Structure, Vol. 13, 1849–1858, December, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.08.014

The Structure of the Human ERCC1/XPF Interaction Domains Reveals a Complementary Role for the Two Proteins in Nucleotide Excision Repair

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

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

[2]Department of Cell Biology and Genetics
Erasmus Medical Center
3000 DR Rotterdam
The Netherlands

Summary

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

Introduction

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

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

Inherited defects in the NER process cause the serious prototype repair disorders xeroderma pigmentosum (XP) and Cockayne syndrome (CS), highlighting an extreme risk of UV-induced skin cancer and many features of accelerated ageing, respectively (de Boer and Hoeijmakers, 2000). Mice targeted for knockout (KO) versions of genes involved in NER usually reflect the human pathology of XP and CS to a considerable extent. In contrast, ERCC1- or XPF-targeted KO mice display a much more severe phenotype with multiorgan involvement, severe runting, and death before weaning (Tian et al., 2004; Weeda et al., 1997). This observation points to other functions of this endonuclease outside the context of NER. In line with this finding, XPF mutations associated with XP are relatively rare and always hypomorphic, whereas a first case of human ERCC1 deficiency remains to be reported.

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

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

For the action of ERCC1/XPF, heterodimer formation is essential, where the actual nuclease domain is

*Correspondence: r.boelens@chem.uu.nl

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

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

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

Results

Purification and Biophysical Characterization of the C-Terminal Parts of XPF and ERCC1

Initial attempts to overexpress the C-terminal part of ERCC1 in *E. coli* cells resulted in insoluble and aggregated protein. Attempts to resolubilize the protein domain in various manners failed. This might be explained by our previous findings indicating that ERCC1 can only form a functional fold in complex with XPF. Therefore, the C-terminal parts of XPF and ERCC1 were coexpressed by using a bicistronic expression vector encoding the two subunits, with XPF being the first in order and ERCC1 containing a His_6 tag at the C terminus. We obtained soluble proteins with very high expression yield, and the elution profile in the gel filtration resulted in one species with the predicted size of the heterodimeric complex. Furthermore, the protein fingerprint spectrum (^{15}N-^{1}H HSQC) revealed the expected number of residues, confirming complex formation and a compact fold. These data are in accordance with the majority of the biochemical observations that illustrate the significance of the close partnership for the smooth functioning of ERCC1 and XPF proteins (Sijbers et al., 1996a).

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

Structure Determination

The solution structure of the 19 kDa complex between the C-terminal domain of ERCC1 (residues 220–297) and the C-terminal domain of XPF (residues 823–905) was solved by heteronuclear double- and triple-resonance NMR spectroscopy by using uniformly ^{15}N- and $^{15}N/^{13}C$-labeled protein. The structure was determined on the basis of 4547 experimental NMR restraints and 197 dihedral angle restraints. A large number of intermolecular distance restraints was collected (442), which permitted the positioning of the two subunits with respect to each other. No distance violations larger than 0.4 Å or dihedral angle violations larger than 5° were found. A summary of the structural and restraint statistics is given in Table 1. Figure 2A shows an overlay of the 20 lowest-energy conformers obtained after the structure calculations.

Structure Description

The structure of the core of the complex, comprising residues 234–294 of ERCC1 and residues 836–895 of XPF, exhibits a global pseudosymmetry with a very similar architecture of the associated partners. The fold of both subunits is built up by helical secondary elements. In particular, the fold of ERCC1 is a characteristic example of the double helix-hairpin-helix $(HhH)_2$ motif (Shao and Grishin, 2000), which is present in a variety of protein families involved in non-sequence-specific DNA binding such as DNA-polymerases, ligases, and nucleases. Each motif forms into a pair of antiparallel α helices connected by a hairpin-like loop. The ERCC1 helices α and β form the first HhH element, and helices δ and ε form the second one connected by a short helix, γ. The angle between the helical vectors on the first and second pair is 68 and 49°, respectively. On the other side, the XPF fold is almost the same, except for a minor, but significant, difference. Although the relative orientation of the XPF helices agrees well with the $(HhH)_2$ fold, only the first hairpin is present. It is formed by helices α' and β' with an angle of 51° (helices of XPF are indicated by a prime sign). The hairpin between helices δ' and ε' has been replaced by a β turn of only three residues. These helices are inclined at an angle of 55°. Although XPF retains only one complete HhH element, the global fold still resembles the $(HhH)_2$ structure (Figures 2B and 2C). The rmsd of the individual ERCC1

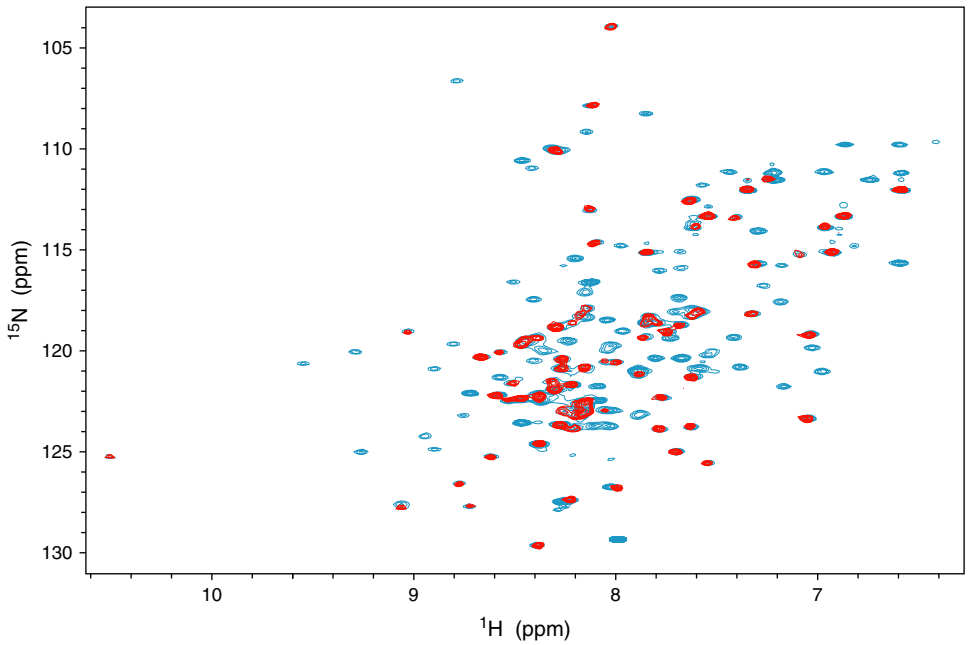

Figure 1. Refolding of the ERCC1/XPF Complex

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

and XPF structures is 2.5 Å (65 Cα atoms), with a Z-score of 8.1.

Protein-Protein Interactions

The two domains of ERCC1 and XPF have a large interaction surface with an area of 1534 Å^2. The orientation of helices α and γ from both subunits resembles a four-helical bundle, although the γ helices are very short. Only helix γ' of XPF deviates from the parallel orientation, and it is tilted by 35° with respect to its ERCC1 mate (Figure 3B). Both ERCC1 and XPF helical pairs form relatively flat and extremely hydrophobic surfaces that are shielded in the protein-protein interface. This central hydrophobic core seems to be the main contributor of the association, since it confers more than half of the interaction surface area. A key determinant of the nonbonded interactions is Phe840 of XPF, which protrudes in the middle of the symmetric helical array (Figures 3A and 3B). The aromatic ring of Phe840 makes contacts with Cys238, Thr241, Leu260, and Leu289 of ERCC1, which are located in helices α, γ, and ε. The rest of the intermolecular contacts are restricted to residues that belong to symmetry-related helices. Lys843 in helix α' of XPF interacts via its aliphatic side chain with Glu237, Cys238, and Thr241 of ERCC1 helix α. In half of the structures in the ensemble, the carboxyl of Lys843 of XPF forms bifurcated hydrogen bonds to the guanidinium group of ERCC1 Arg234. Moreover, the side chain amide group of the same Lys forms a salt bridge with the carboxyl side chain of Glu237 of ERCC1. Regarding the short γ helices, hydrophobic interactions are present between Ile264 of ERCC1 and Ala863 and Ala866 of XPF. The hydrophobic core additionally contains Phe889 of XPF and Val288 and Leu289 of ERCC1.

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

Helix-Hairpin-Helix Motifs and DNA Binding

A DALI search in the Protein Data Bank using the individual protomers as queries detected the same set of homologous structures for both ERCC1 and XPF. In all cases, ERCC1 showed better scores, because it exhibits the canonical $(HhH)_2$ motif. The $(HhH)_2$ domain of ERCC1 closely resembles the $(HhH)_2$ domain of the archaeal homodimeric XPF from *Aeropyrum pernix* (PDB code 2BGW; rmsd 1.9 Å) (Newman et al., 2005) and the

Table 1. Structural Statistics of the Structure Ensemble of the Heterodimeric ERCC1/XPF Protein

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

Residues 227–294 of ERCC1 and residues 831–896 of XPF were used.

holiday junction-recognizing $(HhH)_2$ domain of RuvA (PDB code 1C7Y; rmsd 1.9 Å) (Ariyoshi et al., 2000).We also noticed significant similarity to the C-terminal domain of the bacterial repair protein UvrC (PDB code 1KFT; rmsd 2.1 Å) (Singh et al., 2002), which exhibits the same double HhH organization (Figure 4B). We have shown before that the UvrC $(HhH)_2$ domain is required and sufficient for recognition of the single-to-double strand junction. To address whether the ERCC1/XPF HhH domains mediate DNA binding, we performed electrophoretic mobility shift assays.

Weak binding (K_d of 39 ± 13 μM) was detected on a stem-loop DNA substrate that was shown to be the optimal substrate for ERCC1/XPF in nuclease activity assays (de Laat et al., 1998a). We found that binding is specific for probes that contain single-double strand junctions, as complexes were also formed with a bubble and fork substrate, whereas no binding was detected on dsDNA or ssDNA probes (Figure S1; see the Supplemental Data available with this article online). Though weaker, the binding preference of the isolated $(HhH)_2$ interacting domains is similar to the observed in vitro incision preference of full-length recombinant ERCC1/XPF protein (de Laat et al., 1998a), which shows a K_d of 100 nM for the same stem-loop substrate (unpublished data). In this respect, it is likely that other domains of the native complex contribute to DNA binding as well. Furthermore, it has been shown before that recruitment of ERCC1/XPF in the context of NER is stabilized by various other proteins, including XPA-RPA proteins and the presence of TFIIH and XPG in the preincision complex (Riedl et al., 2003; Volker et al., 2001).

To determine the interaction surface for DNA binding, ^{1}H-^{15}N HSQC spectra of the protein were recorded with successive additions of a stem-loop DNA substrate. Although the chemical shift perturbations that we observed upon titration are relatively small, the most prominent effects are clustered in the second hairpin of ERCC1, with a few in its first hairpin (Figure 4C and Figure S2). In particular, the maximum amide chemical shift perturbations were detected in Gly276 and Gly278 along with the intervening Leu277 forming the classical GhG hairpin. In the crystal structures of protein-DNA complexes containing these hairpin motifs, the glycines form hydrogen bonds with adjacent phosphate oxygens of the DNA, and a positive residue after the second Gly makes polar contacts with the phosphate backbone. Lys281 of ERCC1 is also shifted significantly, suggesting the involvement of this residue in a protein-DNA contact. The first hairpin of ERCC1 is more distinct in terms of sequence composition and lacks the GhG signature. Unfortunately, due to overlap, we were not able to follow the shifts of three critical residues of the first hairpin, Ser244, Lys247, and Thr248. However, the shifts at Val245 and Asn246 demonstrate that this hairpin is also in contact with DNA. There are also randomly distributed shifts both on the ERCC1 and the XPF side. They probably reflect indirect perturbation effects possibly involving small rearrangements in the hydrophobic packing induced by DNA binding. No shifts, however, are observed in the hairpin of XPF, strongly suggesting that this subunit does not directly participate in DNA recognition under the present in vitro conditions. Furthermore, the positive charge mainly in the hairpins of ERCC1 agrees well with the ERCC1 residues found in the proximity of DNA in our experiments (Figures 4D and 4E).

Discussion

ERCC1 Forms an Obligate Heterodimer with XPF

In this study, we show that the driving force of complex formation is mainly hydrophobic interaction, and we show that this yields a compact architecture that enhances the stability of each subunit. Both domains adopt similar structures in the final complex related by a pseudo-2-fold symmetry axis, with ERCC1 displaying the canonical $(HhH)_2$ fold and XPF displaying a very similar fold. The DNA binding domain from ERCC1 resembles the archaeal XPF more closely than the eukaryotic XPF does. The opposite is true for the flanking endonuclease domain, which is inactivated in ERCC1 protein (Figure 4A). In archaebacteria, ERCC1 is absent, while in all sequenced eukaryotes, ranging from fungi, parasites, plants, and invertebrates to vertebrates, ERCC1 and XPF homologs have been found. Furthermore, the archaeal orthologous repair protein, like the human protein, forms a dimeric structure. The structural resemblances of human ERCC1 and archaeal XPF with regard to the $(HhH)_2$ domains agree with the suggested common origin for the eukaryotic ERCC1 and XPF proteins (Gaillard and Wood, 2001).

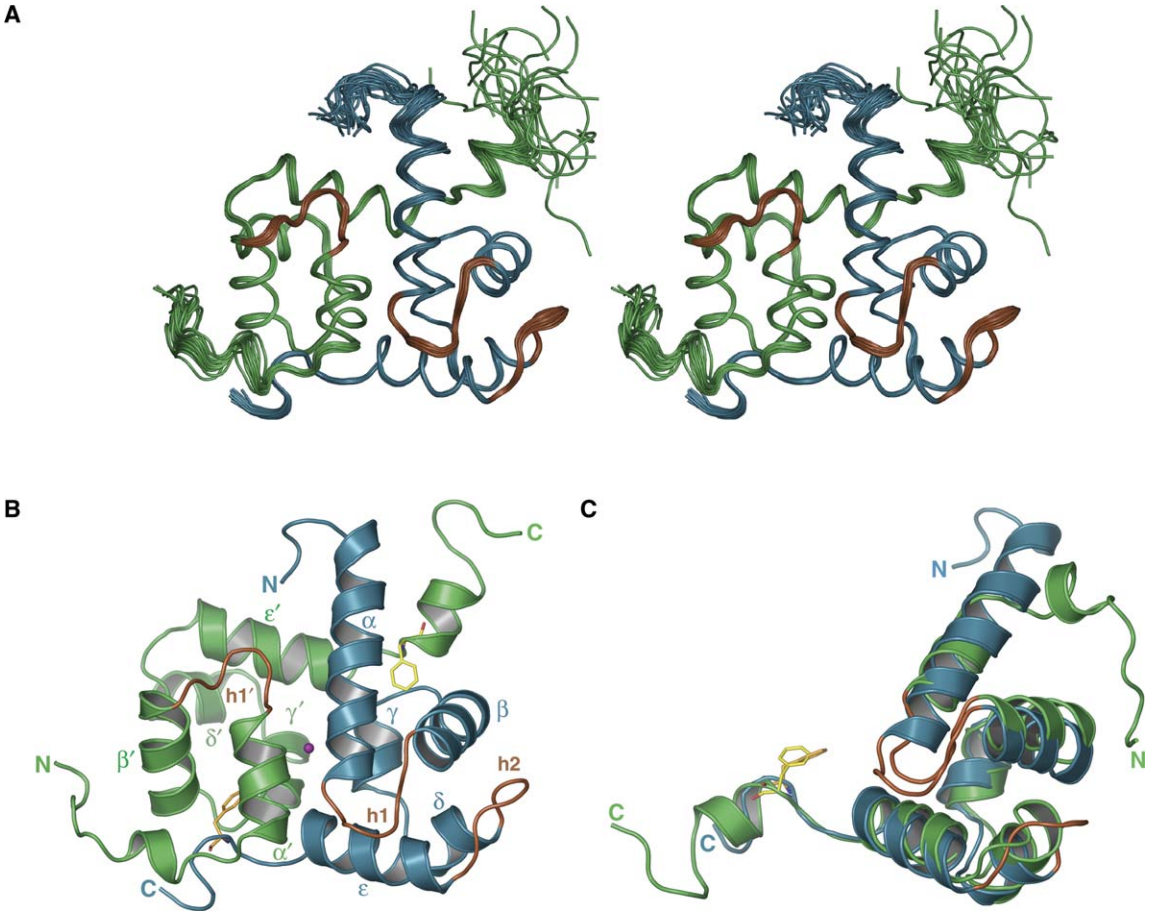

Figure 2. Three-Dimensional Structure of the C-Terminal Domains of the Human ERCC1/XPF Complex

(A) Backbone stereoview of the ensemble of the final 20 structural conformers. ERCC1 is colored blue, XPF is colored green, and the hairpins in both subunits are brown.

(B) Cartoon representation of the lowest-energy model. The purple sphere depicts the center of the pseudo-2-fold symmetry axis. Helices are denoted with Greek letters, from α to ε for ERCC1 and from α' to ε' for XPF, while hairpins are indicated as h1, h2, and h1′. Also shown are Phe residues at the C termini of each subunit. Phe293 of ERCC1 is colored orange, and Phe894 of XPF is colored yellow.

(C) ERCC1 and XPF are superimposed, showing the overall fold similarity. Color conventions are as in (B).

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

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

DNA Binding Is Mediated by the HhH Domain of ERCC1

The observed small chemical shift changes in the core of the ERCC1/XPF complex upon addition of DNA (Figure 4) suggest that small, local conformational changes are required to accommodate the DNA substrate for the human ERCC1/XPF heterodimer. In this respect, we note that the $(HhH)_2$ motif of ERCC1 resembles more closely the corresponding domain of the archaeal homodimeric XPF bound to DNA than that of the free archaeal XPF structure (Newman et al., 2005). It is known that HhH domains recognize various substrates, but

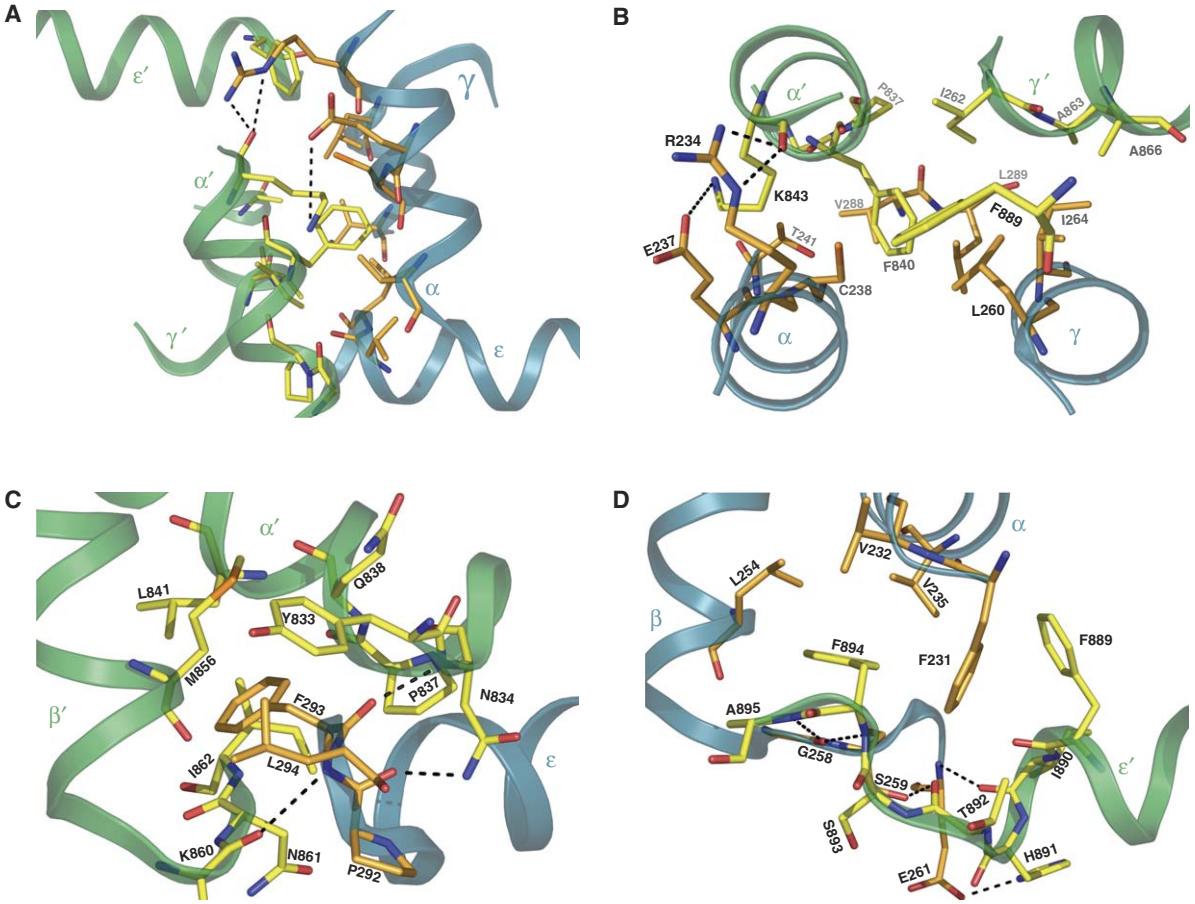

Figure 3. Protein-Protein Interactions along the C-Terminal Domains of ERCC1 and XPF

(A) View of the central interacting core. Helices α in the front, helices γ at the back, and the tips of helices ε that run perpendicularly build up the interacting surface. Residues participating in contacts are depicted in stick representation. The backbone of ERCC1 is shown in blue, XPF is shown in green, and interacting amino acids from either side are shown in orange and yellow, respectively. Intermolecular hydrogen bonds are indicated with black dotted lines.

(B) Top view of the central core. For the sake of clarity, helices ε have been removed.

(C and D) (C) The tight hydrophobic packing of ERCC1 Phe293 to its XPF cavity and (D) the same for Phe894 of XPF in the hydrophobic pocket of ERCC1. In (B), (C), and (D), amino acids are numbered according to the native sequences.

what really determines the specificity remains largely unclear. The crystal structures of the archaeal XPF homodimer and bacterial RuvA bound to DNA have revealed a similar pattern of contacts between GhG hairpins and the minor groove of DNA. Also, in bacterial UvrC (Singh et al., 2002), similar or identical residues that form the double HhH motif were found to be crucial for DNA binding. We observed the same residue contacts for the second hairpin of ERCC1 in our chemical shift perturbation experiments. Even on the first hairpin, which is less homologous with the classic GhG hairpins, the perturbed residues agree very well with the crystallographic complexes. Sequence analysis of ERCC1 and XPF members revealed that residues critically involved in DNA binding in our NMR data are strictly conserved in the ERCC1 family, but are mostly absent in XPF. Remarkably, most of these ERCC1 residues are present and conserved in archaeal XPF and are shown to be in contact with DNA (Newman et al., 2005).

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

Functional Implications of the Intimate Association between the HhH Domains of ERCC1 and XPF

ERCC1/XPF is the last factor arriving at the DNA damage site, resulting in the mature preincision NER complex. It has been shown that the binding of ERCC1/XPF depends on a multitude of sequential prerequisites, including the TFIIH DNA unwinding step (Mone et al., 2004), the presence of XPG at the opposite cleavage site, and the binding of XPA-RPA (Riedl et al., 2003; Volker et al., 2001). Indeed, the middle domain of ERCC1 can interact specifically with XPA (Bessho et al., 1997; Li et al., 1995). Furthermore, we have shown here that the ERCC1

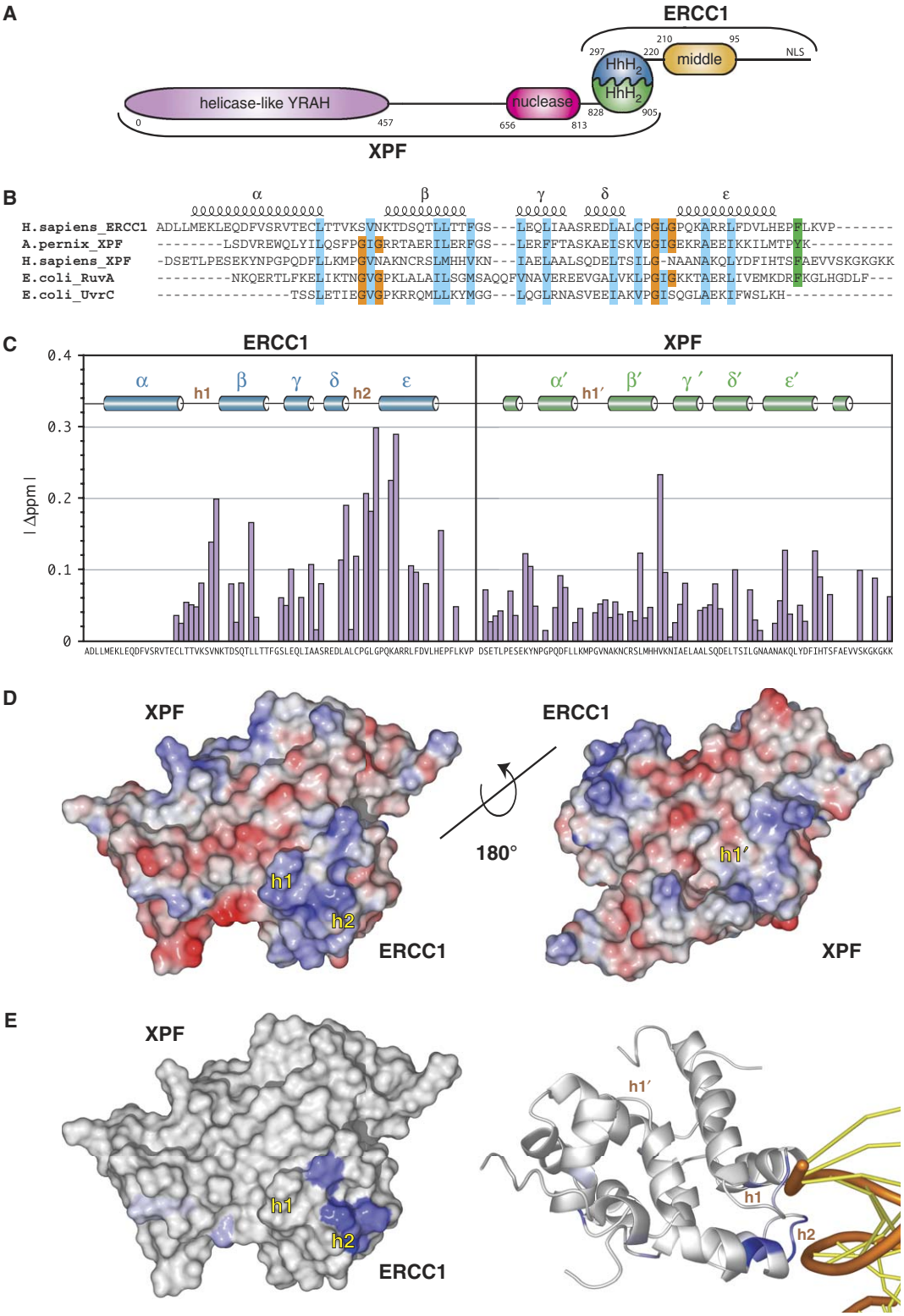

Figure 4. DNA Binding by the ERCC1/XPF Complex

(A) Domain organization of the ERCC1/XPF complex.

(B) Sequence alignment of selected $(HhH)_2$ domains from the three kingdoms. Hydrophobic residues are highlighted in cyan, Gly that belong to hairpin motifs are highlighted in orange, and the aromatic residues at the end of the domains are highlighted in green. The observed secondary elements of ERCC1 are indicated above the sequences.

(C) Normalized chemical shift changes upon DNA titration versus the ERCC1 and XPF sequences. Missing bars indicate either Pro residues or unresolved chemical shifts due to peak overlap.

$(HhH)_2$ domain is sufficient to mediate DNA binding (Figure 4). We propose that ERCC1 combines DNA binding activity and specific interactions with XPA to recruit the endonuclease activity of XPF to the incision complex. This guarantees that XPF cleavage is restricted only to correctly processed DNA templates. This model for ERCC1/XPF is in full accordance with the assembly of NER factors at sites of DNA damage in a sequential and interdependent manner. Our interpretation emphasizes the functional distinction of the tightly associated partners with ERCC1 making all the contacts with DNA and XPA that will target the nuclease of XPF to perform the strand cleavage 5′ upstream of the lesion.

Experimental Procedures

Protein Expression and Purification

Earlier experiments revealed that ERCC1 was expressed in inclusion bodies and that resolubilization of ERCC1 with urea and refolding did not result in a properly folded protein. In addition, we have mapped the interaction domains between ERCC1 and XPF to residues 224–297 and 814–905 for ERCC1 and XPF, respectively (de Laat et al., 1998b). To obtain a stable heterodimeric complex with minimal flanking sequences, putative domain boundaries were determined by using bioinformatics tools as described previously (Folkers et al., 2004). For both domains, several N-terminal deletion constructs around the previously identified domain (ERCC1: 220, 224, 234; XPF: 813, 823, 832) were tested by using in vitro interaction assays (described in de Laat et al., 1998b). From these interaction experiments, we concluded that the minimal interaction domain is composed of residues 220–297 and 823–905 for ERCC1 and XPF, respectively. Therefore, we constructed a bicistronic expression vector in which we first cloned the XPF sequence 823–905 (all oligonucleotide sequences used are available upon request) in the NcoI-BamH1 site of pET28b. Subsequently, in the BamH1-XhoI site of this XPF expression construct, the ERCC1 fragment 220–297, extended with a 27 base pair linker containing an optimal internal ribosome binding site, was cloned, resulting in a bicistronic expression construct in which only the ERCC1 domain contains the His_6 tag.

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

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

DNA Binding

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

For DNA titrations, we used 0.125 mM of ^{15}N-labeled ERCC1/XPF complex in NMR buffer and added increasing amounts of a hairpin sequence in an identical buffer, with 22 unpaired bases (5′-GCCAGC GCTCGGTTTTTTTTTTTTTTTTTTTTTTCCGAGCGCTGGC-3′) until a final concentration of 0.25 mM was reached. ^{15}N-^{1}H HSQC spectra were recorded for the different titration points. Normalized chemical shift changes (δ) for the nonoverlapping residues were calculated by using the equation: $\delta = ([5\delta_{HN}]^2 + [\delta_N]^2)^{0.5}$ (Grzesiek et al., 1996).

NMR Spectroscopy

All NMR spectra were recorded on Bruker AVANCE 700 and AVANCE 900 spectrometers equipped with triple-resonance gradient probes at 295.5 K, using [U-^{13}C, ^{15}N]- and [U-^{15}N]-labeled proteins. Backbone and side chain ^{1}H, ^{15}N, and ^{13}C resonances for ERCC1/XPF were assigned by using the following set of 3D triple-resonance experiments: HNCACB, CBCA(CO)NH, HN(CA)HA, HBHA(CBCACO)NH, HNCO, HN(CA)CO, C(CO)NH-TOCSY, H(CCO)NH-TOCSY, H(C)CH-TOCSY, and (H)CCH-TOCSY spectra acquired at 700 MHz. A full set of 3D heteronuclear-edited NOE spectra were then recorded at 900 MHz for structure determination: 3D NOESY-(^{13}C,^{1}H)-HSQC (τ_{mix} = 75 ms; tuned once for aliphatic and once for aromatic ^{13}C), 3D NOESY-(^{15}N,^{1}H)-HSQC (τ_{mix} = 80 ms), 3D (^{13}C)-HMQC-NOESY-(^{13}C,^{1}H)-HSQC (τ_{mix} = 75 ms), and 3D (^{13}C)-HMQC-NOESY-(^{15}N-^{1}H)-HSQC (τ_{mix} = 80 ms, tuned for aliphatic ^{13}C). These were complemented by homonuclear 2D NOE spectra recorded without (τ_{mix} = 80 ms) and with H[^{15}N] suppression in F2 (τ_{mix} = 80 ms). The triple-resonance and heteronuclear 3D experiments were performed essentially as described by (Cavanagh et al., 1996). All spectra were processed by using the NMRPipe software package (Delaglio et al., 1995) and were analyzed with Sparky (Goddard and Kneller, 2004).

Structure Calculations

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

(D) Two views of surface representation rotated by 180° and colored according to their electrostatic surface potential at ±8 kB T/e for positive (blue) or negative (red) charge potential by using the program GRASP (Nicholls, 1993).
(E) Observed chemical shift changes on the protein surface and model for the interaction of ERCC1/XPF with DNA by fitting ERCC1 on the archaeal $(HhH)_2$ domain. The blue parts denote shifting residues in both representations, and the intensity of the color corresponds to the absolute shift value (cut-off value of 0.1).

refinement run by using CNS (Brünger et al., 1998) according to the standard RECOORD protocol (Nederveen et al., 2005). The final structures were validated with WHATIF (Vriend, 1990) and PROCHECK (Laskowski et al., 1993; Morris et al., 1992). Molecular images were generated with PyMol/NUCCYL (Delano, 2002).

Supplemental Data

Supplemental Data including Figures S1 and S2 are available at http://www.structure.org/cgi/content/full/13/12/1849/DC1/.

Acknowledgments

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

Received: June 24, 2005
Revised: August 26, 2005
Accepted: August 28, 2005
Published: December 13, 2005

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Goddard, T.D., and Kneller, D.G. (2004). SPARKY 3 (computer program). University of California, San Francisco.

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

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

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

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

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

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

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

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

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., et al. (2005). RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins *59*, 662–672.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Zhu, X.D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J.H., and de Lange, T. (2003). ERCC1/XPF removes the 3′ overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Mol. Cell *12*, 1489–1498.

Accession Numbers

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