

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Structural and biochemical analysis of the
UvrA-binding module of the bacterial
transcription-repair coupling factor Mfd**



Nora Aßenmacher

aus

Paderborn

München, 2006

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 09.10.2006

.....
(Nora Aßenmacher)

Dissertation eingereicht am 09.10.2006

1. Gutachter Herr Prof. Dr. Karl-Peter Hopfner

2. Gutachter Herr Prof. Dr. Patrick Cramer

Mündliche Prüfung am 18.12.2006

The presented thesis was prepared in the time from June 2002 to June 2006 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-University of Munich (LMU).

Parts of this PhD thesis have been published:

Assenmacher, N. and Hopfner K.-P. (2004).

MRE11/RAD50/NBS1: complex activities (Review). *Chromosoma* **113**(4): 157-166.

Assenmacher, N., Wenig, K., Lammens, A. and Hopfner, K.-P. (2006).

Structural basis for transcription coupled repair: the N-terminus of Mfd resembles UvrB with degenerate ATPase motifs. *Journal of Molecular Biology* **355**(4): 675-683.

"[One] topic we touched on was mutation ... We totally missed the possible role of ... [DNA] repair although ... I later came to realise that DNA is so precious that probably many distinct repair mechanisms would exist. Nowadays, one could hardly discuss mutation without considering repair."

Francis Crick in "The double helix: a personal view" (Crick, 1974).

Table of contents

1	<u>INTRODUCTION</u>	1
1.1	DNA REPAIR	1
1.1.1	<i>Nucleotide excision repair</i>	2
1.1.1.1	NER in bacteria	3
1.1.1.2	Comparison of NER in eukaryotes to the bacterial system	5
1.1.2	<i>Transcription-coupled DNA repair</i>	5
1.1.2.1	Transcriptional arrest and rescue	6
1.1.2.2	The Mfd protein is the bacterial transcription-repair coupling factor	6
1.1.2.3	Domain architecture and biochemical properties of Mfd	8
1.1.2.4	UvrA binding	10
1.1.2.5	Eukaryotic TCR	11
1.2	STRUCTURE DETERMINATION BY X-RAY CRYSTALLOGRAPHY	12
1.2.1	<i>Structural biology</i>	12
1.2.2	<i>Structure determination by X-ray crystallography</i>	13
1.2.2.1	Theory of X-ray diffraction	13
1.2.2.2	Structure factors and electron density	14
1.2.2.3	Phasing by use of anomalous dispersion	15
1.3	OBJECTIVES	18
2	<u>MATERIALS AND METHODS</u>	20
2.1	MATERIALS	20
2.2	MOLECULAR BIOLOGY METHODS	20
2.2.1	<i>Cloning</i>	20
2.2.2	<i>Site-directed mutagenesis</i>	21
2.3	MICROBIOLOGY METHODS	23
2.3.1	<i>Transformation of E.coli</i>	24
2.3.2	<i>Protein expression</i>	25
2.3.3	<i>Selenomethionine-labelling</i>	25
2.4	PROTEINCHEMICAL METHODS	27
2.4.1	<i>Protein purification</i>	27
2.4.2	<i>Protein-protein interaction assay</i>	28
2.5	PROTEIN ANALYSIS	29
2.5.1	<i>Analytical size exclusion chromatography</i>	29
2.5.2	<i>Limited proteolysis</i>	29
2.5.3	<i>Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)</i>	30
2.5.4	<i>Protein sequencing (Edman, 1950)</i>	30
2.5.5	<i>Matrix assisted laser desorption ionisation Time-of-Flight analysis</i>	30

2.6	FUNCTIONAL ASSAYS.....	31
2.6.1	<i>ATPase activity assay</i>	31
2.6.2	<i>DNA binding assay</i>	32
2.7	STRUCTURAL ANALYSIS OF MFD-N2.....	33
2.7.1	<i>Protein crystallization by sitting drop vapour diffusion</i>	33
2.7.2	<i>Crystallization of Mfd-N2</i>	35
2.7.3	<i>Data collection, structure determination, model building and refinement</i>	35
3	<u>RESULTS</u>	37
3.1	FULL-LENGTH <i>E. COLI</i> MFD.....	37
3.1.1	<i>Purification and crystallization of full-length Mfd</i>	37
3.1.2	<i>Limited proteolysis</i>	38
3.2	PURIFICATION, CRYSTALLIZATION AND STRUCTURE DETERMINATION OF MFD-N2.....	41
3.2.1	<i>Purification of Mfd-N2</i>	41
3.2.2	<i>Crystallization</i>	42
3.2.3	<i>Data collection</i>	44
3.2.4	<i>Structure determination and refinement</i>	45
3.2.5	<i>Mfd-N2 crystallized with two molecules in the asymmetric unit</i>	49
3.3	STRUCTURE OF MFD-N2.....	51
3.3.1	<i>Mfd-N2 crystal structure</i>	51
3.3.2	<i>Conservation of the Mfd N-terminus</i>	52
3.3.3	<i>Comparison of Mfd-N2 to UvrB</i>	53
3.3.4	<i>Domain 2</i>	57
3.3.4.1	<i>Superposition of domain 2 of Mfd and UvrB</i>	57
3.3.4.2	<i>Potential interaction sites</i>	58
3.3.4.3	<i>Interaction of Mfd mutants with UvrA</i>	60
3.3.5	<i>Functional sites</i>	64
3.3.5.1	<i>The Mfd N-terminus does not bind to DNA</i>	64
3.3.5.2	<i>The Mfd N-terminus contains a degenerated ATPase motif</i>	65
4	<u>DISCUSSION</u>	70
4.1	<i>THE MFD N-TERMINUS RESEMBLES UVRB</i>	70
4.2	<i>THE ROLE OF MFD IN RECRUITMENT OF THE UVRA-UVRB COMPLEX</i>	73
5	<u>SUMMARY</u>	78
6	<u>REFERENCES</u>	79

<u>7</u>	<u>SUPPLEMENTARY MATERIAL.....</u>	<u>86</u>
7.1	STABLE FRAGMENTS OF MFD	86
7.2	ABBREVIATIONS.....	89
7.3	AMINOACIDS AND NUCLEOTIDES.....	91
<u>8</u>	<u>CURRICULUM VITAE</u>	<u>92</u>
<u>9</u>	<u>ACKNOWLEDGEMENTS.....</u>	<u>93</u>

1 Introduction

1.1 DNA repair

DNA, the carrier of genetic information, is constantly threatened by a variety of damaging agents. Sources of DNA damage can be either exogenous (like chemicals or radiation) or endogenous (reactive metabolites like oxygen radicals or replication errors). They affect either the nucleobases or the backbone of the DNA helix (Lodish *et al.*, 2000; Hoeijmakers, 2001). Examples for common DNA lesions are listed in table 1.1.

Table 1.1: DNA damage types (according to Lodish *et al.*, 2000; Hoeijmakers, 2001).

DNA damage types	Examples	Caused by
Base modifications	Oxidation: 8-oxoguanine	Oxygen radicals
	Alkylation: 7-methylguanine	Alkylating reagents
	Deamination of cytosine to uracil	Spontaneous deamination
Mismatches	G/T or A/C pairs	Replication errors
Breaks in the backbone	Single strand breaks (SSBs)	Ionizing radiation or chemicals
	Double strand breaks (DSBs)	
Bulky photoadducts	Cyclobutane-pyrimidine dimers (CPDs), 6-4-Photoproducts	UV radiation
Cross-links	Intrastrand cross-links	Cross-linking agents (bifunctional alkylating agents)
	Interstrand cross-links	

If left unrepaired, these DNA lesions can lead to mutations – which may in higher organisms result in cancer – or cell death.

DNA damaging agents are often used as chemotherapeutics in cancer therapy in order to inhibit DNA replication and therefore stop cell division. In particular, DNA cross-linking agents, e.g. *cis*-diammine dichloroplatinum(II) (cisplatin) or mitomycin C, are applied (Jamieson and Lippard, 1999; Siede *et al.*, 2005).

Cells have evolved multiple repair mechanisms, which use different enzymes and deal with different kinds of lesions (see table 1.2) (reviewed in Lindahl and Wood, 1999; Hoeijmakers, 2001; Alberts *et al.*, 2002; Siede *et al.*, 2005; Friedberg *et al.*, 2006). In humans, several inherited disorders were found to be associated with defects in DNA

damage repair genes (see chapters 1.1.1.2. and 1.1.2.5.). Many of these syndromes are characterized by premature ageing and cancer predispositions (Hoeijmakers, 2001).

Table 1.2: DNA repair systems (Friedberg *et al.*, 2006)

Repair mechanism	Repair systems	Applied to
Direct damage reversal	Photoreactivation	Photoproducts (CPDs)
	Oxidative demethylation	Alkylated bases
	Ligation of SSBs	SSBs
Damage removal (Excision repair)	Nucleotide excision repair (NER)	Bulky, helix-distorting lesions like photoproducts, cisplatin-adducts, or cross-links
	- Global genome repair (GGR)	
	- Transcription coupled repair (TCR)	
	Base excision repair (BER)	Modified bases
Double strand break (DSB) repair	Mismatch repair (MMR)	Single-base mispairs
	Homologous recombination (HR)	Double strand breaks
Damage tolerance	Non-homologous end-joining (NHEJ)	
	Trans-lesion synthesis (TLS)	

In the following chapters, the repair systems of nucleotide excision repair (NER) and transcription-coupled repair (TCR) will be described in more details.

1.1.1 Nucleotide excision repair

Nucleotide excision repair (NER) is a functionally conserved DNA repair system which can be found in all kingdoms of life (Sancar, 1996; Ogrunc *et al.*, 1998; Batty and Wood, 2000). NER deals with a broad range of chemically and structurally unrelated helix-distorting DNA lesions like UV-induced photoproducts, bulky chemical adducts as well as inter- and intrastrand cross-links (Sancar and Rupp, 1983; Batty and Wood, 2000; Van Houten *et al.*, 2005). The basic NER mechanisms have been strongly conserved throughout evolution, although the enzymes involved in the process differ between prokaryotes and eukaryotes (Batty and Wood, 2000). Interestingly, some mesophilic Archaea use the bacterial system, while in most Archaea, proteins homologous to eukaryotic nucleotide excision repair factors are found (Kelman and White, 2005).

1.1.1.1 NER in bacteria

In bacteria (and in some archaea as well), nucleotide excision repair is mediated by the UvrABC system (reviewed in Batty and Wood, 2000; Van Houten *et al.*, 2005; Truglio *et al.* 2006).

Upon ATP-binding, UvrA dimerizes (Mazur and Grossman, 1991) and forms a complex with UvrB which contains either the UvrA₂-UvrB₁ heterotrimer (Orren and Sancar, 1989) or the UvrA₂-UvrB₂ heterotetramer (Verhoeven *et al.*, 2002). This so-called UvrAB damage-recognition complex binds to DNA and scans the molecule for sites of helix-distorting DNA lesions.

The role of the second UvrB subunit is still being discussed. The UvrA-dimer seems to interact directly with only one UvrB molecule, while the second UvrB binds to the first one. The UvrB-dimer is proposed to function in damage recognition in both DNA strands. The second UvrB dissociates upon UvrC-binding (Hildebrand and Grossman, 1999; Verhoeven *et al.*, 2002).

After damage verification, UvrB is loaded onto the damaged DNA, and UvrA dissociates from the lesion site (Orren and Sancar, 1989; Sancar and Hearst, 1993). DNA becomes wrapped around UvrB (Verhoeven *et al.*, 2001), and UvrB inserts a hairpin motif ("β-hairpin") into the DNA duplex (Truglio *et al.*, 2006b). This step is energy-dependent and requires ATP hydrolysis both by UvrA and UvrB (Van Houten *et al.*, 1988; Moolenaar *et al.*, 2000). UvrB possesses cryptic helicase activity (Orren and Sancar, 1989; Theis *et al.*, 2000; Verhoeven *et al.*, 2002) which is proposed to function in destabilization of the double-helix, so that UvrB can insert the β-hairpin between the two strands (Skorvaga *et al.*, 2004; Truglio *et al.*, 2006a). A kinetic analysis has shown that the formation of the UvrB-DNA pre-incision complex (PIC) is the rate-limiting step of the NER process (Orren and Sancar, 1990).

UvrC is then recruited to the lesion. The UvrB C-terminus interacts with a homologous region located in the N-terminal half of UvrC (Moolenaar *et al.*, 1998; Sohi *et al.*, 2000) UvrC mediates two incisions in the damaged strand (Lin and Sancar, 1992; Verhoeven *et al.*, 2000): The first incision takes place 3 or 4 nucleotides 3' to the lesion. This step requires ATP-binding by UvrB (Orren and Sancar, 1990; Moolenaar *et al.*, 2000; Goosen and Moolenaar, 2001; Truglio *et al.*, 2005). The second incision by UvrC is independent of UvrB. It occurs at the eighth phosphodiesterbond 5' to the damage site (Moolenaar *et al.*, 1995). The first incision is performed by the N-terminal part of UvrC, while the C-terminal

half mediates the 5'-incision. UvrD (helicase II) mediates removal of UvrC and the excised 12- or 13-mer oligonucleotide. DNA polymerase I displaces UvrB and fills the gap. The DNA is finally sealed by DNA ligase (Caron *et al.*, 1985; Husain *et al.*, 1985).

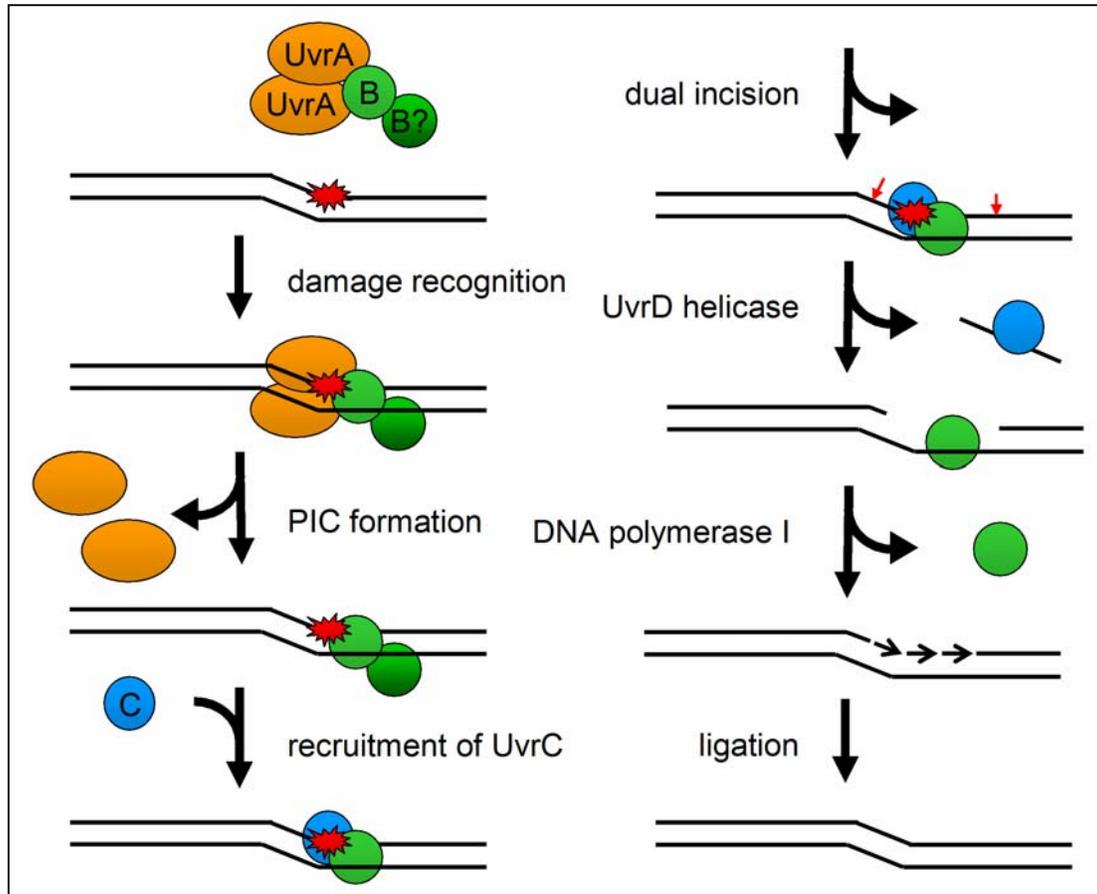


Figure 1.1: Bacterial nucleotide excision repair is mediated by the UvrABC system (adopted from Goosen and Moolenaar, 2001). The mechanism of this repair pathway is described in the text (DNA lesion, red star; UvrA, orange; UvrB, green; UvrC, blue).

A second endonuclease in nucleotide excision repair, Cho (UvrC homologue), was recently identified. Cho is a homologue to the N-terminal part of UvrC. Moolenaar and colleagues pointed out that, in the presence of UvrA and UvrB, Cho can perform the 3' incision to damaged DNA. The incision takes place 4 nucleotides further downstream than the UvrC-mediated cleavage. The incised DNA is further processed by UvrC (Moolenaar *et al.*, 2002).

1.1.1.2 Comparison of NER in eukaryotes to the bacterial system

Nucleotide excision repair in eukaryotic cells shows many remarkable similarities to the bacterial system. The general mechanism is highly conserved, but different proteins are used in the repair process. While bacteria require only three proteins – UvrA, UvrB and UvrC – for damage recognition and the incision reactions, human NER uses up to 19 polypeptides for these steps (table 1.3) (Sancar, 1996; Batty and Wood, 2000; Coin *et al.*, 2006).

Table 1.3: Proteins involved in bacterial and eukaryotic nucleotide excision repair (adopted from Batty and Wood, 2000).

	bacteria	human (yeast)
Damage recognition	UvrA ₂ B _{1/2}	XPC–hHR23B (Rad4–Rad23)
Opening, pre-incision complex	UvrB	XPA (Rad14), RPA, TFIIH
3'-incision	UvrC	XPG (Rad2)
5'-incision	UvrC	ERCC1–XPF (Rad10–Rad1)
Excision, repair synthesis	UvrD, PolI, ligase	PCNA, ligase, RFC, RPA, Pol δ/ε

Humans deficient in nucleotide excision repair suffer from xeroderma pigmentosum (XP), a rare, recessively inherited disease which is mainly characterized by extreme UV sensitivity, parchment skin ("xeroderma") and freckles ("pigmentosum"). XP is associated with an increased risk to develop malignancies, especially skin cancers (Cleaver and Kraemer, 1995; de Boer and Hoeijmakers, 2000; Andressoo and Hoeijmakers, 2005).

1.1.2 **Transcription-coupled DNA repair**

Damage repair in active genes occurs much faster than in the overall genome. This phenomenon is not due to the better accessibility of transcribed DNA regions. It could be shown that the higher repair rate is only true for the transcribed strand, while the non-transcribed strand is repaired at the same rate than the overall genome. These observations lead to the discovery of a mechanism called transcription-coupled repair (TCR) (Mellon *et al.*, 1986; Mellon *et al.*, 1987; Mellon and Hanawalt, 1989). Transcription-coupled repair is present in eukaryotes as well as in prokaryotes.

1.1.2.1 Transcriptional arrest and rescue

RNA polymerase (RNAP) pauses frequently during transcriptional elongation (Fish and Kane, 2002). This transcriptional pausing is a temporary interruption of transcription. Paused polymerase slowly changes its conformation and isomerizes to an arrested state which is accompanied by reverse movement ("backtracking") by several nucleotides (Komissarova and Kashlev, 1997). In this position, the active site is not aligned with the RNA 3'-hydroxyl end, and RNA polymerase cannot resume transcription by itself. In theory, a single RNA polymerase molecule which is stalled irreversibly in an essential gene could cause cell death (Svejstrup, 2002a).

Transcriptional arrest may occur in case of nucleotide starvation or when RNA polymerase encounters a roadblock like a DNA-binding protein (Komissarova and Kashlev, 1997; Fish and Kane, 2002). In addition, intrinsic signals in DNA and RNA have been identified to cause pausing or arrest of RNA polymerase (Artsimovitch and Landick, 2000). Transcription factors (GreA/GreB in bacteria, TFIIS in eukaryotes) are required to reactivate arrested RNA polymerase by inducing internal cleavage of the RNA (Reines *et al.*, 1992; Tornaletti *et al.*, 1999; Kettenberger *et al.*, 2003; Nickels and Hochschild, 2004). Thereby, a new 3'-OH-end is created, and transcription proceeds in the correct DNA-RNA register (Borukhov *et al.*, 1993).

However, one of the most common causes for transcriptional arrest is a non-coding lesion in the transcribed DNA strand (Tornaletti and Hanawalt, 1999), over which transcription cannot continue. Instead, dissociation of the transcription elongation complex and recruitment of damage repair proteins take place. This process is known as transcription-coupled repair (TCR) (reviewed in Svejstrup, 2002a; Mellon, 2005; Saxowsky and Doetsch, 2006).

TCR is mainly considered as a sub-pathway of nucleotide excision repair (NER), but Le Page and colleagues could demonstrate recently that TCR also plays a role in the base excision repair (BER) system (Le Page *et al.*, 2000).

1.1.2.2 The Mfd protein is the bacterial transcription-repair coupling factor

In *Escherichia coli*, the Mfd protein was identified to be responsible for connecting the processes of transcription and DNA repair. Mfd is therefore also termed transcription-repair coupling factor (TRCF) (Selby *et al.*, 1991).

Mfd was already discovered in the 1960s by E. Witkin in a genetic screen. Mutation frequency decline (Mfd) stands for a rapid decrease in the frequency of occurring nonsense suppressor mutations when protein synthesis is transiently inhibited immediately after UV-irradiation (Witkin, 1966). In cells lacking the *mfd* gene product this phenomenon is strongly reduced. Additionally, *mfd*⁻ cells are characterized by a high spontaneous mutation rate, increased sensitivity to UV, and a decreased damage repair rate (Selby and Sancar, 1993). Furthermore, it could be shown that *mfd*⁻ cells are incapable of strand-specific repair. This defect could be complemented by adding purified Mfd protein (Selby *et al.*, 1991).

Recently, Park and colleagues could demonstrate that Mfd is able to rescue stalled RNA polymerase to resume transcription elongation. Mfd binds to template DNA upstream of the transcription machinery. In contrast to GreA/GreB, Mfd acts by translocating the backtracked transcription elongation complex forward. Thereby, the catalytic center is realigned with the original 3'-OH end of the transcript, and RNA synthesis is allowed to resume (Park *et al.*, 2002) (figure 1.2, left panel).

However, if RNA polymerase is blocked by non-coding lesions, productive transcription cannot proceed. In this case, Mfd induces dissociation of the RNA polymerase from template DNA in an ATP-dependent manner (Selby and Sancar, 1995b; Selby and Sancar, 1995a; Park *et al.*, 2002) (figure 1.2, right panel). Mfd is suggested to induce the dissociation of RNA polymerase by pushing it hard over the damage (Park *et al.*, 2002).

Mfd, remaining bound at the lesion, then recruits the nucleotide excision repair machinery. Mfd interacts with UvrA and recruits the UvrA-UvrB complex to the damage site (Selby and Sancar, 1993). Therefore, Mfd can be considered both as a transcription elongation factor (Park *et al.*, 2002; Borukhov *et al.*, 2005) and as a terminator of transcription (Roberts and Park, 2004).

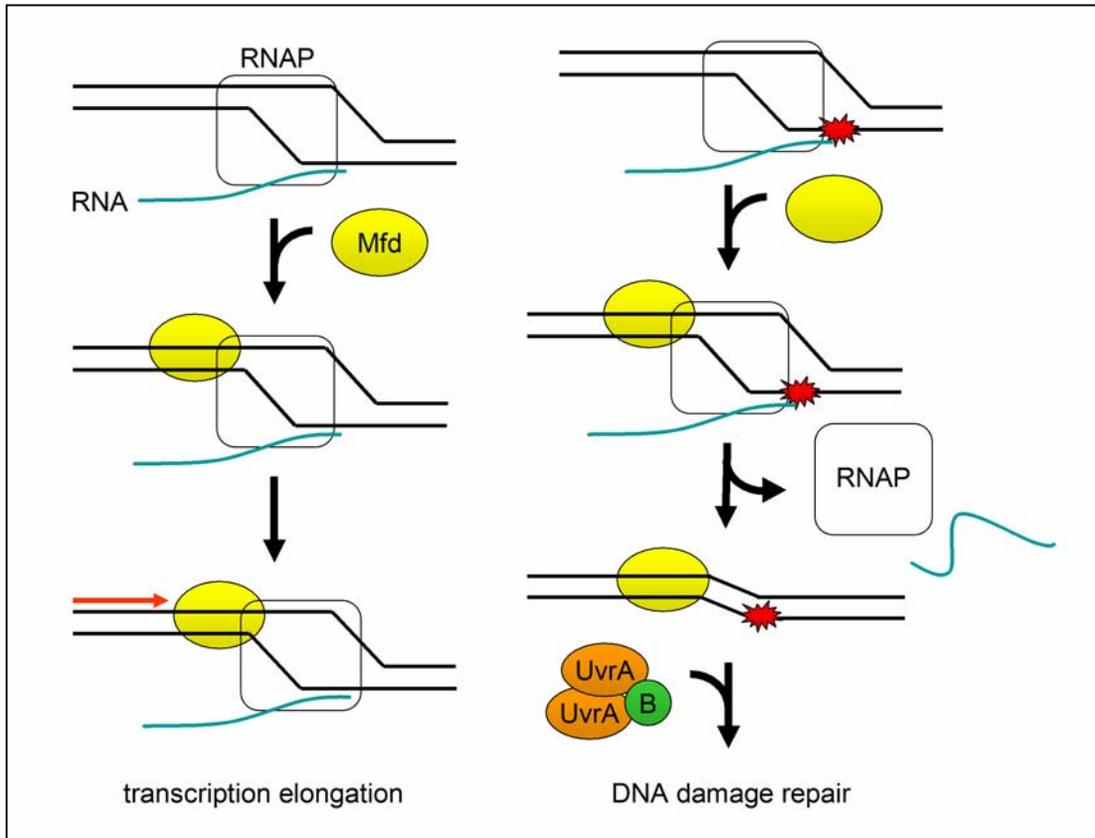


Figure 1.2: Mechanism of bacterial TCR (adopted from Svejstrup, 2002b; Roberts and Park, 2004; Mellon, 2005). Left panel: RNA polymerase (transparent box) is stalled in the backtracked position, the active site and the 3'-end of the RNA (light blue) are not aligned. Mfd (yellow sphere) recognizes stalled RNA polymerase and binds upstream of the transcription elongation complex. Subsequently, Mfd induces forward translocation leading to resumption of transcription (left). Right panel: If RNAP is blocked by DNA damage (red star), transcription elongation cannot resume. Mfd promotes release of RNA polymerase and RNA from the transcribed DNA followed by recruitment of the UvrABC nucleotide excision repair system (UvrA, orange; UvrB, green).

The mechanism of UvrAB recruitment and the subsequent formation of the UvrB-DNA pre-incision complex at stalled transcription sites, however, are not fully understood.

1.1.2.3 Domain architecture and biochemical properties of Mfd

Mfd is a highly conserved monomeric protein. With 1148 residues (130 kDa), it is among the largest 1% of all *Escherichia coli* proteins (Roberts and Park, 2004). Selby and Sancar

could show that Mfd consists of distinct functional regions (figure 1.3) (Selby and Sancar, 1995a; Roberts and Park, 2004).

The N-terminal third of Mfd is involved in UvrA binding. It will be described in more detail in chapter 1.1.2.4.

The RNA polymerase interacting domain of Mfd (RID, residues 379-571) binds to the N-terminus (the first 142 residues) of the RNA polymerase β subunit, near the upstream junction of the transcription bubble (Selby and Sancar, 1995a; Park *et al.*, 2002; Smith and Savery, 2005). In the presence of the $\sigma 70$ factor, RNAP binding by Mfd is blocked (Park *et al.*, 2002).

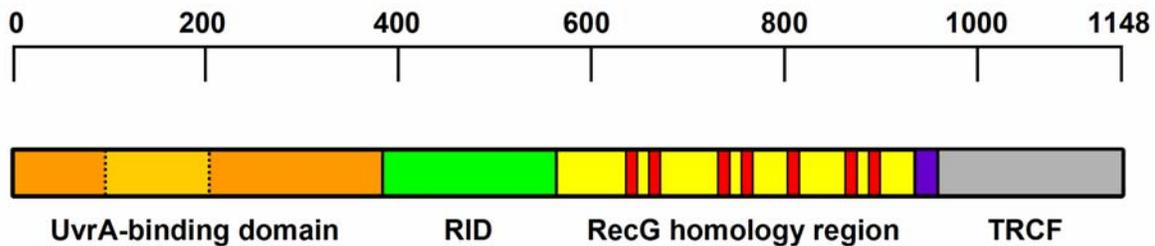


Figure 1.3: Functional domains of the E.coli Mfd protein (adopted from Selby and Sancar, 1995a/b; Roberts and Park, 2004). Residues 1-378, orange: UvrA-binding domain, includes UvrB homology region (residues 82-219, light orange); residues 379-571, green: RNA polymerase interacting domain (RID); residues 598-968, yellow: RecG homology domain, includes dsDNA translocase region with superfamily II helicase motifs (red) and translocation in RecG (TRG) motif (residues 926-965, violet); residues 1005-1113, grey: TRCF domain.

A module related to superfamily II helicases is located in the C-terminal half of Mfd (Gorbalenya *et al.*, 1989; Selby and Sancar, 1993; Mahdi *et al.*, 2003).

This section of the protein bears the DNA binding region and contains the DNA-stimulated ATPase activity of Mfd. Binding to preferentially double stranded polynucleotides requires ATP binding while ATP hydrolysis promotes dissociation (Selby and Sancar, 1995a). Despite its homology to the RecG helicase, Mfd possesses no strand-separating activity. It is rather believed to function as dsDNA translocase (Selby and Sancar, 1995b).

In addition, Mfd and RecG have a helical hairpin motif in common, termed TRG (translocation in RecG) motif which is situated downstream of the translocase domain (Chambers *et al.*, 2003; Mahdi *et al.*, 2003). Mutational analysis confirmed that the TRG

motif in Mfd is required for its RNA polymerase displacement activity (Chambers *et al.*, 2003).

The C-terminal TRCF domain (residues 1005-1113; pfam03461) contains a leucine zipper motif (L1039, L1046, L1053, L1060). It is essential for RNAP release from damaged DNA (Selby and Sancar, 1993).

1.1.2.4 UvrA binding

Mfd acts as a platform for recruiting the nucleotide excision repair machinery to DNA lesions by binding to UvrA (Selby and Sancar, 1993). This process involves the N-terminus of Mfd. A truncated mutant of Mfd lacking the first 378 residues was shown to be defective in UvrA binding (Selby and Sancar, 1995a).

The UvrA-binding region bears a section with close homology to a region in the nucleotide excision repair protein UvrB: Residues 82–219 of Mfd are 22 % identical (62 % homologous) to residues 114-251 of UvrB from the same organism (Selby and Sancar, 1993) (figure 1.4). In both proteins, this section has been shown to play a role in UvrA binding *in vitro* (Hsu *et al.*, 1995; Selby and Sancar, 1995a; Truglio *et al.*, 2004).

There is some evidence that Mfd and UvrB utilize a similar mode of binding to UvrA. In addition to the high sequence homology in the domain 2, Mfd is able to displace UvrB from UvrA *in vitro* (Selby and Sancar, 1993).

EcoMfd	82	S	P	H	D	I	L	S	S	R	L	S	T	L	Y	Q	L	P	T	M	Q	R	G	V	L	I	V	P	V	N	T	L	M	Q	R	V	C	P	H	S	F	L	H	G	H	A	L	V	M	K	K	C	O	R	L	S	R	D	A	L	141	
EcoUvrB	114	S	V	N	E	H	I	E	Q	M	R	L	S	A	T	K	A	M	L	E	R	R	D	V	V	V	V	A	S	V	S	A	I	Y	G	L	G	D	D	L	Y	L	K	-	M	M	L	H	L	T	V	G	M	I	I	D	Q	R	A	I	172	
BcaUvrB	115	K	L	N	D	E	I	D	K	L	R	H	S	A	T	S	A	L	F	E	R	R	D	V	I	V	A	S	V	S	C	I	Y	G	L	G	S	P	E	E	Y	R	E	-	L	V	V	S	L	R	V	G	M	E	I	E	R	N	A	L	173	
EcoMfd	142	R	T	Q	L	D	S	A	G	Y	R	H	V	D	Q	V	M	E	H	C	E	Y	A	T	R	G	A	L	L	D	L	F	P	M	G	-	S	E	I	P	Y	R	L	D	F	F	D	E	I	D	S	L	R	V	F	D	V	S	Q	200		
EcoUvrB	173	L	R	R	L	A	E	L	O	Y	A	R	N	D	Q	A	F	O	R	G	T	F	R	V	R	G	E	V	I	D	I	F	P	A	E	S	D	L	A	L	R	V	E	L	F	D	E	E	V	E	R	L	S	L	F	D	P	T	G	232		
BcaUvrB	174	L	R	R	L	V	D	I	Q	Y	D	R	N	D	I	D	F	R	-	G	T	F	R	V	R	G	D	V	V	E	I	F	P	A	S	R	D	E	H	C	I	R	V	E	F	F	G	D	E	I	E	R	I	R	E	V	D	A	L	T	G	232
EcoMfd	201	R	T	L	E	E	V	E	A	I	N	L	L	P	A	H	E	F	P	T	219																																									
EcoUvrB	233	Q	I	V	S	T	I	P	R	F	T	I	V	P	K	T	H	Y	V	T	251																																									
BcaUvrB	233	K	V	L	G	E	R	H	V	A	L	F	P	A	S	H	F	V	T	251																																										

Figure 1.4: Sequence alignment of residues 82-219 of *Escherichia coli* Mfd (*EcoMfd*) with the corresponding regions of the UvrB proteins from *E. coli* (*EcoUvrB*, residues 114-251) and *Bacillus caldotenax* (*BcaUvrB*, residues 116-251). Conserved residues are shaded, dark indicating stronger conservation.

So far, several molecular structures of UvrB are known. While domain 2 was not defined in the first crystal structures solved (Machius *et al.*, 1999; Nakagawa *et al.*, 1999; Theis *et al.*, 1999), it was visible in the UvrB variant Y96A from *Bacillus caldotenax* (Truglio *et al.*, 2004).

Residues 157-245 form a compact globular domain (denoted domain 2). The structure of this domain revealed a new fold according to structural analysis by Distance matrix alignment (Dali) (Holm and Sander, 1995). Furthermore, some highly conserved residues located on its surface could be identified to be essential for UvrA binding (Truglio *et al.*, 2004). Until 2005, no structure of Mfd was known.

1.1.2.5 Eukaryotic TCR

The phenomenon of transcription-coupled repair is conserved. However, TCR in eukaryotic cells is much more complex than in bacteria and less well studied. Thus, many details are not fully understood (Svejstrup, 2002a; Laine and Egly, 2006; Saxowsky and Doetsch, 2006).

In eukaryotic TCR, RNA polymerase II (RNAPII) functions as the damage recognition factor. Stalled RNAPII is recognized by the proteins XPG and CSB which then recruit, among others, TFIIH and CSA. Excision of the damaged oligonucleotide and repair synthesis in TCR share the protein repertoire with global genome NER (see 1.1.1.2).

Table 1.4: Eukaryotic proteins involved in transcription coupled repair (adopted from Svejstrup, 2002a; Laine and Egly, 2006; Saxowsky and Doetsch, 2006). Repair factors marked with a star () are involved both in GGR and in TCR.*

human (yeast) factors	activities	interaction partners
CSA (Rad28)	E3 ubiquitin ligase	CSB, XAB2, TFIIH
CSB (Rad26)	Swi2/Snf2, DNA binding	CSA, XA2B, TFIIH, RNAPII, XPA; XPG
XAB2 (Syf1)		CSA, CSB, RNAPII, XPA
TFIIH (TFIIH) *	10 subunits including 2 helicases and a cyclin-dependent protein kinase	CSA, CSB, RNAPII, XPG
XPG (Rad2) *	endonuclease (3'-incision)	CSB, RNAPII

Recent findings suggest that the damage is repaired without prior removing RNAPII. TFIIH is thought to induce conformational changes by use of the helicase subunits XPB (Rad25) and XPD (Rad3) (Sarker *et al.*, 2005). A very current model proposes removal of RNA polymerase II together with the lesion (Brueckner *et al.*, 2006).

Mutations in TCR genes (mainly in the genes encoding for CSA or CSB) lead to a severe hereditary disorder named Cockayne's syndrome (CS). CS is characterized by photosensitivity, growth retardation, skeletal and retinal abnormalities and progressive neural degradation. In contrast to xeroderma pigmentosum, CS is not associated with an increased risk of skin cancer or other type of malignancy (Nance and Berry, 1992; de Boer and Hoeijmakers, 2000; Andressoo and Hoeijmakers, 2005).

1.2 Structure determination by X-ray crystallography

1.2.1 Structural biology

Proteins consist of one or more chains of amino acids that fold into three-dimensional structures. The structure of a protein is intrinsically related to its function. Therefore, structure determination of biological macromolecules is a powerful tool to gain information on their biological function and on their mechanism. In addition, structural studies play an important role in protein design and drug development.

Several methods have been developed to determine three-dimensional structures of molecular machines with atomic resolution:

Electron microscopy (EM) is a powerful tool for the determination of large structures, e.g. complexes, organelles or cells. At present, electron cryo-microscopy (cryo-EM) and, in particular, the reconstruction of single-particles can in practice reach a resolution of 4-5 Å. However, its application to small molecules is limited. So far, EM can be used only for particles with a size of at least several hundred kDa (R. Beckmann, DNA repair workshop, July 20, 2006).

In contrast, nuclear magnetic resonance spectroscopy (NMR) can only be applied to small proteins (usually 20-30 kDa). Structures are determined in solution. Therefore, NMR allows time-resolved studies (e.g. folding analysis or kinetics).

X-ray crystallography is a very high resolution method. It has no limitation with respect of molecular weight. However, in order to determine a structure by X-ray diffraction, the

molecule of interest needs to be crystallized which is the major obstacle in this technique (see below). Furthermore, a crystal structure can be considered as a "snapshot". Usually, a crystal structure provides only very little insight into dynamics.

The first protein crystal structure solved was that of sperm whale myoglobin in the 1950s (Kendrew *et al.*, 1958). In 1962, Max Ferdinand Perutz and Sir John Cowdery Kendrew were awarded the Nobel Prize in Chemistry “for their studies of the structures of globular proteins” (<http://nobelprize.org/chemistry/laureates/1962/index.html>).

So far, over 37,000 biological macromolecular structures have been deposited in the RSCB Protein Data Bank (PDB). Most of them (> 90 %) were determined using X-ray crystallography (table 1.5).

Table 1.5: Biological macromolecular structures in the RSCB Protein Data Bank (PDB) (source: <http://www.rcsb.org/pdb/holdings.do>; June 20, 2006).

	Proteins	Nucleic acids (NA)	Protein/ NA complexes	Other	Total
X-ray	29258	902	1353	28	31541
NMR	4690	705	121	6	5522
EM	88	9	29	0	126
Other	73	4	3	0	80
Total	34109	1620	1506	34	37269

1.2.2 Structure determination by X-ray crystallography

The theoretical background of structure determination by X-ray diffraction will be briefly described in the following part. More detailed information can be found in textbooks (e.g. Drenth, 1999; McPherson, 2001; Blow, 2002; Massa, 2002).

1.2.2.1 Theory of X-ray diffraction

X-rays are electromagnetic waves with a wavelength in the range of atomic distances (10^{-10} m = 1 Å). For typical X-ray diffraction experiments, wavelengths between 1.6 and 0.8 Å are used. When an electron is hit by an X-ray photon, it is set into vibration at the X-ray frequency. The vibrating electron emits spherical waves of the same wavelength as the original wave (elastic scattering).

The scattering power of a single molecule in solution is insufficient to generate a detectable signal. Therefore, the molecule of interest needs to be crystallized (see 2.7.1). Crystals are highly ordered structures, where a unit cell containing the molecule of interest is periodically repeated in a three dimensional lattice. Waves scattered from different atoms in a crystal may interfere, and, depending on the phase difference, amplify or damp each other. If the phase shift is proportional to 2π (“in-phase”), the signal is enhanced, and diffraction occurs. The conditions for this constructive interference are given by the Laue equations and Bragg's law.

According to Sir W. H. Bragg and Sir W. L. Bragg, X-ray diffraction by a crystal can be considered as reflections at imaginary lattice planes. Lattice planes are characterized by the Miller indices (hkl) which represent their orientation in the crystal lattice and the spacing between parallel planes.

A signal can only be detected if the distance d and the angle θ between the planes and the incident beam follow the rule (“Bragg’s law”)

$$n \cdot \lambda = 2 \cdot d \cdot \sin \theta$$

where n is an integer, and λ is the wavelength of the X-rays.

Each diffraction spot (h,k,l) corresponds to a reflection at a set of parallel lattice planes (hkl).

1.2.2.2 Structure factors and electron density

The structure factor F is a mathematical description of how the crystal scatters incident radiation. $F(h,k,l)$ is the sum of the scattering contributions of all N atoms in the unit cell to a reflection (h,k,l) .

$$F(h,k,l) = \sum_{j=1}^N f_j \cdot \exp[2\pi \cdot i(hx_j + ky_j + lz_j)] \cdot \exp[-B_i \cdot \sin^2 \theta / \lambda^2]$$

The atomic scattering factor (or form factor) f_j describes the scattering power of an atom j and is dependent on the atom type. The last term of structure factor, the Debye-Waller factor (B-factor), represents the effect of thermal disorder.

The electron density ρ of molecules in a crystal is a three-dimensional repetitive structure. It represents the scattering power of all atoms in the unit cell. Electron density and structure factor are related by Fourier transform (FT): The variation of electron density in a crystal can be used to determine the relative amplitudes and phases of the Fourier

coefficients, the structure factors, by direct FT. Reversely, structure factors can be used to calculate electron density by inverse FT.

Fourier integral (direct FT):

$$F(h,k,l) = V \cdot \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(x,y,z) \cdot \exp[2\pi \cdot (hx + ky + lz)] dx dy dz$$

Fourier series (inverse FT):

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} F(h,k,l) \cdot \exp[-2\pi \cdot i(hx + ky + lz)]$$

The structure factor $F(h,k,l)$ is a complex number which is formed by the amplitude $|F(h,k,l)|$ and the phase $\phi(h,k,l)$ of a scattered wave.

$$F(h,k,l) = |F(h,k,l)| \cdot \exp[i \cdot \phi(h,k,l)]$$

Both amplitude and phase are required for the calculation of an electron density from structure factors. While the amplitude can be derived from the reflection intensity ($I \sim |F|^2$), the phase cannot be directly observed from a diffraction pattern. This is referred to as the “phase problem” of crystallography.

Phase angles can be obtained by several approaches:

If a model for a related molecule is available, it can be used for phase determination by molecular replacement (MR). For *de novo* phasing, heavy atom methods like isomorphous replacement (SIR/MIR) or anomalous dispersion (SAD/MAD) are generally used. Direct methods can be applied only to very small molecules or substructures.

In this PhD thesis, a multiple-wavelength anomalous diffraction (MAD) experiment was carried out using selenium.

1.2.2.3 Phasing by use of anomalous dispersion

If the incident beam possesses an energy close to the eigenfrequency of an atom, some photons are absorbed and re-emitted either at lower energy (fluorescence) or at the same energy with a phase-delay (anomalous dispersion). In case of anomalous dispersion, the atomic scattering factor f_{ano} gains an anomalous contribution which is composed of a real

part Δf and an imaginary part if'' . The phase of the imaginary part if'' is always shifted by $+90^\circ$ (figure 1.5).

$$f_{ano} = f_0 - \Delta f + if'' = f' + if''$$

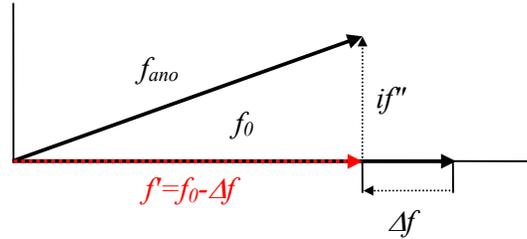


Figure 1.5: The anomalous atomic scattering factor f_{ano} . The anomalous contribution to the atomic scattering factor is composed of a real part Δf and an imaginary component if'' .

Light atoms normally occurring in biological molecules (e.g. carbon, nitrogen or oxygen) do not have transitions in the range of energies that are used in X-ray diffraction experiments. Therefore, heavy metal atoms (e.g. selenium, mercury, platinum etc.) are introduced into proteins. These elements show detectable anomalous scattering at X-ray wavelengths. In this case, scattering can be described as a sum F_{PH} of the normal scattering from light protein atoms F_P and scattering from heavy atoms F_H with a normal (F_{HN}) and an anomalous part (F_{HA}):

$$F_{PH} = F_P + F_H = F_P + F_{HN} + F_{HA}$$

The anomalous signal is dependent on the energy of the X-rays and hence depends on the applied wavelength. Thus, differences in the reflection intensities recorded at different wavelengths close to the absorption edge can be observed.

A consequence of anomalous scattering is the violation of the Friedel's law: In normal scattering, the structure factors describing the reflections (h, k, l) and $(-h, -k, -l)$ have the same amplitudes and opposite phases. $F_P(h, k, l)$ and $F_P(-h, -k, -l)$ are called Friedel mates. In contrast, amplitudes of anomalous structure factors $F_{PH}(h, k, l)$ and $F_{PH}(-h, -k, -l)$ (now named a Bijvoet pair) do not have same magnitudes.

$$d_{ano} = (|F_{PH}(h, k, l)| - |F_{PH}(-h, -k, -l)|) \cdot \frac{f'}{2f''}$$

Using the Bijvoet differences d_{ano} , the heavy atom substructure can be localized by Patterson methods and/or direct methods.

The Patterson function $P(u, v, w)$ does not require phase information. It represents a Fourier transform of squared reflection amplitudes $|F(h, k, l)|$ (\approx intensities). By use of this function, a map showing interatomic distance vectors (u, v, w) is obtained.

$$P(u, v, w) = \sum_{hkl} |F_{hkl}|^2 \exp[-2\pi \cdot i(hu + kv + lw)]$$

The Patterson function calculated with differences in anomalous amplitudes results in a map showing only vectors between the anomalous scatterers. Using this map, heavy atoms can be located in the unit cell. With their coordinates, the contribution of the heavy atoms to the structure factors can be determined. Finally, protein phase angles can be derived, and an electron density can be calculated:

$$F_{PH} = |F_P| \cdot \phi_P + |F_H| \cdot \phi_H$$

The anomalous contribution to diffraction is generally very small. An MAD experiment requires synchrotron radiation which is brighter and less noisy than radiation generated by home sources. In addition, synchrotron sources produce continuous X-ray spectra, and the monochromatic beam is tuneable to the required wavelengths.

In a typical MAD experiment, datasets at three wavelengths are recorded:

A "peak" dataset at the wavelength with maximum f'' , a dataset at the "inflection point" with minimal f' , and a "high energy remote" dataset, where f' is close to normal.

The resonance wavelength of a certain atom is defined by the atom type. But as it can differ slightly dependent on the chemical environment, the exact wavelengths are determined experimentally (figure 1.6). The f'' value can be directly obtained by a fluorescence scan on the crystal, while f' can be derived from f'' by the Kramer-Kronig equation:

$$f' = \frac{2}{\pi} \int_0^{\infty} \frac{\omega' f''(\omega') \delta\omega'}{\omega^2 - \omega'^2}$$

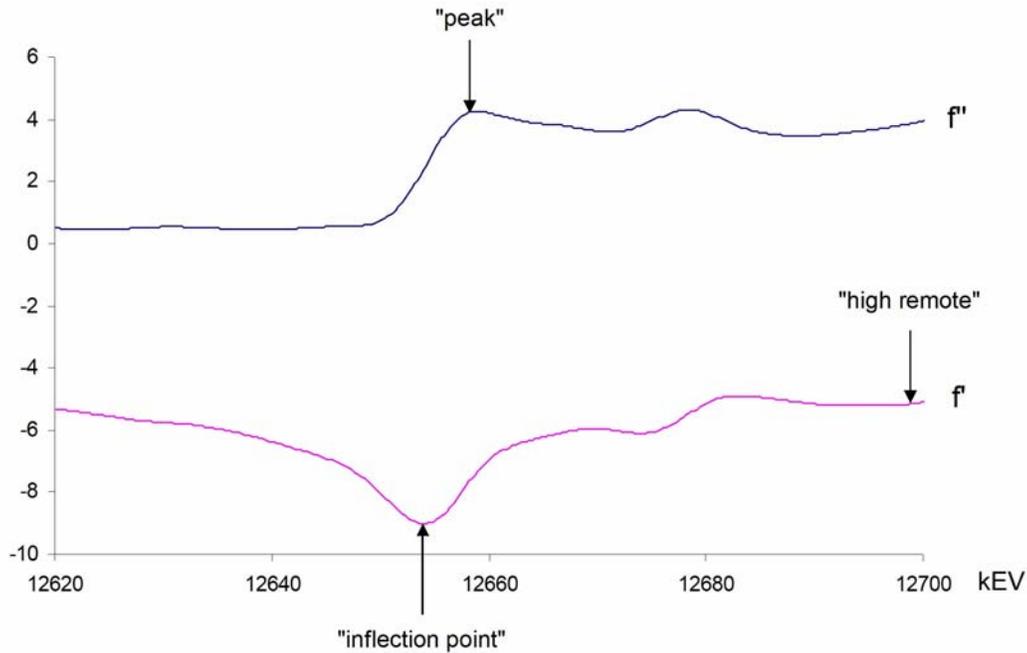


Figure 1.6: Fluorescence scan on selenomethionine containing crystal of *Mfd-N2* around the selenium K-edge (12.6578 kEV). Values of f' and f'' are plotted against the X-ray energy (kEV). Wavelengths used in an MAD experiment are indicated.

Today, many structures are solved using anomalous data collected at only one wavelength (single-wavelength anomalous diffraction, SAD).

In recent years, experimental phase information could even be obtained directly from native protein crystals using anomalous scattering from sulfur atoms. In this approach, data is typically collected at one wavelength far from the absorption edge (5.02 Å). Because of the very weak anomalous contributions, highly accurate and redundant data are required.

1.3 Objectives

DNA damage repair plays an essential role in the maintenance of genomic integrity. Nucleotide excision repair (NER) is a functionally conserved repair system which is present in prokaryotes, archaea, and eukaryotes. NER can deal with a large variety of chemically and structurally unrelated helix-distorting DNA lesions.

However, transcribing RNA polymerase frequently encounters DNA damage, before it has been repaired, and the transcription machinery becomes arrested. Arrested RNA polymerase is a severe threat to the cell, as it can neither resume transcription elongation

nor dissociate from the template DNA by itself. It therefore prevents replication, transcription and repair of this gene.

DNA damage in the transcribed strand of active genes is repaired by a special mode of NER, called transcription-coupled repair (TCR). In prokaryotes, TCR is mediated by the Mfd (mutation frequency decline) protein. Mfd releases arrested transcription elongation complexes, followed by delivery of the UvrABC nucleotide excision repair machinery to the lesion site. The mechanistic details of this process, however, are still poorly understood.

Aim of this PhD thesis was to gain insight into the mechanism of bacterial transcription-coupled repair.

Knowledge of their three-dimensional structure can provide important information on the biological function of proteins, and on their mode of operation. Therefore, X-ray crystallography should be used in order to obtain structural information on the *Escherichia coli* Mfd protein.

Furthermore, the protein should be characterized in a functional way. In particular, the interaction between Mfd and the nucleotide excision repair protein UvrA was investigated. UvrA-binding is mediated by the N-terminus of Mfd, and is required for recruitment of the NER machinery.

The major objective of these studies was to reveal new aspects of this important event during transcription-repair coupling in order to suggest a possible mechanism for this recruitment step.

2 Materials and methods

2.1 Materials

All common chemicals were reagent-grade reagents purchased from Merck (Darmstadt, Germany), Sigma (Deisenhofen, Germany) or Carl Roth (Karlsruhe, Germany), unless otherwise stated. Crystallization screens and crystallization tools were from Hampton Research (Aliso Viejo, USA), Nextal Biotechnologies (Montreal, Canada; now QIAGEN, Hilden, Germany) or Jena Bioscience (Jena, Germany). RP-HPLC purified oligonucleotides were ordered from Thermo Electron Corporation (Ulm, Germany). Enzymes and nucleotides for molecular biology were from Fermentas (St- Leon-Rot, Germany).

2.2 Molecular biology methods

Common molecular biology procedures like polymerase chain reaction (PCR), cleavage of phosphodiester bonds in DNA by restriction endonucleases, dephosphorylation of DNA ends by alkaline phosphatase, ligation of DNA ends, amplification of plasmid DNA and agarose gel electrophoresis were carried out according to standard protocols (Sambrook, 1989).

Bacterial genomic DNA was prepared with DNAzol® reagent (Molecular Research Center, Cincinnati, USA) according to the manufacturer's instructions. Plasmid DNA was isolated using the NucleoSpin®-Plasmid Quick Pure Kit (Macherey-Nagel, Dueren, Germany). DNA fragments were extracted from agarose gels with the NucleoSpin®-Extract II Kit (Macherey-Nagel, Dueren, Germany). DNA-sequencing was performed by Medigenomix (Martinsried, Germany).

2.2.1 Cloning

Genes of interest were amplified by PCR from *Escherichia coli* K12 (XL1-Blue) genomic DNA with ACCUZYME™ DNA polymerase (Bioline, Luckenwalde, Germany) and cloned into the pET-21b, pET-28b or pET-29b vectors (both from Novagen,

Schwalbach/Ts., Germany) or the pTYB1 vector (New England Biolabs, Frankfurt/Main, Germany) according to table 2.3. Oligonucleotides used as PCR-primers were designed using GeneRunner (<http://www.generunner.com/>).

Table 2.1: List of oligonucleotides used for cloning. Underlined regions mark the recognition sites for restriction endonucleases. Sequences are given in 5'-3' direction.

Oligo name	Sequence
Mfd N ₀ for <i>NheI</i>	AAAAGCTAGCATGCCTGAACAATATCGTTATACGC
Mfd N ₀ for <i>NdeI</i>	AAAACATATGCCTGAACAATATCGTTATACGC
Mfd C ₀ rev <i>NotI</i>	AAAAGCGGCCGCAGCGATCGCGTTCTCTTCC
Mfd N333-STOP rev <i>NotI</i>	TTTTTGCGGCCGCTAGTTTTTCAGCTCTGAGAAGAGC
Mfd N333 rev <i>NotI</i>	TTTTTGCGGCCGCGTTTTTCAGCTCTGAGAAGAGC
Mfd N433-STOP rev <i>NotI</i>	AAAAAGCGGCCGCTAACCATGTTTCGGCAGCGCC
Mfd N433 rev <i>NotI</i>	AAAAGCGGCCGCACCATGTTTCGGCAGCGCC
Mfd D586-STOP rev <i>NotI</i>	TTTTGCGGCCGCCTAATCGTGTTTAAACGCGAAGCCCTC
Mfd D586 rev <i>NotI</i>	TTTTGCGGCCGCATCGTGTTTAAACGCGAAGCCCTC
Mfd S964 for <i>NdeI</i>	AAAAACATATGAGCGGCTCAATGGAAACCATCGG
UvrA N ₀ for <i>NdeI</i>	AAAAACATATGGATAAGATCGAAGTTCGGGG
UvrA C ₀ rev <i>XhoI</i>	AAAAACTCGAGCAGCATCGGCTTAAGGAAGCG
UvrB N ₀ for <i>NdeI</i>	AAAAACATATGAGTAAACCGTTCAAACCTGAATTCC
UvrB G583 rev <i>HindIII</i>	AAAAAAAGCTTTCCGTTTCCCTCGTTGTACTTCTGC

2.2.2 Site-directed mutagenesis

Point mutations were introduced by PCR-based site-directed mutagenesis (Ho et al., 1989). Complementary oligonucleotides containing the desired mutation were used in a first PCR. In this reaction, two PCR-products with overlapping ends were generated. These DNA fragments were used as templates in a subsequent reaction, the overlap extension. Here, the overlapping ends were annealed, allowing a 3'-extension of the complementary strand. After 3 cycles, the flanking primers were added, and the fusion product was further amplified by PCR.

Table 2.2: List of oligonucleotides used for site-directed mutagenesis. Complementary regions are underlined, nucleotides coloured in blue correspond to the mutated codons. Sequences are given in 5'-3' direction.

Desired mutation	Sequences (forward / reverse primers)
Mfd R165A	<u>CACG</u> <u>GCG</u> GGCGCGTTGCTGGATCTCTTCC CGCGCC <u>CG</u> CGTGGCGTATTTCGCCGTGC
Mfd R181A	GCTGCCTTAT <u>GCG</u> CCTGATTTCTTTGATGATGAAATC CAAAGAAATCAAG <u>CG</u> CATAAGGCAGCTCACTCCCCATCG
Mfd R181A/D183A	GCTGCCTTAT <u>GCG</u> CCT <u>GCG</u> TTCTTTGATGATGAAATCGACAGC CAAAGAA <u>CG</u> CAAG <u>CG</u> CATAAGGCAGCTCACTCCCCATCG
Mfd F185A	<u>CGT</u> CCTGATTTTC <u>GCG</u> GATGATGAAATCGACAGCCTGC CATCATCC <u>CG</u> CGAAATCAAGACGATAAGGCAGC
Mfd E188A	CTTTGATGAT <u>GCA</u> ATCGACAGCCTGCGGGTG GGCTGTCGAT <u>TGC</u> ATCATCAAAGAAATCAAGACG
Mfd D190A	GATGATGAAAT <u>CGC</u> GAGCCTGCGGGTGTGTTGACG GCAGGCT <u>CGC</u> GATTTTCATCATCAAAGAAATCAAGACG
Mfd Δ2 (AA 124 – 213)	CCACGGTCAT <u>GGCACTAGTTCC</u> CCCCGCGCACGAATTTCCG <u>GGAACTAGTGCC</u> ATGACCGTGGAGAAAAGTGTGTGG
Mfd C118A	CTGTGTGG <u>GGCA</u> CACGTTGCATAAGCGTATTCACCGGAACAATC GCAACGTGTT <u>GCC</u> CCACACAGTTTTTCTCCACGG
Mfd C445A	GATCGCTTTC <u>GGC</u> AATCAGCGCCAGATTACGC GCTGATT <u>GCC</u> GAAAGCGATCTGCTCGGTG

Table 2.3: Expression plasmids

#	Insert (construct)	Vector	Restriction sites	Tag	Remarks
1	Mfd (WT) full-length	pET-21b	<i>NheI/NotI</i>	–	
2	Mfd (WT) full-length	pET-21b	<i>NheI / NotI</i>	C-HIS	"Mfd-FL"
3	Mfd (WT) AA 1 – 333	pET-21b	<i>NdeI / NotI</i>	–	"Mfd-N1"
4	Mfd (WT) AA 1 – 333	pET-21b	<i>NdeI / NotI</i>	C-HIS ₆	"Mfd-N2"
5	Mfd (WT) AA 1 – 433	pET-21b	<i>NdeI / NotI</i>	–	"Mfd-N3"
6	Mfd (WT) AA 1 – 433	pET-21b	<i>NdeI / NotI</i>	C-HIS ₆	"Mfd-N4"

7	Mfd (WT) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-N5"
8	Mfd (WT) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	C-HIS ₆	"Mfd-N6"
9	Mfd (WT) AA 964-1148	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-C1"
10	Mfd (WT) AA 964-1148	pET-28b	<i>NdeI</i> / <i>NotI</i>	N-HIS ₆	"Mfd-C2"
11	UvrA (WT) full-length	pTYB-1	<i>NdeI</i> / <i>XhoI</i>	C-INTEIN	
12	UvrA (WT) full-length	pET-29b	<i>NdeI</i> / <i>XhoI</i>	C-HIS	
13	UvrB (WT) AA 1 – 583	pET-21b	<i>NdeI</i> / <i>HindIII</i>	C-HIS	"UvrB-N"
14	Mfd (R165A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Mut1"
15	Mfd (R181A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Mut2"
16	Mfd (R181A/D183A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Mut3"
17	Mfd (F185A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Mut4"
18	Mfd (E188A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Mut5"
19	Mfd (E190A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Mut6"
20	Mfd (WT) AA 1 – 123 / 214 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"Mfd-Δ2"
21	Mfd (C118A/C445A) AA 1 – 586	pET-21b	<i>NdeI</i> / <i>NotI</i>	–	"CYS-DM"

2.3 Microbiology methods

Bacteria were grown in shaking cultures in liquid LB medium or on LB agar plates containing the appropriate antibiotics. For storage, cells were kept in 40 % glycerol at -80°C.

Table 2.4: Bacterial Strains

Strain	Genotype	Source
XL1-Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F'proAB lacI ^q ZAM15 Tn10 (Tet ^R)]	Stratagene, La Jolla, USA
Rosetta (DE3)	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm lacY1 (DE3) pRARE (Cm ^R)	Novagen, Schwalbach/Ts., Germany
B834 (DE3)	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm met (DE3), transformed with pRARE (Cm ^R) isolated from Rosetta (DE3) cells	Novagen, Schwalbach/Ts., Germany

Table 2.5: Composition of Luria-Bertani (LB)-broth (Miller, 1972)

Bacto-Tryptone	1.0 % (w/v)
Yeast-Extract	0.5 % (w/v)
NaCl	1.0 % (w/v)
pH 7.0	

1.5–2 % (w/v) of Bacto-Agar were added to the medium to prepare LB-agar plates.

Table 2.6: Antibiotics and supplements

Supplement	stock solution	in media
Ampicillin	100 mg/ml (H ₂ O)	100 µg/ml
Chloramphenicol	34 mg/ml (Ethanol)	34 µg/ml
Kanamycin	50 mg/ml (H ₂ O)	50 µg/ml
Tetracycline	10 mg/ml (Ethanol)	10 µg/ml
IPTG	0.5 M (H ₂ O)	0.15 mM

2.3.1 Transformation of *E.coli*

Buffer TfBI

30 mM potassium acetate pH 5.8
 100 mM KCl
 50 mM MnCl₂
 10 mM CaCl₂
 15% Glycerol

Buffer TfBII

10 mM MOPS pH 7.0
 10 mM KCl
 75 mM CaCl₂
 15% Glycerol

The preparation of transformation competent bacteria according to (Hanahan, 1983) was conducted by successive incubations in buffers *TfBI* and *TfBII* on ice. Aliquots of cells in *TfBII* were snap-frozen in liquid nitrogen and stored at -80°C .

For transformation ca. 100 ng of ligated DNA or 10 ng of plasmid DNA were added to 75 μl of competent cells. Cells were incubated on ice for 20 minutes, and a heat step at 42°C was carried out for 45-60 seconds. After addition of 800 μl of LB medium, cells were incubated for 45-60 minutes at 37°C . The suspension was plated on LB-agar plates containing the corresponding antibiotics and incubated over night at 37°C .

2.3.2 Protein expression

Proteins were overexpressed recombinantly in *E.coli* Rosetta (DE3) cells.

Competent cells were transformed with plasmids containing the gene of interest. Cells were grown in LB medium supplemented with the corresponding antibiotics at 37°C . At an OD_{600} of 0.4-0.6, the cultures were cooled down to 18°C . Gene expression was induced by addition of 0.15 mM IPTG, and protein production was carried out overnight at 18°C . Cells were harvested by centrifugation and snap-frozen in liquid nitrogen. Cell pellets were stored at -80°C .

2.3.3 Selenomethionine-labelling

Selenomethionine-substituted protein was produced using the methionine auxotrophic B834 (DE3) strain. Cells were transformed with the pRARE (Cm^{R}) plasmid isolated from the Rosetta (DE3) strain, and with a plasmid containing the gene of interest.

Cells were grown in LB medium containing the appropriate antibiotics at 37°C to an OD_{600} of 0.4. Bacteria were harvested and resuspended in the same amount of LeMaster's medium containing selenomethionine (table 2.7) (LeMaster and Richards, 1985). Appropriate antibiotics were added.

Cells were grown at 37°C for one generation time to deplete the medium of residual methionine. The cultures were cooled on ice, and protein expression was induced by the addition of 0.15 mM IPTG. Selenomethionine-containing protein was produced over night at 18°C . The modified protein was purified as described above. 1 mM DTT was added to each buffer to prevent oxidation of selenomethionine.

Table 2.7: *LeMaster's medium (LeMaster and Richards, 1985)*

autoclavable portion A for LeMaster's medium (g / 2000 ml)			
alanine	1.0	serine	4.166
arginine hydrochloride	1.16	threonine	0.46
aspartic acid	0.8	tyrosine	0.34
cystine	0.066	valine	0.46
glutamic acid	1.5	adenine	1.0
glutamine	0.666	guanosine	1.34
glycine	1.08	thymine	0.34
histidine	0.12	uracil	1.0
isoleucine	0.46	sodium acetate	3.0
leucine	0.46	succinic acid	3.0
lysine hydrochloride	0.84	ammonium chloride	1.5
phenylalanine	0.266	sodium hydroxide	1.7
proline	0.2	dibasic potassium phosphate	21.0

All amino acids were reagent-grade L-enantiomers purchased from Sigma (Deisenhofen, Germany).

After autoclaving of solution A, the solution was cooled down to 37°C. Subsequently, filter-sterilized solution B (200 ml of solution B / 2000 ml of solution A) was added.

non-autoclavable solution B	
glucose	20.0 g
magnesium sulfate heptahydrate	0.5 g
iron sulfate	8.4 mg
sulfuric acid (concentrated)	16.0 µl
thiamin	10.0 mg

Selenomethionine (Calbiochem, Schwalbach/Ts., Germany) was dissolved in sterile H₂O and added to the medium (100 mg / 2200 ml).

2.4 Proteinchemical methods

Physical and chemical parameters like molecular weight, (theoretical) isoelectric point (pI), extinction coefficient etc. for the recombinant proteins were calculated with the ProtParam Tool (Gasteiger *et al.*, 2003) from the ExPASy Proteomics Server (<http://www.expasy.org/>). Protein secondary structure prediction was carried out by the PSIPRED Protein Structure Prediction Server (<http://bioinf.cs.ucl.ac.uk/psipred/>) (Jones, 1999; McGuffin *et al.*, 2000; Bryson *et al.*, 2005). Sequence alignment was performed with ClustalW (<http://align.genome.jp/>) and edited manually using GeneDoc (Nicholas and Nicholas, 1997)

2.4.1 Protein purification

Buffer Ni²⁺-A1

50 mM NaH₂PO₄ pH 8.0
200 mM NaCl

Buffer Ni²⁺-HS

50 mM NaH₂PO₄ pH 8.0
2 M NaCl

Buffer Ni²⁺-A2

50 mM NaH₂PO₄ pH 8.0
200 mM NaCl
10 mM imidazole

Buffer Ni²⁺-B

50 mM NaH₂PO₄ pH 8.0
200 mM NaCl
250 mM imidazole

Dilution Buffer

20 mM TRIS/HCl pH 8.0
1 mM EDTA
10 % Glycerol

SourceQ Buffer

20 mM TRIS/HCl pH 8.0
50 – 500 mM NaCl
0.1 mM EDTA

Size Exclusion Buffer

20 mM TRIS/HCl pH 8.0
200 mM NaCl
0.1 mM EDTA
1 mM DTT

Cell pellets were resuspended in buffer Ni²⁺-A1 supplemented with 200 μ M PMSF and disrupted by sonification. The lysate was cleared by centrifugation and loaded onto Ni²⁺-NTA column (QIAGEN, Hilden, Germany) pre-equilibrated with Ni²⁺-A1. The column was washed subsequently with buffers Ni²⁺-HS and with Ni²⁺-A2. Buffer Ni²⁺-B was used for protein elution. Elution fractions were analyzed using the Bio-Rad's protein assay (Bio-Rad, Munich, Germany), and protein-containing fractions were pooled.

After dilution with dilution buffer 1:5, the eluate was loaded onto a Resource Q anion exchange column (Amersham Bioscience, Freiburg, Germany) equilibrated with SourceQ buffer containing 50 mM NaCl. The protein was eluted with a linear gradient of 10 column volumes from 50 mM to 500 mM NaCl in the same buffer. Peak fractions were pooled, and, after concentration, applied onto a Superdex 200 size-exclusion column (Amersham Bioscience, Freiburg, Germany). Peak fractions were pooled and concentrated to the desired concentration.

2.4.2 Protein-protein interaction assay

Lysis Buffer

50 mM NaH₂PO₄ pH 8.0
100 mM NaCl

High Salt Buffer

50 mM NaH₂PO₄ pH 8.0
1 M NaCl

Wash Buffer

50 mM NaH₂PO₄ pH 8.0
100 mM NaCl
10 mM imidazole

Elution Buffer

50 mM NaH₂PO₄ pH 8.0
100 mM NaCl
250 mM Imidazol

The interaction of UvrA with Mfd mutants was analyzed analogously to Truglio *et al.*, 2004. In a first step, UvrA was immobilized on agarose beads. Subsequently, Mfd mutants were added in excess, and the resin was washed gently in order to avoid disruption of the salt-sensitive complexes. Proteins were eluted and analyzed by SDS-PAGE.

As a first step, full-length UvrA was coupled to Ni²⁺-NTA agarose beads (QIAGEN, Hilden, Germany) via its C-terminal hexahistidine tag. Cells from 600 ml expression culture were resuspended in lysis buffer containing 200 μM PMSF. They were lysed by sonification, and the lysate was cleared by centrifugation. The supernatant was mixed with 1 ml of Ni²⁺-NTA resin and rotated end-over-end for 2 h at 4°C. The resin was washed successively with lysis buffer, high salt buffer and wash buffer. After re-equilibration with lysis buffer, the resin was distributed to 8 Poly-Prep Chromatography Columns (Bio-Rad, Munich, Germany).

For the interaction analysis, lysates from cells expressing untagged mutants of Mfd-N5 (residues 1-586) were added to the resin. Cells from 200 ml expression culture were each resuspended in lysis buffer containing 200 μM PMSF and lysed by sonification. The obtained lysates were cleared by centrifugation, and each lysate was added to one of the UvrA-Ni²⁺-columns and allowed to flow through the resin. The columns were washed

successively with lysis buffer and wash buffer. Elution buffer was used for protein elution. Protein-containing elution fractions were determined using the Bio-Rad's protein assay (Bio-Rad, Munich, Germany), and complexes of UvrA and Mfd-N5 were analyzed by SDS-PAGE. Scanned gels were evaluated using the ImageJ software (<http://rsb.info.nih.gov/ij/>).

Parts of this assay were carried out by Gabriela Guédez-Rodríguez, a Master of Biochemistry student at the University of Munich, Germany.

2.5 Protein analysis

2.5.1 Analytical size exclusion chromatography

Size Exclusion Buffer

20 mM TRIS/HCl pH 8.0
200 mM NaCl
0.1 mM EDTA
1 mM DTT

In order to determine the molecular weight of proteins, analytical gel filtration was performed using a Superdex 200 10/300 GL column (Amersham, Freiburg, Germany). The column was calibrated using the Gel Filtration Standard (Bio-Rad, Munich, Germany) in the same buffer.

2.5.2 Limited proteolysis

In order to discriminate between stable and flexible regions within a protein of interest, limited proteolysis was performed. The reaction was carried out in size exclusion buffer in a total volume of 50 μ l.

30 μ g of purified protein were incubated with different amounts of Proteinase K (Fermentas St- Leon-Rot, Germany) (0.005/0.05/0.5/5 μ g) for 45 minutes at room temperature. The reaction was stopped by addition of 2 μ l of PMSF (saturated solution in 2-propanol). Proteolytic digest was also carried out in the presence of 20 μ M Adenosine 5'-[γ -thio]triphosphate (ATP- γ S, Sigma, Deisenhofen, Germany) and 20 μ M ATP- γ -S plus 20 μ M DNA (dsHOL-1), respectively. 1 mM MgCl₂ were added to the reaction buffer.

-dsHOL-1: 5' -GGCGACGTGATCACCAGATGATGCTAGATGCTTTCCGAAGAGAGAGC
CCGCTGCACTAGTGGTCTACTACGATCTACGAAAGGCTTCTCTCTCG-3'

2.5.3 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were analyzed by SDS-PAGE (Laemmli, 1970) using the vertical Mini-PROTEAN 3 System (Bio-Rad, Munich, Germany). Protein bands were stained with Coomassie Brilliant Blue R-250 (Carl Roth, Karlsruhe). Bands of interest were excised from the gel using sterile disposable scalpels and analyzed by EDMAN-sequencing and/or mass spectrometry.

2.5.4 Protein sequencing (Edman, 1950)

For N-terminal sequencing, proteins were blotted onto a piece of PVDF membrane by passive adsorption.

Excised bands were dried in a Speed-Vac and, after drying, reswollen in 35 µl of 200 mM TRIS/HCl pH 8.5, 2 % SDS. After swelling, a concentration gradient was set up by addition of 150 µl of distilled water. A small piece of PVDF membrane (Carl Roth, Karlsruhe, Germany) was activated in methanol and added to the gel band. Once the solution started to turn blue, 20 µl of methanol were added.

After 1-2 days, the solution had become clear and the membrane had turned blue. The membrane was washed 5 times with 1ml of 10 % Methanol with vortexing for 30 seconds each time.

After air drying, the protein was N-terminally sequenced from the membrane in a PROCISE 491 sequencer (Applied Biosystems, Darmstadt, Germany). The sequencing reaction was carried out by Stefan Benkert (Gene Center, Munich, Germany).

2.5.5 Matrix assisted laser desorption ionisation Time-of-Flight analysis

For mass spectrometric analysis, protein bands of interest were tryptically digested. The obtained peptides were crystallized on a sample plate using an organic matrix and analyzed by Matrix-assisted laser desorption ionisation - Time-of-Flight (MALDI-ToF) mass spectrometry (MS).

Tryptic digest was performed using a modified protocol for the Montage In-Gel Digest 96 Kit (Millipore, Billerica, USA) (J. Rauch, personal communication).

Excised protein bands were cut into 1x1 mm pieces and transferred to 1.5 ml Eppendorf-tubes (Eppendorf AG, Hamburg, Germany: #3810). Bands were washed with ultrapure water from a Milli-RO 60 water purifying system (Millipore GmbH, Schwalbach/Ts., Germany) and twice alternately with 100% acetonitrile (ACN) and 50 mM ammonium bicarbonate. After washing with 50% acetonitrile, the proteins were digested each with 3 µg of sequencing grade modified porcine trypsin (Promega, Mannheim, Germany) in 30 µl 50 mM ammonium bicarbonate over night at 30°C.

The gel pieces were incubated twice with 100 µl of 75% ACN, 12.5 mM ammonium bicarbonate for 30 minutes in order to extract the peptides from the gel. The extracts were pooled, and the obtained peptide mixtures were concentrated in a Speed-Vac.

Lyophilized peptides were resolved in 10 µl 0.1 % of trifluoroacetic acid (TFA) and mixed 1+1 with freshly prepared matrix solution. 1 µl was spotted on a matrix assisted laser desorption ionization (MALDI) AnchorChip sample plate (Bruker Daltonik, Bremen, Germany).

Peptide mass fingerprinting analysis was performed on a Bruker Reflex III MALDI-ToF mass spectrometer (Bruker Daltonik, Bremen, Germany) by Jens Rauch (Klinikum Grosshadern, Munich, Germany) and Thomas Knöfel (GSF, Munich, Germany). The list of peptide masses was aligned with the MASCOT Search Engine (<http://www.matrixscience.com>) using the SwissProt Database.

-Matrix solution: α -cyano-4-hydroxycinnamic acid (CHCA) (Bruker Daltonik, Bremen, Germany) prepared as a saturated solution in 50% ACN, 0.1% TFA

2.6 Functional assays

2.6.1 ATPase activity assay

ATPase Buffer

40 mM HEPES, pH 7.8

100 mM KCl

8 mM MgCl₂

4% Glycerol

5 mM DTT

ATPase activity was tested by thin layer chromatography (TLC).

5 μM of purified protein were incubated with 10 μM ATP (containing 1/3000 Redivue $\gamma^{32}\text{P}$ -ATP (Amersham Biosciences, Freiburg, Germany)) at 37°C for 20 minutes in a total volume of 20 μl . ATPase activity was determined in the absence and in the presence of 50 μM (10-fold excess) of dsDNA (dsHOL-1).

Aliquots of 1 μl were spotted on a polyethyleneimine (PEI) cellulose plate (MERCK, Darmstadt, Germany). TLC plates were developed in 0.5 M LiCl, 1 M formic acid, dried and analyzed with a STORM Phosphor-Imager and ImageQuant Software (both Amersham Biosciences, Freiburg, Germany).

The upper spot corresponds to liberated γ - ^{32}P and the lower spot to non-hydrolyzed ATP.

-dsHOL-1: 5' -GGCGACGTGATCACCAGATGATGCTAGATGCTTTCCGAAGAGAGAGC
CCGCTGCACTAGTGGTCTACTACGATCTACGAAAGGCTTCTCTCTCG-3'

2.6.2 DNA binding assay

Annealing Buffer

100 mM TRIS/HCl pH 7.5

100 mM NaCl

10 mM MgCl₂

1 mM DTT

1xTB Buffer

90 mM TRIS

90 mM boric acid

Binding buffer

40 mM HEPES, pH 7.8

100 mM KCl

8 mM MgCl₂

4 % Glycerol

5 mM DTT

100 $\mu\text{g/ml}$ BSA

2 mM ATP- γ -S

DNA binding activities of Mfd constructs were analyzed using the electrophoretic mobility shift assay (EMSA). Complexes of protein and DNA migrate through a native polyacrylamide (PAA) gel more slowly than free oligonucleotides.

One DNA strand was radioactively labelled. 5'-labelling of single stranded DNA was performed using T4 polynucleotide kinase (PNK) (Fermentas, St- Leon-Rot, Germany) according to manufacturer's instructions with Redivue $\gamma^{32}\text{P}$ -ATP (Amersham Biosciences, Freiburg). Unincorporated radionucleotides were removed using the Nucleotide Removal Kit (QIAGEN, Hilden, Germany). In order to generate dsDNA, two complementary oligonucleotide strands were annealed.

The gel shift assay was performed using a modified protocol according to (Selby and Sancar, 1995a).

Proteins were incubated in binding buffer for 20 minutes on ice. Labelled dsDNA was added, and the samples were incubated for further 20 minutes at 4°C.

The final concentration of the oligonucleotides was 0.2 nM, proteins were each present at 1.5 µM in a total reaction volume of 20 µl.

In order to separate free probe from protein-bound polynucleotide, 15 µl from the binding reaction were analyzed on a native 6% polyacrylamide gel in 0.2x TRIS-borate (TB) buffer. After gel drying on a Model 583 Gel Dryer (Bio-Rad, Munich, Germany), radioactivity was recorded with the STORM Phosphorimager (Amersham Biosciences, Freiburg, Germany). The lower visible band corresponds to protein-free probe, the upper band to the shifted protein-bound oligonucleotide.

-dsDNA-1: 5'-AAAAGCAAATTGCCTT-3'
3'-TTCGTTTAACGG-5'

dsDNA-2: 5'-AAAAGCAAATTGCCGAAGACGAACGCGTT-3'
3'-TTCGTTTAACGGCTTCTGCTTGC GC-5'

2.7 Structural analysis of Mfd-N2

2.7.1 Protein crystallization by sitting drop vapour diffusion

In order to determine the three-dimensional structure of a molecule by X-ray diffraction, high quality crystals are required. X-ray scattering from one molecule in solution would not generate a signal strong enough for detection. In a crystal, the molecule of interest is periodically repeated, all molecules having the same relative position and orientation. If scattered waves from these molecules interfere in a constructive manner, they give rise to a diffraction pattern (see 1.2.3).

The quality of a crystal is influenced by many parameters like structural flexibility, solvent content, impurities, defects in the crystal lattice, mosaicity etc. The shape and size of a crystal also play an important role in its diffraction properties.

Typical crystals used in X-ray diffraction experiments have a size of 100-300 µm in all three dimensions. Today, strong synchrotron radiation also allows structure determination with smaller crystals. In this case, crystals with a size of 40 µm x 40 µm x 110 µm were sufficient.

Protein crystallization requires high amounts of pure and homogeneous protein. The protein is slowly concentrated to a supersaturated state. During this process, crystals may appear. There are several crystallization techniques, including microdialysis, batch crystallization, hanging drop or sitting drop vapour diffusion (for an overview see for example <http://www.hamptonresearch.com/support/Growth101Lit.aspx>). Today, the vapour diffusion techniques are the most popular ones. In this PhD thesis, crystals were grown by the sitting drop vapour diffusion method (figure 2.1).

In this technique, crystallization reagent is given into the reservoir of a crystallization plate. Reservoir solutions typically consist of buffer solution, precipitant, and salt. A small droplet of concentrated protein sample mixed with reservoir solution is set on a platform in vapour equilibration with the reservoir. As the drop contains a lower reagent concentration than the reservoir, water vapour leaves the drop. Thereby, the protein drop is slowly concentrated to a supersaturated state. In most cases, the protein will form aggregates and precipitate out of solution. Under certain conditions, stable nuclei may form, and crystals growth may take place.

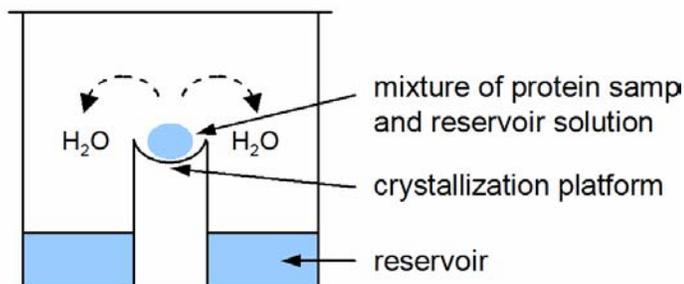


Figure 2.1: Principle of protein crystallization by sitting drop vapour diffusion (adopted from <http://www.hamptonresearch.com>).

In order to find a condition where nucleation and crystal growth are favored over precipitation, many crystallization reagents need to be tested. For this purpose, commercial sparse matrix screens are available, e.g from Hampton Research (Aliso Viejo, USA), Nextal Biotechnologies (Montreal, Canada) (now QIAGEN, Hilden, Germany) or Jena Bioscience (Jena, Germany).

In order to improve the size and quality of obtained crystals, initial crystallization conditions are subsequently modified. Common approaches are variations of drop size, protein concentration, pH and ionic strength of the reservoir solution, and the use of additives like alcohols, sugars or salts.

2.7.2 Crystallization of Mfd-N2

Mfd-N2 was crystallized by sitting drop vapour diffusion methods using 96-well crystallization plates (Corning, Schiphol-Rijk, The Netherlands).

Crystals of Mfd-N2 were obtained with 50 μ l of reservoir solution containing 0.08 M sodium citrate pH 5.6, 0.16 M ammonium sulfate, 20% PEG-4000, 0.8 M sodium formate. After mixing 1 μ l of protein solution (4 mg/ml in 20 mM TRIS/HCl, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT) with 1 μ l of reservoir solution, crystals appeared overnight at 20°C.

For initial crystal setups with commercial screens, the Hydra II semi-automatic protein crystallization robot (Matrix Technologies Apogent Discoveries, Hudson, USA) was used to set 0.5 + 0.5 μ l drops. In order to improve size and quality of obtained crystals, initial crystallization conditions were refined manually. 1 + 1 μ l drops were set in the same plates. The reservoir solution composition and reservoir volume as well as the protein concentration were varied. Screening of additives was performed as well.

Crystals were transferred to mother liquid containing 20% PEG-400 and snap-frozen in liquid nitrogen.

2.7.3 Data collection, structure determination, model building and refinement

All diffraction data were collected at beamline ID14-4 (ESRF, Grenoble, France) with an ADSC Q4 CCD detector using ProDC (<http://www.esrf.fr/computing/bliss/gui/prodc/>). MOSFLM (Powell, 1999) was used to set up a data collection strategy in order to achieve high redundancy and completeness. Prior to data collection the optimal wavelengths for the MAD experiment were determined with a fluorescence scan on the selenium containing crystal.

Data were processed with DENZO and SCALEPACK (Otwinowski and Minor, 1997) or with XDS and XSCALE (Kabsch, 1993).

SOLVE (Terwilliger, 2002) was used to locate heavy atom sites, and phases were calculated with SHARP (de la Fortelle and Bricogne, 1997). Initial phases were improved with SOLOMON (Abrahams and Leslie, 1996). Automated and manual model building were carried out using ARP/wARP (Morris *et al.*, 2003) and MAIN (Turk, 1992),

respectively. CNS v.1.1 (Brunger *et al.*, 1998) was used for refinement. Coordinates as well as topology and parameter files for hetero-compounds were retrieved from the HIC-Up server (Kleywegt and Jones, 1998). Stereochemistry of the final model was analyzed with PROCHECK (Laskowski *et al.*, 1993).

3 Results

3.1 Full-Length *E.coli* Mfd

3.1.1 Purification and crystallization of full-length Mfd

The gene encoding full-length Mfd was amplified from *Escherichia coli* XL1 Blue genomic DNA and cloned into the pET-21b vector (see table 2.3). The protein was recombinantly overexpressed in Rosetta (DE3) cells with a C-terminal HIS₆-tag (“Mfd-FL”).

Mfd-FL was purified as described in 2.4.1. First, lysate from Mfd-FL-expressing cells was loaded onto a Ni²⁺-NTA column. The protein was further purified using a Resource Q anion exchange column and by size exclusion chromatography with a HiLoad 26/60 Superdex 200 pg column (figure 3.1).

Using this purification protocol, about 6 mg of highly pure and homogeneous protein could be obtained from 12 l of expression culture. Protein identity was confirmed by EDMAN-sequencing.

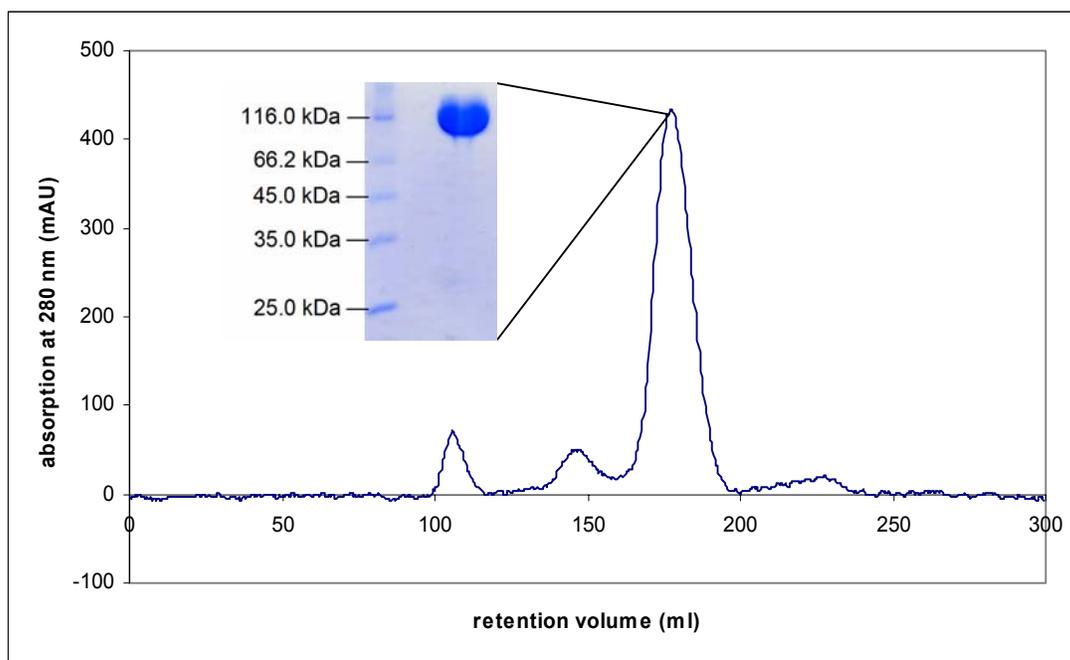


Figure 3.1: Elution profile of Mfd-FL from the HiLoad 26/60 Superdex 200 pg size exclusion column. The major peak corresponds to purified Mfd-FL. SDS-PAGE analysis of the peak fractions is shown as inset.

Crystallization setups were carried out with 5 mg/ml protein. No three-dimensional crystals could be grown. Using Nextal classic screen condition #34 (0.1 M tri-sodium citrate pH 5.6, 0.2 M potassium/sodium tartrate, 2.0 M ammonium sulfate), bushes of needles were obtained. However, they were not reproducible and could not be improved. Without any tag, no better results were achieved (data not shown).

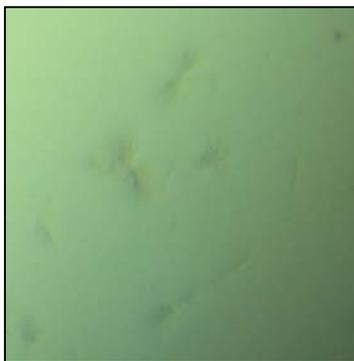


Figure 3.2: Crystals of Mfd-FL were obtained from Nextal classic screen condition #34 (0.2 M potassium/sodium tartrate, 0.1 M tri-sodium citrate pH 5.6, 2.0 M ammonium sulfate) using 5 mg/ml protein.

3.1.2 Limited proteolysis

Highly mobile regions in proteins can inhibit crystallization. In order to identify those regions in Mfd, limited proteolysis was performed. Proteolysis of natively folded proteins occurs mainly at highly flexible parts, like loops, while globular domains are rather rigid and more resistant to proteolysis (Fontana *et al.*, 1986; Fontana *et al.*, 2004). Therefore, stable fragments in the proteolysis experiment may correspond to regions which are compact and hence serve as good candidates for crystallization.

The protein was incubated with different amounts of Proteinase K under native conditions. Compared to other proteolytic enzymes, endopeptidase Proteinase K possesses broader substrate specificity as it cleaves peptide bonds after aliphatic, aromatic or hydrophobic amino acids. Thus, it is more likely to cleave within in a distinct region.

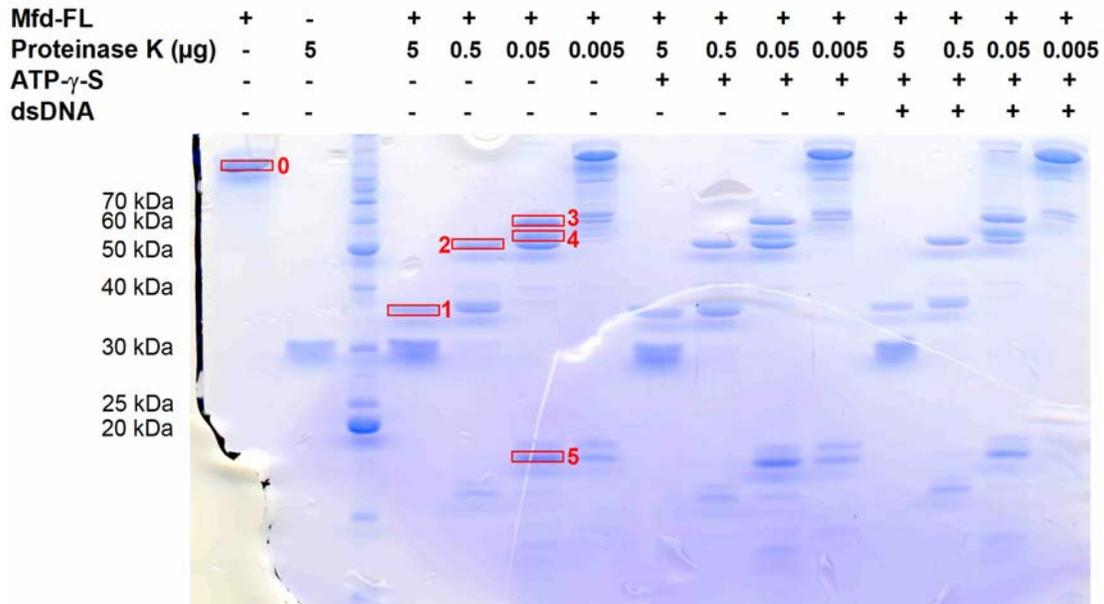


Figure 3.3: Limited proteolysis of Mfd-FL. From left to right: Cleavage pattern without ATP or DNA, with ATP, and with ATP and dsDNA. Bands marked with red boxes were analyzed by EDMAN-sequencing and mass spectrometry.

Mfd contains a dsDNA translocase domain (residues 598-968) which is involved in DNA binding and ATP hydrolysis (Selby and Sancar, 1995a; Selby and Sancar, 1995b) (see 1.1.2.3). Binding of ATP or DNA may lead to conformational changes in this module. Therefore, the experiment was also carried out in the presence of ATP- γ -S, a non-hydrolyzable ATP analogon, and in the presence of ATP- γ -S and dsDNA, respectively. However, the addition of ATP- γ -S or dsDNA did not result in any detectable difference in the pattern (figure 3.3). In none of the experiments, stable fragment of this part could be found (see below). This indicates high flexibility in this region.

Stable cleavage products (figure 3.3, red boxes) were analyzed by EDMAN-sequencing and peptide mass fingerprinting analysis. In peptide mass fingerprinting analysis, the fragments of interest were tryptically digested under denaturing conditions. Peptides were identified by MALDI-ToF mass spectrometry. As reference, Mfd-FL (band 0) was used. Mfd-FL could be identified in the SwissProt database using the MASCOT Search Engine. Sequence coverage of 48 % (556/1162) was achieved.

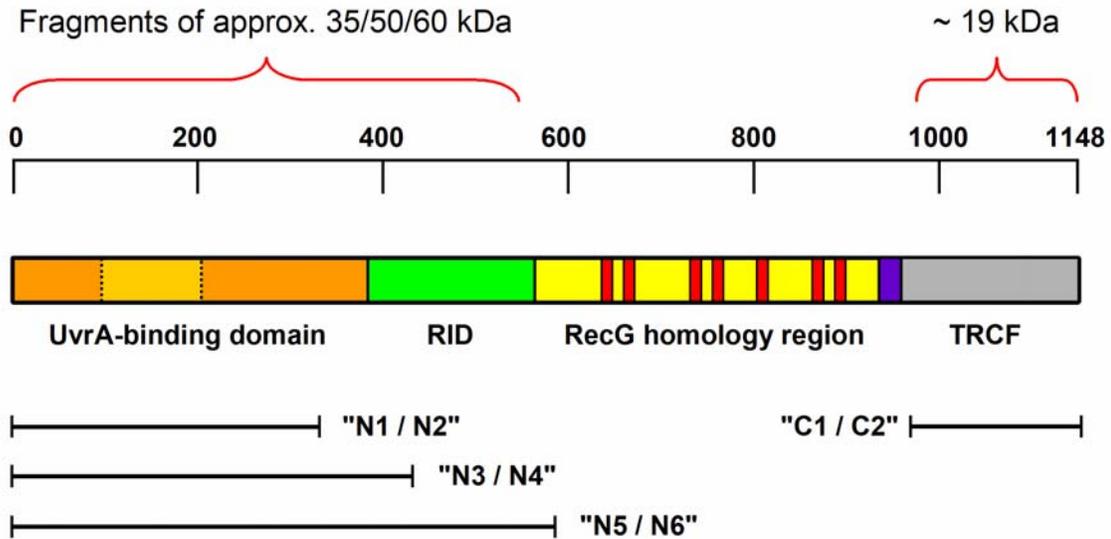


Figure 3.4: Stable regions of Mfd identified by limited proteolysis. While the N- and C-termini appeared rather stable, no fragments of the catalytic domain could be found. N- and C-terminal constructs are schematically shown.

By mass spectrometry, only peptides in the range between 800 and 3300 Da were analyzed. Thus, peptides shorter than 8 residues were neglected. However, in some regions of Mfd, tryptic cleavage sites are located closer to each other. These peptides could not be detected in any of the samples. For instance, the N-terminal 40 residues were found in none of the samples, although EDMAN-sequencing of the full-length protein as well as of fragments 1-4 could identify the native N-terminus. Therefore, the list of identified peptides for a certain fragment was always compared with that for the full-length protein and/or longer fragments.

A detailed list of peptides identified for each fragment can be found in the supplementary material (7.1).

Bands 1-4 were found to be N-terminal fragments of 35 kDa (1), 50 kDa (2/4) and 60 kDa (3). Band 5 corresponds to the C-terminal 163 residues. No stable fragments of the catalytic domain could be found (figure 3.4).

Based on these results, several N- and C-terminal constructs were cloned (table 3.1). For construct design, protein secondary structure prediction from the PSIPRED method (Jones, 1999) was also taken into account.

Table 3.1: N- and C-terminal constructs of *E.coli* Mfd which were designed based on the results of the limited proteolysis assay.

construct	no tag	molecular weight	with HIS ₆ -tag	molecular weight
M1 – N333	"Mfd-N1"	38.1 kDa	"Mfd-N2"	39.3 kDa
M1 – N433	"Mfd-N3"	49.2 kDa	"Mfd-N4"	50.5 kDa
M1 – D586	"Mfd-N5"	66.1 kDa	"Mfd-N6"	67.3 kDa
S964 – A1148	"Mfd-C1"	21.4 kDa	"Mfd-C2"	23.6 kDa

3.2 Purification, crystallization and structure determination of Mfd-N2

3.2.1 Purification of Mfd-N2

For structure determination, an N-terminal construct of Mfd comprising the first 333 residues of *E.coli* Mfd with a C-terminal hexahistidine tag ("Mfd-N2") was used.

Mfd-N2 was purified as described in 2.4.1. Briefly, Mfd-N2 overexpressing Rosetta (DE3) cells were lysed by sonification, and the clarified lysate was loaded onto a Ni-NTA column. After elution, the protein was further purified by anion exchange chromatography using a Resource Q column, and by size exclusion chromatography on a HiLoad 16/60 Superdex 200 pg column (figure 3.5). Peak fractions from the size exclusion were pooled. Protein was concentrated to 5 mg/ml and used for crystallization trials. From 6 l of expression culture, 20 mg of highly pure Mfd-N2 could be obtained.

In the original purification protocol, elution fractions from the Ni²⁺-NTA column were directly applied to the size exclusion column. Crystals could be obtained from this protein batch, but they could not be reproduced or refined. Therefore, the anion exchange step was introduced between the affinity chromatography and the size exclusion column in order to further improve purity and homogeneity.

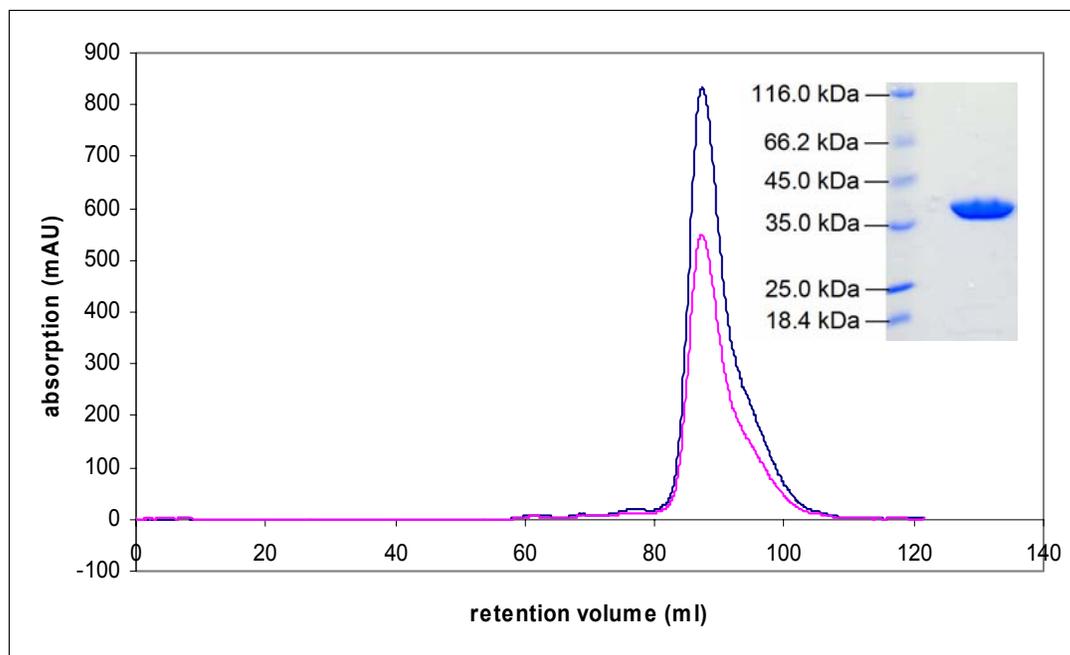


Figure 3.5: Elution profile of Mfd-N2 from S200 16/60 size exclusion column (absorption at 280 nm, blue; absorption at 260 nm, pink). SDS-PAGE analysis of pooled and concentrated peak fractions is shown as insert.

Other N-terminal constructs, "Mfd-N4" and "Mfd-N6" (table 3.1), were purified analogously.

3.2.2 Crystallization

For initial crystallization trials with commercial screens, a crystallization robot was used to set 0.5 + 0.5 μ l drops with 50 μ l reservoir volume in 96-well plates. First crystals were obtained with 5 mg/ml at 20°C.

Small crystals with hexagonal morphology appeared over night in the crystallization conditions Hampton Index #6 (0.1 M TRIS/HCl pH 8.5, 2.0 M ammonium sulfate) and Jena Biosciences Screen 3 #C6 (0.1 M sodium citrate pH 5.6, 0.2 M ammonium sulfate, 25% PEG-4000).

20% PEG-400 was added as cryoprotectant and crystals were snap-frozen in liquid nitrogen. Diffraction to ~ 9 Å at beamline PX (SLS, Villigen, Switzerland) could be detected using crystals from the Jena Biosciences condition.

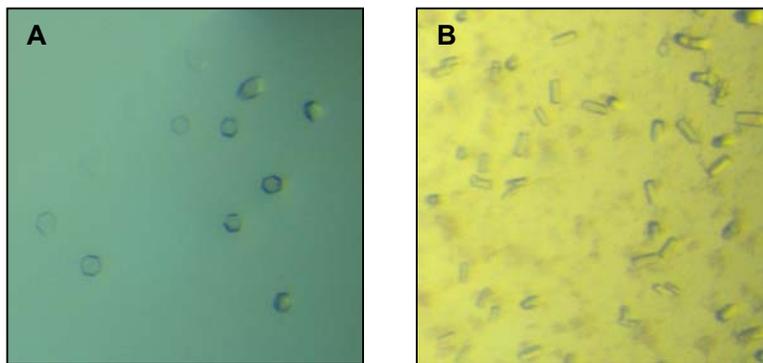


Figure 3.6: Initial crystals of Mfd-N2 with a diameter of $< 10 \mu\text{m}$ were obtained from A) Hampton Index reagent #6 (0.1 M TRIS/HCl pH 8.5, 2.0 M ammonium sulfate) and B) Jena Biosciences Screen 3 condition #C6 (0.1 M sodium citrate pH 5.6, 0.2 M ammonium sulfate, 25% PEG-4000).

For structure determination, crystal size and quality had to be improved. Therefore, refinements were set up by hand using the condition from the Jena Biosciences screen. A lower protein concentration of 4 mg/ml and a larger drop size of 1 + 1 μl (50 μl reservoir volume) gave slightly bigger crystals. Variation of the reservoir solution composition, however, did not result in any improvement compared to the original condition.

Therefore, extensive additive screening was performed. 80 % (40 μl) of the initial crystallization condition (0.1 M sodium citrate pH 5.6, 0.2 M ammonium sulfate, 25% PEG-4000) were mixed with 20 % (10 μl) of a crystal screen reagent. Significantly bigger crystals could be observed when using condition #25 from Hampton Index (3.5 M sodium formate) or Nextal Classic condition #48 (4.0 M sodium formate) as additive.

By variation of the formate concentration in the additive stock solution and the proportion between initial condition and additive, a final optimized crystallization condition was achieved. Using 0.08 M sodium citrate pH 5.6, 0.16 M ammonium sulfate, 20% PEG-4000, 0.8 M sodium formate, crystals with a maximum size of 40 μm x 40 μm x 110 μm were obtained (figure 3.7). Selenium-containing crystals could be grown in the same conditions.

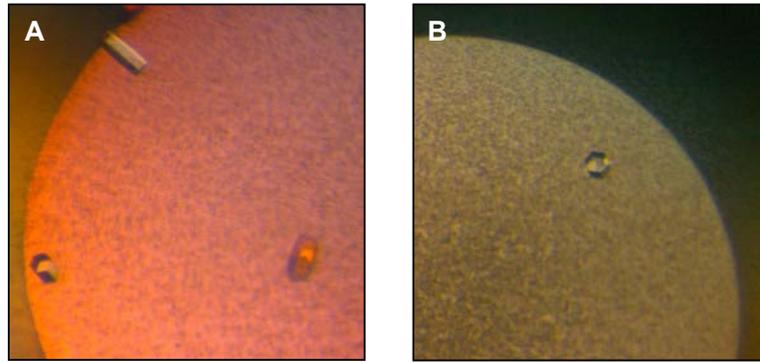


Figure 3.7: Refined crystals were obtained with 0.08 M sodium citrate pH 5.6, 0.16 M ammonium sulfate, 20% PEG-4000 and 0.8 M sodium formate at 4 mg/ml. Panel A) shows native crystals, B) shows a crystal of selenomethionine containing Mfd-N2. The crystals were of hexagonal morphology and had a diameter of ca. 40 μm .

Crystals belonged to space group $P6_522$ with unit cell constants of $a=b=112.56 \text{ \AA}$, $c=213.50 \text{ \AA}$, $\alpha=\beta=90^\circ$ $\gamma=120^\circ$ (native crystal). The selenium containing crystals were isomorphous with a larger cell volume by only 0.3 %.

The crystals showed an extremely low mosaicity of 0.24 (derivative, determined with DENZO) and 0.09, respectively (native, calculated by XDS). The asymmetric unit contained two molecules of Mfd-N2. This results in a solvent content of 50.45 % and a Matthews volume of $2.48 \text{ \AA}^3/\text{dalton}$ of protein (Matthews, 1968; Kantardjieff and Rupp, 2003).

For data collection, crystals were transferred to reservoir solution supplemented with 20% PEG-400 and snap-frozen in liquid nitrogen.

3.2.3 Data collection

All diffraction data used for structure determination were collected at beamline ID14-4 (ESRF, Grenoble, France) with an ADSC Q4 CCD detector. 140 images (1° oscillation each) were recorded for each dataset.

A native dataset at 0.97395 \AA was collected to 2.1 \AA resolution.

For phase determination, a three wavelength anomalous dispersion experiment was carried out at the selenium K edge using one selenium containing crystal. The optimal wavelengths were determined experimentally with a fluorescence scan (see figure 1.7).

Datasets for the peak wavelength at 0.97788 Å (12678.23 keV: f' -5.32, f'' 4.27) and the inflection point at 0.979804 Å (12653.96 keV: f' -9.01, f'' 2.55) were collected to 2.6 Å. Due to radiation damage, diffraction data for the high remote wavelength at 0.97395 Å (12707.6 keV: f' -4.17, f'' 4.00) could only be obtained to 2.8 Å.

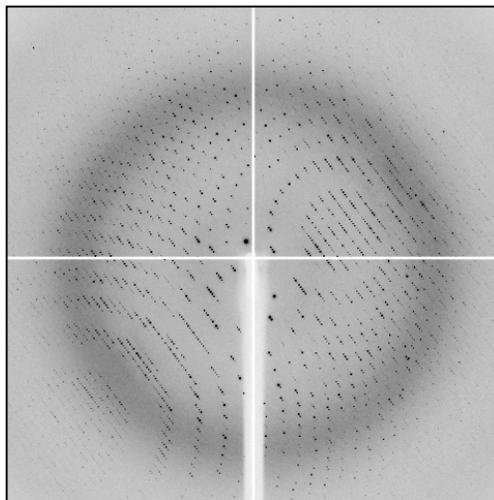


Figure 3.8: *Diffraction image of Mfd-N2 recorded at beamline ID14-4 (ESRF, Grenoble, France). Native crystals diffracted to a resolution of 2.1 Å.*

3.2.4 Structure determination and refinement

Anomalous data were processed with DENZO and SCALEPACK (Otwinowski and Minor, 1997), while the native dataset was processed using XDS and XSCALE (Kabsch, 1993). Data were first indexed and scaled in space group P622. By means of systematic absences in the dataset (figure 3.9), the potential space groups could be limited to the enantiomorphous P6₁22 and P6₅22 (International Tables for Crystallography, Volume A: Space-group symmetry, 2002). The correct space group was P6₅22, which was revealed after map calculation with the two possible screw axes.

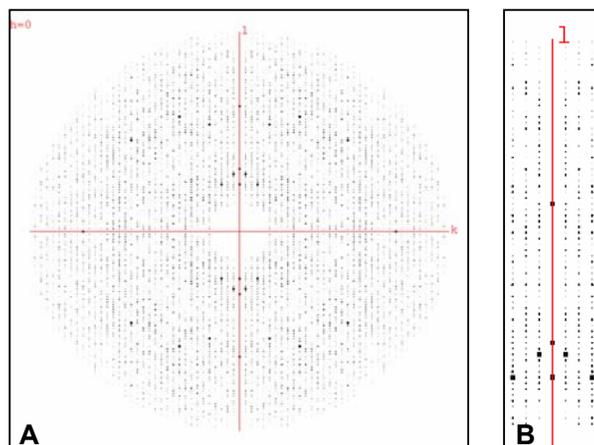


Figure 3.9: A) Presentation of the native dataset (*native.mtz*) with *hklview* (CCP4) on section *0kl*. Data was processed in space group *P622*. B) Zoomed section of A. On the *l*-axis ($h=0, k=0$), only every sixth reflection was present. This corresponds to a 6_1 or 6_5 screw axis.

18 selenium sites could be located by SOLVE (Terwilliger, 2002). Initial phases were calculated with SHARP (de la Fortelle and Bricogne, 1997), and were improved by solvent flattening with SOLOMON (Abrahams and Leslie, 1996).

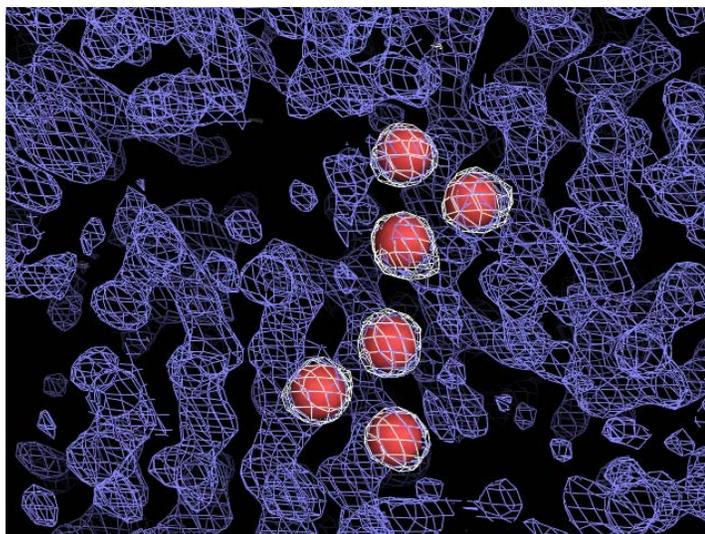


Figure 3.10: Section of 1σ contoured MAD map at 2.8 \AA resolution (blue mesh) with anomalous difference density contoured at 2.5σ (white mesh). Six selenium atoms are shown as red spheres.

The resolution of 2.8 Å allowed automated model building. ARP/wARP (Morris *et al.*, 2003) built 567 out of 688 possible amino acids in the asymmetric unit. The model was completed manually with MAIN (Turk, 1992). After bulk solvent correction and anisotropic overall B-value correction, the model was refined against the 2.1 Å native data by rigid body refinement with CNS v.1.1 (Brunger *et al.*, 1998). The refinement was continued by iterative cycles of simulated annealing, positional refinement and individual B-factor refinement with CNS, followed by manual model building with MAIN. Initial NCS restraints were gradually removed in the final cycles of the refinement.

In the refined structure, 93.9% of the residues are found in the core of the Ramachandran plot, and none of the residues is found in a disallowed region (Laskowski *et al.*, 1993).

Data collection and model statistics are given in table 3.2.

The final model comprised 617 residues of which 15 were present in two conformations (eight residues in molecule A, seven residues in molecule B). In addition, 440 water molecules, one sulphate and three sodium ions were detectable. Interestingly, also two PEG-400 molecules were visible in the density. PEG-400 was not present in the original crystallization reagent. It was used for cryoprotection. Crystals were transferred into PEG-400 containing solution only for a few seconds directly before freezing.

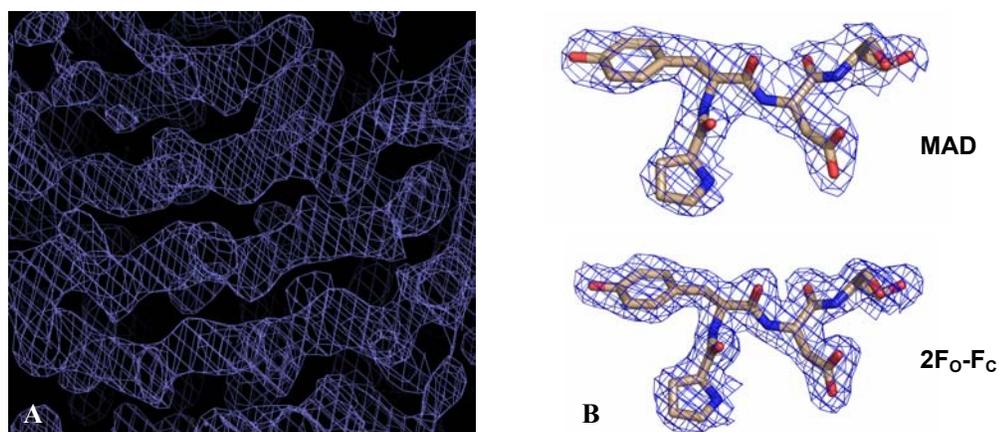


Figure 3.11: A) Zoomed section of 1σ contoured MAD electron density map (blue mesh) showing 4 antiparallel β -sheets. B) Exemplary experimental (2.8 Å, upper panel) and final $2F_o-F_c$ (2.1 Å, lower panel) electron density maps. The final model is shown as colour-coded sticks.

Table 3.2: Crystallographic data collection and model refinement

Crystal	SeMet			native
Space Group	P6 ₅ 22 ¹			P6 ₅ 22 ²
Dataset	peak	inflection point	high remote	native
Wavelength (Å)	0.97788	0.9798	0.97395	0.97395
Data range (Å)	20.0-2.6	20.0-2.6	20.0-2.8	20.0-2.1
Observations (unique)	416744 (46059)	417006 (46109)	259154 (36731)	1018169 (47377)
Completeness (%) (last shell)	98.6 (96.8) ³	98.6 (97.2) ³	98.3 (96.5) ³	99.5 (98.7)
R _{sym} ⁴ (last shell)	0.098 (0.273)	0.095 (0.271)	0.120 (0.324)	0.082 (0.362)
R _{meas} ⁵ (last shell)	0.104 (0.290)	0.101 (0.287)	0.128 (0.344)	0.084 (0.372)
I/σI (last shell)	23.0 (7.1)	19.6 (7.9)	16.8 (6.9)	33.9 (10)

Refinement	native
Data range (Å)	19.86-2.1
Reflections F>0 (cross validation)	47432
Protein atoms (solvent molecules)	5064 (442)
R _{work} ⁶ / R _{free} ⁷ (%)	19.8 / 23.0
rmsd bond length (Å) / bond angles (°)	0.0084 / 1.34
Core (disallowed) in Ramachandran plot (%)	93.9 (0)

¹ unit cell (P6₅22) (Å/°): a=b=112.56 c=213.50 α=β=90 γ=120, two molecules per asymmetric unit

² unit cell (P6₅22) (Å/°): a=b=112.65 c=213.86 α=β=90 γ=120, two molecules per asymmetric unit

³ anomalous completeness

⁴ R_{sym} is the unweighted R value on I between symmetry mates

⁵ R_{meas} is the weighted R value on I between symmetry mates (Diederichs and Karplus, 1997).

⁶ R_{work} = $\sum_{hkl} \left| |F_{obs}(hkl)| - |F_{calc}(hkl)| \right| / \sum_{hkl} |F_{obs}(hkl)|$ for reflections in the working data set

⁷ R_{free} = $\sum_{hkl \text{ testset}} \left| |F_{obs}(hkl \text{ testset})| - |F_{calc}(hkl \text{ testset})| \right| / \sum_{hkl} |F_{obs}(hkl \text{ testset})|$ for 5% of reflections against which the model was not refined

Near the interface of the two molecules in the asymmetric unit, six methionine residues (three residues of each molecule) are located very close to each other (figure 3.10). The intramolecular distances between the sulfur atoms are in the range of 4-5 Å (table 3.3). Small distances between the heavy atoms in the selenium-containing crystal can cause difficulties in the separation of the anomalous peaks during phase determination. This may

explain why other phasing attempts, e.g. by SAD with SHELXD (Ness *et al.*, 2004), had failed.

Table 3.3: Intramolecular distances between the sulfur atoms of methionine residues M66, M68, and M101.

Methionine residues	molecule A	molecule B
M66 – M68	4.19 Å	4.20 Å
M66 – M101	4.20 Å	4.47 Å
M68 – M101	4.89 Å	5.28 Å

3.2.5 Mfd-N2 crystallized with two molecules in the asymmetric unit.

One asymmetric unit contained two molecules of Mfd-N2 (figure 3.12). The two molecules were highly similar (RMSD of 0.93 Å). They shared a buried surface of 1834.19 Å² (Brunger *et al.*, 1998).

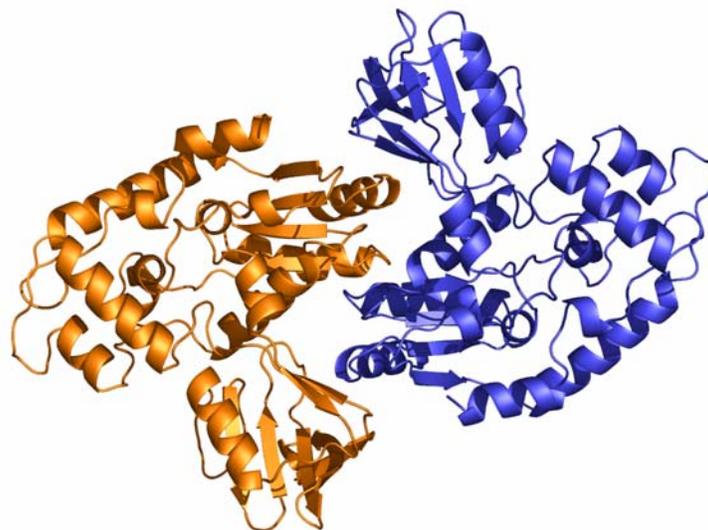


Figure 3.12: Mfd-N2 crystallized with two molecules in the asymmetric unit (molecule A, orange; molecule B, blue).

According to literature, Mfd is functional as a monomeric protein (Selby and Sancar, 1993; Roberts and Park, 2004). In order to distinguish between a biological dimer and a crystallographic complex, analytical size exclusion chromatography using a Superdex 200 10/300 GL column was performed (figure 3.13).

The retention volume of macromolecules from size exclusion columns depends on the hydrodynamic radius which corresponds to the (approximate) molecular weight for globular proteins. The column had previously been calibrated with a gel filtration standard. Thus, the apparent molecular weight (kDa) could be calculated by use of the retention volume (ml).

From a retention peak volume of 15.17 ml, a molecular weight of 32.2 kDa could be calculated. This is consistent with the size of a monomer (39.3 kDa).

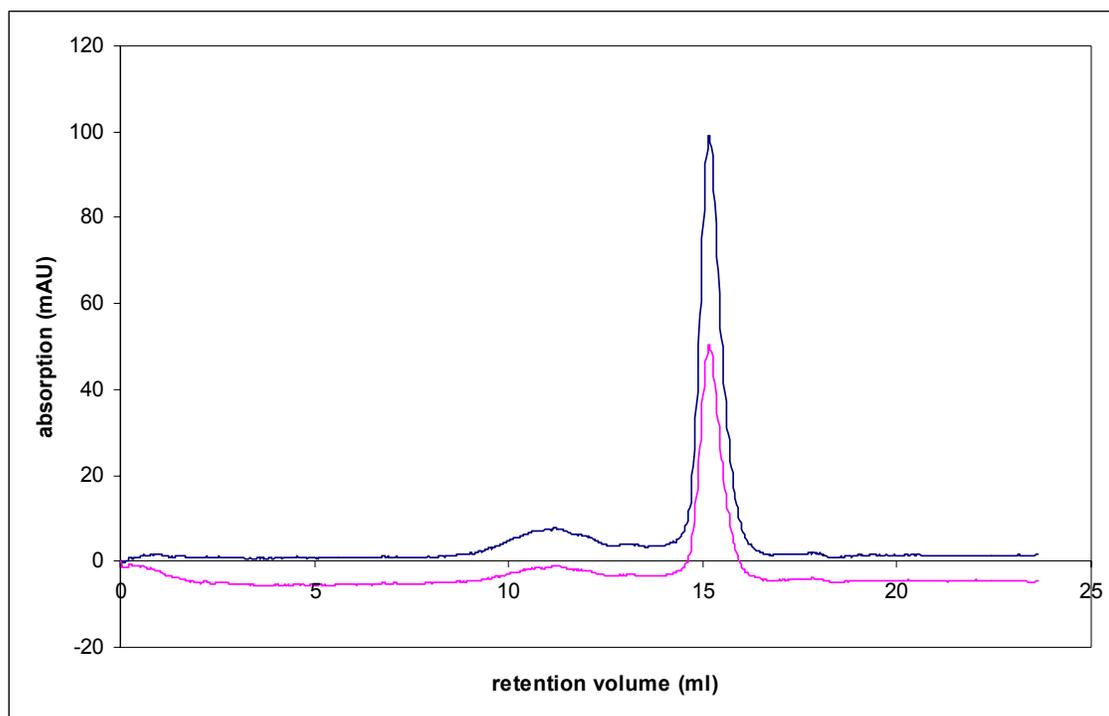


Figure 3.13: Analytical size exclusion chromatography of Mfd-N2 (absorption at 280 nm, blue; absorption at 260 nm, pink). The retention volume of 15.17 ml indicates that Mfd-N2 is monomeric.

3.3 Structure of Mfd-N2

3.3.1 Mfd-N2 crystal structure

Mfd-N2 is a triangular molecule of approximately $60 \text{ \AA} \times 60 \text{ \AA} \times 30 \text{ \AA}$ dimensions. It consists of three structural domains denoted domains 1A, 1B and 2 according to the UvrB nomenclature (Theis *et al.*, 1999). Each domain forms one corner of the triangle (figure 3.14).

Domain 1A (residues 26-70, 85-114, 266-286, 324-333, coloured orange in figure 3.14) contains both the N- and the C-terminus of Mfd-N2 and forms the structural framework of the molecule. A central parallel β -sheet (strands $\beta 1$, $\beta 2$, $\beta 3$, $\beta 11$) is sandwiched between two layers of α -helices (αA , αB and αC , αD , αK , αN). Domain 1A possesses the typical α/β fold of RecA-like ATPases.

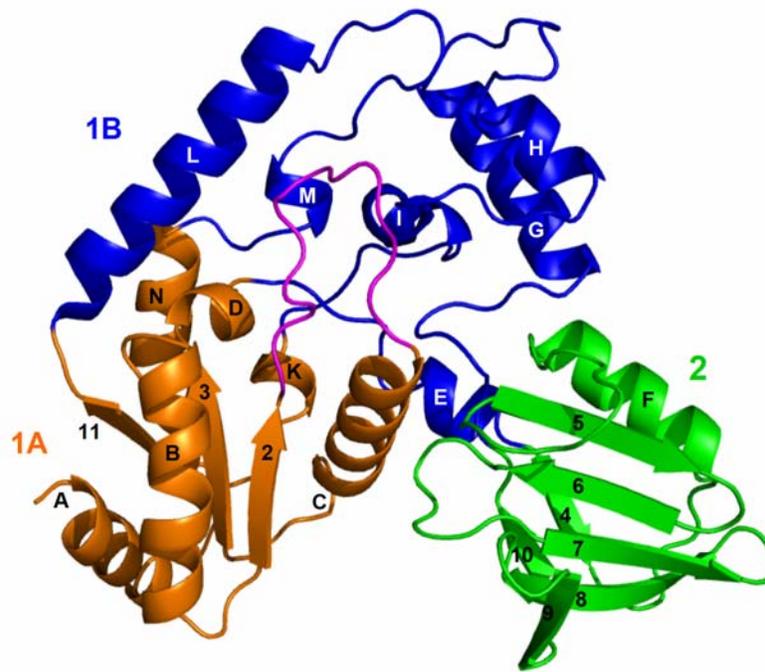


Figure 3.14: Ribbon presentation of Mfd-N2 with annotated secondary structure (α -helices, letters A-N; β -sheets, numbers 1-11). Mfd-N2 consists of three domains (domain 1A, orange; domain 1B, blue with loop, magenta; domain 2, green) that form a triangularly shaped structure.

Domain 1B (residues 115-127, 212-265, 287-323, coloured in blue) is formed by three segments inserted into the primary sequence of domain 1A. Domain 1B is situated "on top" of domain 1A and is mainly α -helical. The body of domain 1B consists of two large

α -helical lobes, i.e. a helix-loop-helix protrusion (α L, α M) and a three helix bundle (α G, α H, α I). A prominent loop (residues 71-84, magenta in figure 3.14), joining β 2 and α C of domain 1A, binds along the interface of the two α -helical lobes and completes domain 1B. Domain 1A and 1B form a structural unit. Domain 1B is assembled by three insertions into domain 1A. In addition, the two domains share an extensive, hydrophobic interface. Therefore, the mutual orientation of 1A and 1B appears rather stable.

The compact globular domain 2 (residues 127–212, green in figure 3.14) is formed by a single insertion between helices α E and α G of domain 1B. Domain 2 is situated at the side of the interface of domains 1A and 1B. Domain 2 has a $\beta\alpha\beta\beta\beta\beta\beta$ topology. A double layer of β -sheets packs on one side against a single α -helix (α F). The two antiparallel strands of the inner β -sheet (β 4, β 10) connect domain 2 to domain 1B. The five strands of the outer β -sheet (β 5- β 9) pack against this inner β -sheet and form a flat, slightly twisted solvent-exposed surface.

In contrast to the intimate interaction of domains 1A and 1B, domain 2 appears less firmly attached. The interface of domain 2 with the remainder of Mfd-N2 is formed mainly by two loops, the loop between α F and β 5 with domain 1B, and the loop between β 6 and β 7 with α C of domain 1A. The two β -strands of the inner sheet covalently attach domain 2 to domain 1B. The interface between domain 2 and the remainder of Mfd-N2 is rather hydrophilic in nature. It contains a number of hydrogen bonds, but does not possess a large hydrophobic component. From a structural point of view, this interaction could allow some movement of domain 2.

3.3.2 Conservation of the Mfd N-terminus

Sequence conservation between Mfd proteins from different organisms was mapped onto the molecular surface of Mfd-N2 (figure 3.15).

Two conserved surface patches on Mfd-N2 are revealed: One patch is formed by a loop located across the interface of domains 1A and 1B. This loop corresponds to the “ β -hairpin” motif in UvrB which functions in DNA binding in the nucleotide excision repair protein (see below) (Skorvaga *et al.*, 2002; Truglio *et al.*, 2006b).

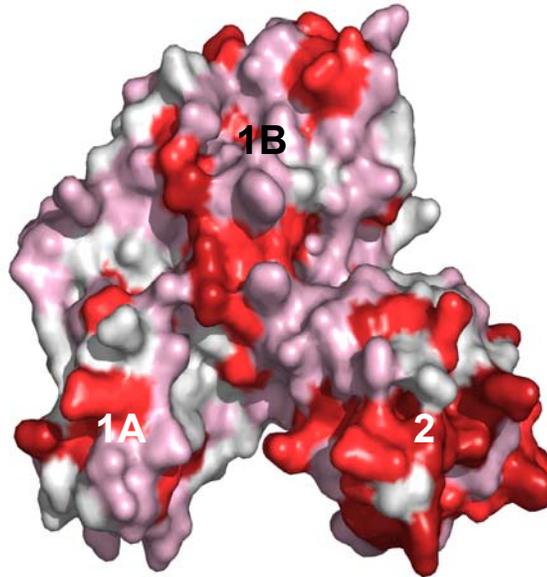


Figure 3.15: Molecular surface of Mfd-N2 (oriented as in figure 3.14) with annotated structural domains. The colour coding corresponds to Mfd sequence conservation (dark red, conserved; white, unconserved). The most strongly conserved region of the Mfd N-terminus is the putative UvrA binding domain 2.

The most strongly conserved region in Mfd-N2 is found on the surface of domain 2 (Selby and Sancar, 1993). Domain 2 is believed to function in UvrA-binding implicating a role of the conserved residues (Truglio *et al.*, 2004). This would also be consistent with the strong conservation in domain 2 between Mfd and UvrB. Domain 2 will be discussed in more detail in chapter 3.3.4.

3.3.3 Comparison of Mfd-N2 to UvrB

Both in Mfd and UvrB, the N-terminal region functions in UvrA binding (Hsu *et al.*, 1995; Selby and Sancar, 1995a; Truglio *et al.*, 2004).

By sequence comparison between UvrB and Mfd proteins, high homology was observed for domain 2 and, therefore, structural similarity has been proposed for this region (Selby and Sancar, 1993; Truglio *et al.*, 2004).

The structure and topology of Mfd-N2 was compared with the crystal structure of *Bacillus caldotenax* UvrB variant Y95A (PDB-ID 1T5L; Truglio *et al.*, 2004) (figures 3.16 and 3.18). Surprisingly, not only domain 2 is structurally similar to the corresponding part of UvrB. In fact, all three domains of Mfd-N2 resemble the fold and mutual arrangement of the three N-terminal domains of UvrB.

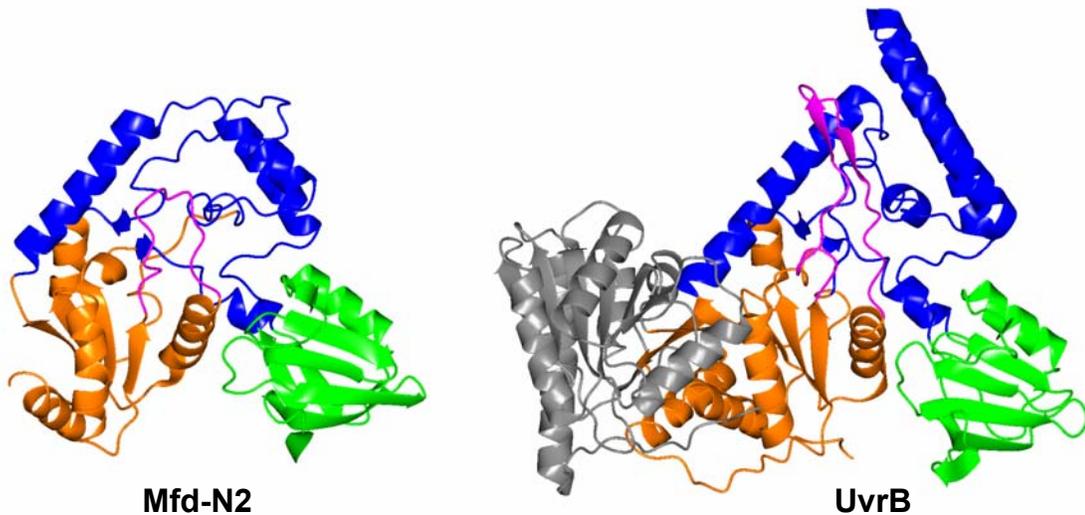
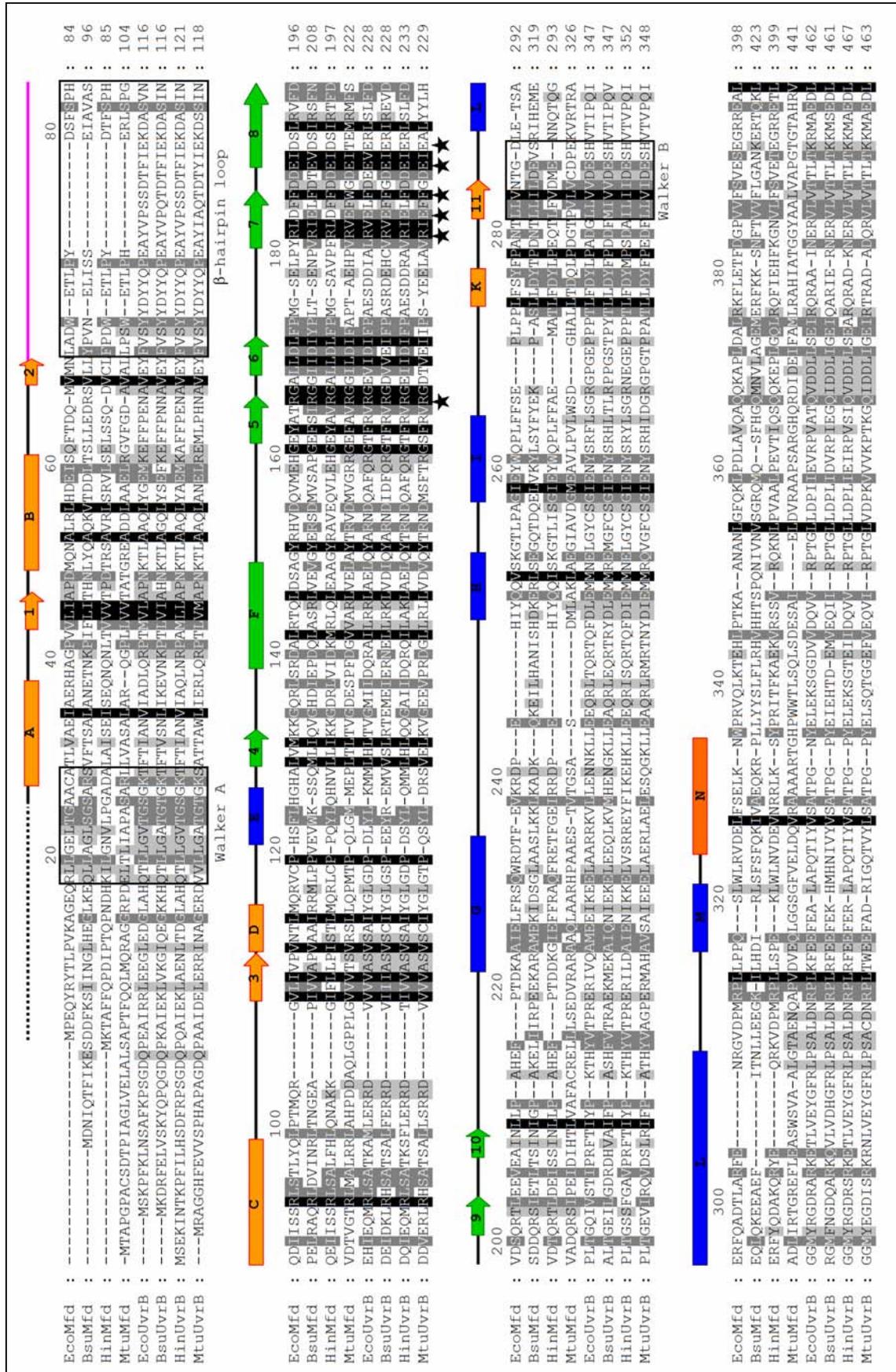


Figure 3.16: Comparison of the Mfd-N2 structure with *B.caldotenax* UvrB (PDB-ID 1T5L; Truglio *et al.*, 2004). The orientation and colour coding is as in figure 3.14 (domain 1A, orange; domain 1B, blue with loop, magenta; domain 2, green; UvrB domain 3, grey).

A sequence alignment of UvrB and Mfd N-terminal regions was generated with ClustalW (<http://align.genome.jp/>) and edited manually using GeneDoc (Nicholas and Nicholas, 1997) based on *B.caldotenax* UvrB and *E.coli* Mfd-N2 crystal structures.

The sequences of *E.coli* Mfd-N2 (residues 26-333) and the corresponding region of *B.caldotenax* UvrB (residues 45-390) have low pair-wise sequence identity of 17 %. However, high structural similarity (RMSD of 2.7 Å (Potterton *et al.*, 2002; Potterton *et al.*, 2004)) can be observed.

Figure 3.17 (next page): Structure-based sequence alignment of Mfd and UvrB N-terminal regions. The secondary structure of Mfd-N2 is shown on top of the alignment and annotated according to figure 3.14. Conserved residues between Mfd and UvrB are shaded, dark indicating stronger conservation. Functional motifs mentioned in the text are boxed. Residues used in mutational studies are indicated with stars. Abbreviations: *Eco*, Escherichia coli; *Bsu*, Bacillus subtilis; *Hin*, Haemophilus influenza; *Mtu*, Mycobacterium tuberculosis.



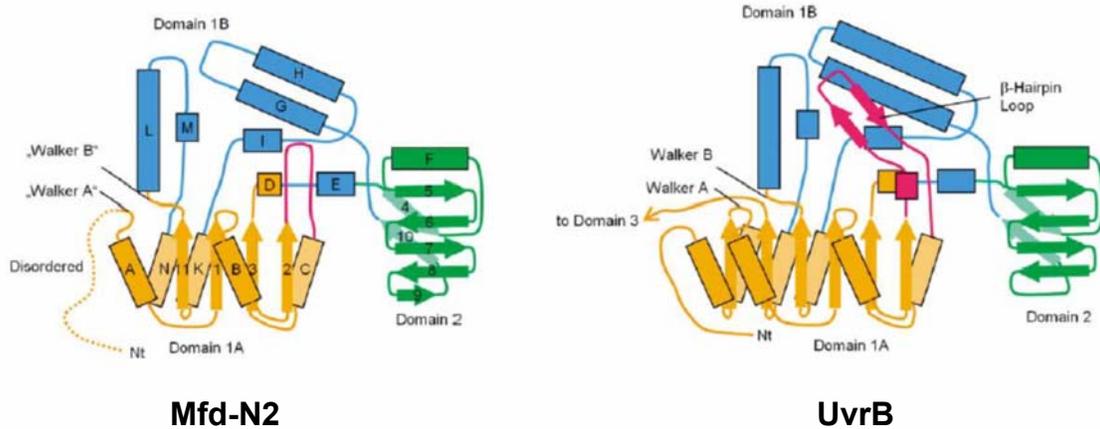


Figure 3.18: Topology diagrams of domains 1A, 1B and 2 of Mfd-N2 (left panel) and the corresponding domains of UvrB (right panel), using the color code and secondary structure annotation of figure 18. Walker A and B motifs of Mfd-N2 are indicated with quotation marks, because they are degenerated from the canonical sequences.

Domain 1A of Mfd possesses the typical α/β fold of RecA-type ATPase domains and corresponds to the first RecA-like domain of UvrB. Domain 1B is more compact in Mfd than in UvrB but also shares the basic architecture. Interestingly, the β -hairpin that is implicated in DNA binding in UvrB is missing in Mfd (see 3.3.5.2). Domain 2, as expected from sequence similarity, is structurally very similar to UvrB domain 2 (see 3.3.4).

However, while this domain by itself matches well, it is rotated with respect to domains 1A and 1B (figure 3.19). This was also found in the full-length *E.coli* Mfd structure (Deaconescu *et al.*, 2006). As suggested above, the hydrophilic interface between domain 2 and the remainder of Mfd-N2 might allow some movement of domain 2 with respect to domains 1A and 1B. The crystal structure of *B.caldotenax* UvrB mutant Y95A solved by Truglio and colleagues was the first structure of an UvrB protein where domain 2 was clearly defined. In several other crystal structures of UvrB proteins, no clear electron density of domain 2 was visible (Machius *et al.*, 1999; Nakagawa *et al.*, 1999; Theis *et al.*, 1999; Truglio *et al.*, 2004). This indicates a high mobility of domain 2 also in UvrB.

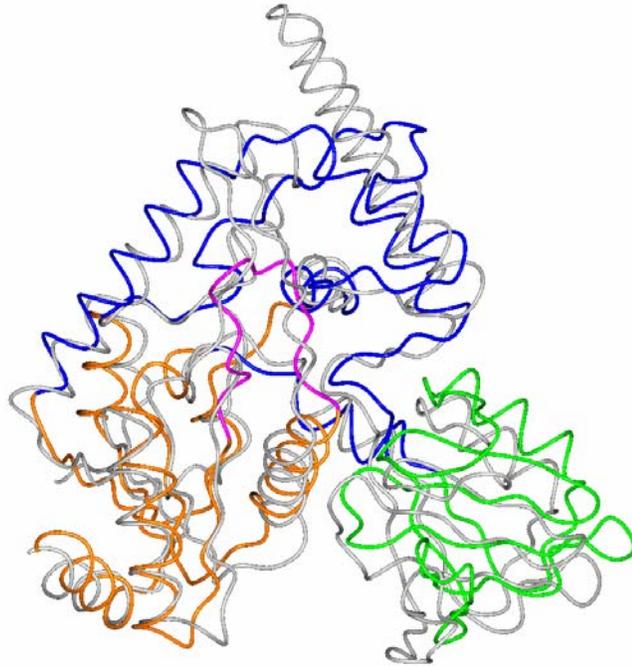


Figure 3.19: Backbone worms of domains 1A, 1B and 2 of Mfd-N2 (colour code of figure 3.14) and B.caldotenax UvrB (light gray) superimposed on domains 1A and 1B. Domain 2 is rotated with respect to the remainder of the molecule.

UvrB possesses a second RecA-like domain, domain 3 (coloured grey in figure 3.16). Together with domain 1A, domain 3 forms the ATP-dependent "helicase" motor in UvrB (Hsu *et al.*, 1995). The corresponding region is missing in the Mfd-N2 construct. It was shown recently that Mfd does not possess an equivalent domain to UvrB domain 3 (Deaconescu *et al.*, 2006).

3.3.4 Domain 2

3.3.4.1 Superposition of domain 2 of Mfd and UvrB

As shown above, the architecture of Mfd-N2 very much resembles that of UvrB. The region with the highest structural similarity between Mfd and UvrB was found to be domain 2. Residues 126-213 of *E.coli* Mfd superimpose well (RMSD of 1.35 Å) with domain 2 (residues 157–245) of *B.caldotenax* UvrB (figure 3.20).

Domain 2 is the most conserved region of Mfd-N2 among Mfd proteins (see above). In addition, high sequence homology between Mfd and UvrB can be found (figure 3.17).

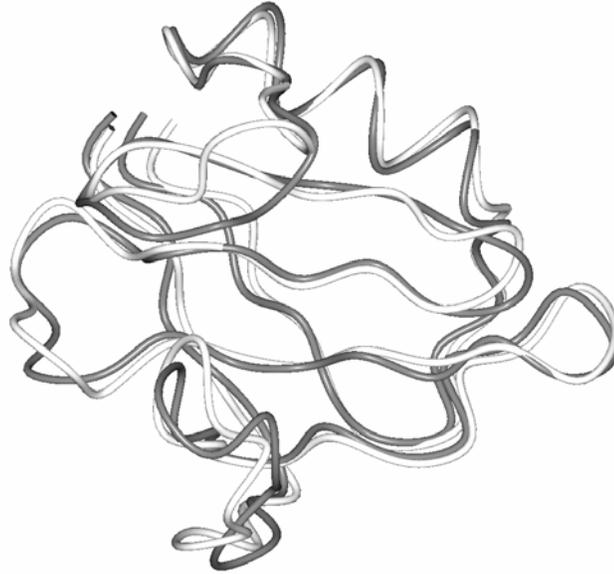


Figure 3.20: Superposition of domain 2 of *E. coli* Mfd (grey) and *B. caldovenax* UvrB (white). This region is strongly conserved between Mfd and UvrB.

3.3.4.2 Potential interaction sites

Both in Mfd and UvrB, the N-termini function in UvrA binding. (Hsu *et al.*, 1995; Selby and Sancar, 1995a; Truglio *et al.*, 2004). In UvrB, this region could be further limited down to domain 2, which was shown to be essential for a productive UvrA-UvrB interaction *in vitro* (Hsu *et al.*, 1995). Due to high sequence homology and structural similarity (see above), this is also expected to be true in Mfd (Truglio *et al.*, 2006a).

Many residues that are highly conserved between Mfd and UvrB cluster at the outside of the five-stranded twisted β -sheet of domain 2. Some of these residues in UvrB were found to be critical for the interaction with UvrA (Truglio *et al.*, 2004). Most of these conserved residues possess a charged character which is consistent with the salt-sensitivity of the UvrA-UvrB/Mfd interaction (Selby and Sancar, 1993; Selby and Sancar, 1994; Hsu *et al.*, 1995; Truglio *et al.*, 2004).

An unusually solvent-exposed hydrophobic amino acid can be found at the "tip" of domain 2. This phenylalanine 185 is extremely conserved, and might function as an "anchor" in the complex.

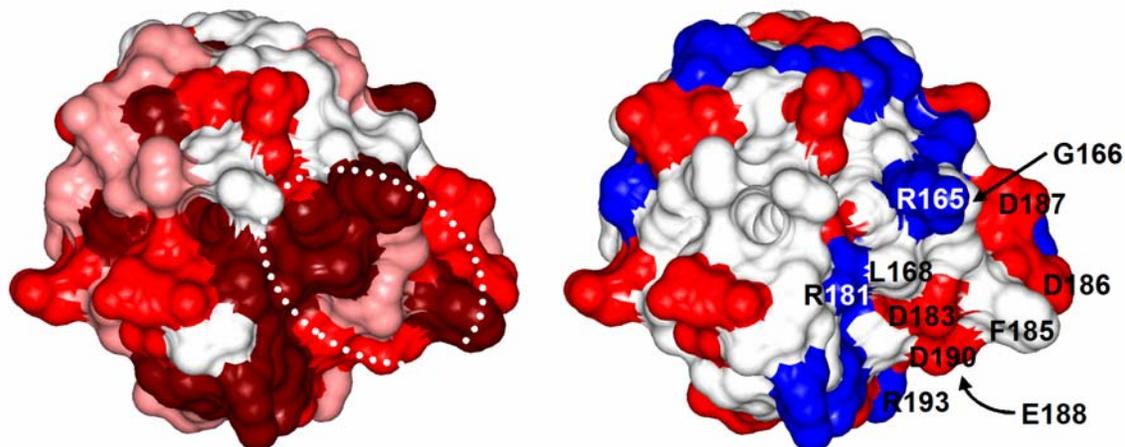


Figure 3.21: Molecular surface of Mfd domain 2 (orientation of figure 3.20). The left panel shows sequence conservation between Mfd and UvrB (dark red, conserved; white, unconserved), the conserved patch which is implicated to function in UvrA binding (see below) is encircled with a dotted line. The right panel is coloured according to residue types (negatively charged residues, red; positively charged residues, blue). Strongly conserved amino acids of Mfd and UvrB are mapped on the surface.

In order to investigate the role of these residues in UvrA binding, UvrA-binding properties of Mfd mutants was assayed (table 3.4).

Table 3.4: Mfd domain 2 mutants used for interaction studies and the corresponding residues in Bacillus caldotenax UvrB. Residues marked with a star () were shown to play a role in UvrA binding in UvrB (Truglio et al., 2004).*

Mfd (<i>E.coli</i>)	UvrB (<i>B.caldotenax</i>)
R165A	R196*
R181A	R213*
R181A/D183A	R213/E215*
F185A	F217
E188A	E220
D190A	E222

3.3.4.3 Interaction of Mfd mutants with UvrA

In order to analyse the affinities of Mfd mutants to UvrA, wild-type UvrA was immobilized on agarose beads. After addition of Mfd mutants, complexes were eluted and analyzed by SDS-PAGE.

The gene encoding wild-type UvrA was cloned into the pET-29 vector (table 2.3), and UvrA was expressed with a C-terminal HIS₆-tag. An untagged Mfd construct comprising the first 586 residues ("Mfd-N5") was used as binding partner. UvrA and Mfd possess a similar molecular weight. Therefore, the C-terminally truncated construct of Mfd was used. Mfd-N5 consists of the UvrA binding region and the RNA polymerase interacting domain. The affinity of this construct to UvrA had been confirmed previously. Mutants were generated by PCR-based site-directed mutagenesis from the wild-type gene and cloned into the pET-21b vector (table 2.3).

In a first step, UvrA was bound to Ni²⁺-NTA resin. After extensive washing, lysates of cells expressing Mfd-N5 mutants were added. Lysate from untransformed cells was used as negative control. The immobilized proteins were washed carefully with low salt buffer, and complexes of UvrA and Mfd-N5 were eluted afterwards. Elution fractions were analyzed by SDS-PAGE, and Mfd-N5 containing protein bands were quantified using the Image J software (<http://rsb.info.nih.gov/ij/>). The levels of UvrA-bound mutants were compared to the amount of retained wild-type protein (100 %).

The experiment was carried out analogously to Truglio *et al.*, 2004, who had analyzed the binding properties of UvrB mutants to UvrA. However, a different coupling method had to be used. For UvrA-UvrB interaction studies, proteins from *Bacillus caldotenax* were used. *B. caldotenax* UvrA could be expressed with a C-terminal intein-tag (Chong *et al.*, 1997; Chong *et al.*, 1998) and was immobilized on Chitin Beads (NEB, Frankfurt/Main, Germany). This system could not be applied for the *E. coli* protein. The fusion-protein of *E. coli* UvrA and the intein-tag could not be expressed, probably due to its molecular weight of ~160 kDa (103.8 kDa for UvrA, 55 kDa for the tag). Interestingly, the *B. caldotenax* protein which is about the same size as *E. coli* UvrA (105.6 kDa), could be produced in *E. coli* with the large intein-tag. Therefore, *E. coli* UvrA was immobilized on Ni²⁺-NTA agarose by use of a C-terminal HIS₆-tag. Upon addition of Mfd, UvrA "leakage" from the resin could be observed (figure 3.22). Therefore, the amount of retained Mfd had to be compared to the amount of bound UvrA.

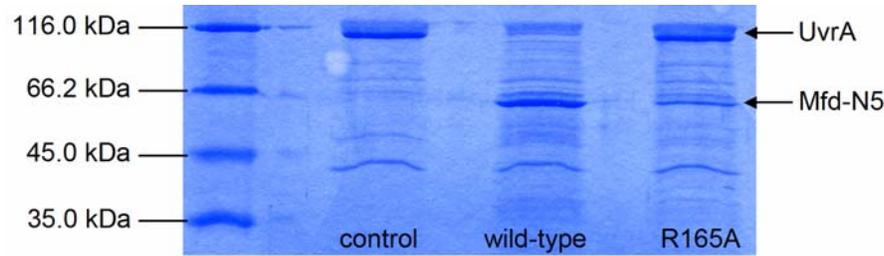


Figure 3.22: UvrA-binding properties of Mfd wild-type protein and domain 2 mutants. An exemplary section of an SDS-PAGE analysis is shown. From left to right: protein molecular weight marker; negative control (no Mfd); wild-type Mfd-N5, Mfd mutant R165A. Less UvrA remained bound to the Ni^{2+} -NTA resin when wild-type Mfd was added. The mutant R165A possessed dramatically decreased affinity to UvrA than wild-type Mfd.

Another problem during evaluation was the strong background level. The interaction between UvrA and Mfd (as well as UvrB) is salt-labile (Selby and Sancar, 1993; Hsu *et al.*, 1995; Truglio *et al.*, 2004). Salt-sensitivity of the complex was tested previously (data not shown). Therefore, columns were washed carefully with 100 mM NaCl. Not all impurities could be removed (figure 3.22)

Thus, exact quantification of protein bands was difficult, and values obtained from multiple experiments resulted in high standard deviations (see figure 3.23). Nevertheless, clear differences between the mutants could be observed.

While mutants E188A and D190A bound to UvrA at (approximately) wild-type level, the mutants R165A, R181A, R181A/D183A and F185 showed a drastically decreased affinity to UvrA (figure 3.23 and table 3.5).

Of all mutants analyzed, mutant R165A possesses the lowest affinity to UvrA. Compared to wild-type Mfd, only a fifth (18.6%) of protein amount was retained at the UvrA column. Residue R165 is absolutely conserved in all known UvrB and Mfd proteins. The homologous residue in *B.caldotenus* UvrB, R196, was also shown to play a role in UvrA binding (Truglio *et al.*, 2004).

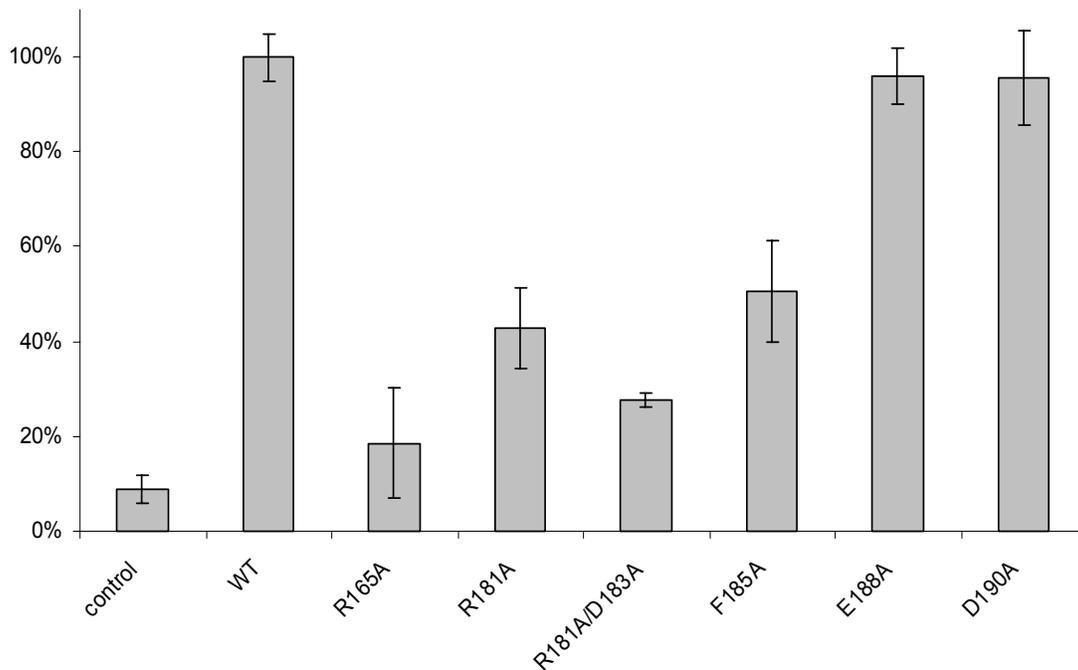


Figure 3.23: Affinities of different Mfd domain 2 mutants to wild-type UvrA. Shown are protein levels of Mfd-N5 which were retained at the UvrA column \pm standard deviation from four independent experiments.

Likewise, residues R181 and D183 seem to be involved in the UvrA-Mfd interaction. Of mutant R181A, protein was bound at less than half the amount (42.9%) of wild-type protein, the double mutant R181A/D183A was retained at a very low level level of 27.6%. R181 (R213 in UvrB) is strictly conserved among UvrB and Mfd. In some Mfd proteins, the residue corresponding to D183 (E215) is a glutamate instead of an aspartate, but the negatively charged character of this position is maintained in all Mfd and UvrB proteins (see figure 3.21). In UvrB, the double mutant R213A/E215A possessed an even lower affinity of 12% (table 3.5; Truglio *et al.*, 2004).

However, affinities of UvrB and Mfd mutants to UvrA were determined by quantification of SDS-PAGE bands. Therefore, the exact values obtained should be treated carefully (see above). Nevertheless, qualitative results from both experiments were in agreement with each other. UvrA-binding seems to require the same residues in Mfd as in UvrB and occurs electrostatically in both complexes.

Table 3.5: Protein levels of Mfd mutants bound to UvrA and comparison to the corresponding UvrB mutants (Truglio *et al.*, 2004); n.d. = not determined

<i>E.coli</i> Mfd	retention	<i>B.caldotenax</i> UvrB	retention
R165A	18.6%	R194A/R196A	40%
R181A	42.9%	R213A	n.d.
R181A / D183A	27.6%	R213A/E215A	12%
F185A	50.5%	F217	n.d.
$\Delta 2$	n.d.	$\Delta 2$	no binding

The amino acid F185 is an extremely conserved hydrophobic residue. It is located at the "tip" of domain 2 and is unusually solvent-exposed. The F185A mutant showed decreased affinity by ca. 50% to UvrA. These findings suggest that F185 might function as an "anchor" for the interaction with UvrA. No results are known for the corresponding UvrB mutant, but they are expected to be comparable.

All residues mentioned above are located very close to each other and form a conserved, mainly charged "patch" on the surface of Mfd domain 2 (figure 3.21).

Residues E188 and D190 are located slightly outside this region. Interestingly, the mutants E188A and D190A bound to UvrA at wild-type level (96% and 95.6%, respectively). Despite the high conservation and the charged character of these solvent exposed residues, they do not seem to be of importance for the interaction, probably due to their position.

An UvrB $\Delta 2$ mutant, where the complete domain 2 (residues 158-244) was replaced by a short linker, does not bind to UvrA at all (Truglio *et al.*, 2004). The same is expected for the corresponding Mfd mutant. UvrA-binding properties of an Mfd " $\Delta 2$ " variant will be analyzed.

In order to quantify binding affinities more exactly, a different method should be used. By Co-IP, e.g. using anti-HIS-tag antibodies coupled to protein A or protein G beads, UvrA "leakage" from the column material could be avoided. In addition, this approach may reduce strong background signals. Careful elution would be required to prevent elution of the antibody chains (J. Rauch, personal communication).

Surface plasmon resonance (SPR) is a very powerful method to determine binding affinities. SPR is used e.g. in the BIAcore system (BIAcore, Uppsala, Sweden). In order to immobilize one binding partner on a chip, chip surfaces with different functional groups are available. In this case, covalent coupling of UvrA would be most adequate. Ni²⁺-coated

surface for HIS-tagged proteins would not be recommended as the addition of Mfd lead to UvrA leakage (see above).

3.3.5 Functional sites

3.3.5.1 The Mfd N-terminus does not bind to DNA

Comparison of the crystal structures of Mfd-N2 and *B.caldotenax* UvrB revealed a striking difference between the two molecules:

UvrB domain 1B contains a β -hairpin (residues 90-116, magenta in figure 3.16) (Theis *et al.*, 1999; Skorvaga *et al.*, 2002; Truglio *et al.*, 2006b). This hairpin is conserved among UvrB proteins and is implicated in DNA-binding (Skorvaga *et al.*, 2002). Truglio and coworkers could show recently that UvrB binds to DNA by inserting the hairpin between the strands of the double helix. This leads to a destabilization of the damaged duplex and allows formation of the pre-incision complex. (Truglio *et al.*, 2006b). The β -hairpin is essential for discriminating between damaged and non-damaged DNA (Moolenaar *et al.*, 2001).

Interestingly, an equivalent to this β -hairpin is missing in Mfd (figure 3.16). At the corresponding position, between $\beta 2$ and αC of domain 1A, only a short, non-functional loop (residues 71-84) is present. This loop region is highly conserved among Mfd (see 3.3.2) and may represent the "base" of a degenerated hairpin.

DNA binding of full-length Mfd as well as of N-terminal constructs was tested by electrophoretic mobility shift assay (EMSA). Proteins were incubated with radioactively labelled double stranded oligonucleotides. Subsequently, protein-bound DNA was separated from free probe by polyacrylamide gel electrophoresis under native conditions. DNA-binding by Mfd is dependent on the presence of ATP, whereas ATP-hydrolysis disrupts the interaction (Selby and Sancar, 1995a; Selby and Sancar, 1995b). Therefore, the experiment was carried out in the presence of ATP- γ -S, a non-hydrolyzable ATP-analogue.

Full-length Mfd interacts with dsDNA, while the N-terminus of Mfd does not possess any DNA binding properties (figure 3.26).

This is consistent with the findings by Selby and Sancar who showed that ATP-dependent interaction between Mfd and double stranded polynucleotides is mediated by the C-

terminal dsDNA translocase domain of Mfd (see 1.1.2.3). No DNA-binding was reported for other regions of Mfd (Selby and Sancar, 1995a).

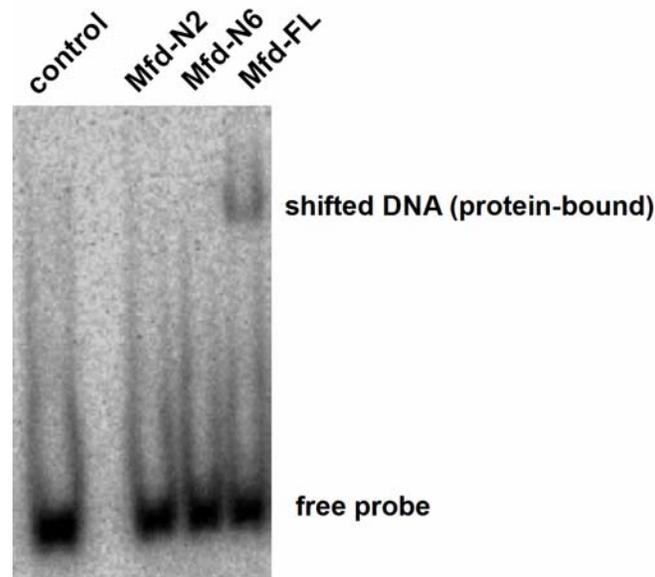


Figure 3.26: Electrophoretic mobility shift assay (EMSA) of Mfd N-terminal constructs and full-length protein. An exemplary PAGE-analysis from one out of three independent experiments is shown. The lower band corresponds to free DNA probe, the upper band to protein-bound oligonucleotides. While full-length Mfd could bind to the probe, the constructs showed no affinity to DNA.

In nucleotide excision repair, UvrB uses the β -hairpin to bind to damaged DNA (Skorvaga *et al.*, 2002; Truglio *et al.*, 2006b). It is essential for the discrimination between damaged and non-damaged DNA (Moolenaar *et al.*, 2001). In Mfd, DNA damage recognition occurs by recognition of stalled RNA polymerase (Selby and Sancar, 1993). This is mediated by the RNA polymerase interacting domain which is located C-terminal of the UvrA-binding region (Selby and Sancar, 1995a). Binding to and translocation along double stranded DNA is mediated by the C-terminal translocase module.

3.3.5.2 The Mfd N-terminus contains a degenerated ATPase motif

The fold of Mfd domain 1A is that of RecA-like domains. However, it lacks the functional motifs of active RecA-like ATPases (table 3.6): Walker A motif (identical to helicase motif I) residues are involved in nucleotide binding and positioning of the γ -phosphate.

Residues of the Walker B motif (DExx, helicase motif II) coordinate the magnesium ion and polarize the water molecule for nucleophilic attack (Walker et al., 1982).

In UvrB domain 1A, the functional ATPase motifs are present. UvrB is an ATP-dependent helicase. By itself, it only possesses cryptic ATPase activity which is activated in the presence of UvrA and damaged DNA. Full-length UvrB possesses a C-terminal regulatory domain 4 which inhibits DNA binding and ATPase activity in the absence of UvrA (Caron and Grossman, 1988; Wang *et al.*, 2006). C-terminally truncated UvrB lacking this autoinhibitory domain shows increased binding to DNA and enhanced ATPase activity (Wang *et al.*, 2006).

Table 3.6: Alignment of Walker A and Walker B motifs from UvrB proteins and the corresponding regions in Mfd. "o" stands for hydrophobic residues, residues coloured in green are in accordance with the canonical sequences.

Walker A		Walker B	
consensus	G G GK(S/T)	consensus	oooDESH
EcoUvrB	QTLLGVTGS GKT	EcoUvrB	LLVVDESHV
BsuUvrB	QTLLGATGT GKT	BsuUvrB	MLVVDESHV
HinUvrB	QTLLGVTGS GKT	HinUvrB	LLIIDEHSV
MtuUvrB	VVLLGATGT GKS	MtuUvrB	LLVIDESHV
BcaUvrB	QTLLGATGT GKT	BcaUvrB	LLIVDESHV
EcoMfd	RLLGELTGAACA	EcoMfd	LLVNTG-DL
BsuMfd	QLLAGLSG SARS	BsuMfd	LLILDEVSR
HinMfd	KILGNVLPGADA	HinMfd	LFVDME--N
MtuMfd	DETLIAPASAR	MtuMfd	FVLVCDPEK

In *B.caldotenax* UvrB, residues Thr41, Gly42, Thr43 and Lys45 of helicase motif I are involved in phosphate binding (Theis *et al.*, 1999). In Mfd-N2, the corresponding region was not ordered in the crystal structure. Sequence alignment (table 3.6) indicates that the canonical Walker A residues are not present in Mfd.

Side chains of the conserved amino acids Glu338 and Asp339 of helicase motif II point toward the Mg²⁺-ion in UvrB (Theis et al., 1999). The corresponding region in Mfd lacks these acidic residues (figure 3.24). In addition, the typical three-dimensional conformation of active ATPases cannot be found in Mfd.

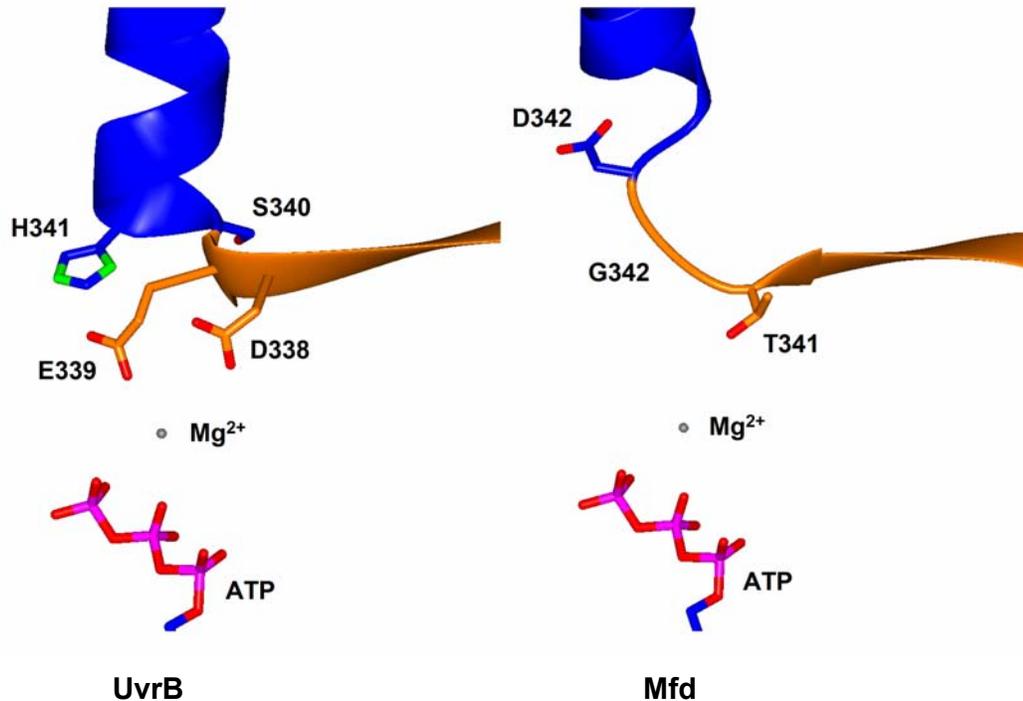


Figure 3.24: Detailed view of the Walker B (DESH) motif of ATP-bound UvrB (PDB-ID 1D9Z; Theis et al., 1999) and the corresponding region of Mfd (domain 1A, orange; domain 1B, blue; non-carbon atoms are coloured according to the atom type: magnesium, grey; nitrogen, green; oxygen, red; phosphorus, magenta).

Interestingly, the "Walker A" and "Walker B" motifs of *Bacillus subtilis* Mfd show higher similarity to the canonical sequences (table 3.6). Therefore the *B.subtilis* Mfd N-terminus might be able to bind and/or hydrolyze ATP.

ATPase activity of different N-terminal constructs was tested by thin layer chromatography. As controls, Mfd-FL (see 3.1) and UvrB-N were used.

UvrB-N comprises the first 583 residues of *E.coli* UvrB. It lacks the regulatory C-terminal domain 4. The gene encoding UvrB residues 1-583 was amplified by PCR from genomic *E.coli* XL1 Blue DNA and cloned into the pET-21b vector (table 2.3). UvrB-N was expressed with a C-terminal hexahistidine-tag and purified analogously to Mfd-N2.

ATPase activity was determined in the absence (light blue) and in the presence (violet) of dsDNA (figure 3.25).

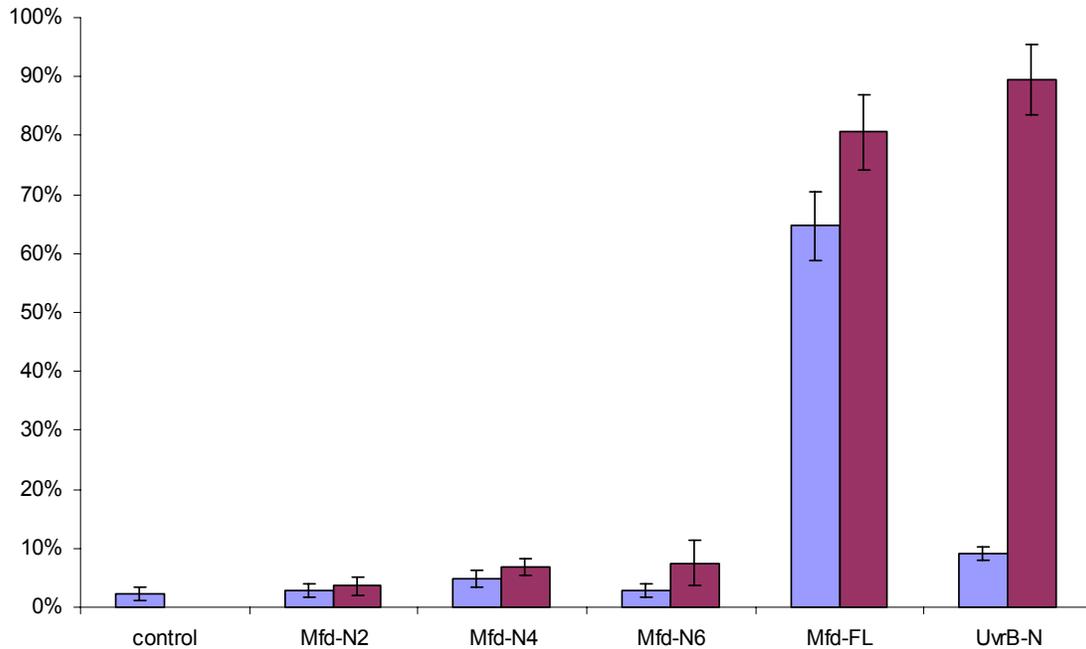


Figure 3.25: ATPase activity assay of Mfd-N2 (1-333), Mfd-N4 (1-433), Mfd-N6 (1-586), Mfd-FL and UvrB-N (1-583) without (light blue) and with (violet) dsDNA. Shown are mean values and standard deviation of three independent experiments.

It could be shown that the Mfd N-terminus does not possess ATPase activity. Full-length Mfd is able to hydrolyze ATP by use of its C-terminal dsDNA translocase domain (Selby and Sancar, 1995b) (see 1.1.2.3). Constructs Mfd-N2, Mfd-N4 and Mfd-N6 lacking the C-terminus are not able to hydrolyze ATP. UvrB-N contains functional Walker A and Walker B motifs and possesses DNA-stimulated ATPase activity (Caron and Grossman, 1988; Theis *et al.*, 1999).

ATP binding and hydrolysis by UvrB are required in two steps of nucleotide excision repair: UvrA-mediated loading of UvrB onto damaged DNA does not only require ATP hydrolysis by UvrA but also by UvrB (Van Houten *et al.*, 1988; Moolenaar *et al.*, 2000). UvrB probably destabilizes the DNA duplex in an energy-consuming reaction in order to insert the β -hairpin between the strands (Goosen and Moolenaar, 2001; Truglio *et al.*, 2006b). UvrC-mediated incision into the damaged DNA requires binding of a new ATP molecule by UvrB (Orren and Sancar, 1990; Moolenaar *et al.*, 2000; Truglio *et al.*, 2005).

This induces a conformational change in the DNA, resulting in higher sensitivity to nucleases (Moolenaar *et al.*, 2000).

In contrast, Mfd acts on sites of transcription. Here, the DNA double helix has already been opened (Selby and Sancar, 1994). In the incision reactions, Mfd is not involved. Therefore, functional ATPase motifs are not essential in this region of Mfd.

Full-length Mfd does possess ATPase activity. Mfd translocates along double stranded DNA in an ATP-dependent manner. ATPase activity is also used for the dissociation of stalled RNA polymerase. ATP hydrolysis is carried out by the C-terminal translocase domain (Selby and Sancar, 1995a; Selby and Sancar, 1995b).

4 Discussion

Cells frequently encounter DNA damage caused by variable exo- or endogenous sources. DNA lesions are a severe threat to genomic integrity and can lead to mutations, cancer, or cell death. Therefore, different DNA repair mechanisms exist in order to deal with all different types of lesions (reviewed in Lindahl and Wood, 1999; Lodish et al., 2000; Hoeijmakers, 2001; Friedberg et al., 2006)

Bulky DNA lesions that affect structure of the DNA double helix are repaired by a mechanism called nucleotide excision repair (NER). NER can be found in all kingdoms of life and is functionally conserved throughout evolution (Sancar, 1996; Ogrunc *et al.*, 1998; Batty and Wood, 2000). In bacteria, nucleotide excision repair is performed by the UvrABC system (Van Houten *et al.*, 2005; Truglio *et al.*, 2006a).

DNA damage in active genes is repaired by a special mode of NER, called transcription-coupled repair (TCR). Non-coding lesions in the transcribed strand cause transcription elongation complexes to arrest. A "transcription-repair coupling factor" functions in the release of arrested RNA polymerase and delivers the NER machinery to the lesion-site.

Aim of this work was to gain structural insights into the mechanism of bacterial transcription-coupled repair. During this PhD thesis, the crystal structure of the N-terminal 333 residues of the *Escherichia coli* transcription-repair coupling factor, the Mfd protein, was solved. The Mfd N-terminus binds to the nucleotide excision repair protein UvrA and is involved in the recruitment of the UvrABC repair system to DNA lesions at stalled transcription sites. The interaction between Mfd and UvrA was further analyzed biochemically in order to reveal mechanistic details of this process.

4.1 The Mfd N-terminus resembles UvrB

The Mfd N-terminus has a triangular structure consisting of three domains (domains 1A, 1B and 2). The structure of the Mfd N-terminus very much resembles the architecture of the three N-terminal domains of UvrB (also denoted domain 1A, 1B and 2, respectively). However, it lacks functional elements that are implicated in ATP-driven damage recognition of UvrB (Theis *et al.*, 1999; Moolenaar *et al.*, 2000; Truglio *et al.*, 2006b).

Both in Mfd and UvrB, the N-terminus is implicated in UvrA-binding. In Mfd, this is the only function associated with this region (Hsu *et al.*, 1995; Selby and Sancar, 1995a; Selby and Sancar, 1995b). Both in Mfd and in UvrB, the interaction to UvrA is mediated by domain 2 (Hsu *et al.*, 1995; Truglio *et al.*, 2004, and this work). Domain 2 of Mfd possesses high sequence homology to UvrB as well as close structural similarity (Selby and Sancar, 1993; Truglio *et al.*, 2004, and this work). In both proteins, strongly conserved residues are located at the surface of domain 2. Several of these residues could be shown to be essential for UvrA-binding both in UvrB and in Mfd (Truglio *et al.*, 2004, and this work). Therefore, binding to UvrA seems not only to be a conserved function between UvrB and Mfd. In addition, both proteins seem to use a similar mechanism.

In UvrB, two additional activities reside within the N-terminus. The UvrB protein is a weak helicase and possesses cryptic ATPase activity. Functional ATPase motifs are located in domain 1A (Caron and Grossman, 1988; Hsu *et al.*, 1995; Theis *et al.*, 1999). In contrast, the corresponding region of Mfd does not adopt the conformation of active ATPases. Only a degenerated ATPase motif can be found in domain 1A. We could confirm that the Mfd N-terminus is not able to hydrolyze ATP (Selby and Sancar, 1995a; Selby and Sancar, 1995b).

Domain 1B of UvrB contains a conserved β -hairpin motif which functions in DNA binding (Theis *et al.*, 1999; Skorvaga *et al.*, 2002; Truglio *et al.*, 2006b). In Mfd, the corresponding region is more compact and lacks the DNA-binding motif. Instead, a short, non-functional loop can be found which may correspond to a degenerated hairpin. No DNA binding can be found for this part of Mfd.

However, the fold and overall arrangement of domains 1A and 1B is conserved in Mfd. In UvrB as well as in Mfd, the three N-terminal domains form a compact module with a central domain 1A. Domains 1B and 2 are inserted into the primary sequence of domain 1A (Theis *et al.*, 1999; Truglio *et al.*, 2006b, and this work). Hence, domains 1A and 1B may play an important role in forming the architecture of the N-terminal module.

UvrB possesses a second RecA-like domain, domain 3, which is located beside the compact module of domains 1A, 1B and 2 (Theis *et al.*, 1999). Domain 3, together with domain 1A, forms the "helicase" motor of UvrB (Hsu *et al.*, 1995). Domain 3 as well as a regulatory domain 4 (Sohi *et al.*, 2000; Wang *et al.*, 2006) are not present in Mfd

(Deaconescu *et al.*, 2006). These domains do not contribute to the compact structure of domains 1A, 1B and 2. In addition, their primary sequence is found C-terminal of domains 1A/1B/2.

Taken together, these findings suggest that the N-terminal part of Mfd might have evolved from an UvrB molecule or from a common precursor: Mfd and UvrB have a common function by binding to UvrA. In both proteins, this function is mediated by the highly conserved domain 2. No biochemical function is associated with domains 1A and 1B in Mfd. They seem to be involved in maintaining the architecture of the compact module. In UvrB, additional functional motifs are present. Nucleic acid binding and ATP hydrolysis are mediated by the N-terminal region. There is evidence that these activities might have been present in an Mfd precursor, but – as they are not required for the function of Mfd – have been degenerated.

For most other functional domains of Mfd, "counterparts" can be found as well:

The RNA polymerase interacting domain resembles the KOW domain of the bacterial transcription factor NusG which is believed to function in RNA polymerase binding as well (Li *et al.*, 1992; Steiner *et al.*, 2002; Deaconescu *et al.*, 2006).

The RecG homology module located in the C-terminal half of Mfd, possesses motifs related to superfamily II helicases (Selby and Sancar, 1993; Selby and Sancar, 1995b). It possesses high sequence homology to RecG (Chambers *et al.*, 2003; Mahdi *et al.*, 2003), and both proteins share structural similarity (Singleton *et al.*, 2001; Deaconescu *et al.*, 2006). In both proteins, this region was found to function as dsDNA translocase (Singleton *et al.*, 2001; Park *et al.*, 2002).

Swi2/Snf2 ATPases translocate on double stranded DNA by travelling along both minor groove backbone strands (Durr *et al.*, 2005). Structural resemblance of the Mfd translocase domain to the *Sulfolobus solfataricus* Swi2/Snf2 ATPase suggests a similar translocation mechanism for Mfd (Deaconescu *et al.*, 2006).

The domain which is unique to the Mfd protein is the very C-terminal TRCF domain which protects the UvrA-binding surface. This domain revealed a novel protein fold with only very weak similarity to the Rpb1 subunit of eukaryotic RNA polymerase II (Selby and Sancar, 1993; Cramer *et al.*, 2001; Deaconescu *et al.*, 2006).

The crystal structure of the Mfd N-terminus provides a first structural insight into bacterial TCR. The remarkable similarity between Mfd and UvrB indicates an evolutionary connection between global genome and transcription-coupled nucleotide excision repair. In addition, it suggests a mechanism by which Mfd might form an UvrA recruitment factor at stalled transcription complexes.

4.2 The role of Mfd in recruitment of the UvrA-UvrB complex

In bacterial nucleotide excision repair, DNA damage recognition is performed by the UvrA₂-UvrB_{1/2} complex (Orren and Sancar, 1989; Mazur and Grossman, 1991). Upon damage verification, the UvrB-DNA pre-incision complex (PIC) is formed by the action of UvrA: UvrA loads UvrB onto the damaged DNA in an energy-consuming reaction. This step is the rate-limiting reaction in the nucleotide excision repair process (figure 4.1, left panel) (Van Houten *et al.*, 1988; Orren and Sancar, 1990; Myles *et al.*, 1991).

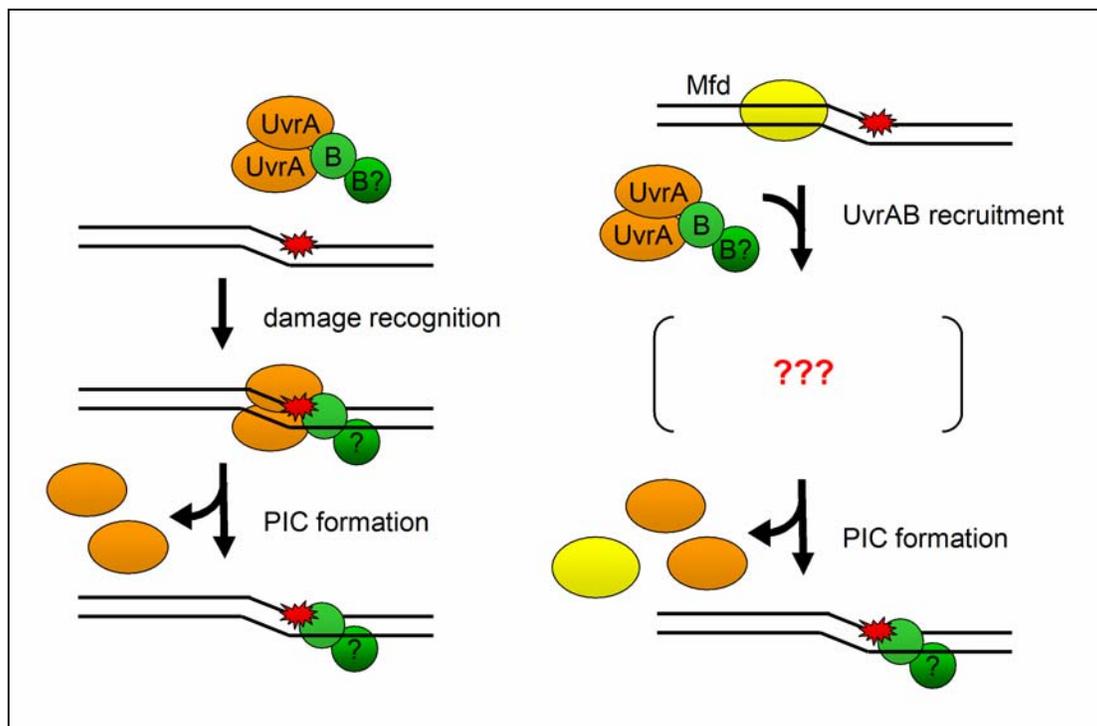


Figure 4.1: Recruitment of the UvrAB complex and formation of the pre-incision complex (PIC) in global genome (left panel) and in transcription-coupled nucleotide excision repair (right panel) (DNA lesion, red star; Mfd, yellow; UvrA, orange; UvrB, green).

In transcription-coupled repair, DNA damage is recognized by elongating RNA polymerase. As consequence, RNA polymerase becomes arrested (Tornaletti and Hanawalt, 1999). Arrested RNA polymerase is recognized and subsequently released by the transcription-repair coupling factor, Mfd (Selby and Sancar, 1995b; Park *et al.*, 2002). Mfd itself remains bound at the lesion site and recruits the nucleotide excision repair system to the lesion site (Selby *et al.*, 1991; Selby and Sancar, 1993; Roberts and Park, 2004). However, the mechanistic details of UvrAB recruitment and the subsequent formation of the UvrB-DNA pre-incision complex at stalled transcription sites are not fully understood (figure 4.1, right panel).

So, what could be the function of Mfd in these steps?

The role of Mfd seems to involve binding to UvrA. Mfd can interact directly with UvrA (Selby and Sancar, 1995a). There is high evidence that Mfd and UvrB use a similar mode of binding to UvrA: High structural similarity of domain 2 can be found, and in both proteins, highly conserved residues are essential for UvrA binding (Truglio *et al.*, 2004, and this work). In addition, UvrB can be displaced from UvrA by Mfd *in vitro* (Selby and Sancar, 1993). These findings suggest that Mfd and UvrB compete for binding to UvrA.

In vitro, Mfd can bind directly to the UvrA dimer (Mazur and Grossman, 1991). However, the existence of a free UvrA₂-Mfd complex *in vivo* is rather unlikely. In full-length Mfd, the UvrA binding region is buried in the interface between domain 2 and the TCRF domain (denoted domain 7), and therefore seems to be protected (Deaconescu *et al.*, 2006). A C-terminally truncated version of Mfd lacking domain 7 causes defects not only in transcription-coupled repair but also in global nucleotide excision repair by interacting with UvrA in an unproductive manner (Selby and Sancar, 1995a). Hence, UvrA and Mfd are not likely to interact independently of transcriptional arrest. Deaconescu and colleagues suggest that, upon binding to RNA polymerase, Mfd undergoes conformational changes by which the UvrA binding domain becomes surface-exposed and thus accessible (Deaconescu *et al.*, 2006). Therefore, in contrast to UvrB, Mfd seems to interact with UvrA only at sites of arrested transcription.

After binding to RNA polymerase and its dissociation, Mfd remains bound at the DNA in close proximity to the damage site (Selby *et al.*, 1991; Selby and Sancar, 1993; Roberts

and Park, 2004). Thus, the exposed UvrA-binding region of Mfd may act as recruitment signal for UvrA and attract the UvrA-UvrB complex.

The subsequent formation of the UvrB-DNA pre-incision complex probably involves Mfd as well (see below).

On the other hand, it is also possible that Mfd only attracts the UvrAB complex to the lesion site and does not take part in the subsequent reaction ("attraction model", figure 4.2).

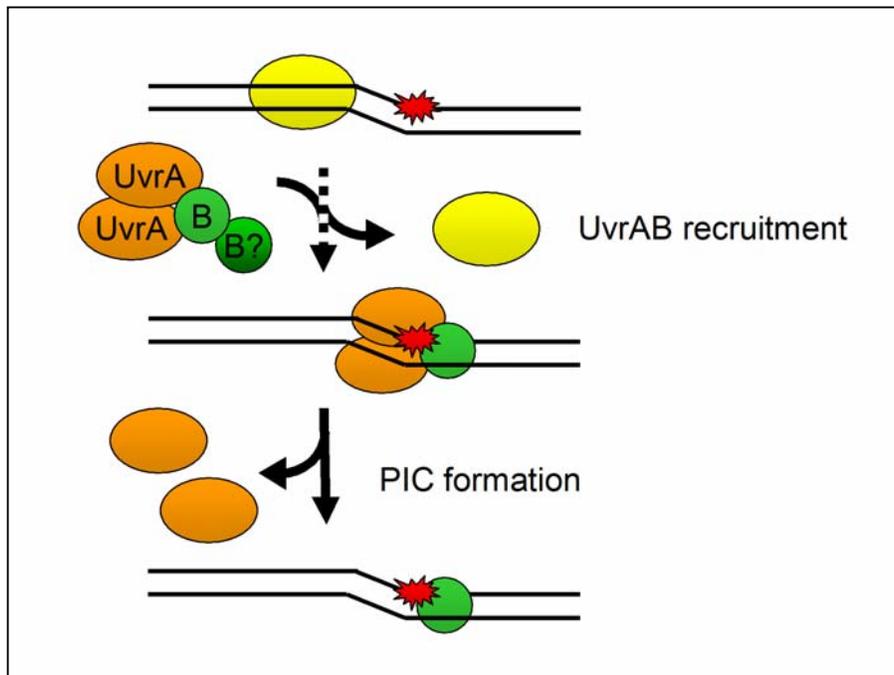


Figure 4.2: *Attraction model: Mfd attracts the UvrAB complex close to the damage site. The formation of the pre-incision complex occurs independently of Mfd (colour code as in figure 4.1).*

Damage repair in active genes occurs at a much higher rate than in the overall genome (Mellon *et al.*, 1986; Mellon *et al.*, 1987; Mellon and Hanawalt, 1989). The rate limiting step in global nucleotide excision repair is the UvrA-dependent loading of UvrB onto the damaged DNA (Orren and Sancar, 1990). Therefore, this step is probably facilitated in transcription-coupled repair (Selby and Sancar, 1995b).

One difference between global and transcription-coupled nucleotide excision repair concerning this reaction is quite evident: During formation of the pre-incision complex, UvrB inserts a β -hairpin motif between the two strands of the DNA double helix (Truglio *et al.*, 2006b). In global NER, the UvrAB complex acts on double stranded DNA, and the

DNA duplex has to be destabilized beforehand. This requires ATP-hydrolysis by UvrB. Before UvrC can incise the damaged DNA, a new ATP molecule has to be bound by UvrB resulting in conformational changes in the DNA molecule (Moolenaar *et al.*, 2000; Goosen and Moolenaar, 2001). In contrast, TCR takes place at transcription bubbles where the DNA double helix has already been opened (Selby and Sancar, 1994). UvrB can insert the hairpin without hydrolysis and rebinding of ATP. Mfd possesses affinity to transcription-bubble like DNA structures *in vitro* (Selby and Sancar, 1995a). It therefore is imaginable that Mfd forms a "placeholder", until the UvrA-UvrB complex approaches the damage site. This feature probably has an effect on the repair rate. However, by itself, it is presumably insufficient to explain the dramatical discrepancies in the repair rates between global and transcription-coupled repair.

Mfd might therefore also be actively involved in the formation of the UvrB-DNA complex. Mfd is able to displace UvrB from UvrA *in vitro* (Selby and Sancar, 1993). Hence, Mfd could also help to release UvrA from UvrB *in vivo*.

Recognition of lesion-bound Mfd may promote the dissociation of the UvrAB complex and, in consequence, enhances formation of the UvrB-DNA pre-incision complex. UvrB-loading is an energy-consuming reaction (Myles *et al.*, 1991). Thus, it would be interesting to investigate, if ATPase-activity of UvrA can be stimulated by the presence of Mfd.

In this scenario, Mfd directly interacts with UvrA which is still involved in the UvrA₂-UvrB complex. As mentioned above, Mfd and UvrB seem to interact with UvrA in a highly similar way (see above). It is therefore quite likely that Mfd and UvrB compete for the same binding site on UvrA ("competitive model"). On the other hand, it is also imaginable that Mfd induces conformational changes in UvrA leading to UvrB-release. This would require a second, allosteric binding site on UvrA.

Although much is known about the UvrA-binding domains of UvrB and Mfd, the UvrB-/Mfd-interacting region of UvrA is only poorly understood. The UvrB-binding site of UvrA could be narrowed down to the first 230 amino acids (Claassen and Grossman, 1991). For binding of UvrB or Mfd, dimerization of UvrA is essential (Claassen and Grossman, 1991; Myles *et al.*, 1991). This is achieved at high concentrations of UvrA or in the presence of ATP (Myles *et al.*, 1991). At physiological concentrations, most UvrA is present as a dimer in the UvrA₂-UvrB_{1/2} complex (Orren and Sancar, 1989).

UvrA contains two ABC ATPase domains (Doolittle *et al.*, 1986). The N-terminal ATPase motifs are involved in dimer formation. They are located within the UvrB-binding region (Myles *et al.*, 1991). It is therefore imaginable that two equivalent binding sites for UvrB/Mfd are present in one UvrA dimer.

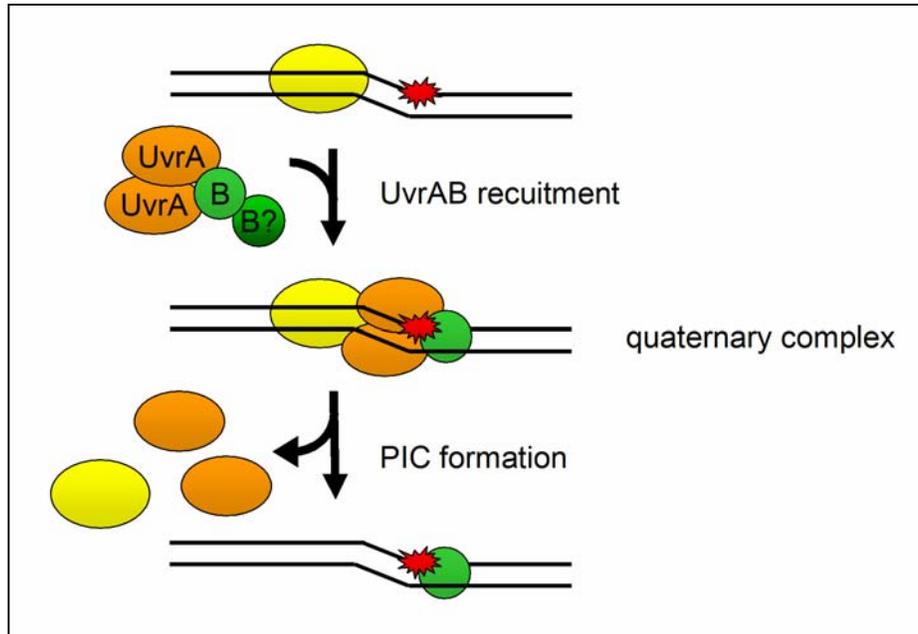


Figure 4.3: Allosteric model: According to this model, Mfd directly takes part in the release of UvrB from UvrA and in the formation of the PIC. This reaction may involve a transient quaternary complex.

Properties of the UvrA-UvrB complex have been intensively studied. There is no evidence that two UvrB molecules can bind to different sites on UvrA simultaneously. Due to intrinsic asymmetry, only one binding site may therefore be occupied at one time. In this case, binding at one site by Mfd may induce conformational changes in UvrA leading to the release of UvrB at the other binding site. This displacement-mechanism certainly would accelerate the process of pre-incision complex formation. It would require the presence of a short-living quaternary complex containing both UvrB and Mfd ("allosteric model", figure 4.3). In order to confirm the existence of such a complex, fluorescence methods, like fluorescence correlation spectroscopy (FCS), could be applied.

According to the latter models, Mfd is directly involved in the dissociation of the UvrA-UvrB complex and hence in the formation of the UvrB-DNA pre-incision complex.

5 Summary

The Mfd (mutation frequency decline) protein is responsible for connecting the cellular processes of transcription and DNA repair in bacteria. Mfd, also termed transcription-repair coupling factor (TRCF), recognizes arrested transcription elongation complexes and catalyzes their dissociation from damaged template DNA in an ATP-dependent manner. Subsequently, Mfd recruits the UvrABC nucleotide excision repair machinery to the damage site. The mechanistic details of this process are not fully understood.

X-ray crystallography was used in order to give structural insights into the mechanism of bacterial transcription-coupled repair. During this PhD thesis, the crystal structure of the N-terminus (residues 1-333) of *Escherichia coli* Mfd ("Mfd-N2") was solved. The Mfd N-terminus is implicated to function in UvrA-binding. It bears a region with high homology to the nucleotide excision repair protein UvrB.

Mfd-N2 is a triangularly shaped molecule of approximately 60×60×30 Å dimensions which contains three structural domains (domains 1A, 1B and 2). Interestingly, the structure of Mfd-N2 very much resembles that of the three N-terminal domains of UvrB.

Mfd domain 1A adopts a typical RecA fold. However, it lacks the functional motifs of active ATPases, and we could confirm that the Mfd N-terminus does not possess any ATPase activity. Domain 1B matches the damage-binding domain of the UvrB. Interestingly, Mfd is bereft of the damage-binding motif of UvrB domain 1B, and no DNA binding is associated with this part of Mfd. Domain 2, which possesses the highest sequence homology to UvrB, closely matches the three-dimensional structure of the implicated UvrA-binding domain of UvrB.

Highly conserved amino acids between Mfd and UvrB can be found on the surface of domain 2. Using site-directed mutagenesis, several of these residues could be shown to function in the UvrA-Mfd interaction. Remarkably, the corresponding residues in UvrB are required for productive interaction between UvrA and UvrB as well.

Taken together, these results suggest that Mfd and UvrB interact with UvrA in a similar manner. Mfd may form an UvrA-recruitment factor at stalled transcription complexes that resembles UvrB architecturally but not catalytically. The molecular similarity between Mfd and UvrB indicates an evolutionary connection between global genome and transcription-coupled nucleotide excision repair in bacteria.

6 References

- Abrahams, J.P. and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1): 30-42.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). *Molecular Biology of the Cell*. New York, Garland Science.
- Andressoo, J.O. and Hoeijmakers, J.H. (2005). Transcription-coupled repair and premature ageing. *Mutat Res* **577**(1-2): 179-194.
- Artsimovitch, I. and Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**(13): 7090-7095.
- Batty, D.P. and Wood, R.D. (2000). Damage recognition in nucleotide excision repair of DNA. *Gene* **241**(2): 193-204.
- Blow, D. (2002). *Outline of Crystallography for Biologists*. New York, Oxford University Press.
- Borukhov, S., Lee, J. and Laptenko, O. (2005). Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* **55**(5): 1315-1324.
- Borukhov, S., Sagitov, V. and Goldfarb, A. (1993). Transcript cleavage factors from *E. coli*. *Cell* **72**(3): 459-466.
- Brueckner, F., Hennecke, U., Carell, T. and Cramer, P. (2006). CPD damage recognition by transcribing RNA polymerase II. *Submitted. Cited with permission of F.B.*
- 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 Crystallogr D* **54**(5): 905-921.
- Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S. and Jones, D.T. (2005). Protein structure prediction servers at University College London. *Nucleic Acids Res* **33**(Web Server issue): W36-W38.
- Caron, P.R. and Grossman, L. (1988). Involvement of a cryptic ATPase activity of UvrB and its proteolysis product, UvrB* in DNA repair. *Nucleic Acids Res* **16**(20): 9651-9662.
- Caron, P.R., Kushner, S.R. and Grossman, L. (1985). Involvement of helicase II (uvrD gene product) and DNA polymerase I in excision mediated by the uvrABC protein complex. *Proc Natl Acad Sci U S A* **82**(15): 4925-4929.
- Chambers, A.L., Smith, A.J. and Savery, N.J. (2003). A DNA translocation motif in the bacterial transcription--repair coupling factor, Mfd. *Nucleic Acids Res* **31**(22): 6409-6418.
- Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H. and Xu, M.Q. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**(2): 271-281.
- Chong, S., Montello, G.E., Zhang, A., Cantor, E.J., Liao, W., Xu, M.Q. and Benner, J. (1998). Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res* **26**(22): 5109-5115.

- Claassen, L.A. and Grossman, L. (1991). Deletion mutagenesis of the Escherichia coli UvrA protein localizes domains for DNA binding, damage recognition, and protein-protein interactions. *J Biol Chem* **266**(17): 11388-94.
- Cleaver, J.E. and Kraemer, K.H. (1995). Xeroderma pigmentosum and Cockayne syndrome. In: Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (editors). *The Metabolic Basis of Inherited Disease* (seventh ed.). New York, McGraw-Hill. 4393-4419.
- Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M. and Egly, J.M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell* **21**(2): 215-26.
- Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**(5523): 1863-1876.
- Crick, F. (1974). The double helix: a personal view. *Nature* **248**(5451): 766-769.
- de Boer, J. and Hoeijmakers, J.H. (2000). Nucleotide excision repair and human syndromes. *Carcinogenesis* **21**(3): 453-360.
- de la Fortelle, E. and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* **276**: 472-494
- Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J. and Darst, S.A. (2006). Structural basis for bacterial transcription-coupled DNA repair. *Cell* **124**(3): 507-520.
- Diederichs, K. and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**(4): 269-275.
- Doolittle, R.F., Johnson, M.S., Husain, I., Van Houten, B., Thomas, D.C. and Sancar, A. (1986). Domainal evolution of a prokaryotic DNA repair protein and its relationship to active-transport proteins. *Nature* **323**(6087): 451-3.
- Drenth, J. (1999). Principles of Protein X-Ray Crystallography. Heidelberg, Springer-Verlag.
- Durr, H., Korner, C., Muller, M., Hickmann, V. and Hopfner, K.P. (2005). X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**(3): 363-73.
- Edman, P. (1950). Method for determination of the amino acid sequence in peptides. *Acta Chem Scand* **4**: 283-293.
- Egly, J.M. (2001). The 14th Datta Lecture. TFIIH: from transcription to clinic. *FEBS Lett* **498**(2-3): 124-8.
- Fish, R.N. and Kane, C.M. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* **1577**(2): 287-307.
- Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., Picotti, P. and Zambonin, M. (2004). Probing protein structure by limited proteolysis. *Acta Biochim Pol* **51**(2): 299-321.
- Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M. and Zambonin, M. (1986). Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **25**(8): 1847-1851.
- Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A. and Ellenberger, T. (2006). DNA repair and mutagenesis. Washington D.C., ASM Press.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**(13): 3784-3788.
- Goosen, N. and Moolenaar, G.F. (2001). Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Res Microbiol* **152**(3-4): 401-409.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17**(12): 4713-4730.

- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**(4): 557-580.
- Hildebrand, E.L. and Grossman, L. (1999). Oligomerization of the UvrB nucleotide excision repair protein of *Escherichia coli*. *J Biol Chem* **274**(39): 27885-90.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1): 51-59.
- Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**(6835): 366-374.
- Holm, L. and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20**(11): 478-80.
- Hsu, D.S., Kim, S.T., Sun, Q. and Sancar, A. (1995). Structure and function of the UvrB protein. *J Biol Chem* **270**(14): 8319-8327.
- Husain, I., Van Houten, B., Thomas, D.C., Abdel-Monem, M. and Sancar, A. (1985). Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proc Natl Acad Sci U S A* **82**(20): 6774-6778.
- International Tables for Crystallography, Volume A: Space-group symmetry. (2002).
- Jamieson, E.R. and Lippard, S.J. (1999). Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem Rev* **99**(9): 2467-98.
- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**(2): 195-202.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**(6): 795-800.
- Kantardjieff, K.A. and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**(9): 1865-1871.
- Kelman, Z. and White, M.F. (2005). Archaeal DNA replication and repair. *Curr Opin Microbiol* **8**(6): 669-76.
- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**(4610): 662-666.
- Kettenberger, H., Armache, K.J. and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**(3): 347-357.
- Komissarova, N. and Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A* **94**(5): 1755-1760.
- Laine, J.P. and Egly, J.M. (2006). When transcription and repair meet: a complex system. *Trends Genet.*
- 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 Cryst* **26**(2): 283-291.
- Le Page, F., Kwoh, E.E., Avrutskaya, A., Gentil, A., Leadon, S.A., Sarasin, A. and Cooper, P.K. (2000). Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* **101**(2): 159-171.
- LeMaster, D.M. and Richards, F.M. (1985). ¹H-¹⁵N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**(25): 7263-7268.
- Li, J., Horwitz, R., McCracken, S. and Greenblatt, J. (1992). NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. *J Biol Chem* **267**(9): 6012-9.

- Lin, J.J. and Sancar, A. (1992). Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. *J Biol Chem* **267**(25): 17688-17692.
- Lindahl, T. and Wood, R.D. (1999). Quality control by DNA repair. *Science* **286**(5446): 1897-905.
- Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. and Darnell, J. (2000). *Molecular Cell Biology*. N.Y., W.H. Freeman and Company.
- Machius, M., Henry, L., Palnitkar, M. and Deisenhofer, J. (1999). Crystal structure of the DNA nucleotide excision repair enzyme UvrB from *Thermus thermophilus*. *Proc Natl Acad Sci U S A* **96**(21): 11717-11722.
- Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q. and Lloyd, R.G. (2003). A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. *Embo J* **22**(3): 724-734.
- Massa, W. (2002). *Kristallstrukturbestimmung*. Stuttgart, B.G. Teubner.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J Mol Biol* **33**(2): 491-497.
- Mazur, S.J. and Grossman, L. (1991). Dimerization of *Escherichia coli* UvrA and its binding to undamaged and ultraviolet light damaged DNA. *Biochemistry* **30**(18): 4432-4443.
- McGuffin, L.J., Bryson, K. and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**(4): 404-405.
- McPherson, A. (2001). *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press.
- Mellon, I. (2005). Transcription-coupled repair: a complex affair. *Mutat Res* **577**(1-2): 155-161.
- Mellon, I., Bohr, V.A., Smith, C.A. and Hanawalt, P.C. (1986). Preferential DNA repair of an active gene in human cells. *Proc Natl Acad Sci U S A* **83**(23): 8878-8882.
- Mellon, I. and Hanawalt, P.C. (1989). Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342**(6245): 95-98.
- Mellon, I., Spivak, G. and Hanawalt, P.C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51**(2): 241-249.
- Miller, J.H. (1972). *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Moolenaar, G.F., Bazuine, M., van Knippenberg, I.C., Visse, R. and Goosen, N. (1998). Characterization of the *Escherichia coli* damage-independent UvrBC endonuclease activity. *J Biol Chem* **273**(52): 34896-34903.
- Moolenaar, G.F., Franken, K.L., Dijkstra, D.M., Thomas-Oates, J.E., Visse, R., van de Putte, P. and Goosen, N. (1995). The C-terminal region of the UvrB protein of *Escherichia coli* contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. *J Biol Chem* **270**(51): 30508-30515.
- Moolenaar, G.F., Herron, M.F., Monaco, V., van der Marel, G.A., van Boom, J.H., Visse, R. and Goosen, N. (2000). The role of ATP binding and hydrolysis by UvrB during nucleotide excision repair. *J Biol Chem* **275**(11): 8044-8050.
- Moolenaar, G.F., Hoglund, L. and Goosen, N. (2001). Clue to damage recognition by UvrB: residues in the beta-hairpin structure prevent binding to non-damaged DNA. *Embo J* **20**(21): 6140-6149.
- Moolenaar, G.F., van Rossum-Fikkert, S., van Kesteren, M. and Goosen, N. (2002). Cho, a second endonuclease involved in *Escherichia coli* nucleotide excision repair. *Proc Natl Acad Sci U S A* **99**(3): 1467-1472.

- Morris, R.J., Perrakis, A. and Lamzin, V.S. (2003). ARP/wARP and automatic interpretation of protein electron density maps *Methods Enzymol* **374**: 229-244.
- Myles, G.M., Hearst, J.E. and Sancar, A. (1991). Site-specific mutagenesis of conserved residues within Walker A and B sequences of Escherichia coli UvrA protein. *Biochemistry* **30**(16): 3824-3834.
- Nakagawa, N., Sugahara, M., Masui, R., Kato, R., Fukuyama, K. and Kuramitsu, S. (1999). Crystal structure of Thermus thermophilus HB8 UvrB protein, a key enzyme of nucleotide excision repair. *J Biochem (Tokyo)* **126**(6): 986-990.
- Nance, M.A. and Berry, S.A. (1992). Cockayne syndrome: review of 140 cases. *Am J Med Genet* **42**(1): 68-84.
- Ness, S.R., de Graaff, R.A., Abrahams, J.P. and Pannu, N.S. (2004). CRANK: new methods for automated macromolecular crystal structure solution. *Structure* **12**(10): 1753-1761.
- Nicholas, K.B. and Nicholas, H.B. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignment. Distributed by the author.
- Nickels, B.E. and Hochschild, A. (2004). Regulation of RNA polymerase through the secondary channel. *Cell* **118**(3): 281-284.
- Ogrunc, M., Becker, D.F., Ragsdale, S.W. and Sancar, A. (1998). Nucleotide excision repair in the third kingdom. *J Bacteriol* **180**(21): 5796-5798.
- Orren, D.K. and Sancar, A. (1989). The (A)BC excinuclease of Escherichia coli has only the UvrB and UvrC subunits in the incision complex. *Proc Natl Acad Sci U S A* **86**(14): 5237-5241.
- Orren, D.K. and Sancar, A. (1990). Formation and enzymatic properties of the UvrB.DNA complex. *J Biol Chem* **265**(26): 15796-15803.
- Otwinowski, Z. and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode *Methods Enzymol* **276**(A): 307-326.
- Park, J.S., Marr, M.T. and Roberts, J.W. (2002). E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**(6): 757-767.
- Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. and Noble, M. (2002). The CCP4 molecular-graphics project. *Acta Crystallogr D* **58**(Pt 11): 1955-7.
- Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G.N., Cohen, S., Perrakis, A. and Noble, M. (2004). Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D* **60**(Pt 12 Pt 1): 2288-94.
- Powell, H. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D* **55**(10): 1690-1695.
- Reines, D., Ghanouni, P., Li, Q.Q. and Mote, J., Jr. (1992). The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *J Biol Chem* **267**(22): 15516-15522.
- Roberts, J. and Park, J.S. (2004). Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr Opin Microbiol* **7**(2): 120-125.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
- Sancar, A. (1996). DNA excision repair. *Annu Rev Biochem* **65**: 43-81.
- Sancar, A. and Rupp, W.D. (1983). A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. *Cell* **33**(1): 249-260.
- Sarker, A.H., Tsutakawa, S.E., Kostek, S., Ng, C., Shin, D.S., Peris, M., Campeau, E., Tainer, J.A., Nogales, E. and Cooper, P.K. (2005). Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Mol Cell* **20**(2): 187-198.

- Saxowsky, T.T. and Doetsch, P.W. (2006). RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chem Rev* **106**(2): 474-488.
- Selby, C.P. and Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. *Science* **260**(5104): 53-58.
- Selby, C.P. and Sancar, A. (1994). Mechanisms of transcription-repair coupling and mutation frequency decline. *Microbiol Rev* **58**(3): 317-329.
- Selby, C.P. and Sancar, A. (1995a). Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *J Biol Chem* **270**(9): 4882-4889.
- Selby, C.P. and Sancar, A. (1995b). Structure and function of transcription-repair coupling factor. II. Catalytic properties. *J Biol Chem* **270**(9): 4890-4895.
- Selby, C.P., Witkin, E.M. and Sancar, A. (1991). Escherichia coli mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: in vitro complementation with purified coupling factor. *Proc Natl Acad Sci U S A* **88**(24): 11574-11578.
- Siede, W., Kow, Y.W. and Doetsch, P.W. (2005). DNA Damage Recognition. New York, Taylor & Francis Group.
- Singleton, M.R., Scaife, S. and Wigley, D.B. (2001). Structural analysis of DNA replication fork reversal by RecG. *Cell* **107**(1): 79-89.
- Skorvaga, M., DellaVecchia, M.J., Croteau, D.L., Theis, K., Truglio, J.J., Mandavilli, B.S., Kisker, C. and Van Houten, B. (2004). Identification of residues within UvrB that are important for efficient DNA binding and damage processing. *J Biol Chem* **279**(49): 51574-51580.
- Skorvaga, M., Theis, K., Mandavilli, B.S., Kisker, C. and Van Houten, B. (2002). The beta-hairpin motif of UvrB is essential for DNA binding, damage processing, and UvrC-mediated incisions. *J Biol Chem* **277**(2): 1553-1559.
- Smith, A.J. and Savery, N.J. (2005). RNA polymerase mutants defective in the initiation of transcription-coupled DNA repair. *Nucleic Acids Res* **33**(2): 755-764.
- Sohi, M., Alexandrovich, A., Moolenaar, G., Visse, R., Goosen, N., Vernede, X., Fontecilla-Camps, J.C., Champness, J. and Sanderson, M.R. (2000). Crystal structure of Escherichia coli UvrB C-terminal domain, and a model for UvrB-uvrC interaction. *FEBS Lett* **465**(2-3): 161-164.
- Steiner, T., Kaiser, J.T., Marinkovic, S., Huber, R. and Wahl, M.C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *Embo J* **21**(17): 4641-53.
- Svejstrup, J.Q. (2002a). Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol* **3**(1): 21-29.
- Svejstrup, J.Q. (2002b). Transcription repair coupling factor: a very pushy enzyme. *Mol Cell* **9**(6): 1151-1152.
- Terwilliger, T.C. (2002). Automated structure solution, density modification and model building. *Acta Crystallogr D* **58**(11): 1937-1940.
- Theis, K., Chen, P.J., Skorvaga, M., Van Houten, B. and Kisker, C. (1999). Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *Embo J* **18**(24): 6899-6907.
- Theis, K., Skorvaga, M., Machius, M., Nakagawa, N., Van Houten, B. and Kisker, C. (2000). The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutat Res* **460**(3-4): 277-300.
- Tornaletti, S. and Hanawalt, P.C. (1999). Effect of DNA lesions on transcription elongation. *Biochimie* **81**(1-2): 139-146.

- Tornaletti, S., Reines, D. and Hanawalt, P.C. (1999). Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. *J Biol Chem* **274**(34): 24124-24130.
- Truglio, J.J., Croteau, D.L., Skorvaga, M., DellaVecchia, M.J., Theis, K., Mandavilli, B.S., Van Houten, B. and Kisker, C. (2004). Interactions between UvrA and UvrB: the role of UvrB's domain 2 in nucleotide excision repair. *Embo J* **23**(13): 2498-2509.
- Truglio, J.J., Croteau, D.L., Van Houten, B. and Kisker, C. (2006a). Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* **106**(2): 233-252.
- Truglio, J.J., Karakas, E., Rhau, B., Wang, H., DellaVecchia, M.J., Van Houten, B. and Kisker, C. (2006b). Structural basis for DNA recognition and processing by UvrB. *Nat Struct Mol Biol* **13**(4): 360-364.
- Truglio, J.J., Karakas, E., Rhau, B., Wang, H., Dellavecchia, M.J., Van Houten, B. and Kisker, C. (2006b). Structural basis for DNA recognition and processing by UvrB. *Nat Struct Mol Biol*.
- Truglio, J.J., Rhau, B., Croteau, D.L., Wang, L., Skorvaga, M., Karakas, E., Dellavecchia, M.J., Wang, H., Van Houten, B. and Kisker, C. (2005). Structural insights into the first incision reaction during nucleotide excision repair. *Embo J* **24**(5): 885-894.
- Turk, D. (1992). Weiterentwicklung eines Programms fuer Molekülgraphik und Elektrondichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. PhD Thesis, Technical University, Munich.
- Van Houten, B., Croteau, D.L., Dellavecchia, M.J., Wang, H. and Kisker, C. (2005). 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* **577**(1-2): 92-117.
- Van Houten, B., Gamper, H., Hearst, J.E. and Sancar, A. (1988). Analysis of sequential steps of nucleotide excision repair in *Escherichia coli* using synthetic substrates containing single psoralen adducts. *J Biol Chem* **263**(32): 16553-16560.
- Verhoeven, E.E., van Kesteren, M., Moolenaar, G.F., Visse, R. and Goosen, N. (2000). Catalytic sites for 3' and 5' incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *J Biol Chem* **275**(7): 5120-5123.
- Verhoeven, E.E., Wyman, C., Moolenaar, G.F. and Goosen, N. (2002). The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *Embo J* **21**(15): 4196-4205.
- Verhoeven, E.E., Wyman, C., Moolenaar, G.F., Hoeijmakers, J.H. and Goosen, N. (2001). Architecture of nucleotide excision repair complexes: DNA is wrapped by UvrB before and after damage recognition. *Embo J* **20**(3): 601-611.
- Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* **1**(8): 945-951.
- Wang, H., DellaVecchia, M.J., Skorvaga, M., Croteau, D.L., Erie, D.A. and Van Houten, B. (2006). UvrB domain 4, an autoinhibitory gate for regulation of DNA binding and ATPase activity. *J Biol Chem* **281**(22): 15227-15237.
- Witkin, E.M. (1966). Radiation-induced mutations and their repair. *Science* **152**(727): 1345-1353.

7 Supplementary material

7.1 Stable fragments of Mfd

Band 0: Mfd-FL (control)

MASMPEQRYRTLVPVKAGEQRLLGELTGAACATLVAEIAERHAGPVVLIAPDMQNALRLHDE
 ISQFTDQMVMNLADWETLPYDSFSPHQDI ISSRLSTLYQLPTMQRGVLI VPVNTLMQRVCP
 HSFLHGHALVMKKGQRLSRDALRTQLDSAGYRHVDQVMEHGEYATRGALLDLFPMGSELPY
 RLDFFDDEIDSLRVFDVDSQRTLEEVEAINLLPAHEFPTDKAAIELFRSQWRDTFEVKRDP
 EHIYQQVSKGTLPAGIEYWQPLFFSEPLPPLFSYFPANTLLVNTGDLETS AERFQADTLAR
 FENRGVDPMRPLLPQSLWLRVDELFSSELKNWPRVQLKTEHLPTKAANANLGFQKLPDLAV
 QAQQKAPLDALRKFLFETFDGPPVFSVESEGRREALGELLARIKIAPQRIMRLDEASDRGRY
 LMIGAAEHGFVDTVRLALICESDLLGERVARRRQDSRRTINPDTLIRNLAELHIGQPVVH
 LEHGVGRYAGMTTLEAGGITGEYLMITYANDAKLYVPVSSLHLISRYAGGAEENAPLHKL
 GDAWSRARQKAAEKVRDVAEELDIYAQRAAKEGF AFKHDREQYQLFCDSFPFETTPDQAQ
 AINAVLSDMCQPLAMDRLVCGDVGFGKTEVAMRAAFLAVDNHKQVAVLVPTTLLAQQHYDN
 FRDRFANWPVRIEMISRFSAKEQTQILAEVAEGKIDILIGTHKLLQSDVKFKDLGLLIVD
 EEHRFGVRHKERIKAMRANVDILTATPIPRTLNMAMSGMRDLSI IATPPARRLAVKTFV
 REYDSMVVREAILREILRGGQVYYLYNDVENIQKAAERLAELVPEARIAIGHGQMRERELE
 RVMNDFHHQRFNVLVCTTIIETGIDIPTANTII IERADHFGLAQLHQLRGRVGRSHHQAYA
 WLLTPHPKAMTTDAQRLEAIASLEDLGAGFALATHDLEIRGAGELLGEEQSGSMETIGFS
 LYMELLENAVDALKAGREPSLEDLTSQQTEVELRMPSSLPDDFIPDVNTRLSFYKRIASAK
 TENEELEEKVELIDRFGLLPDPARTLLDIARLRQQAQKLGIRKLEGNEKGGVIEFAEKNHV
 NPAWLIGLLQKQPQHYRLDGPTRLKFIQDLSEKTRIEWVRQFMRELEENAI AAAALEHHH
 HHH&

Band 1 (approximately 35 kDa)

MASMPEQRYRTLVPVKAGEQRLLGELTGAACATLVAEIAERHAGPVVLIAPDMQNALRLHDE
 ISQFTDQMVMNLADWETLPYDSFSPHQDI ISSRLSTLYQLPTMQRGVLI VPVNTLMQRVCP
 HSFLHGHALVMKKGQRLSRDALRTQLDSAGYRHVDQVMEHGEYATRGALLDLFPMGSELPY
 RLDFFDDEIDSLRVFDVDSQRTLEEVEAINLLPAHEFPTDKAAIELFRSQWRDTFEVKRDP
 EHIYQQVSKGTLPAGIEYWQPLFFSEPLPPLFSYFPANTLLVNTGDLETS AERFQADTLAR
 FENRGVDPMRPLLPQSLWLRVDELFSSELKNWPRVQLKTEHLPTKAANANLGFQKLPDLAV
 QAQQKAPLDALRKFLFETFDGPPVFSVESEGRREALGELLARIKIAPQRIMRLDEASDRGRY
 LMIGAAEHGFVDTVRLALICESDLLGERVARRRQDSRRTINPDTLIRNLAELHIGQPVVH
 LEHGVGRYAGMTTLEAGGITGEYLMITYANDAKLYVPVSSLHLISRYAGGAEENAPLHKL
 GDAWSRARQKAAEKVRDVAEELDIYAQRAAKEGF AFKHDREQYQLFCDSFPFETTPDQAQ
 AINAVLSDMCQPLAMDRLVCGDVGFGKTEVAMRAAFLAVDNHKQVAVLVPTTLLAQQHYDN
 FRDRFANWPVRIEMISRFSAKEQTQILAEVAEGKIDILIGTHKLLQSDVKFKDLGLLIVD
 EEHRFGVRHKERIKAMRANVDILTATPIPRTLNMAMSGMRDLSI IATPPARRLAVKTFV
 REYDSMVVREAILREILRGGQVYYLYNDVENIQKAAERLAELVPEARIAIGHGQMRERELE
 RVMNDFHHQRFNVLVCTTIIETGIDIPTANTII IERADHFGLAQLHQLRGRVGRSHHQAYA
 WLLTPHPKAMTTDAQRLEAIASLEDLGAGFALATHDLEIRGAGELLGEEQSGSMETIGFS
 LYMELLENAVDALKAGREPSLEDLTSQQTEVELRMPSSLPDDFIPDVNTRLSFYKRIASAK
 TENEELEEKVELIDRFGLLPDPARTLLDIARLRQQAQKLGIRKLEGNEKGGVIEFAEKNHV
 NPAWLIGLLQKQPQHYRLDGPTRLKFIQDLSEKTRIEWVRQFMRELEENAI AAAALEHHH
 HHH&

Band 2 (approximately 50 kDa)

MASMPEQYRYTLPVKAGEQRLLGELTGAACATLVAEIAERHAGPVVLIAPDMQNALRLHDE
ISQFTDQMVMNLADWETLPYDSFSPHQDI ISSRLSTLYQLPTMQRGVLI VPVNTLMQRVCP
HSFLHGHALVMKKGQRSLRDALRTQLDSAGYRHVDQVMEHGEYATR GALLDLFPMGSEL PY
RLDFFDDEIDSLRVFDVDSQRTLEEVEAINLLPAHEFPTDKAAIELFRSQWRDTFEVKRDP
EHIYQQVSKGTLPAGIEYWQPLFFSEPLPPLFSYFPANTLLVNTGDLETS AERFQADTLAR
FENRGVDPMRPLLPPQSLWLRVDELFSSELKNWPRVQLKTEHLPTKAANANLGFQKLPDLAV
QAQQKAPLDALRKFL E TFDGPPVFSVESEGRREALGELLARIKIAPQRIMRLDEASDRGRY
LMIGAAEHGFVDTVRNLALICESDLLGERVARRRQDSRRTINPDTLIRNLAELHIGQPVVH
LEHGVGRYAGMTTLEAGGITGEYLMMLTYANDAKLYVPVSSLHLISRYAGGAENAPLHKL G
GDAWSRARQKAAEKVRDVAEELLDIYAQRAAKEGF AFKHDREQYQLFCDSFPFETTPDQAQ
AINAVLSDMCQPLAMDRLVCGDVGF GKTEVAMRAAFLAVDNHKQVAVLVPTTLLAQQHYDN
FRDRFANWPVRIEMISRFRSAKEQTQILAEVAEGKIDILIGTHKLLQSDVKFKDLGLLIVD
EEHRFGVRHKERIKAMRANVDILT LTATPI PRTLNMAMSGMRDLSI IATPPARRLAVKTFV
REYDSMVVREAILREILRGQVYYLYNDVENIQKAAERLAELVPEARIAIGHGQMRERELE
RVMNDFHHQRFNVLVCTTIIETGIDIPTANTII IERADHFGLAQLHQLRGRVGRSHHQAYA
WLLTPHPKAMTTDAQKRLEAIASLEDLGAGFALATHDLEIRGAGELLGEEQSGSMETIGFS
LYMELLENAVDALKAGREPSLEDLTSQQTEVELRMP SLLPDDFIPDVNTRLSFYKRIASAK
TENELEEIKVELIDRFGLLPDPARTLLDIARLRQQAQKLGIRKLEGNEKGGVIEFAEKNHV
NPAWLIGLLQKQPQHYRLDGPTRLKFIQDLSERKTRIEWVRQFMRELEENAIAAAALEHHH
HHH&

Band 3 (approximately 60 kDa)

MASMPEQYRYTLPVKAGEQRLLGELTGAACATLVAEIAERHAGPVVLIAPDMQNALRLHDE
ISQFTDQMVMNLADWETLPYDSFSPHQDI ISSRLSTLYQLPTMQRGVLI VPVNTLMQRVCP
HSFLHGHALVMKKGQRSLRDALRTQLDSAGYRHVDQVMEHGEYATR GALLDLFPMGSEL PY
RLDFFDDEIDSLRVFDVDSQRTLEEVEAINLLPAHEFPTDKAAIELFRSQWRDTFEVKRDP
EHIYQQVSKGTLPAGIEYWQPLFFSEPLPPLFSYFPANTLLVNTGDLETS AERFQADTLAR
FENRGVDPMRPLLPPQSLWLRVDELFSSELKNWPRVQLKTEHLPTKAANANLGFQKLPDLAV
QAQQKAPLDALRKFL E TFDGPPVFSVESEGRREALGELLARIKIAPQRIMRLDEASDRGRY
LMIGAAEHGFVDTVRNLALICESDLLGERVARRRQDSRRTINPDTLIRNLAELHIGQPVVH
LEHGVGRYAGMTTLEAGGITGEYLMMLTYANDAKLYVPVSSLHLISRYAGGAENAPLHKL G
GDAWSRARQKAAEKVRDVAEELLDIYAQRAAKEGF AFKHDREQYQLFCDSFPFETTPDQAQ
AINAVLSDMCQPLAMDRLVCGDVGF GKTEVAMRAAFLAVDNHKQVAVLVPTTLLAQQHYDN
FRDRFANWPVRIEMISRFRSAKEQTQILAEVAEGKIDILIGTHKLLQSDVKFKDLGLLIVD
EEHRFGVRHKERIKAMRANVDILT LTATPI PRTLNMAMSGMRDLSI IATPPARRLAVKTFV
REYDSMVVREAILREILRGQVYYLYNDVENIQKAAERLAELVPEARIAIGHGQMRERELE
RVMNDFHHQRFNVLVCTTIIETGIDIPTANTII IERADHFGLAQLHQLRGRVGRSHHQAYA
WLLTPHPKAMTTDAQKRLEAIASLEDLGAGFALATHDLEIRGAGELLGEEQSGSMETIGFS
LYMELLENAVDALKAGREPSLEDLTSQQTEVELRMP SLLPDDFIPDVNTRLSFYKRIASAK
TENELEEIKVELIDRFGLLPDPARTLLDIARLRQQAQKLGIRKLEGNEKGGVIEFAEKNHV
NPAWLIGLLQKQPQHYRLDGPTRLKFIQDLSERKTRIEWVRQFMRELEENAIAAAALEHHH
HHH&

Band 4 (approximately 50 kDa)

M**A****S****M****P****E****Q****Y****R****T****L****P****V****K****A****G****E****Q****R****L****L****G****E****L****T****G****A****A****C****A****T****L****V****A****E****I****A****E****R****H****A****G****P****V****V****L****I****A****P****D****M****Q****N****A****L****R****L****H****D****E**
I**S****Q****F****T****D****Q****M****V****M****N****L****A****D****W****E****T****L****P****Y****D****S****F****S****P****H****Q****D****I****S****S****R****L****S****T****L****Y****Q****L****P****T****M****Q****R****G****V****L****I****V****P****V****N****T****L****M****Q****R****V****C****P**
H**S****F****L****H****G****H****A****L****V****M****K****K****G****Q****R****L****S****R****D****A****L****R****T****Q****L****D****S****A****G****Y****R****H****V****D****Q****V****M****E****H****G****E****Y****A****T****R****G****A****L****L****D****L****F****P****M****G****S****E****L****P****Y**
R**L****D****F****F****D****D****E****I****D****S****L****R****V****F****D****V****D****S****Q****R****T****L****E****E****V****E****A****I****N****L****L****P****A****H****E****F****P****T****D****K****A****A****I****E****L****F****R****S****Q****W****R****D****T****F****E****V****K****R****D****P**
E**H****I****Y****Q****Q****V****S****K****G****T****L****P****A****G****I****E****Y****W****Q****P****L****F****F****S****E****P****L****P****P****L****F****S****Y****F****P****A****N****T****L****L****V****N****T****G****D****L****E****T****S****A****E****R****F****Q****A****D****T****L****A****R**
F**E****N****R****G****V****D****P****M****R****P****L****L****P****P****Q****S****L****W****L****R****V****D****E****L****F****S****E****L****K****N****W****P****R****V****Q****L****K****T****E****H****L****P****T****K****A****A****N****A****N****L****G****F****Q****K****L****P****D****L****A****V**
Q**A****Q****Q****K****A****P****L****D****A****L****R****K****F****L****E****T****F****D****G****P****V****V****F****S****V****E****S****E****G****R****R****E****A****L****G****E****L****L****A****R****I****K****I****A****P****Q****R****I****M****R****L****D****E****A****S****D****R****G****R****Y**
L**M****I****G****A****A****E****H****G****F****V****D****T****V****R****N****L****A****L****I****C****E****S****D****L****L****G****E****R****V****A****R****R****R****Q****D****S****R****R****T****I****N****P****D****T****L****I****R****N****L****A****E****L****H****I****G****Q****P****V****V****H**
L**E****H****G****V****G****R****Y****A****G****M****T****T****L****E****A****G****G****I****T****G****E****Y****L****M****L****T****Y****A****N****D****A****K****L****Y****V****P****V****S****S****L****H****L****I****S****R****Y****A****G****G****A****E****N****A****P****L****H****K****L****G**
G**D****A****W****S****R****A****R****Q****K****A****A****E****K****V****R****D****V****A****A****E****L****D****I****Y****A****Q****R****A****A****K****E****G****F****A****F****K****H****D****R****E****Q****Y****Q****L****F****C****D****S****F****P****F****E****T****T****P****D****Q****A****Q**
A**I****N****A****V****L****S****D****M****C****Q****P****L****A****M****D****R****L****V****C****G****D****V****G****F****G****K****T****E****V****A****M****R****A****A****F****L****A****V****D****N****H****K****Q****V****A****V****L****V****P****T****T****L****L****A****Q****Q****H****Y****D****N**
F**R****D****R****F****A****N****W****P****V****R****I****E****M****I****S****R****F****R****S****A****K****E****Q****T****Q****I****L****A****E****V****A****E****G****K****I****D****I****L****I****G****H****K****L****L****Q****S****D****V****K****F****K****D****L****G****L****L****I****V****D**
E**E****H****R****F****G****V****R****H****K****E****R****I****K****A****M****R****A****N****V****D****I****L****T****L****T****A****T****P****I****P****R****T****L****N****M****A****M****S****G****M****R****D****L****S****I****I****A****T****P****P****A****R****R****L****A****V****K****T****F****V**
R**E****Y****D****S****M****V****R****E****A****I****L****R****E****I****L****R****G****G****Q****V****Y****L****N****D****V****E****N****I****Q****K****A****A****E****R****L****A****E****L****V****P****E****A****R****I****A****I****G****H****G****Q****M****R****E****R****E****L****E**
R**V****M****N****D****F****H****H****Q****R****F****N****V****L****V****C****T****T****I****I****E****T****G****I****D****I****P****T****A****N****T****I****I****E****R****A****D****H****F****G****L****A****Q****L****H****Q****L****R****G****R****V****G****R****S****H****H****Q****A****Y****A**
W**L****L****T****P****H****P****K****A****M****T****T****D****A****Q****K****R****L****E****A****I****A****S****L****E****D****L****G****A****G****F****A****L****A****T****H****D****L****E****I****R****G****A****G****E****L****L****G****E****E****Q****S****G****S****M****E****T****I****G****F****S**
L**Y****M****E****L****L****E****N****A****V****D****A****L****K****A****G****R****E****P****S****L****E****D****L****T****S****Q****Q****T****E****V****E****L****R****M****P****S****L****L****P****D****D****F****I****P****D****V****N****T****R****L****S****F****Y****K****R****I****A****S****A****K**
T**E****N****E****L****E****E****I****K****V****E****L****I****D****R****F****G****L****L****P****D****P****A****R****T****L****L****D****I****A****R****L****R****Q****A****Q****K****L****G****I****R****K****L****E****G****N****E****K****G****G****V****I****E****F****A****E****K****N****H****V**
N**P****A****W****L****I****G****L****L****Q****Q****P****Q****H****Y****R****L****D****G****P****T****R****L****K****F****I****Q****D****L****S****E****R****K****T****R****I****E****W****R****Q****F****M****R****E****L****E****N****A****I****A****A****A****A****L****E****H****H****H**
H**H****H****&**

Band 5 (approximately 18 kDa)

M**A****S****M****P****E****Q****Y****R****T****L****P****V****K****A****G****E****Q****R****L****L****G****E****L****T****G****A****A****C****A****T****L****V****A****E****I****A****E****R****H****A****G****P****V****V****L****I****A****P****D****M****Q****N****A****L****R****L****H****D****E**
I**S****Q****F****T****D****Q****M****V****M****N****L****A****D****W****E****T****L****P****Y****D****S****F****S****P****H****Q****D****I****S****S****R****L****S****T****L****Y****Q****L****P****T****M****Q****R****G****V****L****I****V****P****V****N****T****L****M****Q****R****V****C****P**
H**S****F****L****H****G****H****A****L****V****M****K****K****G****Q****R****L****S****R****D****A****L****R****T****Q****L****D****S****A****G****Y****R****H****V****D****Q****V****M****E****H****G****E****Y****A****T****R****G****A****L****L****D****L****F****P****M****G****S****E****L****P****Y**
R**L****D****F****F****D****D****E****I****D****S****L****R****V****F****D****V****D****S****Q****R****T****L****E****E****V****E****A****I****N****L****L****P****A****H****E****F****P****T****D****K****A****A****I****E****L****F****R****S****Q****W****R****D****T****F****E****V****K****R****D****P**
E**H****I****Y****Q****Q****V****S****K****G****T****L****P****A****G****I****E****Y****W****Q****P****L****F****F****S****E****P****L****P****P****L****F****S****Y****F****P****A****N****T****L****L****V****N****T****G****D****L****E****T****S****A****E****R****F****Q****A****D****T****L****A****R**
F**E****N****R****G****V****D****P****M****R****P****L****L****P****P****Q****S****L****W****L****R****V****D****E****L****F****S****E****L****K****N****W****P****R****V****Q****L****K****T****E****H****L****P****T****K****A****A****N****A****N****L****G****F****Q****K****L****P****D****L****A****V**
Q**A****Q****Q****K****A****P****L****D****A****L****R****K****F****L****E****T****F****D****G****P****V****V****F****S****V****E****S****E****G****R****R****E****A****L****G****E****L****L****A****R****I****K****I****A****P****Q****R****I****M****R****L****D****E****A****S****D****R****G****R****Y**
L**M****I****G****A****A****E****H****G****F****V****D****T****V****R****N****L****A****L****I****C****E****S****D****L****L****G****E****R****V****A****R****R****R****Q****D****S****R****R****T****I****N****P****D****T****L****I****R****N****L****A****E****L****H****I****G****Q****P****V****V****H**
L**E****H****G****V****G****R****Y****A****G****M****T****T****L****E****A****G****G****I****T****G****E****Y****L****M****L****T****Y****A****N****D****A****K****L****Y****V****P****V****S****S****L****H****L****I****S****R****Y****A****G****G****A****E****N****A****P****L****H****K****L****G**
G**D****A****W****S****R****A****R****Q****K****A****A****E****K****V****R****D****V****A****A****E****L****D****I****Y****A****Q****R****A****A****K****E****G****F****A****F****K****H****D****R****E****Q****Y****Q****L****F****C****D****S****F****P****F****E****T****T****P****D****Q****A****Q**
A**I****N****A****V****L****S****D****M****C****Q****P****L****A****M****D****R****L****V****C****G****D****V****G****F****G****K****T****E****V****A****M****R****A****A****F****L****A****V****D****N****H****K****Q****V****A****V****L****V****P****T****T****L****L****A****Q****Q****H****Y****D****N**
F**R****D****R****F****A****N****W****P****V****R****I****E****M****I****S****R****F****R****S****A****K****E****Q****T****Q****I****L****A****E****V****A****E****G****K****I****D****I****L****I****G****H****K****L****L****Q****S****D****V****K****F****K****D****L****G****L****L****I****V****D**
E**E****H****R****F****G****V****R****H****K****E****R****I****K****A****M****R****A****N****V****D****I****L****T****L****T****A****T****P****I****P****R****T****L****N****M****A****M****S****G****M****R****D****L****S****I****I****A****T****P****P****A****R****R****L****A****V****K****T****F****V**
R**E****Y****D****S****M****V****R****E****A****I****L****R****E****I****L****R****G****G****Q****V****Y****L****N****D****V****E****N****I****Q****K****A****A****E****R****L****A****E****L****V****P****E****A****R****I****A****I****G****H****G****Q****M****R****E****R****E****L****E**
R**V****M****N****D****F****H****H****Q****R****F****N****V****L****V****C****T****T****I****I****E****T****G****I****D****I****P****T****A****N****T****I****I****E****R****A****D****H****F****G****L****A****Q****L****H****Q****L****R****G****R****V****G****R****S****H****H****Q****A****Y****A**
W**L****L****T****P****H****P****K****A****M****T****T****D****A****Q****K****R****L****E****A****I****A****S****L****E****D****L****G****A****G****F****A****L****A****T****H****D****L****E****I****R****G****A****G****E****L****L****G****E****E****Q****S****G****S****M****E****T****I****G****F****S**
L**Y****M****E****L****L****E****N****A****V****D****A****L****K****A****G****R****E****P****S****L****E****D****L****T****S****Q****Q****T****E****V****E****L****R****M****P****S****L****L****P****D****D****F****I****P****D****V****N****T****R****L****S****F****Y****K****R****I****A****S****A****K**
T**E****N****E****L****E****E****I****K****V****E****L****I****D****R****F****G****L****L****P****D****P****A****R****T****L****L****D****I****A****R****L****R****Q****A****Q****K****L****G****I****R****K****L****E****G****N****E****K****G****G****V****I****E****F****A****E****K****N****H****V**
N**P****A****W****L****I****G****L****L****Q****Q****P****Q****H****Y****R****L****D****G****P****T****R****L****K****F****I****Q****D****L****S****E****R****K****T****R****I****E****W****R****Q****F****M****R****E****L****E****N****A****I****A****A****A****A****L****E****H****H****H**
H**H****H****&**

Peptides found by mass spectrometry are shaded in light grey. The dotted line corresponds to the next peptide found in the full-length protein and/or in larger fragments. Residues which were identified by EDMAN-sequencing are underscored in bold.

7.2 Abbreviations

Å	ångström (=10 ⁻¹⁰ m)
AA	amino acid or residue
ACN	acetonitrile
ATP	adenosine triphosphate
ATP-γ-S	adenosine 5'-O-(thio-triphosphate)
<i>Bca</i>	<i>Bacillus caldotenax</i>
BER	base excision repair
bp	base pair(s)
BSA	bovine serum albumin
<i>Bca</i>	<i>Bacillus caldotenax</i>
<i>Bsu</i>	<i>Bacillus subtilis</i>
ca.	circa
CHCA	α-cyano-4-hydroxy-cinnamic acid
CS	Cockayne's syndrome
Dali	distance matrix alignment
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DSB	double strand break
<i>Eco / E.coli</i>	<i>Escherichia coli</i>
e.g.	<i>exempli gratia</i> (for example)
EMSA	electrophoretic mobility shift assay
GGR	global genome repair
<i>Hin</i>	<i>Haemophilus influenza</i>
HPLC	high performance liquid chromatography
HR	homologous recombination
IEX	ion exchange chromatography
kb	kilobase pair(s);
LB	Luria-Bertani
M	molar
MAD	multiple-wavelength anomalous dispersion
MALDI	matrix-assisted laser desorption ionisation
Mfd	mutation frequency decline
MMR	mismatch repair
MR	molecular replacement
MS	mass spectrometry
<i>Mtu</i>	<i>Mycobacterium tuberculosis</i>
<i>mut</i>	mutator
NCBI	National Center for Biotechnology Information
n.d.	not determined
NER	nucleotide excision repair
NHEJ	non-homologous end-joining

NMR	nuclear magnetic resonance spectroscopy
o/n	over night
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PEI	polyethyleneimine
pI	isoelectric point
PIC	pre-incision complex
PEG	polyethylene glycol
PMSF	phenyl methyl sulfonyl fluorid
PNK	polynucleotide kinase
PVDF	polyvinyliden fluorid
RID	RNA polymerase interacting domain
RMSD	root mean square deviation
RNA	ribonucleic acid
RNAP	RNA polymerase
RNAPII	eukaryotic RNA polymerase II
RPA	replication protein A
RP-HPLC	reversed phase HPLC
SAD	single-wavelength anomalous diffraction
SDS	sodium dodecyl sulfate
SPR	surface plasmon resonance
ssDNA	single stranded DNA
TB	tris-borate
TCR	transcription-coupled repair
TEC	transcription elongation complex
TF	transcription factor
TFA	trifluoroacetic acid
TLC	thin layer chromatography
TLS	translesion synthesis
ToF	Time-of-Flight
TRCF	transcription-repair coupling factor
TRG	translocation in RecG
v	volume
w	weight
w/o	without
wt	wild-type
XP	xeroderma pigmentosum

7.3 Aminoacids and nucleotides

One letter code	Three letter code	Amino acid
A	Ala	alanine
B	Asx	asparagine or aspartic acid
C	Cys	cystein
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	Phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
L	Leu	leucine
K	Lys	lysine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
X	Xaa	unknown or other
Y	Tyr	tyrosine
Z	Glx	glutamine or glutamic acid

One-letter code	Nucleobase
A	adenine
C	cytosine
G	guanine
T	thymine

8 Curriculum vitae

Personal Details

Name Nora Maria Aßenmacher
Date of Birth May 12, 1979
Place of Birth Paderborn, Germany
Nationality German
Marital Status single

Education

1985-1989 Grundschule Benhausen
1989-1992 Pelizaeus-Gymnasium, Paderborn
1992-1997 Jugenddorf-Christophorusschule Braunschweig

Academic Training

1997 – 2002 University of Leipzig
Studies in Biochemistry (Diplom)

2001 – 2002 University of Leipzig, Institute of Biochemistry
Diploma Thesis: "Determination of nucleotide affinities of RNase T1 variants using the BIAcore system"
Supervisors: Prof. Dr. Uli Hahn and Dr. Marc Struhalla

since June 2002 LMU Munich, Gene Center / Institute of Biochemistry
PhD Thesis in structural biology
Supervisor: Prof. Dr. Karl-Peter Hopfner

9 Acknowledgements

I would like to thank all the people who have contributed to this work:

First of all, I am very grateful to my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work in his laboratory, for his guidance and invaluable advice.

I want to thank current and former members of the Hopfner lab. Many thanks to Derk Bemeleit, Katharina Büttner, Dr. Cui Sheng, Dr. Harald Dürr, Sophia Hartung, Annette Karcher, Axel Kirchhofer, Alfred Lammens, Alexandra Schele, and Dr. Gregor Witte for their scientific support as well as for the social and enjoyable lab atmosphere.

In particular, I would like to thank Dr. Katja Lammens for her patient teaching about X-ray crystallography, both in theoretical and practical aspects.

Furthermore, I am grateful for the help I got from Gabriela Guédez-Rodríguez, a Master student who did a practical course in the lab.

I would like to extend my personal thanks to all people on the forth floor for the nice atmosphere, for helpful advice, and for frequent exchange of materials.

Many thanks also to all service people at the Gene Center who make work much easier and more comfortable.

I am very grateful to have a wonderful, loving family as well as great and – especially during the last weeks – understanding friends.

Besonders bedanken möchte ich bei meinen Eltern und bei meiner Oma für ihre stetige Unterstützung und ihr vollstes Vertrauen.

Heartfelt thanks to my boyfriend, Jens Rauch, for providing me with loving support, patience, and with bright optimism.