ni^{2+}-loaded NTA surface of the SPR chip. By fitting response values, corresponding to bound ERCC1-XPF, against the concentration of homopolymeric 10 nucleotide DNA sequences (Supplementary Figure S8A), we demined that poly-dG_{10} fragment binds to the heterodimer with a K_d of 2.5 ± 0.4 μM and poly-dC_{10} a K_d of 63 ± 14 μM. The binding affinity of poly-dC_{10} and poly-dA_{10} for the ERCC1-XPF heterodimer were about 2- and 4- fold lower than that of poly-dC_{10}.

Binding of the HhH domain of XPF to telomere sequences

Because XPF prefers binding to a poly-G sequence, we hypothesized that XPF might bind specifically to the telomeric hexanucleotide repeat sequence, TTAGGG. Competition experiments were performed using a single telomere sequence TTAGGG, the C-rich complementary strand CCCTAA, a split telomere probe GGGTTA, an inverted telomere probe GGGATT and a randomized sequence GAGCTA. Even in the presence of 25 μM ssDNA only marginal competition was observed for any of the 5 sequences (Supplementary Figure S3).

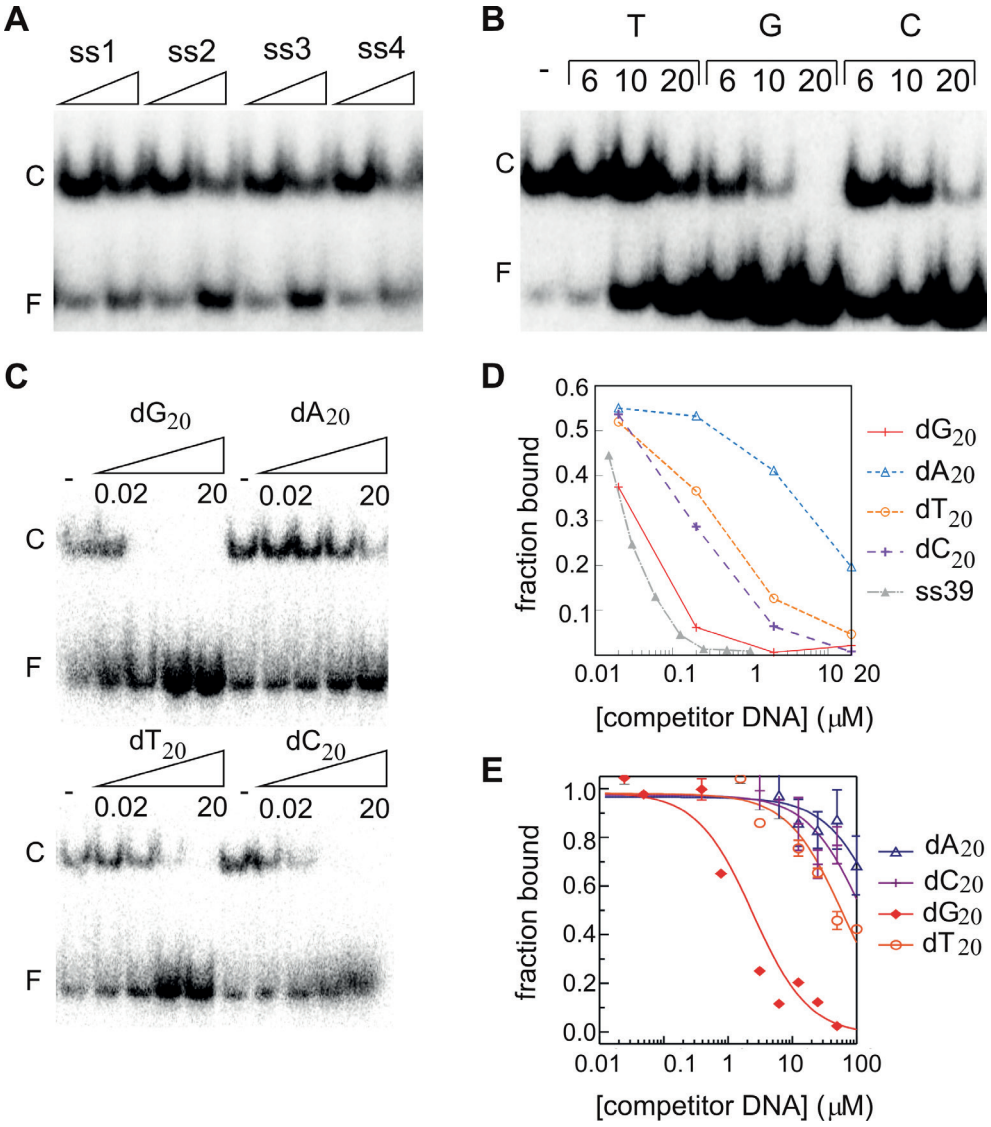


Figure 3. The HhH domain of XPF binds preferentially to guanine-rich substrate. *A*, Binding of 1.25 μM XPF in the presence of 0.2 or 2.0 μM of four distinct 39nt ssDNA probes. *B*, Competition of XPF binding to B10 substrate in the presence of 10 μM of oligo dT, dG and dC of 6, 10 or 20 nucleotides. *C*, Competition of XPF binding to ss39 in the absence (-) or presence of 0.02, 0.2, 2 and 20 μM poly dA, dT, dG and dC. *D*, Quantification of a representative competition experiment as described in C. *E*, SPR experiment showing the apparent binding affinities (K_D^{app}) of the ERCC1-XPF HhH domain for poly dA, dT, dG and dC of 10 nucleotides. The response values after association after 60 s of loading (R_{60}) were divided by the R_{60} value of the ERCC1-XPF HhH domain in absence of DNA (F), plotted against the total concentration of the DNA and fitted considering a 1:1 binding model. Experimental sensorgrams are shown in Supplementary Figure S8A.

However, the DNA binding affinity of XPF to two copies of the telomere sequences was somewhat higher than to the random nucleotide sequence. We subsequently introduced either a poly-T stretch (12 nt) or two copies of the G-rich telomere sequence into either the bubble or the splayed arm DNA probes (Supplementary Figure S1). Competition experiments revealed that the corresponding dT12 substrates were equally effective as the probes containing double telomere sequences (Supplementary Figure S3). These results further confirm that the bubble DNA is a better substrate for homodimeric XPF than the splayed-arm DNA. Together, these binding studies indicate that XPF does not bind specifically to telomere sequences.

XPF binds ssDNA in the heterodimeric XPF-ERCC1 complex.

The previous finding that the ERCC1-XPF heterodimer has preference for a ssDNA/dsDNA junction containing substrate (33), combined with the finding that homodimeric XPF binds preferentially to single strand DNA, strongly suggests that in the ERCC1-XPF heterodimeric complex XPF can also bind ssDNA.

We first established whether heterodimeric ERCC1-XPF could bind to the 10 nt ssDNA sequence used for the determination of the XPF-ssDNA structure. Whereas in EMSA experiments we cannot establish low-affinity binding, NMR titration experiments allowed us to demonstrate binding of the ERCC1-XPF complex to the 10 nt ssDNA fragment (Figure 4, Supplementary Figure S6), significant chemical shift changes for 10 nt ssDNA sequences were only obtained under low salt conditions (<50 mM), in agreement with *in vitro* DNA-binding studies (data not shown).

By analyzing the amide chemical shift changes for the free and ssDNA-bound XPF homodimer and ERCC1-XPF complex, we found a similar surface of XPF to be affected by the addition of the 10 nt ssDNA sequence, involving helix β and the following loop. The same region was also affected in the XPF HhH homodimeric protein upon addition of this DNA fragment (Supplementary Figure S4). Importantly, the previously determined DNA-binding surface of ERCC1 (Supplementary Figure S4), was not affected by the addition of ssDNA, since only a few isolated ERCC1 residues were influenced by the addition of ssDNA (Figure 4, Supplementary Figure S6). In addition to amide proton chemical shift changes in the ^{15}N-^{1}H HSQC spectra of ERCC1-XPF, also the ^{31}P-NMR spectrum of the ssDNA reveals significant chemical-shift changes upon addition of ERCC1-XPF, demonstrating complex formation (Supplementary Figure S5). The importance of the determined ssDNA binding surface of XPF for DNA binding by the ERCC1-XPF complex was underscored by the 3- and 2-fold decrease in binding affinity upon mutation of H857 and K860 to alanine (see below). These experiments clearly demonstrate the ability of XPF to also bind ssDNA in the heterodimeric ERCC1-XPF complex and show the importance of this binding surface for substrate recognition.

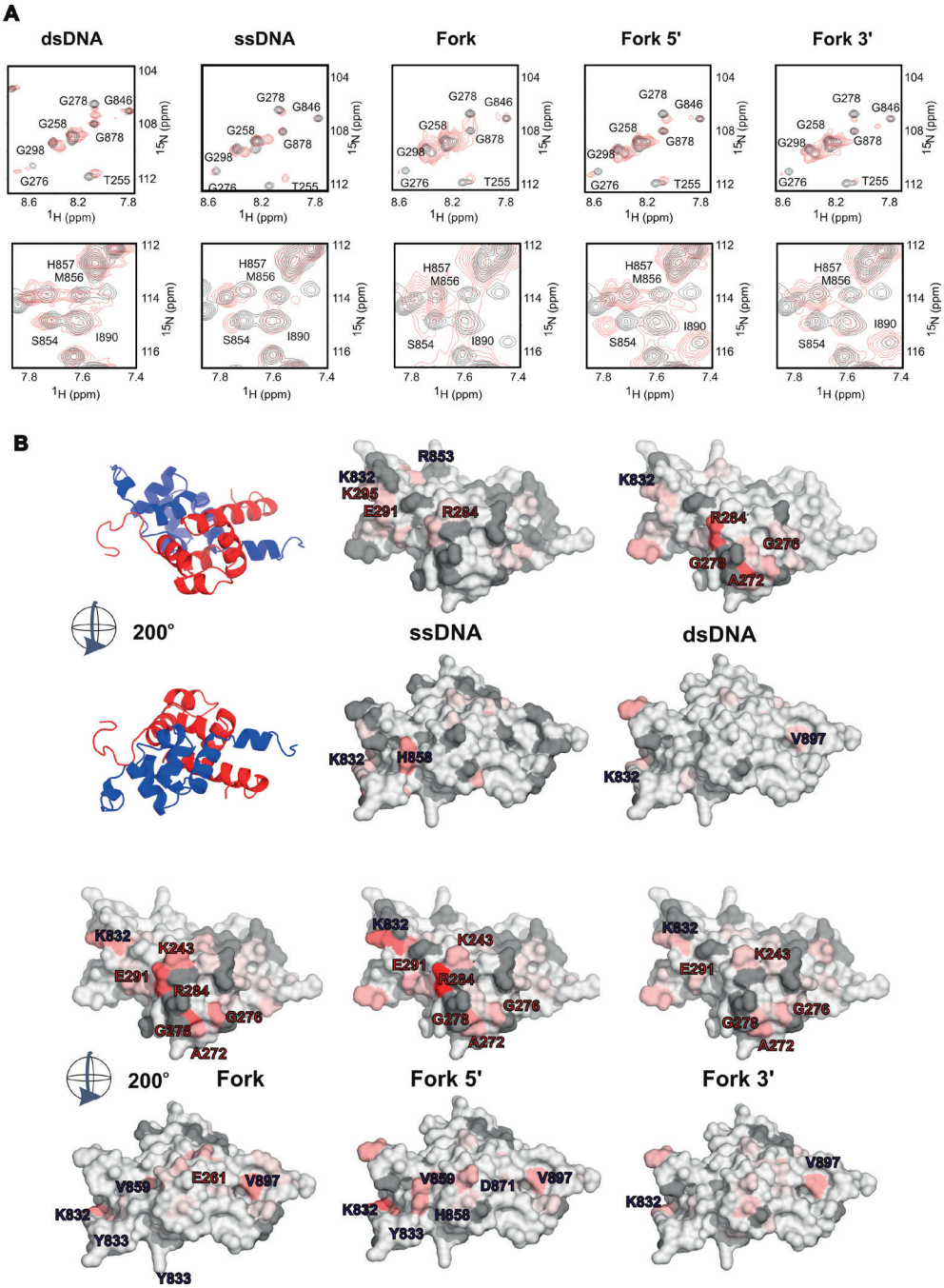
The HhH domain of ERCC1 specifically recognizes dsDNA

The above results (Figure 1-4) suggest that ERCC1 and XPF have complementary roles in dsDNA and ssDNA recognition that could dictate the high selectivity of ERCC1-XPF in binding ss/dsDNA junction substrates. To test this hypothesis, we performed NMR titration experiments with dsDNA (10 or 20 bp) and with a 20 or 30 bp splayed-arm probe containing the 10 nt ssDNA sequence that was used to determine the XPF-ssDNA structure. The binding surfaces for these probes were determined by following the chemical shift changes upon addition of DNA under various salt conditions (Supplementary Table 1). The results of these experiments are summarized in Figure 4, showing a representative set of ^{15}N-^{1}H HSQC spectra for a few of the most affected residues (Figure 4A). By calculating the average of three to five independent titration experiments (Supplementary Figure S6, Supplementary Table 1), the most affected residues were identified and plotted on the surface of the ERCC1-XPF structure (Figure 4B).

Importantly, upon addition of dsDNA, chemical shift changes were observed on the ERCC1 surface, while the XPF surface remained mostly unaffected. Mainly residues located in the second hairpin of ERCC1 show pronounced shifts, while a few residues from the first hairpin and surrounding helices are somewhat affected (Figure 4, Supplementary Figure S6). The established dsDNA-binding surface of ERCC1 is similar to that found before using hairpin DNA (Supplementary Figure S4, (33)). NMR studies using the splayed arm showed that in addition to the dsDNA-binding surface of the ERCC1 protein also residues in XPF are affected by the addition of this ssDNA-containing sequence. Most pronounced shifts in ERCC1 were found in the second hairpin regions including G276 and G278 while the first hairpin region encompassing K243-T248 was affected to a lesser extent. For XPF the regions 832-833 and 852-859 were mostly affected by the addition of splayed arm DNA. These experiments clearly establish that the hairpin regions of ERCC1 are involved in dsDNA binding, while the previously determined ssDNA binding surface of XPF is also involved in ssDNA binding in the heterodimeric complex.

Figure 4. Separate DNA-binding surfaces in ERCC1 and XPF for dsDNA and ssDNA, respectively. *A*, Representative ^{15}N-^{1}H HSQC spectrum of affected residues of the ERCC1-XPF complex showing distinct chemical shift changes upon addition of various DNA sequences (Supplementary Table 1). Free ERCC1-XPF spectrum in black, spectrum in the presence of a four-fold excess of DNA in red. For this experiment to 80 µM ERCC1-XPF 320 µM of the indicated DNA fragments was added in a buffer containing 5 mM phosphate buffer and 100 mM NaCl. *B,* The determined average CSP plus standard deviation of three to five independent titration experiments (Supplementary Figure S6) is plotted on the surface of the ERCC1-XPF structure in two different views rotated by 200°. All residues (~25) that were significantly affected (composite average chemical shift >0.2ppm) were colored. The most affected residues are in red (>0.8ppm), the other residues are colored relatively to this maximum chemical shift in red shades. Missing or unambiguous residues are depicted in gray. The position of the most affected residues is labeled on the surface.

To independently show that ERCC1 can bind dsDNA, we took advantage of a recent observation, revealing that the ERCC1-XPF complex dissociates during SPR experiments leaving ERCC1 bound to the chip (manuscript in preparation). Subsequent addition of 30 nt ssDNA (6.25 µM) did not lead to an appreciable change in mass, while addition of the same concentration



of 30 nt dsDNA led to a significant increase in signal (Supplementary Figure S7). The on-rate and the off-rate for ds30 binding to immobilized ERCC1 were determined to be $6 \pm 3*10^3$ $M^{-1}s^{-1}$ and the off-rate as $5.5 \pm 0.1* 10^{-2}$ s^{-1} respectively, giving a K_d of 9 ± 3 μM binding affinity. This relatively low binding affinity agrees well with the observed binding affinities of the ERCC1-XPF heterodimer for dsDNA in EMSA (data not shown, (33)), NMR (Figure 4) and SPR experiments (Supplementary Figure S8). While ERCC1 only binds dsDNA, both, ssDNA (ss30) and dsDNA (ds30), can bind to ERCC1-XPF with a K_D of 1 and 2 μM, respectively, as determined by SPR competition experiments (Supplementary Figure S8). These results indicate that the two independent DNA-binding surfaces present in ERCC1 and XPF together contribute to both substrate specificity and overall binding affinity of the complex.

XPF binds the non-damaged strand in the ERCC1-XPF complex

Whereas the presence of two independent DNA-binding surfaces, which are probably occupied concurrently permits positioning of the ERCC1-XPF heterodimer on ss/dsDNA junctions, it does not provide an explanation for the polarity of the cleavage. The binding preference of XPF could define this polarity by preferentially recognizing either the non-damaged (5´overhang) or the damaged (3´overhang) single strand.

Based on the orientation of the ssDNA sequence in the solution structure of XPF bound to ssDNA (36) and the observation that ERCC1 binds dsDNA, we can predict how the two DNA sequences can be linked. If we assume that ERCC1 binds dsDNA in similar manner as the structural homologs RuvA to a Holliday junction (38) and archaeal XPF to a FLAP substrate (39), it is possible to predict how the dsDNA and ssDNA fragments are connected and provide a model whether XPF binds to the damaged or the non-damaged DNA strand.

In vitro DNA-binding studies using splayed-arm substrates with either one or two 10 nt ssDNA fragments revealed that a substrate containing two arms was a better substrate than a probe containing one ssDNA sequence (data not shown). We did not observe significant differences in the affinity of the ERCC1-XPF complex for splayed-arm substrates with either 3´or 5´ssDNA sequence. We next performed NMR titrations using a 10 or 20 bp stem substrate with the ssDNA sequence connected to either the 3´or 5´of the stem. Both probes, irrespective of salt concentration or stem length, caused chemical shift perturbations in both ERCC1 and XPF upon binding of DNA, involving the two above-mentioned ssDNA- and dsDNA-binding surfaces (Figure 4). Although the overall CSPs were similar, the residues that were significantly affected are not identical arguing that the two probes bind in a different way. In comparison with the splayed-arm sequence, the DNA fragment containing the 5´ssDNA sequence binds to the ssDNA- and the dsDNA-binding surfaces highly similarly, also residues outside these main binding surfaces show similar CSPs (Figure 4). This suggests that most CSPs for the splayed-arm substrate come from the binding to the 5´ssDNA extension.

Mutant XPF and ERCC1 proteins bind DNA with lower affinity

The full-length ERCC1-XPF complex processes ss/ds DNA junctions with high selectivity (31,34,40). Above, we described two DNA-binding surfaces for XPF and ERCC1 proteins and explained the distinct roles of the XPF and ERCC1 helix-hairpin-helix domains in the ss/ds-DNA junction recognition. To validate this model, we mutated residues that would be in contact with DNA, and that would therefore affect the ability to bind the ss/ds-DNA substrate. For the XPF mutants K860A and H857A, residues that are in direct contact with ssDNA (36), the binding affinity to bubble substrate (B10) was significantly reduced as compared to wild type ERCC1-XPF (Figure 5). However, mutation of the hydrophobic residue I876A did not affect binding significantly. Mutations outside the ssDNA-binding surface (N834A, K860A and D871A) cause only a small decrease in affinity, whereas a two-fold decrease in binding affinity was noted for the double mutant Q838A and D839A.

Mutation of the positively charged ERCC1 residues (K247E, R283E, R284E)

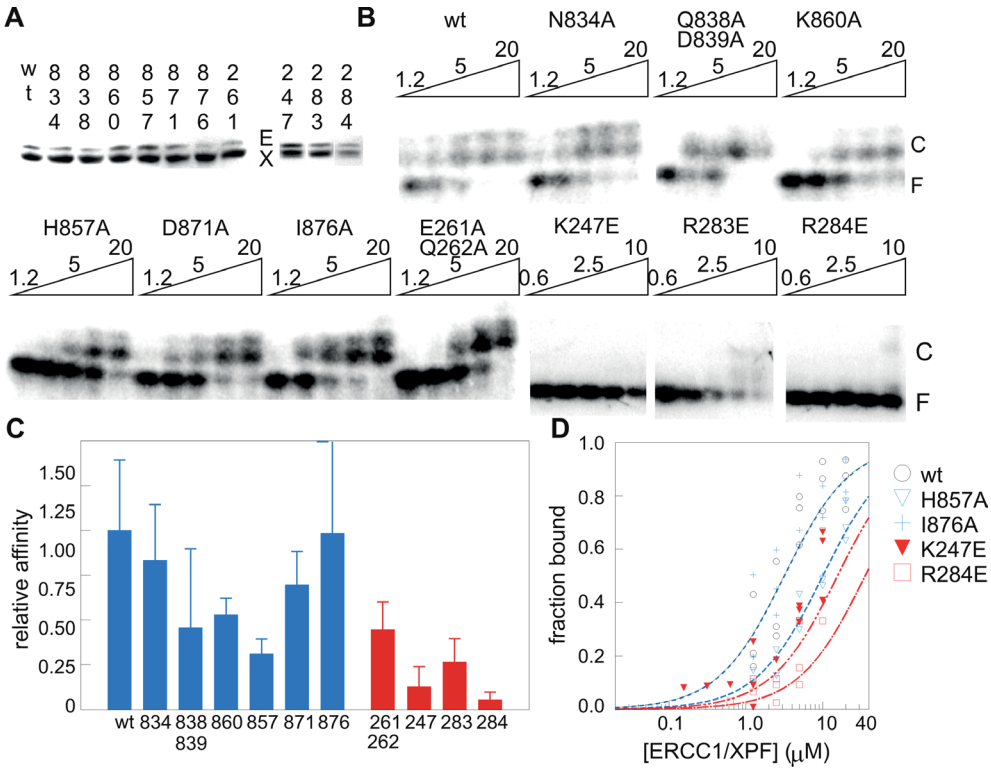


Figure 5. HhH domain surface residues affect binding of ERCC1-XPF to ss/dsDNA junctions. *A,* SDS-PAGE showing the ERCC1-XPF mutants used for binding experiments. *B,* Binding of the indicated amounts of ERCC1 and XPF mutant proteins (μM) to B10 DNA. *C,* Quantification of at least three independent experiments, calculated as average binding (+/- standard deviation) relative to the binding found for wild type ERCC1-XPF. *D,* Binding curves for a few mutants showing the simulated binding curve based on the calculated binding affinity based on three independent experiments (shown by the indicated symbols).

surrounding the hairpin residues that show the largest CSPs, led to highly reduced binding affinities (Figure 5). Mutation of the ERCC1 residues E261 and Q262, both located outside the dsDNA-binding surface, led to a much smaller two-fold decrease in binding affinity. Together, these experiments indicate the presence of two independent, functionally distinct DNA-binding surfaces present in ERCC1 and XPF, that both contribute to specificity and binding affinity. Mutation of residues that in our model are in contact with DNA, significantly affected the ability of ERCC1-XPF complex to bind ss/dsDNA substrates, underscoring the importance of these residues in substrate recognition.

DISCUSSION

The full-length ERCC1-XPF heterodimer is essential for DNA damage removal 5' to a lesion (12,14). Biochemical and structural studies provided a molecular explanation for this observation as the C-terminal double helix-hairpin-helix region of these two proteins form an obligate heterodimeric complex (33,34). Our functional studies subsequently revealed that ERCC1 is responsible for dsDNA binding. Recently, we demonstrated that the homodimeric XPF binds ssDNA (36), suggesting a role for XPF in substrate recognition by the heterodimeric ERCC1-XPF. Here, we analyzed the XPF DNA-binding preferences and show that it binds ssDNA in a non-sequence specific fashion but with a preference for substrates containing two ssDNA extensions (Figure 1). The homodimeric XPF can bind to Holliday junction substrate (Supplementary Figure S2) and has a preference for guanine-rich sequences (Figure 3), but does not have binding preference for telomere sequences (Supplementary Figure S3). NMR titrations revealed that XPF irrespective whether present as homo- or heterodimeric complex binds ssDNA sequence using the same surface (Figure 4). Importantly, ssDNA probes do not bind to the previously determined dsDNA-binding surface. By using various splayed-arm probes we show that the two nucleic-acid binding surfaces of ERCC1 and XPF within the heterodimeric protein can both be bound concurrently, using approximately the same interaction surfaces as for their preferred substrates (Figure 4). Kinetic experiments support the view that both binding surfaces are required for both specificity and binding affinity.

ssDNA binding by the HhH domain of XPF homodimers

In vitro DNA-binding experiments demonstrated a clear preference of XPF for ssDNA substrates (33). The binding profile as found in NMR titration experiments for a XPF homodimer with a 10 nt ssDNA fragment support this notion (36). The presence of two separate DNA binding domains in XPF facilitate high-affinity DNA binding to either long ssDNA substrates or probes containing two conformationally restricted ssDNA fragments. These data suggest that the two XPF-binding surfaces contact each one ssDNA molecule concurrently. Although all ssDNA sequences tested were suitable substrates for XPF, guanine stretches were more

effective in XPF recognition than thymidine or cytosine, while adenosine was a poor substrate for XPF (Figure 3). Thus DNA binding affinity and thereby the ability to repair damaged DNA is not entirely sequence-independent. Despite the absence of evidence *in vivo*, few *in vitro* studies suggest as well that cleavage is not completely sequence independent: cleavage of an identical sequence where an acetylaminofluorene adduct positioned at three distinct guanines led to significant variation in incision efficiency (41). Similar differences were found for benzo[a]pyrenyl-guanine lesion placed at various positions (42,43). In addition de Laat *et al* (31) presented previously evidence for sequence specificity by showing that splayed-arm substrates with distinct sequence composition around the junction were cleaved at distinct positions in the stem sequence arguing that the incision position is somewhat dictated by the DNA sequence. A recent study by Bowles et al (44) provides further support for differences in cleavage rate depending on the stem-loop sequence, although these studies show that the DEAH helicase-like domain is critically required for these effects. This all argues that the incision position is weakly dictated by the DNA sequence, we propose that this is related to the observed preference of XPF for G-rich sequences.

ss/dsDNA junction recognition by the HhH domains of ERCC1-XPF

Tsodikov et al (34) suggested that both HhH domains of ERCC1 and XPF contain ssDNA binding surfaces and that each specifically binds to one of the two arms of the DNA substrate. Our NMR titration studies using hairpin 22 (33), ssDNA (36), dsDNA and splayed-arm substrates (Figure 4 and Supplementary Figure S6) argue against such a model. We show that ssDNA is preferentially bound by XPF and not by ERCC1, while dsDNA substrates are specifically recognized by ERCC1 (Figure 4). ss/dsDNA containing probes make contact to both the XPF ssDNA-binding surface and the ERCC1-dsDNA-binding surface (Figure 4). The importance of these binding domains for recognition of ss/dsDNA sequences was confirmed by mutagenesis (Figure 5).

The crystal structures of the *A.pernix* XPF bound to DNA provide detailed insight into incision by XPF in the archaea bacteria (39). Extrapolation of this structural information to eukaryotic repair factors is complicated however, considering the distinct subunit composition and the different substrate specificities of both complexes. *In vitro* cleavage assays reveal that archaeal XPF has preference for flap substrates, while ERCC1-XPF most effectively processes splayed-arm substrates (31,39). How this heterodimeric protein complex recognizes ss/dsDNA junctions is an unanswered question. Our previous DNA-binding studies revealed that the HhH domains of ERCC1-XPF show a similar substrate specificity as the native complex, showing that the HhH domain region is required and sufficient for substrate recognition (33). Using site directed mutagenesis (Figure 5), SPR (Supplementary Figure S7) and NMR spectroscopy (Figure 4) we demonstrate that a dsDNA-binding site is located near the tip of the two hairpin structures in ERCC1. This conserved

DNA-binding surface is similar to that in other HhH domain proteins (45,46) including that in archaeal XPF (39). For the archaeal XPF homodimer it was proposed that the HhH domains of the two monomers bind the two dsDNA sequences of a flap substrate (39). The structural homology suggests functional similarity supporting the notion that ERCC1 recognizes dsDNA. Using the proposed model based on the *A. pernix* XPF (39), and knowing the polarity of the ERCC1-XPF heterodimer with respect to the damage, it can be expected that ERCC1 binds to the upstream dsDNA sequence placing the catalytic domain of XPF in close proximity to cleave the damaged DNA strand (39,47). This model does not provide an explanation for the substrate preference; however, the ssDNA-binding surface of XPF can contribute to this specificity.

We propose that ERCC1 binds dsDNA and XPF ssDNA, thereby positioning the ERCC1-XPF complex at the ss/dsDNA junction. This does not exclude that also other regions of ERCC1-XPF or other repair proteins can further substantiate substrate specificity. Indeed we and others noted that also the central domain of ERCC1 contains an ssDNA binding surface (47,48). In addition, XPA and RPA, that bind to ssDNA, and also interact with respectively ERCC1 (48) and XPF (49), can contribute to ERCC1-XPF positioning on the damaged DNA. The presence of multiple weak DNA-binding surfaces within this DNA repair-complex together facilitates the correct positioning of the nuclease domain with respect to the damage and prevents inappropriate DNA-binding and incision. Additional support for this model comes from a recent study by Su et al (50) which shows that mutations of the individual DNA binding domains in full length ERCC1 and XPF lead both *in vitro* and *in vivo* to a decrease in cleavage efficiency.

XPF HhH domain binds to the non-damaged strand

To determine which strand is bound by XPF, we modeled the ssDNA sequence into the ERCC1-XPF heterodimer structure based on the previously determined solution structure of homodimeric XPF bound to ssDNA (36). The dsDNA is positioned based on homology with the archaeal XPF-DNA structure (39,47). Assuming that the proposed models for dsDNA and ssDNA binding to ERCC1-XPF are correct, the gap between the dsDNA and the ssDNA can be filled by connecting the dsDNA fragment to either the 5' or 3' end of the ssDNA. As a result XPF can bind the damaged or the non-damaged strand, respectively. If we assume that XPF binds the non-damaged strand, the most conserved residues are located near the 3' end of the ssDNA including the second distorted hairpin and several residues in between the ssDNA and dsDNA binding region (T240, N246). Chemical shift changes for several of these surrounding residues were observed upon addition of a splayed arm substrate (e.g. K243, M856). Also the significant decrease in DNA binding by the Q838A/D839A ERCC1-XPF mutant and the overall positive charge of this surface argue that the 3´end of the ssDNA (the non-damaged strand) is connecting to dsDNA (Figure 6 upper panel). On the other hand

if the dsDNA would connect to the 5´end of the ssDNA, the distance to dsDNA would be substantially larger. Furthermore this region is poorly conserved, only small chemical shift perturbations are found upon addition of splayed arm substrate in the region between the two DNA binding domains and only limited effects on binding affinity by mutagenesis in this region was found. Combined with the overall negative charge for this side of the ERCC1-XPF molecule (Figure 6 lower panel) this would make the connection between the dsDNA and the 5' end of the ssDNA (being the damaged strand in our model) *via* this side unlikely.

By combining the subtle CSP differences for the two splayed-arm substrate with the substrates containing one ssDNA strand (Figure 4, Supplementary Figure S6), the effect of mutagenesis on the ability of ERCC1-XPF to bind DNA (Figure 5), and the charge and the sequence conservation (Figure 6), we propose that XPF recognizes the non-damaged strand. This agrees well with the previously reported binding and incision preference for full length ERCC1-XPF in the presence of RPA (51).

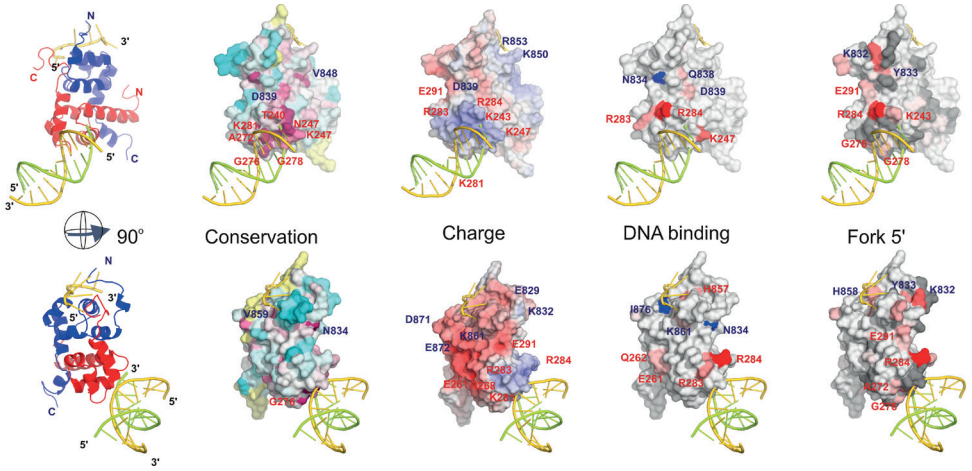


Figure 6. **Model for binding of ERCC1-XPF HhH domains to an ss/dsDNA junction.** The left panel shows a model for ssDNA and dsDNA binding to respectively XPF (blue) and ERCC1 (red) based on structural models for archaeal XPF bound to dsDNA (2BGW) (39) and human XPF bound to ssDNA (2KN7) (36). The other panels show surface representations plotting sequence conservation, binding interface, effect of mutations on binding and charge distribution. Conservation: colored according to sequence conservation calculated using the Consurf server (60), based on the complete multiple sequence alignment for the eukaryotic repair proteins, plotted according to the default coloring: most conserved red, least conserved cyan. Fork 10: colored according to chemical shift perturbation by a Fork substrate (as shown in Figure 4). Mutation: colored according to the contribution of a residue to the ability of the ERCC1-XPF heterodimeric HhH domain to bind to B10 DNA when mutated. Most affected red, least affected light red. Residues that could be mutated without significant effect on binding are depicted in blue. Charge: colored according to electrostatic surface potential calculated using the APBS software (61) (blue positive, red negative).

Role of XPF in sequence-dependent incision

Following damage recognition by XPC, acting as a damage sensor by recognizing helical distortions on the non-damaged strand (52), the ATP-dependent DNA unwinding by TFIIH (53), creates a DNA topology suitable for binding of RPA and XPA to the non-damaged and damaged DNA respectively (54). Binding of these proteins further opens the damaged DNA and serves through multiple interactions as a platform for XPG and ERCC1-XPF that subsequently perform the 3' and 5' incisions, respectively (55,56). This well orchestrated cleavage process results both *in vitro* and *in vivo* in the removal of 24 to 32 nucleotides. The substantial variation in both cleavage position with respect to the damage and the length of the removed sequence suggests some heterogeneity in the cleavage mechanism (5,57). In analogy to the archaeal XPF-DNA complex, we propose a model for the positioning of the ERCC1-XPF protein at the ss/dsDNA junction. The previously determined solution structure of the XPF-ssDNA complex explains preference for ssDNA over dsDNA, which is clearly distinct from the dsDNA preference of ERCC1. The noted preference of XPF for G-rich sequences (Figure 3), which is consistent with the structure of the XPF ssDNA complex (36) may dictate the binding of the ERCC1-XPF complex. The presence of one or a few specific nucleotides within the accessible ssDNA sequence (of the non-damaged strand) can in turn, through the positioning of ERCC1-XPF on the DNA, determine where cleavage will occur (31). We propose that both the DNA sequence-dependent differences in cleavage efficiency and the heterogeneity in the cleavage position by the ERCC1-XPF complex are the result of the deoxyguanosine preference of the ssDNA-binding domain of XPF.

Material and Methods

Protein expression and purification

The HhH domain of the ERCC1-XPF heterodimer was expressed and purified as described before (33). The homodimeric XPF HhH domain expression and purification have been described before (35). The ERCC1-XPF mutants were prepared using the QuikChange protocol (Stratagene) and expressed as the wild-type ERCC1-XPF complex (33).

Electrophoretic Mobility Shift Assay

EMSA experiments were performed as described before (33,58) using the radio labeled bubble 10 probe or Holliday junction as substrate in a buffer containing 10 mM Tris pH 7.5, 100 mM NaCl, 10 % glycerol, 1 mM DTT and BSA (final concentration 20 μg/ml). All oligonucleotides were purchased from Operon or Eurogentec and annealed by incubating the 2 (or 4) mixed strands (final concentration: 50 μM) for 5 minutes at 95 °C followed by a cooling step for one hour in a solution containing 10 mM Tris pH 8.0 and 100 mM NaCl. For the competition experiments the indicated amount of competitor (Supplementary Figure S1) and the radio labeled probe were mixed in a tube and subsequently the protein

containing solution (~1 μM) was added to this mixture. After incubation for 30 minutes on ice, samples were loaded on a 0.5x TBE buffered 5% acrylamide gel and electrophoresis was carried out for 2.5 hours at 160 V at room temperature. Analysis and quantification was performed as described before (58).

Surface plasmon resonance measurements

SPR experiments were performed in 10 mM Hepes, pH 7.5, 50mM NaCl and 0.005% (w/v) Tween20 (SPR buffer) at 10 μl/min at 12°C using a Biacore® X system (Biacore AB) (59). ERCC1-XPF HhH domain was dialyzed to the SPR buffer using Zeba Desalt Spin Columns (Thermo Scientific). Low-binding tubes and tips were used to prevent the loss of the sample during the incubations and dilutions. Before each experiment, 5 μl of 0.3 M Ni^{2+} was loaded on flow cell 2 of the NTA sensor chip (Biacore AB) while the flow cell 1 was used as a reference surface. 50 nM of ERCC1-XPF HhH domain was incubated in the SPR buffer on ice for 20 min, in the absence or presence of different oligonucleotides (concentration ranging from 0.01 to 100 μM). Then, it was injected over the NTA sensor chip, followed by the association for 60 s and the dissociation for 120 s. The flow cell 1 (without Ni^{2+}) baseline curves were subtracted from the flow cell 2 experimental curves using the Biaevaluation 3.2 software. Between consecutive injections, the chip was regenerated with 10 μl of 0.25 M EDTA in 3.5 M guanidium, pH 8. All the experiments were performed at least in duplicate.

For ERCC1 DNA binding studies, 6.25 μM of ssDNA or dsDNA was loaded on the immobilized ERCC1 after the dissociation of the XPF from the ERCC1-XPF complex. Since the addition of DNA prevented the binding of the His-tagged ERCC1-XPF HhH domain on the Ni-NTA surface, the relative amount of (DNA-free) protein (F) was determined as a response values at the end of the loading at the time 60 s (R_{60}) divided by the R_{60} value of the ERCC1-XPF HhH domain in absence of DNA. To calculate the apparent dissociation constant (K_D^{app}) for binding of each oligonucleotide to the ERCC1-XPF HhH domain the relative amount of the DNA-free protein (F) was fitted against the total oligonucleotide concentration and fitted according to a simple 1:1 model of interaction using GraphPad Prism: (1-F) = [DNA]/(K_D^{app} + [DNA]), where F represents the relative amount of unbound ERRC1-XPF and K_D^{app} the apparent equilibrium binding constant.

NMR experiments

NMR titrations were followed by recording ^{15}N-^{1}H HSQC spectra of ERCC1-XPF by adding small volumes of a concentrated solution of commercially purchased DNA oligonucleotides (Eurogentec or Operon). The ^{15}N labeled ERCC1-XPF protein and unlabeled DNA were dissolved in the same buffer containing 5-50 mM sodium phosphate buffer and ~25-100 mM NaCl at pH 7.0. All NMR data were collected at 22 °C on a Bruker DRX600 spectrometer equipped with a z-gradient triple-resonance cryoprobe or a Bruker Avance

900 spectrometer equipped with a 5 mm z-gradient triple-resonance probe. A set of ^{15}N-^{1}H HSQC spectra was acquired with the successive addition of ssDNA, dsDNA and splayed-arm DNA substrates to 40-100 μM ^{15}N labeled ERCC1-XPF. The NMR data were processed and analyzed as described before (33). To compare the chemical shift changes on the DNA backbone of a 10 nt ssDNA fragment (36) upon addition of protein, ^{1}H decoupled 1D ^{31}P spectra of the free and bound ssDNA were acquired on a Bruker DRX500 spectrometer equipped with a QXI probe.

ACKNOWLEDGEMENTS

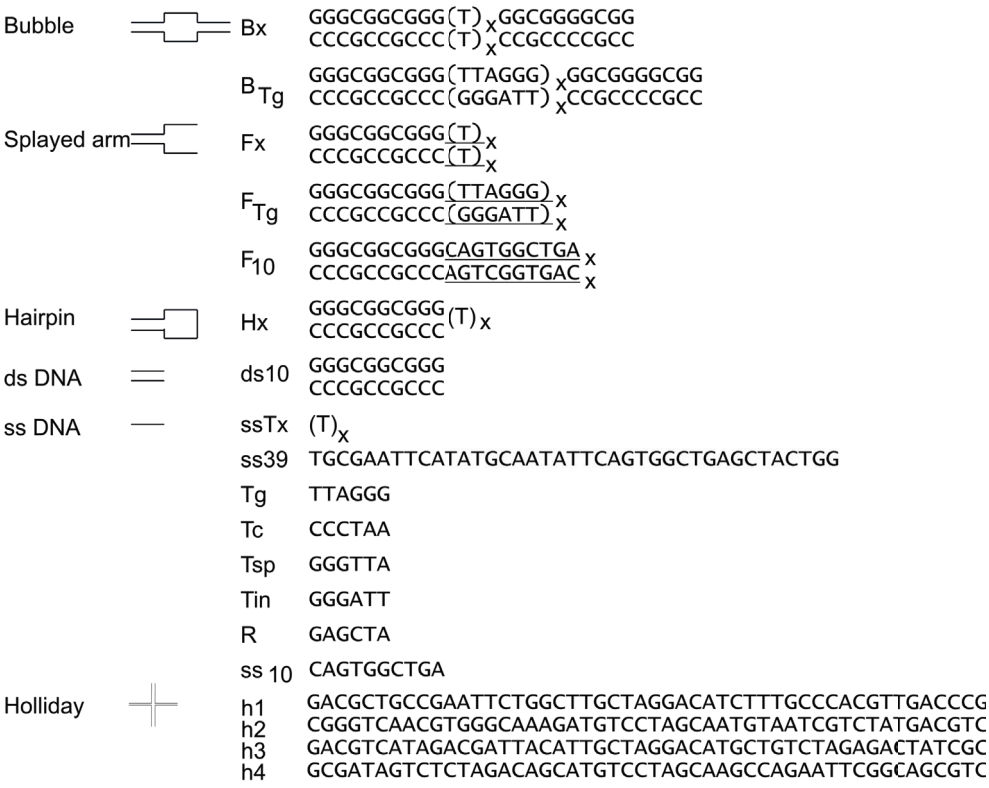
The authors thank Annet Vliegenthart, Jolinde Soesbergen, Mehmet Demirci and Donal van Uunen for their various contributions during the project. This work was financially supported by the Netherlands Foundation for Chemical Research (NWO-CW), the Center for Biomedical Genetics (CBG), the European Commission (project 031220/Spine2-complexes and project 261863/Bio-NMR), the Slovenian Infrastructural center for analysis of molecular interaction, and the Slovenian Research Agency (Z1-4071).

REFERENCES

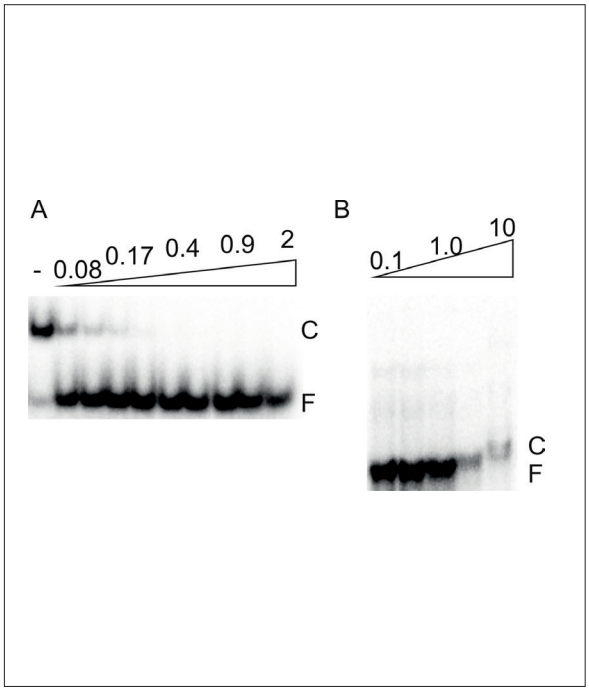
1. Sancar, A. (1996) *Annual Review of Biochemistry* **65**, 43-81
2. Cleaver, J. E., and Crowley, E. (2002) *Front Biosci* **7**, d1024-1043
3. Friedberg, E. C. (2003) *Nature* **421**(6921), 436-440
4. Hanawalt, P. C. (2002) *Oncogene* **21**(58), 8949-8956
5. Araujo, S. J., and Wood, R. D. (2000) *Mutation Research-DNA Repair* **459**(2), 171-172
6. Cleaver, J. E., Thompson, L. H., Richardson, A. S., and States, J. C. (1999) *Human Mutation* **14**(1), 9-22
7. Cleaver, J. E. (2005) *Nat Rev Cancer* **5**(7), 564-573
8. Cleaver, J. E., and Revet, I. (2008) *Mech Ageing Dev*
9. 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., Theil, A. F., Vermeulen, W., van der Horst, G. T. J., Meinecke, P., Kleijer, W. J., Vijg, J., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (2006) *Nature* **444**(7122), 1038-1043
10. 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., Rademakers, S., de Rooij, J., Jaspers, N. G. J., Hoeijmakers, J. H. J., and Wood, R. D. (1996) *Cell* **86**(5), 811-822
11. Bessho, T., Sancar, A., Thompson, L. H., and Thelen, M. P. (1997) *J Biol Chem* **272**(6), 3833-3837
12. de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G., and Hoeijmakers, J. H. (1998) *Nucleic Acids Res* **26**(18), 4146-4152
13. Nishigori, C., Ishizaki, K., Takebe, H., Imamura, S., and Hayakawa, M. (1986) *Arch Dermatol* **122**(5), 510-511
14. Gaillard, P. H., and Wood, R. D. (2001) *Nucleic Acids Res* **29**(4), 872-879
15. Sijbers, A. M., van der Spek, P. J., Odijk, H., van den Berg, J., van Duin, M., Westerveld, A., Jaspers, N. G. J., Bootsma, D., and Hoeijmakers, J. H. J. (1996) *Nucleic Acids Res* **24**(17), 3370-3380
16. Hayashi, T., Takao, M., Tanaka, K., and Yasui, A. (1998) *Mutation Research-DNA Repair* **407**(3), 269-276
17. Yagi, T., Wood, R. D., and Takebe, H. (1997) *Mutagenesis* **12**(1), 41-44
18. Tian, M., Shinkura, R., Shinkura, N., and Alt, F. W. (2004) *Mol Cell Biol* **24**(3), 1200-1205
19. 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. J. (1997) *Curr Biol* **7**(6), 427-439
20. McWhir, J., Selfridge, J., Harrison, D. J., Squires, S., and Melton, D. W. (1993) *Nat Genet* **5**(3), 217-224
21. Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H., and Niedernhofer, L. J. (2008) *Mol Cell Biol* **28**(16), 5082-5092
22. Al-Minawi, A. Z., Saleh-Gohari, N., and Helleday, T. (2008) *Nucleic Acids Research* **36**(1), 1-9
23. Kuraoka, I., Kobertz, W. R., Ariza, R. R., Biggerstaff, M., Essigmann, J. M., and Wood, R. D. (2000) *J Biol Chem* **275**(34), 26632-26636
24. Fisher, L. A., Bessho, M., and Bessho, T. (2008) *J Biol Chem* **283**(3), 1275-1281
25. Zhu, X. D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H. J., and de Lange, T. (2003) *Mol Cell* **12**(6), 1489-1498
26. Munoz, P., Blanco, R., Flores, J. M., and Blasco, M. A. (2005) *Nat Genet* **37**(10), 1063-1071
27. Niedernhofer, L. J., Essers, J., Weeda, G., Beverloo, B., de Wit, J., Muijtjens, M., Odijk, H., Hoeijmakers, J. H. J., and Kanaar, R. (2001) *Embo J* **20**(22), 6540-6549
28. Fekairi, S., Scaglione, S., Chahwan, C., Taylor, E. R., Tissier, A., Coulon, S., Dong, M. Q., Ruse, C., Yates, J. R., 3rd, Russell, P., Fuchs, R. P., McGowan, C. H., and Gaillard, P. H. (2009) *Cell* **138**(1), 78-89
29. Svendsen, J. M., Smogorzewska, A., Sowa, M. E., O'Connell, B. C., Gygi, S. P., Elledge, S. J., and Harper, J. W. (2009) *Cell* **138**(1), 63-77
30. Munoz, I. M., Hain, K., Declais, A. C., Gardiner, M., Toh, G. W., Sanchez-Pulido, L., Heuckmann, J. M., Toth, R., Macartney, T., Eppink, B., Kanaar, R., Ponting, C. P., Lilley, D. M., and Rouse, J. (2009) *Mol Cell* **35**(1), 116-127
31. de Laat, W. L., Appeldoorn, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998) *J Biol Chem* **273**(14), 7835-7842
32. Park, C. H., Bessho, T., Matsunaga, T., and Sancar, A. (1995) *J Biol Chem* **270**(39), 22657-22660

33. Tripsianes, K., Folkers, G., AB, E., Das, D., Odijk, H., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., and Boelens, R. (2005) *Structure* **13**(12), 1849-1858
34. Tsodikov, O. V., Enzlin, J. H., Scharer, O. D., and Ellenberger, T. (2005) *Proc Natl Acad Sci U S A* **102**(32), 11236-11241
35. Das, D., Tripsianes, K., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R., Boelens, R., and Folkers, G. E. (2008) *Proteins* **70**(4), 1551-1563
36. Das, D., Folkers, G. E., van Dijk, M., Jaspers, N. G., Hoeijmakers, J. H., Kaptein, R., and Boelens, R. (2012) *Structure* **20**(4), 667-675
37. Brookman, K. W., Lamerdin, J. E., Thelen, M. P., Hwang, M., Reardon, J. T., Sancar, A., Zhou, Z. Q., Walter, C. A., Parris, C. N., and Thompson, L. H. (1996) *Molecular and Cellular Biology* **16**(11), 6553-6562
38. Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2000) *Proc Natl Acad Sci U S A* **97**(15), 8257-8262
39. Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P. P., White, M. F., and McDonald, N. Q. (2005) *Embo J* **24**(5), 895-905
40. Bowles, M., Lally, J., Fadden, A. J., Mouilleron, S., Hammonds, T., and McDonald, N. Q. *Nucleic Acids Res*
41. Mu, D., Bertrand-Burggraf, E., Huang, J. C., Fuchs, R. P., Sancar, A., and Fuchs, B. P. (1994) *Nucleic Acids Res* **22**(23), 4869-4871
42. Cai, Y., Patel, D. J., Geacintov, N. E., and Broyde, S. (2007) *J Mol Biol* **374**(2), 292-305
43. Cai, Y., Patel, D. J., Geacintov, N. E., and Broyde, S. (2009) *J Mol Biol* **385**(1), 30-44
44. Bowles, M., Lally, J., Fadden, A. J., Mouilleron, S., Hammonds, T., and McDonald, N. Q. (2012) *Nucleic Acids Res*
45. Shao, X., and Grishin, N. V. (2000) *Nucleic Acids Res* **28**(14), 2643-2650
46. Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2005) *Structure* **13**(8), 1183-1192
47. Tripsianes, K., Folkers, G. E., Zheng, C., Das, D., Grinstead, J. S., Kaptein, R., and Boelens, R. (2007) *Nucleic Acids Research* **35**(17), 5789-5798
48. Tsodikov, O. V., Ivanov, D., Orelli, B., Staresincic, L., Shoshani, I., Oberman, R., Scharer, O. D., Wagner, G., and Ellenberger, T. (2007) *Embo Journal* **26**(22), 4768-4776
49. Fisher, L. A., Bessho, M., Wakasugi, M., Matsunaga, T., and Bessho, T. (2011) *J Mol Biol* **413**(2), 337-346
50. Su, Y., Orelli, B., Madireddy, A., Niedernhofer, L. J., and Scharer, O. D. (2012) *J Biol Chem* **287**(26), 21846-21855
51. de Laat, W. L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998) *Genes Dev* **12**(16), 2598-2609
52. Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., and Hanaoka, F. (2005) *Mol Cell Biol* **25**(13), 5664-5674
53. Coin, F., Oksenych, V., and Egly, J. M. (2007) *Molecular Cell* **26**(2), 245-256
54. Missura, M., Buterin, T., Hindges, R., Hubscher, U., Kasparkova, J., Brabec, V., and Naegeli, H. (2001) *Embo Journal* **20**(13), 3554-3564
55. Mocquet, V., Laine, J. P., Riedl, T., Yajin, Z., Lee, M. Y., and Egly, J. M. (2008) *Embo Journal* **27**(1), 155-167
56. Riedl, T., Hanaoka, F., and Egly, J. M. (2003) *Embo J* **22**(19), 5293-5303
57. de Laat, W. L., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1999) *Genes Dev* **13**(7), 768-785
58. de Jong, R. N., Truffault, V., Diercks, T., Ab, E., Daniels, M. A., Kaptein, R., and Folkers, G. E. (2008) *Structure* **16**(1), 149-159
59. Nieba, L., Nieba-Axmann, S. E., Persson, A., Hämäläinen, M., Edebratt, F., Hansson, A., Lidholm, J., Magnusson, K., Karlsson, Å. F., and Plückthun, A. (1997) BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Analytical Biochemistry* **252**(2), 217-228
60. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) *Nucleic Acids Res* **38**(Web Server issue), W529-533
61. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc Natl Acad Sci U S A* **98**(18), 10037-10041

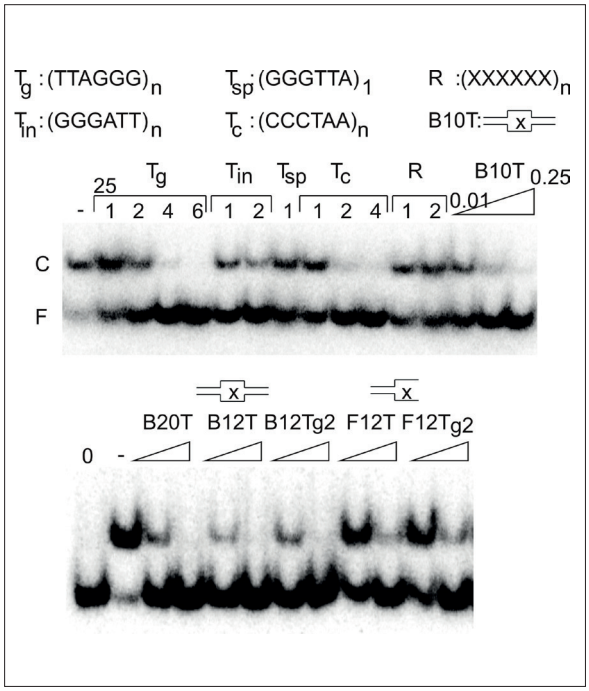
SUPPLEMENTAL INFORMATION



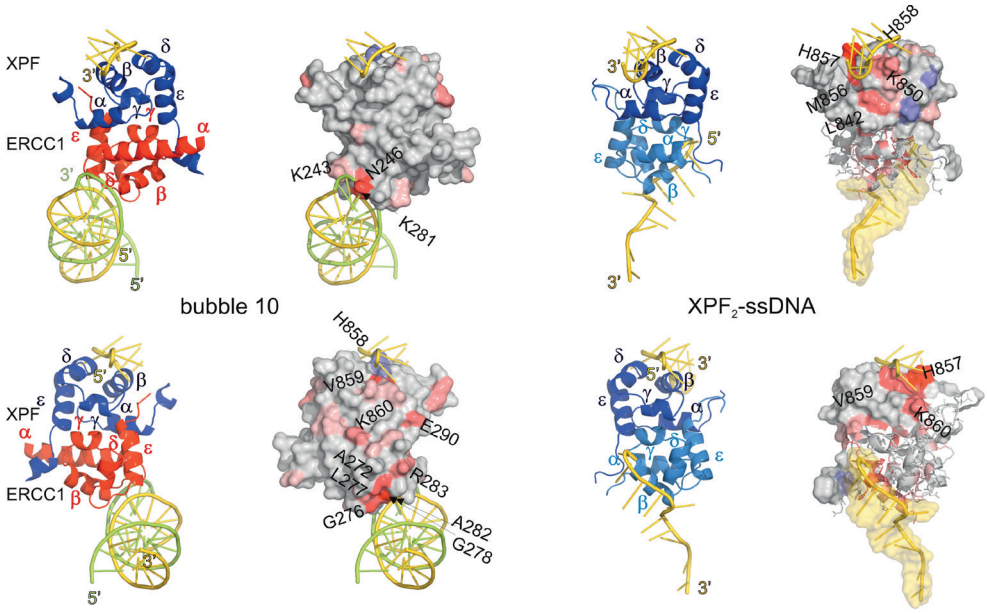
Supplementary Figure S1. Probes used for DNA binding to the HhH domain of XPF. The figure shows the names, symbols and abbreviations and sequences of the various probes used in this study. The dsDNA probes are depicted as duplexes where the sequence of the upper strand is given 5' to 3' and the lower strand 3' to 5'. All ssDNA sequences and the 4 strands that together form the Holliday junction are all given 5' to 3'. ss_{10} refers to the DNA sequence used for the structure determination of the homodimeric HhH domain XPF structure bound to ssDNA and in NMR titration experiments (36).



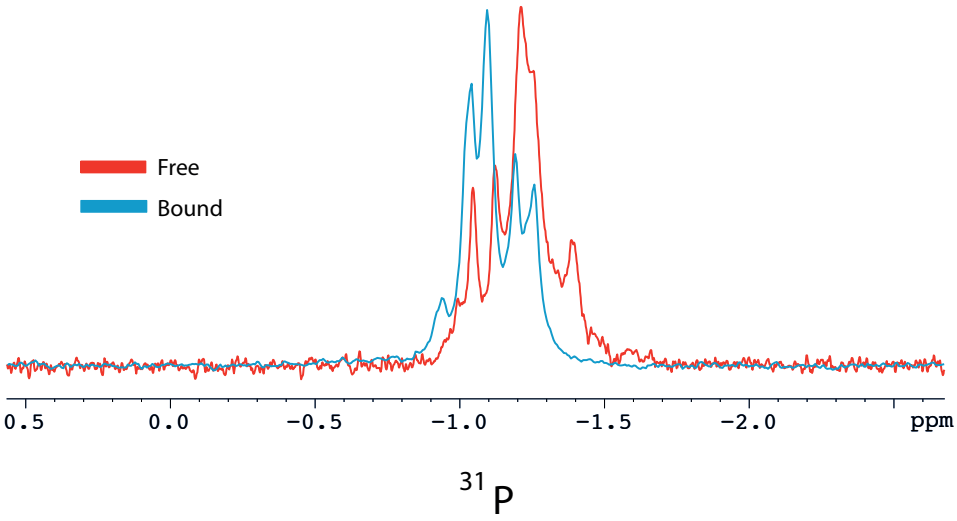
Supplementary Figure S2. The HhH domain of XPF binds to Holliday junction substrate. *A,* Competition of XPF binding (1.25 μM) to 0.02 μM B10 substrate using non-labeled Holliday junction substrate (μM) as described in Figure 1C. *B,* Binding of XPF to gel purified Holliday junction substrate, concentrations indicated above in μM.



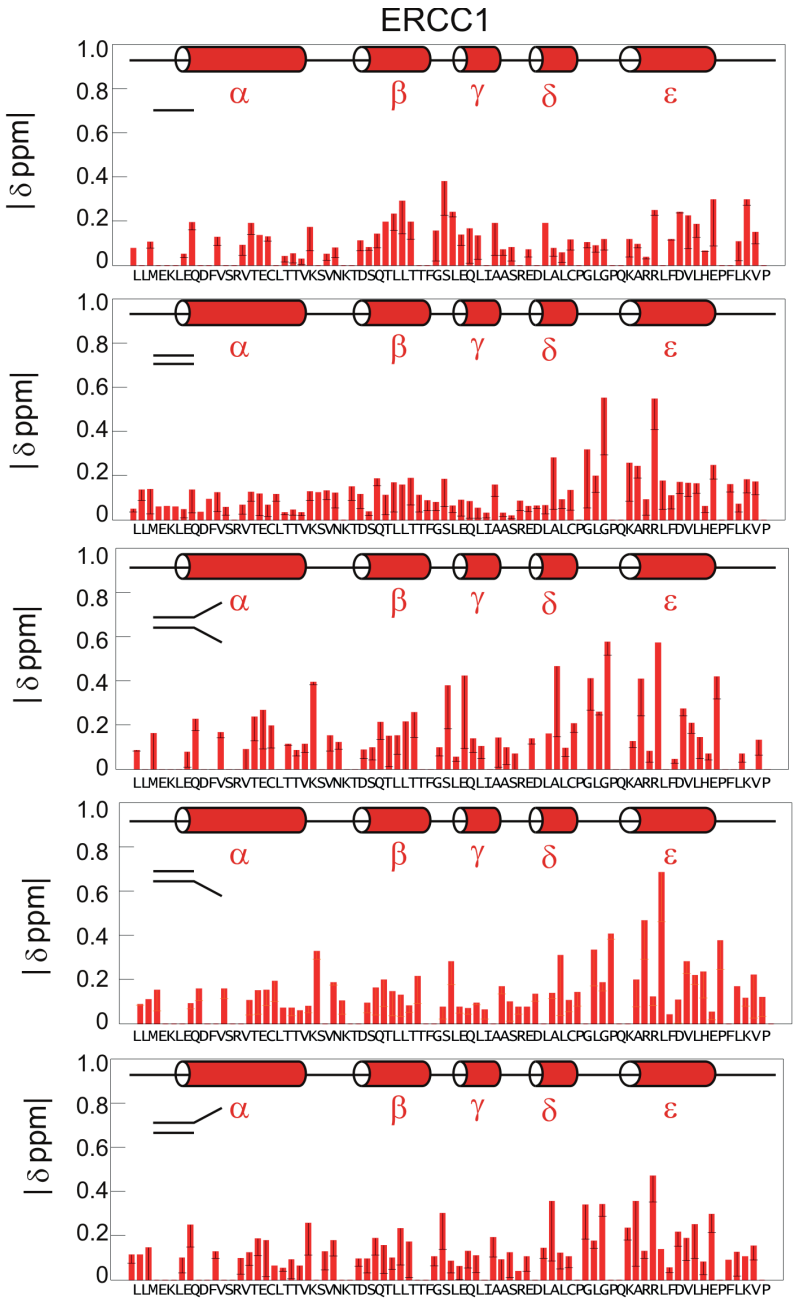
Supplementary Figure S3. The HhH domain of XPF binds non-specifically to telomere sequences. The competition assay is as described in Figure 1, with 25 μM of the G-rich telomere probe (T_G: TTAGGG), the inverted telomere probe (T_{in}: GGGATT), a split telomere probe (T_{sp}: GGGTTA), the C-rich telomere sequence (T_C: CCCTAA) or a randomized sequence (R: GAGCTA) of the indicated number of repeats. For comparison competition by 0.01, 0.05 and 0.25 μM B10 substrate is shown. B) Competition by 0.02 and 0.2 μM bubble (B) or splayed arm (F) substrates containing either 12 thymines (12T) or 2 copies of the G-rich telomere sequence (T_{G2}: $TTAGGG_2$). For comparison competition by equivalent amounts of B20T is shown.



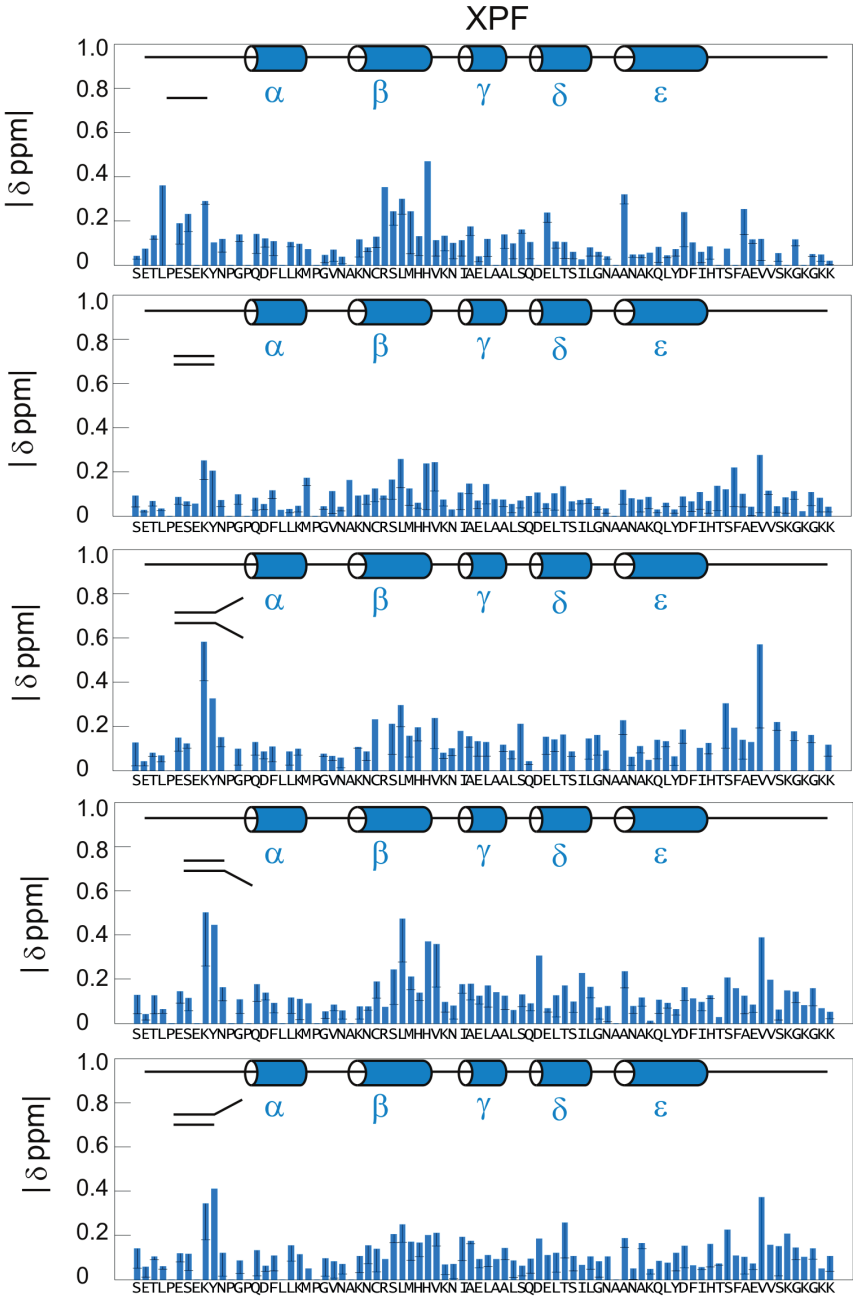
Supplementary Figure S4. Previously identified dsDNA and ssDNA interfaces in ERCC1 and XFP. CSP results as described before for hairpin 20 binding to the HhH domain of the ERCC1-XPF heterodimer (33) and ssDNA to the HhH domain of the XPF homodimer (35) plotted on the surface as described in Figure 4.



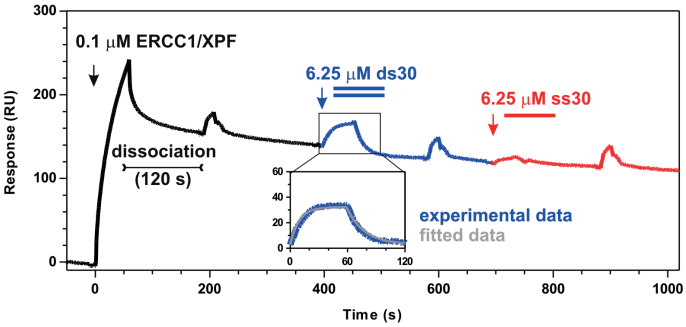
Supplementary Figure S5. Phosphorous NMR spectra show binding of ERCC1-XPF to various DNA substrates. A comparison of proton decoupled 1D ^{31}P spectra of the 10mer (CAGTGGCTGA) ssDNA phosphate backbone free (red) and bound (blue) to the ERCC1-XPF protein.



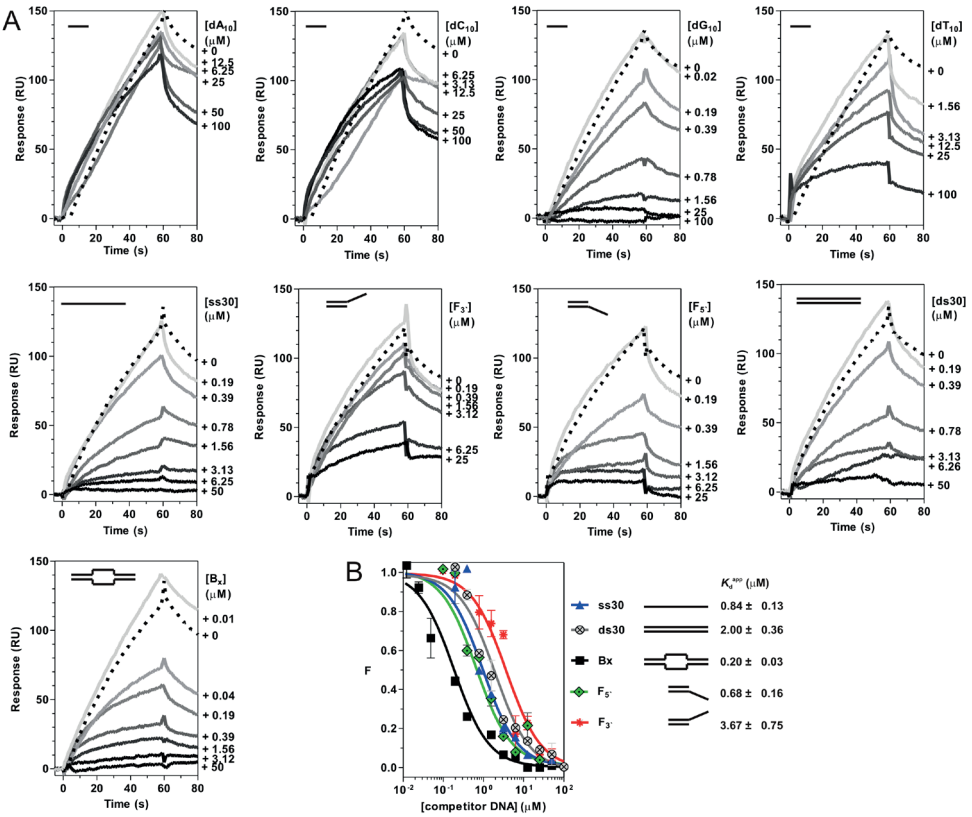
Supplementary Figure S6. Binding of ERCC1-XPF to DNA substrates containing combinations of ssDNA and dsDNA sequences. The figure shows normalized chemical shift changes for ERCC1 (red) and XPF (blue) by the addition 3-4 fold excess of DNA (details on sequences used is presented in the Supplementary Table 1) to 25-100 μM ERCC1-XPF complex under various salt and buffer conditions. The previously determined secondary structure elements are depicted above. Missing bars are either proline residues or residues that could not be unambiguously assigned due to (exchange) line broadening or peak overlap.



Supplementary Figure S6. (Continued) Binding of ERCC1-XPF to DNA substrates containing combinations of ssDNA and dsDNA sequences.



Supplementary Figure S7. Substrate preference of ERCC1 for dsDNA established by SPR. On the graph the experimental curve of sub-sequential loading of (black arrow) 0.1 μM ERCC1-XPF complex to the Ni^{2+}-loaded NTA SPR chip, followed by XPF dissociation (dissociation), loading (blue arrow) and dissociation of ds30 probe and (red arrow) loading of ss30 probe is shown. From the ds30 experimental curve the ds30 non-specific binding to the chip surface was subtracted, fitted according to the Langmuir 1:1 binding model and shown as an inset.



Supplementary Figure S8. ERCC1-XPF HhH domain binding to the various ssDNA or dsDNA probes followed by SPR competition experiments. *A,* The representative sensorgrams of 50 nM ERCC1-XPF HhH domain in an absence or a presence of various DNA probes were measured on the NTA SPR chip. *B,* In order to calculate the apparent DNA binding affinity (K_D^{app}) the response values at the end of the loading (R_{60}) were divided by the R_{60} value of the ERCC1-XPF HhH domain in absence of DNA and plotted against the total concentration of the DNA and fitted considering a 1:1 binding model.

Supplementary Table 1. DNA sequences used for NMR titration experiments. The table shows the various DNA probes used in the NMR titration experiments and the corresponding sequences used to prepare these. The ERCC1-XPF HhH domain protein concentration is indicated. The molar excess DNA used for final evaluation of the chemical shift differences; the NaCl concentration and the phosphate buffer concentration for each probe is indicated.

DNA	Sequence	Sequence	[ERCC1-XPF] (μM)	Molar excess	[NaCl] (mM)	[$NaHPO_4$] (mM)
ssDNA	CAGTGGCTGATT		100	4	10	5
	CAGTGGCTGATT		40	2.5	25	12.5
	CAGTGGCTGATT		100	4	50	25
dsDNA	gggcggcgggaatcagccactgcc	ggcagtggctgattcccgccgccc	150	3	100	50
	ggcggggcgggggcggcggg	cccgccgcccccgccccgcc	150	2	100	50
	gggcggcggg	cccgccgccc	40	2.5	25	12.5
	gggcggcggg	cccgccgccc	80	4	100	5
	gggcggcggg	cccgccgccc	100	4	50	25
Fork	gggcggcgggTTTTTTCAGTGGCTGA	CAGTGGCTGATTTTTTcccgccgccc	40	2.5	25	12.5
	gggcggcgggTTTTTTCAGTGGCTGA	CAGTGGCTGAcccgccgccc	80	4	100	5
	gggcggcgggTTTTTTCAGTGGCTGA	CAGTGGCTGAcccgccgccc	100	4	50	25
Fork 3'	ggcggggcgggggcggcgggTTCAGTGGCTGA	cccgccgccccgccccgcc	150	2	100	50
	gggcggcgggTTTTTTCAGTGGCTGA	cccgccgccc	40	2.5	25	12.5
	gggcggcgggTTTTTTCAGTGGCTGA	cccgccgccc	80	4	100	5
	gggcggcgggTTTTTTCAGTGGCTGA	cccgccgccc	100	4	50	25
Fork 5'	CAGTGGCTGATTcccgccgcccccgccccgcc	ggcggggcgggggcggcggg	150	2	100	50
	CAGTGGCTGATTTTTTcccgccgccc	gggcggcggg	40	2.5	25	12.5
	CAGTGGCTGAcccgccgccc	gggcggcggg	80	4	100	5
	CAGTGGCTGAcccgccgccc	gggcggcggg	100	4	50	25

Chapter 5

Summary

A multitude of events provides the energy and molecules to build and maintain the cell. In this, proteins perform an important function. They can function as enzymes that catalyze chemical reactions in and around a cell, they can be involved in cell mobility or molecular transport through membranes, they can regulate gene expression, create protection, or repair defects. The amazing versatility of protein function highly depends on their structure and their interactions to other molecules. Consequently, a small change in structure, interaction specificity or affinity can disturb the balance in protein function, thus changing the fate of the cells and frequently leading to disease.

The proteins in the cell are encoded by DNA, the carrier of genetic information, and it is essential to maintain the genomic code. Despite all kinds of regulatory precautions DNA damage occurs and it occurs continuously and frequently. In fact, the changes in DNA sequence are even the driving force for evolution and for the diversity of living organisms, but it can also lead to disease and loss of genetic information. To ensure the integrity and stability of the genome, cells have developed a number of DNA repair machineries. One of these is the human Nucleotide Excision Repair machinery that, for example, can repair DNA damage caused by UV light. The heterodimeric complex ERCC1-XPF is part of this machinery.

The two proteins, ERCC1 and XPF that associate as a heterodimeric complex, are partially homologous. As a complex ERCC1-XPF functions as a structure-specific endonuclease. It performs an incision at single strand/double strand DNA junctions on the 5'-side of damaged and distorted DNA. Though ERCC1-XPF is best known for its function in Nucleotide Excision Repair, it has also been linked to several other pathways such as DNA Interstrand Cross-Link repair. Much of our knowledge on ERCC1-XPF comes through observations in diseases that its malfunction has triggered. Interest in understanding the role of ERCC1-XPF in DNA repair has recently increased due to the link found in cancer patients between resistance to chemotherapy and ERCC1-XPF expression during tumor treatment.

In the research described in this thesis a variety of spectroscopic, functional and biophysical techniques have been used to answer some of the unresolved questions with respect to the structure, function and stability of the human ERCC1-XPF complex.

Chapter 1 provides a comprehensive review of the current knowledge about the human heterodimeric ERCC1-XPF complex, including the DNA repair pathways it is involved in, the genetic diseases that link it to these DNA repair pathways, and a description of the structural motifs and functional domains of the proteins forming the complex.

The heterodimeric interaction regions of ERCC1 and XPF lie in the C-termini of the two proteins. The proteins contact each other by a core interaction interface and pseudosymmetric C-terminal phenylalanine anchors that are locked in N-terminally located

cavities of the partner protein. In **Chapter 2**, structural and functional consequences of a point-mutation in the interface of the ERCC1-XPF heterodimer are described. Until now, only two patients have been reported who carry this mutation in either one allele or both alleles, and both suffered from severe disorders and died in early life. The F231L mutation is positioned on the tandem Helix-hairpin-Helix $(HhH)_2$ domain of ERCC1 and localizes at its hydrophobic interaction interface with XPF. Comparison of the NMR structures of the wildtype and mutant protein complexes showed that while F231 is part of ERCC1's N-terminal cavity, its replacement to leucine disrupts this cavity and alters the orientation of the important interaction anchor residue F894 of XPF. Further biophysical analysis indicates that these structural alterations reduce the stability of the mutant ERCC1-XPF complex. Previously, it has been observed that *in vivo* and *in vitro* ERCC1 is not stable in its monomeric state. The observed accelerated heterodimer dissociation could diminish the cellular concentration of the heterodimeric complex, which agrees with the impairment of DNA repair and the accumulation of DNA repair intermediates that cause the phenotype of this mutation.

In vitro, XPF can form homodimers and these are found to be more stable than the ERCC1-XPF heterodimer. However, *in vivo* and *in vitro* the formation of the ERCC1-XPF $(HhH)_2$ heterodimer is preferred. In **Chapter 3**, we report that the dissociation of the complex into its monomeric components is very fast. Actually, the dissociation is so fast that ERCC1-XPF heterodimer should not exist *in vitro*. The fact that ERCC1-XPF is quite stable suggests that these monomers rapidly re-associate as a heterodimer. It was observed that ERCC1 secondary structure elements have lower stability than those of XPF. Besides that, when ERCC1 unfolds at conditions such as elevated temperatures, it self-aggregates and precipitates, while XPF remains stable until very high temperatures. Upon ERCC1 precipitation at high temperature, XPF monomers become available for homodimer formation. *In vivo*, XPF homodimerization must be slower than the re-association of monomeric XPF with free monomeric ERCC1. Since homodimeric XPF has a structure different from XPF in the heterodimer, a reason for the undetected XPF homodimer can either be that the rate-limiting step for its formation is the interconversion rate between these two folds, or that the homodimeric XPF fold corresponds to a less populated state. From the hydrogen deuterium exchange rates of homodimeric and heterodimeric XPF as monitored by NMR we conclude that homodimeric XPF adapts a fold with lower intrinsic stability as compared to its heterodimeric form. This supports the hypothesis that the XPF fold in the homodimer corresponds to a less populated state. In conclusion, a correct fold of XPF is essential for its association with ERCC1 into a functional ERCC1-XPF complex.

A functional property of the ERCC1-XPF complex is its capacity to bind near damaged DNA. The preferential binding to single strand/double strand DNA junctions comes from the C-terminal $(HhH)_2$ domains of ERCC1-XPF. This mediates the incision function of the XPF endonuclease domain at the 5'-side of the damaged strand. Earlier chemical shift

perturbation studies have shown that the two hairpins of ERCC1 are able to interact with a 22-nucleotide dsDNA substrate and that ERCC1 experiences small local conformational changes to accommodate this substrate. However, in these earlier studies no role in DNA recognition could be found for the $(HhH)_2$ domain of XPF. Therefore, it was initially thought that XPF could not bind DNA, perhaps due to the short second hairpin of its second HhH motif. Later, it was shown that the $(HhH)_2$ motif of homodimeric XPF was able to bind to ssDNA. It appears that in the first HhH motif of homodimeric XPF charged amino acid sidechains of the second helix interact with a 10-nucleotide ssDNA phosphate backbone. In fact, a guanine base located at the fifth nucleotide is accommodated into a pocket formed by residues from the two HhH motifs of homodimeric XPF, suggesting even specificity of XPF for guanine bases. Therefore, an actual model of ERCC1-XPF binding to ss/dsDNA junctions lies in the dual DNA binding capacity of ERCC1 and XPF.

In **Chapter 4,** we further examined the DNA binding preferences of the XPF homodimer and present experimental proof for the current model for DNA binding of the ERCC1-XPF heterodimer. We show that the homodimeric XPF $(HhH)_2$ domain binds preferably to guanine-rich ssDNA and ss/dsDNA Holliday junction sequences. Moreover, it is shown that the XPF $(HhH)_2$ domain in the ERCC1-XPF heterodimer uses the same surface as XPF in the homodimer to bind ssDNA, whereas the ERCC1 surface with previously determined affinity for dsDNA shows no affinity for these ssDNA substrates. To investigate whether all DNA binding properties of ERCC1 and XPF could converge into a single model for binding of heterodimeric ERCC1-XPF to ss/dsDNA junctions and whether this binding is sequence-specific or not, we analyzed the affinities of the ERCC1 and XPF DNA binding surfaces simultaneously by using various Y-shaped DNA substrates. It was found that both binding surfaces are essential for binding affinity as well as specificity. Based on the orientation of the ssDNA sequence in the solution structure of XPF homodimer bound to ssDNA, and assuming that ERCC1 binds dsDNA as its structural homologs RuvA and archaeal XPF, we propose that the HhH domain of XPF binds to the non-damaged single strand of a repair bubble DNA.

Taken together, the studies described in this thesis identified considerable dynamics in the ERCC1-XPF complex formation, demonstrated the effects of a disease-related interface mutation in ERCC1 on the structure and stability of the complex, and substantiated our model of ERCC1-XPF specificity for a ds/ssDNA junction, where ERCC1 binds dsDNA and XPF ssDNA.

Samenvatting

Een groot aantal processen draagt bij aan de energieproductie en de assemblage van moleculen betrokken bij de opbouw en het onderhoud van de cel. Eiwitten spelen hierin een belangrijke rol. Ze functioneren als enzymen die chemische reacties katalyseren in en om de cel, ze spelen een rol in celmobiliteit of moleculair transport door membranen, ze kunnen genexpressie reguleren, bescherming creëren of fouten herstellen. De indrukwekkende diversiteit van de functie van eiwitten is sterk afhankelijk van hun structuur en interacties met andere moleculen. Een kleine verandering in structuur, interactie-specificiteit of affiniteit van eiwitten kan dus het evenwicht in hun functie verstoren, en zodoende de ontwikkeling van cellen veranderen. Dit ligt vaak ten grondslag aan ziektes.

Eiwitten zijn in de cel gecodeerd door DNA, de drager van genetische informatie, en het is van essentieel belang om het genoom intact te laten. Ondanks veel voorzorgsmaatregelen, ontstaat DNA-schade en dit ontstaat continu en frequent. In feite zijn de veranderingen in DNA-sequentie zelfs de drijvende kracht voor evolutie en voor de grote verscheidenheid van levende organismen. Het kan echter ook leiden tot ziekte en verlies van genetische informatie. Om de integriteit en de stabiliteit van het genoom te bewaren, hebben cellen een aantal DNA-reparatieprocessen. Een van deze is het Nucleotide Excisie Reparatie (NER)-mechanisme van de mens dat, bijvoorbeeld, door UV-licht veroorzaakte DNA-schade kan herstellen. Het heterodimere complex ERCC1-XPF maakt deel uit van dit herstelmechanisme.

De twee eiwitten, ERCC1 en XPF, die associëren tot een heterodimeercomplex, zijn deels homoloog. Wanneer het een complex vormt, functioneert ERCC1-XPF als een structuurspecifiek endonuclease. Het voert een incisie uit op enkelstrengse/dubbelstrengse DNA-overgangen aan de 5'-zijde van beschadigd en vervormd DNA. Ook al is ERCC1-XPF vooral bekend om zijn functie in NER, het wordt ook in verband gebracht met verschillende andere trajecten, zoals de reparatie van DNA-tussenstreng kruisverbindingen. Veel van onze kennis over ERCC1-XPF komt uit studies van ziektes waarin dit DNA herstel niet goed functioneert. De belangstelling voor het begrijpen van de rol van ERCC1-XPF in DNA-herstel is recentelijk toegenomen vanwege het gevonden verband bij kankerpatiënten tussen resistentie voor chemotherapie en de expressie van ERCC1-XPF tijdens behandeling van de tumor.

In het onderzoek dat in dit proefschrift wordt beschreven, zijn verschillende spectroscopische, functionele en biofysische technieken gebruikt om een aantal intrigerende vragen over de structuur, functie en stabiliteit van het menselijke ERCC1-XPF-complex op te helderen.

Hoofdstuk 1 is een inleiding tot het onderzoek in dit proefschrift en geeft een overzicht van de huidige kennis over het heterodimere ERCC1-XPF-complex van de mens, inclusief verschillende DNA-herstelprocessen waarin het een aandeel heeft, de genetische ziektes die in verband zijn gebracht met deze DNA-reparatieprocessen, en een beschrijving van de structuurmotieven en functionele domeinen van de eiwitten die dit complex vormen.

De raakvlakken tussen ERCC1 en XPF liggen in de C-termini van beide eiwitten. De

eiwitten maken contact met elkaar door een centraal en pseudo-symmetrisch interactievlak. Verder wordt dit complex aan weerszijden van dit interactievlak gestabiliseerd door twee C-terminale phenylalanine's, die als ankers in N-terminaal gelokaliseerde holtes van het partnereiwit vallen. In **Hoofdstuk 2** zijn de gevolgen van een puntmutatie in deze raakvlakken op de structuur en functie van het ERCC1-XPF heterodimeer beschreven. Tot op heden is van slechts twee patiënten vastgesteld dat die deze mutatie dragen in een of beide allelen. Beide leden aan ernstige ontregelingen en stierven vroeg. De F231L-mutatie is gepositioneerd op het tandem-Helix-haarspeld-Helix $(HhH)_2$-domein van ERCC1 en bevindt zich in het hydrofobe interactievlak met XPF. Een vergelijking van de NMR-structuren van de wildtype en mutant-eiwitcomplexen tonen dat, hoewel F231 deel uitmaakt van ERCC1's N-terminale holte, de vervanging van F231 door een leucine deze holte vervormt en de oriëntatie van het belangrijke interactie-ankerresidu F894 van XPF verandert. Verdere biofysische karakterisering toont aan dat deze structuurveranderingen de stabiliteit van het gemuteerde ERCC1-XPF-complex reduceren. Eerder is geobserveerd dat in vivo en in vitro ERCC1 niet stabiel is in monomere toestand. De waargenomen versnelde heterodimeer-dissociatie zou de cellulaire concentratie kunnen verlagen van het heterodimere complex, wat in overeenstemming is met de vermindering van DNA-reparatie en de ophoping van DNA-reparatie-tussenproducten die tot het fenotype van deze mutatie leiden.

In vitro kan XPF homodimeren vormen en deze zijn over het algemeen stabieler dan het ERCC1-XPF-heterodimeer. Toch is er in vivo en in vitro een voorkeur voor de vorming van het ERCC1-XPF $(HhH)_2$. In **Hoofdstuk 3** laten we zien dat de dissociatie van het complex in zijn monomere componenten erg snel is. De dissociatie is zelfs zo snel dat ERCC1-XPF-heterodimeer in vitro niet zou kunnen bestaan. Het feit dat ERCC1-XPF toch redelijk stabiel is, suggereert dat deze monomeren snel herassociëren tot een heterodimeer. In dit hoofdstuk is ook beschreven dat de ERCC1 secundaire structuurelementen een lagere stabiliteit hebben dan die van XPF. Tevens is waargenomen dat ERCC1, wanneer het ontvouwt bij condities zoals een verhoogde temperatuur, zelf-aggregeert en het neer slaat, en dat XPF stabiel blijft tot hele hoge temperaturen. Daardoor komen bij hoge temperaturen, wanneer ERCC1 precipiteert, XPF-monomeren beschikbaar voor homodimeer-formatie. In vivo moet de XPF homodimerisatie echter langzamer zijn dan de herassociatie van monomeer XPF met vrij monomeer ERCC1. Omdat XPF in het homodimeer een andere structuur heeft dan XPF in het heterodimeer, is een mogelijke reden voor de in vivo afwezigheid van XPF-homodimeer ofwel dat de omzetting tussen de twee conformaties erg langzaam is, ofwel dat de homodimeer-conformatie van XPF correspondeert met een laag gepopuleerde toestand. Uit bestudering van de waterstof-deuterium-uitwisselingssnelheden van homodimeer en heterodimeer XPF met NMR, kunnen we concluderen dat homodimeer XPF een structuur vormt met lagere intrinsieke stabiliteit dan zijn heterodimeervorm. Dit onderbouwt de hypothese dat de XPF-conformatie in de homodimeer correspondeert met een laag gepopuleerde toestand.

We kunnen concluderen dat een correcte vouwing van XPF essentieel is voor de associatie met ERCC1 in een functioneel ERCC1-XPF-complex.

Van belang voor de functie van het ERCC1-XPF-complex is zijn vermogen om in de buurt van beschadigd DNA te binden. De voorkeur voor binding van ERCC1-XPF aan enkelstrengs(ss)/dubbelstrengs(ds) DNA-overgangen wordt bepaald door de C-terminale $(HhH)_2$ domeinen van ERCC1-XPF. Deze voorkeur positioneert de incisiefunctie van het XPF-endonucleasedomein aan de 5'-zijde van de beschadigde DNA-streng. Eerdere studies waarin met NMR de veranderingen in chemische-verschuiving van ERCC1 werden bestudeerd in titraties met een 22-nucleotide dsDNA-substraat, hebben aangetoond dat de twee haarspelden van ERCC1 verantwoordelijk zijn voor interacties met dsDNA en dat ERCC1 kleine lokale conformatieveranderingen ervaart om dit substraat te accommoderen. Echter, in deze eerdere studies werd geen rol voor het $(HhH)_2$-domein van XPF in DNA-herkenning gevonden. Daarom werd in eerste instantie gedacht dat XPF geen DNA kon binden, met als mogelijke reden de afwijkende korte tweede haarspeld van XPF. Later werd echter gevonden dat het $(HhH)_2$-motief van homodimeer XPF in staat was om ssDNA te binden. In het complex van homodimeer XPF met ssDNA zijn geladen aminozurenzijketens van de tweede helix van het eerste HhH-motief verbonden met de fosfaatgroepen van de ruggegraat van ssDNA. Een guanine-base op positie 5 van een 10-nucleotide ssDNA is ingesloten in een holte die gevormd is door de residuen van de twee HhH-motieven van homodimeer XPF. Dit suggereert tevens dat XPF specificiteit heeft voor guanine-basen. Dus, een actueel model van de ERCC1-XPF-binding met ss/dsDNA-overgangen omvat de duale DNA-bindingscapaciteit van ERCC1 en XPF.

In **Hoofdstuk 4**, onderzoeken wij de specificiteit van DNA-binding van het XPF homodimeer in meer detail en leveren wij experimenteel bewijs voor het actuele model voor DNA-binding van het ERCC1-XPF heterodimeer. We laten zien dat het $(HhH)_2$-domein van homodimeer XPF bij voorkeur aan guanine-rijk ssDNA en ss/dsDNA Holliday-structuren bindt. Verder wordt getoond dat het XPF $(HhH)_2$-domein van het ERCC1-XPF-heterodimeer hetzelfde interactie-oppervlak gebruikt om ssDNA te binden als het XPF-homodimeer, terwijl ERCC1 geen affiniteit toont voor deze ssDNA-substraten maar wel, zoals eerder bepaald, affiniteit voor dsDNA heeft. Om vast te stellen of alle DNA-bindingseigenschappen van ERCC1 en XPF kunnen worden samengebracht in een enkel model voor de binding van heterodimeer ERCC1-XPF aan ss/dsDNA-overgangen en of deze binding sequentie-specifiek is of niet, hebben we de affiniteiten geanalyseerd van ERCC1-XPF voor verschillende Y-vormige DNA-substraten, die gelijktijdig aan de ERCC1 en de XPF DNA-bindingsvlakken kunnen binden. In dit hoofdstuk tonen we een model, hoe het HhH-domein van XPF van ERCC1-XPF bindt aan de onbeschadigde enkelstreng van een reparatiebubbel-DNA, gebaseerd op de oriëntatie van de ssDNA-sequentie in de structuur van het ssDNA-complex van XPF-homodimeer, en de structuren van de homologe DNA-complexen van RuvA en archaeaal XPF.

Samenvattend, de studies beschreven in dit proefschrift, laten aanzienlijke dynami-

ca in de ERCC1-XPF-complexformatie zien, demonstreren de effecten van ziektegerelateerde mutaties van ERCC1 in het interactiegebied met XPF op de structuur en stabiliteit van het complex, en onderbouwen ons model van de ERCC1-XPF-specificiteit voor ds/ssDNA-overgangen, waarbij ERCC1 dsDNA bindt en XPF ssDNA.

خلاصه فارسی

هترودیمر توسط تکنیک رزونانس مغناطیسی هسته نشان می‌دهد که ساختار ایکس.پی.اف در همودیمر با پروتئینی با پایداری درونی کمتر نسبت به حالت هترودیمری تطابق دارد. این امر تقویت کننده این نظریه است که حالت همودیمری ایکس.پی.اف با تاخوردگی پروتئینی باسطح انرژی کم‌جمیت‌تر نسبت به هترودیمر تطابق دارد. در نتیجه، تاخوردگی صحیح ایکس.پی.اف برای تشکیل کمپلکس با ای.آر.سی.سی۱ ضرورت دارد.

یکی از ویژگیهای ساختاری ای.آر.سی.سی۱-ایکس.پی.اف قابلیت اتصال آن به نواحی نزدیک به دی.ان.ای آسیب‌دیده است. ترجیح در اتصال به دی.ان.ای در محل الحاق تک‌رشته/دورشته مربوط به انتهای کربونیل یعنی دومین (اچ.اچ.اچ)$_{۲}$ می‌باشد. این نوع برهم‌کنش منجر به عملکرد برشی در سمت ۵' دی.ان.ای آسیب‌دیده توسط دومین اندونوکلئاز می‌شود.مطالعات گذشته بر روی تغییرات شیفت شیمیایی نشان داده است که دو ناحیه هیرپین[۷] در ای.آر.سی.سی۱ با سوبسترای دی.ان.ای دورشته ۲۲ نوکئوتیدی برهم‌کنش می‌کنند و ای.آر.سی.سی۱ هنگام این برهم‌کنش دستخوش تغییرات ناحیه‌ای کوچکی می‌شود. اما در این مطالعات نقشی برای تشخیص دی.ان. ای توسط دومین (اچ.اچ.اچ)$_{۲}$ در ایکس.پی.اف یافت نشده بود.در نتیجه، در آغاز این تصور وجود داشت که ایکس.پی.اف به علت اینکه ناحیه هیرپین دومش کوتاهتر می‌باشد، قادر به اتصال به دی.ان.ای نیست. بعدها، نشان داده شد که ایکس.پی.اف در پروتئین همودیمر قادر به اتصال به دی.ان.ای تک‌رشته می‌باشد. گروههای باردار در اولین موتیف اچ.اچ.اچ قادر به برهم‌کنش با زنجیره اصلی فسفاتی دی.ان.ای تک‌رشته ۱۰ نوکلئوتیدی می‌باشند که نشانگر اختصاصی بودن برهم‌کنش با بازهای گوانین می‌باشد. در نتیجه، یک مدل اتصالی ای.آر.سی.سی۱-ایکس.پی.اف به دی.ان.ای در محل الحاق تک‌رشته/دورشته به ظرفیت اتصال هر دو پروتئین ای.آر.سی.سی۱ و ایکس.پی. اف وابسته است.

در **فصل چهارم**، این فرضیه ترجیح اتصال دی.ان.ای را آزموده‌ایم و درستی این مدل اتصال ای.آر.سی.سی۱ -ایکس.پی.اف به دی.ان.ای را با روشهای آزمایشگاهی اثبات نموده‌ایم. ما نشان می‌دهیم که دومین (اچ.اچ.اچ)$_{۲}$ ایکس.پی.اف همودیمری اتصال به توالی‌های دی.ان.ای تک‌رشته غنی از گوانین و الحاق هالیدی تک‌رشته/دورشته را ترجیح می‌دهد. همچنین، دومین (اچ.اچ.اچ)$_{۲}$ ایکس.پی.اف هترودیمری از سطح مشابه ایکس.پی.اف همودیمری برای اتصال به دی.ان.ای تک‌رشته استفاده می‌کند، در حالیکه ای.آر.سی.سی۱ تمایلی برای اتصال به دی.ان.ای تک‌رشته نشان نمی‌دهد. برای درک این موضوع که آیا این ویژگیها در یک مدل اتصال ای.آر.سی.سی۱-ایکس.پی.اف هترودیمری به محل الحاق تک‌رشته/دورشته خلاصه می‌شود یا نه، و آیا این اتصال ویژه توالی‌های خاص است یا نه، ما تمایل سطوح برهم‌کنش کننده با دی.ان.ای را در ای.آر.سی.سی۱ و ایکس.پی.اف با استفاده از سوبستراهای مختلف مورد بررسی قرار دادیم و نشان دادیم که هر دو سطح در میزان تمایل و اختصاصی بودن اتصال ضروری هستند. بر اساس آرایش توالی دی.ان.ای تک‌رشته در ساختار متصل به دی.ان. ای ایکس.پی.اف همودیمری و با این فرض که ای.آر.سی.سی۱ همانند همولوگهایش به دی.ان.ای دورشته متصل می‌شود، ما پیشنهاد می‌کنیم که دومین اچ.اچ.اچ در ایکس.پی.اف به حباب دی.ان.ای تک‌رشته آسیب‌ندیده متصل می‌شود.

در کل، مطالعات ارائه شده در این رساله، دینامیک قابل توجه در تشکیل کمپلکس ای.آر.سی.سی۱-ایکس.پی.اف و اثرات یک جهش بیماری‌زا در سطح برهم‌کنش ای.آر.سی.سی۱ بر ساختار و پایداری را نشان می‌دهد و درستی مدل ما برای اختصاصیت ای.آر.سی.سی۱-ایکس.پی. اف برای دی.ان.ای در محل الحاق تک‌رشته/دورشته را اثبات می‌کند.

[1] DNA [2] UV [3] ERCC1-XPF [4] ICL-repair [5] $(HhH)_2$ [6] SPR [7] hairpin

وقایع مختلفی انرژی و مولکولهای لازم برای ساخت و بقای سلول تامین می‌نمایند. پروتئینها نقش مهمی در این امر دارند و می‌توانند به عنوان آنزیم واکنشهای شیمیایی را در سلول و محیط اطراف آن کاتالیز نمایند، در حرکت سلولی یا نقل و انتقال از خلال غشا نقش بازی نمایند، بیان ژن را تنظیم نمایند،نقش حفاظتی داشته باشند و یا تخریبها را ترمیم نمایند. تنوع چشمگیر در عملکرد پروتئینها تا حد زیادی وابسته به ساختار و برهم‌کنشهای آنها با مولکولهای دیگر است. در نتیجه، تغییری ناچیز در ساختار، عملکرد اختصاصی یا تمایل پروتئینها به هم تعادل عملکردی را بر هم زده و در نتیجه منجر به تغییر سرنوشت سلول و در بسیاری موارد بروز بیماریها می‌گردد.

پروتئینها در سلول توسط دی.ان.ای[1] کد می‌شوند که حامل اطلاعات ژنتیکی است، و حفظ این اطلاعات ژنتیکی امری بسیار ضروری است. علی‌رغم تمام تمهیدات انجام شده، تخریب دی.ان.ای به صورت مداوم و پی‌درپی رخ می‌دهد. در واقع، همین تغییرات در توالی دی.ان.ای عامل رخداد تکامل و تنوع زیستی در موجودات زنده می‌باشد، ولی در عین حال می‌تواند منجر به اختلالات بیماری‌زا و از دست رفتن اطلاعات ژنتیکی نیز بشود. برای اطمینان از حفظ قوام و پایداری ژنوم، سلولها روش‌هایی برای ترمیم آسیبهای دی.ان.ای گسترش داده اند. یکی از این روشها، ترمیم برداشت نوکلئوتید آسیب‌دیده است، که می‌تواند به عنوان مثال آسیب حاصل از اشعه یو.وی[2] را ترمیم نماید. کمپلکس هترودیمری ای.آر.سی.سی۱-ایکس.پی.اف[3] یکی از پرووتئینهای فعال در این سیستم ترمیمی است.

دو پروتئین ای.آر.سی.سی۱ و ایکس.پی.اف به عنوان کمپلکسی هترودیمری به هم پیوسته است و بخشهایی از این دو پروتئین همولوگ می‌باشد. ای.آر.سی.سی۱-ایکس.پی.اف کمپلکسی است که به عنوان یک اندونوکلئاز ساختار ویژه عمل می‌نماید و برشی در سمت '۵ دی.ان.ای آسیب‌دیده در محل الحاق دی.ان.ای تک رشته/دورشته ایجاد می‌کند. اگرچه ای.آر.سی.سی۱-ایکس.پی.اف بیشتر به خاطر عملکردش در ترمیم برداشت نوکئوتید شناخته شده، عمکردش به مسیرهای ترمیمی دیگری نیز ارتباط دارد از جمله ترمیم پل بین‌رشته‌ای[4] دی.ان.ای. قسمت عمده دانش ما در مورد ای.آر.سی.سی۱-ایکس.پی.اف از تحقیق در مورد بیماریهایی که منجر به نقص می‌شوند نشات می‌گیرد. اخیرا، علاقه دانشمندان به درک نقش ای.آر.سی.سی۱-ایکس.پی.اف در ترمیم دی.ان.ای افزایش یافته که علت عمده آن ارتباط کشف شده بین مقاومت بیماران سرطانی به شیمی درمانی و میزان بیان ای.آر.سی.سی۱-ایکس.پی.اف هنگام درمان تومور می‌باشد.

در این رساله، از روشهای اسپکتروسکوپی و تکنیکهای سنجش عملکرد و بیوفیزیکی برای پاسخ به برخی سوالات در مورد ساختار، عملکرد و پایداری کمپلکس انسانی ای.آر.سی.سی۱-ایکس.پی.اف استفاده شده است.

فصل اول به تفصیل و توضیح دانش کنونی درباره کمپلکس انسانی ای.آر.سی.سی۱-ایکس.پی.اف می‌پردازد که شامل مسیرهای ترمیم دی.ان.ای، بیماریهای ژنتیکی مرتبط با این مسیرها، و توصیف موتیفهای ساختاری و دومین های عملکردی پروتئینهای تشکیل‌دهنده کمپلکس می‌پردازد.

نواحی برهم‌کنش هترودیمری ای.آر.سی.سی۱ و ایکس.پی.اف در انتهای کربونیل دو پروتئین قرار دارد. دو پروتئین از طریق سطح مرکزی برهم‌کنشی با هم در تماسند و زنجیره‌های جانبی آمینواسیدهای فنیل آلانین در انتهای کربونیل نسبت به هم تقارن سودوسیمتریک داشته و به صورت قفل و کلید در حفره های پروتئین پارتنر قرار میگیرند. در **فصل دوم**، عواقب ساختاری و عملکردی یک جهش در این سطح برهم کنش ای.آر.سی.سی۱-ایکس.پی.اف توضیح داده شده است. تا کنون، تنها دو بیمار حامل این جهش در یک یا هر دو الل گزارش شده اند و هر دو این بیمارها دارای نواقص شدیدی بوده و در عنفوان کودکی جان سپردند. جهش فنیل آلانین ۲۳۱ به لوسین در دومین (چ.اچ.اچ)$_{۲}$[5] در ای.آر.سی.سی۱ قرار داشته و متعلق به سطح برهم‌کنش هیدروفوبیک آن با ایکس.پی.اف می‌باشد.مقایسه ساختار کمپلکسها در پروتئین طبیعی و جهش یافته با استفاده از تکنیک رزونانس مغناطیسی هسته نشان می‌دهد که جایگزینی ذکرشده در پروتئین جهش‌یافته منجر به از بین رفتن حفره پروتئینی در این ناحیه و تغییر آرایش آمینواسید کلیدی فنیل‌آلانین ۸۹۴ در این حفره می‌شود. مطالعات بیوفیزیکی نشان می‌دهد که این تغییرات پایداری ای.آر.سی.سی۱-ایکس.پی.اف را کاهش داده و با توجه به اینکه ای.آر.سی.سی۱ در پروتئین طبیعی به صورت مونومری ناپایدار است، افزایش جدایش کمپلکس ای.آر.سی.سی۱-ایکس.پی.اف در پروتئین جهش یافته می‌تواند منجر به کاهش غلظت سلولی کمپلکس هترودیمری بشود که با تداخل در فرآیند ترمیم دی.ان.ای و تجمع پروتئینهای واسطه دیگر این فرآیند تطابق دارد.

در محیط آزمایشگاهی، ایکس.پی.اف قادر به تشکیل همودیمر با خودش است که به نظر می‌آید پایدارتر از هترودیمر ای.آر.سی.سی۱-ایکس.پی.اف باشد. به هر صورت، چه در محیط آزمایشگاهی و چه در بدن موجود زنده، هترودیمر ای.آر.سی.سی۱-ایکس.پی.اف ترجیح داده شده است. در **فصل سوم**، ما جدایش بسیار سریع کمپلکس به مونومرهایش را طی آزمایش اس.پی.آر[6] گزارش می‌کنیم، جدایشی که آنقدر سریع است که به نظر می‌رسد کمپلکس ای.آر.سی.سی۱-ایکس.پی.اف نباید در موجود زنده وجود داشته باشد. این واقعیت که کمپلکس ای.آر.سی.سی۱ -ایکس.پی.اف کمپلکسی پایدار است، پیشنهاد میکند که این کمپلکس پس از جدایش مجددا به صورت هترودیمر تشکیل می‌شود.

همچنین، ایکس.پی.اف تنها در دمای بالا وقتی ای.آر.سی.سی۱ رسوب می‌کند، با خودش تشکیل همودیمر می‌دهد. در موجود زنده، این تشکیل همودیمر باید کندتر از تشکیل مجدد هترودیمر باشد. مطالعه بیشتر ساختار ایکس.پی.اف نشان می‌دهد که این پروتئین در واقع در حالت هترودیمر نسبت به همودیمر ساختاری متفاوت دارد که در مطالعات گذشته این تفاوت چندان مورد توجه قرار نگرفته بوده است. اینکه همودیمر کمتر مشاهده می‌شود می‌تواند به این دلیل باشد که پروتئین با ساختار همودیمریک در سطح انرژی با جمعیت متفاوت قرار دارد و یا اینکه سرعت تبدیل دو ساختار به یکدیگر در این پروتئین مرحله محدود کننده سرعت باشد. مقایسه میزان تبادل هیدروژن با دوتریوم در همودیمر و

Akcnowledgements

Doing a PhD , sometimes feels like a lonesome journey, but I have been surounded by wonderful people who have contributed to this work and has given enormous support during thiscourse of my academic career.

I am deeply thankful to my supervisor and promoter Prof. Rolf Boelens and my co-promoter Dr. Gert Folkers for the excellent remarks and fruitful input during my PhD period. I appreciate all the time you devoted to our fruitful discussions during these last months when your schedule was busier than ever. I am also thank ful for the useful remarks of Prof. s Jan Hoeijmakers and Robert Kaptein on my first paper.

I am grateful to Hans Wienk who was involved in my project with a constructive critical eye and a sense of humor. Hans, I appreciate your scientific contributions and feedback during my PhD years! Ramachandra Dongre, king Rama, I started my PhD witnessing you being commited to conquer all the ups and downs of research in the field of solution NMR, trying to purify the bad behaving IF2-G2-G3 and simultanously set up new NMR experiment. Thanks for late-night brainstorming sessions while performing protein purifications or spectral assignments. I am sure you have attained huge knowledge and experience that is benefiting a lot the people in your current lab! Klaartje Houben, my officemate from the beginning, I am glad that I shared the office with you. I appreciate your willingness to discuss the questions I had in my project every now and then. Also, your attitude towards female scientists is very inspiring to me!

XueLi (Li Xue) and Anna Vangone! My not just colleagues but great friends! It is already about 2 years that we have you in the group! You started at the same time about when Luciana was visiting. Suddenly, there was some color to our group: girls! You showed that it doesn't need to be numbered 50-50, to feel 50-50! It has been wonderful and indeed very fruitful for me to have you around. Scientifically, I love how smooth things has been around you. Anna, I appreciate all the time we have spent in the lab and outside discussing different matters. Thanks for all your support and the effort you are already putting as my paranymph! XueLi, I appreciate how continually you encouraged me through the writing of this dissertation. Thanks for all the input and career advices you have given me. It has been very eye-opening. Luciana, thanks for all the discussions we had in the lab and over the skype. It felt really good to be at similar level with someone. To be both more biochemist than NMR"ist". I very much enjoyed your short stay and wish you had come sooner and stayed longer. Good luck with your PhD!

Mohammed Kaplan, my other paranymph, I really enjoyed our discussions about food and culture and I learned how close is our cultures! Thanks for being a supportive colleague and wonderful friend. Gydo van Zundert, we used to be office mates but I know no one likes that small desk on the way in the office! I knew you will not last that long but you lasted longer than I thought! Thanks for the goodbye chocolate. I ate it all during the late night works when no food was available anymore! And thank you so much for translating

my Samenvatting! Sorry that I did not manage to bring you the Shah's crown. Dr. Abhishek Cukkemane, I really enjoyed our discussions and the Indian Cuisine. Also, thank you for your cool advices.

Hans Ippel, Tobias Madl, Hugo van Ingen and Henry Hocking thanks for the brilliant remarks during the work discussions and supports in setting up NMR experiments. Markus, Elwin and Denni, thanks for the positive words and support during the conversations we had in the lab. Mark Daniels, thanks for your advices besides being a cool lab engineer. I am glad that you finally joined back to the lab. Nuno, Deepak, Sabina, Amanda, Klemen, Alma, Tessa and Eline, I wish you were here now, but I am glad that each of you are not here for good reasons. Ezgi Karaca, João Rodrigues, Adrien Melquiond , Panagiotis Kastritis, Mark van Dijk! you were the core of the longest lasting team of the group that not only shared your moments at work but also at home. The tie between you was very attractive and inspiring. Keep continuing! I especially miss your group when it was the most expanded having Cristophe schmitdz and Mikael Trellet in it, when we used to gatrher during the movie nights! Also, it was nice to have super nice people every now and then in your group.

Barbara Hendricx, thanks for that first dinner at your place that was a good start for me, feeling welcomed to the Netherlands. And thanks for all the administrative work you have done for me including jobs that are not your duty but you willingly have done, especially jobs like proofreading my samenvatting. Many thankful thoughts to former or new group members are still floating in my mind: Munishika Kalia and Marie Renault for their supportive words, Malgorzata Szeląg for our refreshing conversations, Ramon for our discussions on ERCC1-XPF and his endless questions about Iran. Ivan, Cecilia and Cunliang welcome to the group, thanks for your smiling faces and good luck with your PhD. I thank all the present and former colleagues that might have slipped out of my mind during writing these lines.

I extend my thanks to my friends from other groups of Bijvoet Institute, first of all Els Halff. Els joonam, kheili khoshhalam ke Iran umadi va behet khosh gozasht. Thanks for being my practical colleague and friend who supported me unconditionally from the beginning of my stay in the Netherlands. I appreciate all the efforts and lengthy emails that now you send me from UK to inform me on many stuff. I am so rewarded to have friends like you. Marloes joonam, our occasional, sometimes even accidental, meetings are wonderful and I like your sense of humor. Now, the exceptional Bijvoet team stars, former and current, Elif, David Minde, Alfonso, Andrea V., Deniz, Jonas, Martina, Tania, Hedwich, Camilla, Daniel, Luca, Nadia, Viviana and thanks for your lectures and remarks during AIO evenings and collaborations every now and then with the lab facilities. Also, thanks for making my life spicy! A word of gratitude to all my good friends Hajar, Shalaleh, Nahid, Fatemeh, Ana Dovc, Tahmineh, Afrouz, Pasha, Akvile, Shekoofeh and Samad who supported and consoled me during the rough times of this long journey. And thanks to those who brought sunshine to

the cloudy winters of the Netherlands: Mina, Pourya, Khale Mitra and Mr. Sohrab, Naghmeh, Marjolein, Peyman, Iman R., Shima, Amir, Vahid and my artist flatmate Sima.

Last but definitely not the least, I am especially grateful to my wonderful family for their unconditional love from far-away and close-by. My kind father whose compassion and determination to be humane will eternally stay in my memory, my mother, my role model, whose devotion and support has given me the strength in the roughest times during my PhD and beyond, my dear sister, whose love and honest opinions has always warmth my heart and helped me to keep in the track and my brothers for their advices and support.

قدرشناس خانواده عزیزم هستم که غم دوری را به جان خریدند و دورادور از من حمایت نمودهاند. قدردان پدر دلسوز، مهربان و شریفم هستم که گرچه در کنارم نیست ولی یاد و خاطرش ابدی است، مادرم، سرمشق زندگیم، که ایثار و حمایت بی دریغش در سختترین لحظات در طول دوره دکترا و زمانهای قبل از آن به من توان کافی بخشیدو خواهر خوشقلبم که عشق و نظرات صادقانه اش قلبم را گرمی بخشیده و مسیر زندگی را برایم روشن نموده است و قدرشناس برادران مهربانم هستم برای راهکارها و حمایتهای بی دریغشان.

Thanks to all of you,

بی نهایت سپاسگذار همگی هستم

Maryam مریم

Curriculum Vitae

Maryam Faridounnia was born on the 13th of May 1980 in Tehran, Iran. After finishing high school, she participated in the competitive National University Entrance Exam after which she started her bachelor studies in Biology at the Faculty of Science of the University of Guilan, where she graduated in 2004. Thereafter she participated in the postgraduate entrance exam and was admitted to the masters studies in Biophysics of Tarbiat Modares University, Tehran. She obtained her masters degree in 2007 with a thesis entitled 'Purification and biophysical characterization of some peptides from the venom of scorpion *Mesobuthus eupeus*'. She presented her research at the 5th National Biotechnology Congress 2007 in Tehran (Iran) resulting in a 1st prize, and at the international conferences ICACT 2007 (Paris, France) and FEBS 2008 (Athens, Greece). During her studies she received a DAAD traineeship scholarship for a 2-month research project on drug delivery at the Martin-Luther University in Halle, Germany. At the end of 2008, she started her PhD project at Utrecht University in The Netherlands under supervision of Prof. Dr. R. Boelens and Dr. G. Folkers. In Utrecht she studied the structure and stability of ERCC1-XPF DNA repair complexes, using NMR spectroscopy and other biophysical techniques. The results of her work are presented in this thesis.