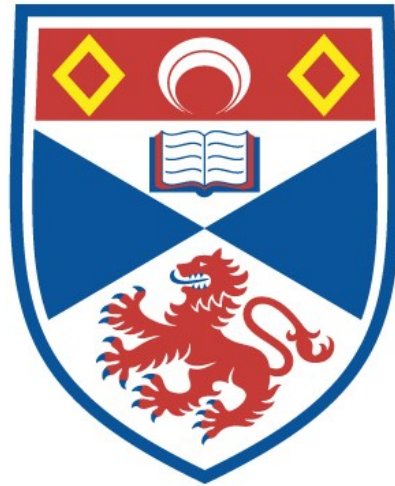


University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

Characterisation of *Sulfolobus solfataricus* XPF

An archaeal DNA repair nuclease

Jennifer A. Roberts

**A thesis submitted for the degree of
Doctor of Philosophy**

January 2005



**Univerisity
of
St Andrews**



CONTENTS

Contents	i
Tables & figures	v
Abbreviations	x
Declaration	xiii
Abstract	xiv
Acknowledgements	xv
1 <u>Introduction</u>	1
1.1 The 3 domains of life	2
1.2 DNA damage	3
1.3 Overview of DNA repair pathways	4
1.4 Nucleotide excision repair	6
1.4.1 NER in bacteria	6
1.4.2 NER in eukarya	7
1.4.3 NER in archaea	10
1.5 XPF/Mus81 protein family	13
1.6 Roles of XPF-ERCC1/Rad1-Rad10	17
1.6.1 NER	17
1.6.2 XPF-ERCC1/Rad1-Rad10 & recombination	17
1.6.3 Interstrand cross-link repair	19
1.6.4 XPF-ERCC1/Rad1-Rad10 summary	22
1.7 Roles of Mus81*	22
1.7.1 Mus81* & repair of stalled replication forks	22
1.7.2 Mus81* & recombination	26
1.8 Role of euryarchaeal Hef	29

2	<u>Materials & Methods</u>	30
2.1	Cloning & protein expression	31
2.1.1	Cloning & vectors	31
2.1.2	Site directed mutagenesis	33
2.1.3	Protein expression	33
2.2	Protein purification	33
2.2.1	Purification of native recombinant <i>Sso</i> XPF & <i>Sso</i> Fen1	33
2.2.2	Purification of N-terminal 1-138 <i>Sso</i> XPF mutant	34
2.2.3	Purification of native recombinant <i>Ape</i> XPF	35
2.2.4	Purification of native recombinant <i>Sso</i> XPB1	35
2.2.5	Purification of his-tagged PCNA subunits & PCNA heterotrimer	35
2.2.6	Purification of GST-fusion proteins	36
2.2.7	Protein analysis & concentration determination	36
2.3	Tryptic digestion of proteins	37
2.3.1	Tryptic digestion of <i>Sso</i> XPF	37
2.3.2	Tryptic digestion of <i>Sso</i> XPB1	37
2.4	Glutaraldehyde cross-linking of proteins	37
2.5	GST-fusion interaction studies	38
2.5.1	GST- <i>Sso</i> XPF & GST-PCNA2 fusion affinity chromatography	38
2.5.2	Small scale GST-fusion affinity assays	38
2.6	Isothermal titration calorimetry	39
2.6.1	The VP-ITC micro calorimeter	39
2.6.2	<i>Sso</i> XPF-metal ion binding	41
2.6.3	<i>Sso</i> XPF-PCNA binding	41
2.7	Catalytic & DNA binding assays	42
2.7.1	DNA substrates	42
2.7.2	<i>Sso</i> XPF endonuclease assays	42
2.7.3	<i>Sso</i> XPB1 electrophoretic mobility shift assays	43
2.7.4	<i>Sso</i> XPB1 helicase assays	44

3	An archaeal endonuclease dependent on heterotrimeric PCNA	45
3.1	Introduction	46
3.2	Expression & purification of <i>S. solfataricus</i> XPF	47
3.3	Mapping <i>Sso</i> XPF domains	48
3.4	<i>Sso</i> XPF forms a homodimer	50
3.5	Metal ion binding	52
3.6	Identifying <i>Sso</i> XPF protein partners	55
3.7	PCNA background	56
3.8	Confirming <i>Sso</i> XPF interacts with PCNA	58
3.9	<i>Sso</i> XPF has a conserved PCNA interaction motif	59
3.10	<i>Sso</i> XPF is a PCNA-dependent, structure specific nuclease	60
3.11	<i>Sso</i> XPF cleaved 4-way DNA junctions	65
3.12	<i>Sso</i> XPF was also active with other metal ions	65
3.13	Summary & conclusions	68
4	<u>SsoXPF nuclease activity</u>	70
4.1	Introduction	71
4.2	Quantification of <i>Sso</i> XPF nuclease activity	72
4.3	<i>Sso</i> XPF substrate specificity	74
4.3.1	3' flaps & nicked structures – importance of a 5' DNA end	74
4.3.2	<i>Sso</i> XPF cleaved flaps & overhangs	77
4.3.3	<i>Sso</i> XPF cleaved dsDNA containing lesions weakly	79
4.3.4	Summary	79
4.4	<i>Sso</i> XPF & rescue of stalled replication forks	82
4.5	<i>Sso</i> XPF & homologous recombination	84

CONTENTS

4.6	<i>Sso</i> XPF can cleave DNA processively in vitro	91
4.7	Summary & conclusions	95
5	<i>Sso</i>XPF interaction with PCNA	97
5.1	Introduction	98
5.2	Recombinant PCNA proteins	98
5.3	<i>Aeropyrum pernix</i> XPF is active with <i>S. solfataricus</i> PCNA	99
5.4	Stoichiometry & affinity of <i>Sso</i> XPF binding to PCNA	101
5.5	<i>Sso</i> XPF requires PCNA subunits 1 & 3 for activity	105
5.6	Loading of PCNA onto DNA	107
5.7	Stimulation of <i>Sso</i> XPF nuclease activity by PCNA	112
5.8	Comparing PCNA stimulation of <i>Sso</i> XPF & <i>Sso</i> Fen1	116
5.9	Summary & conclusions	121
6	<i>Sso</i>XPF interaction with other proteins	122
6.1	Introduction	123
6.2	Effects of <i>S. solfataricus</i> SSB & Alba on <i>Sso</i> XPF activity	124
6.3	PCNA interacts with SSO1289 & <i>Sso</i> XPB1	126
6.4	SSO1289	128
6.5	<i>Sso</i> XPB1	128
6.6	Expression & purification of <i>Sso</i> XPB1	129
6.7	Confirming the <i>Sso</i> XPB1 interaction with PCNA	130
6.8	Mapping <i>Sso</i> XPB1 domains	131
6.9	<i>Sso</i> XPB1 specifically bound splayed duplex DNA	133
6.10	<i>Sso</i> XPB1 helicase activity	135

6.11	Summary & conclusions	136
7	Conclusions & future work	138
	References	143
	Appendix 1:	
	Oligonucleotides used for DNA substrates	162
	Appendix 2:	
	Calculating thermodynamic parameters for <i>Sso</i> XPF interaction with manganese ions from isothermal titration calorimetry data	163

TABLES & FIGURES

Table 3.1	Thermodynamic parameters for <i>Sso</i> XPF manganese ion binding	53
Table 4.1	Relative <i>Sso</i> XPF cleavage rates for different DNA structures	74
Table 4.2	<i>Sso</i> XPF rates for D loop & 25 nt bubble DNA	88
Table 5.1	Isothermal titration calorimetry data for <i>Sso</i> XPF binding to PCNA	103
Table 5.2	Estimated fold stimulation of <i>Sso</i> XPF & <i>Sso</i> Fen1 nuclease activity in the presence of PCNA	120
Figure 1.1	The 3 domains of life	2
Figure 1.2	Summary of nucleotide excision repair in eukarya	9
Figure 1.3	<i>Sso</i> XPF sequence alignment	12
Figure 1.4	Domain organisation of XPF proteins	13

TABLES & FIGURES

Figure 1.5	Active site model for <i>P. furiosus</i> Hef showing conserved residues numbered according to their position in <i>Sso</i> XPF	15
Figure 1.6	Substrate specificity of XPF-ERCC1 & Mus81*	16
Figure 1.7	Overview of recombination mechanisms	18
Figure 1.8	Models for repair of interstrand cross-links	21
Figure 1.9	Processing of replication forks stalled at a leading strand lesion	24
Figure 1.10	Processing of replication forks stalled at a lagging strand lesion	25
Figure 1.11	Mus81* and generation of cross-over recombinants	28
Figure 2.1	Schematic diagram of the VP-ITC Unit	40
Figure 3.1	Purification of recombinant <i>Sso</i> XPF protein	47
Figure 3.2	<i>Sso</i> XPF dimers	48
Figure 3.3	<i>Sso</i> XPF consists of 2 domains	49
Figure 3.4	Cross-linking of full length <i>Sso</i> XPF & nuclease domain	51
Figure 3.5	Isothermal titration calorimetric analysis of <i>Sso</i> XPF manganese ion binding	53
Figure 3.6	Identifying the <i>Sso</i> XPF interaction with PCNA	56
Figure 3.7	Confirmation of the <i>Sso</i> XPF-PCNA interaction	58
Figure 3.8	PCNA subunit specific interaction with <i>Sso</i> XPF	59
Figure 3.9	Identifying the PCNA interaction motif in <i>Sso</i> XPF	61
Figure 3.10	Confirming the PCNA interaction motif	62
Figure 3.11	<i>Sso</i> XPF is a structure specific, PCNA-dependent endonuclease	63
Figure 3.12	Cleavage of a splayed duplex structure with a different DNA sequence	64
Figure 3.13	The D52A <i>Sso</i> XPF mutant interacts with PCNA	64

TABLES & FIGURES

Figure 3.14	<i>Sso</i> XPF cleaved a 4-way DNA junction	66
Figure 3.15	<i>Sso</i> XPF cleaved DNA in the presence of different metal ions	67
Figure 3.16	Schematic diagram of <i>Sso</i> XPF domain organisation	68
Figure 4.1	Quantifying <i>Sso</i> XPF nuclease activity for splayed duplex DNA	73
Figure 4.2	Relative <i>Sso</i> XPF cleavage rates for different DNA structures	75
Figure 4.3	Quantifying <i>Sso</i> XPF activity for nicked duplex DNA with and without a 5' phosphorylated end	76
Figure 4.4	<i>Sso</i> XPF cleaved a 3' flap	77
Figure 4.5	Effect of 3' overhang length on <i>Sso</i> XPF nuclease activity	78
Figure 4.6	<i>Sso</i> XPF cut weakly at bulges and abasic sites in DNA	80
Figure 4.7	<i>Sso</i> XPF cut dsDNA containing a thymine dimer	81
Figure 4.8	Model for <i>Sso</i> XPF cleavage at stalled and reversed replication forks	83
Figure 4.9	Role of <i>Sso</i> XPF in recombination – Model 1: removal of 3' flap intermediates during SDSA	85
Figure 4.10	<i>Sso</i> XPF cleaved bubble DNA lacking an invading strand	86
Figure 4.11	<i>Sso</i> XPF cleaved a D loop structure	87
Figure 4.12	Quantifying <i>Sso</i> XPF activity for D loop & 25 nt bubble DNA	88
Figure 4.13	Quantifying <i>Sso</i> XPF cleavage of D loop b75 strand	89
Figure 4.14	Role of <i>Sso</i> XPF in recombination – Model 2: resolution of D loop structures to create cross-over recombinants	90
Figure 4.15	Processive cleavage of DNA flap substrates by <i>Sso</i> XPF	92
Figure 4.16	Rate of <i>Sso</i> XPF cleavage at individual sites in 3' overhang DNA	93

Figure 4.17	<i>Sso</i> XPF processively cleaved extended dsDNA regions of nicked duplex and 5' overhang DNA	94
Figure 5.1	SDS-PAGE analysis of purified his-tagged PCNA subunits and assembly of PCNA heterodimer	99
Figure 5.2	<i>Aeropyrum pernix</i> XPF	100
Figure 5.3	Isothermal titration calorimetry of <i>Sso</i> XPF binding to PCNA fitted to different binding models	102
Figure 5.4	<i>Sso</i> XPF binding to PCNA subunit 1	105
Figure 5.5	<i>Sso</i> XPF activity with combinations of PCNA subunits	106
Figure 5.6	Comparing <i>Sso</i> XPF activity with PCNA subunits 1, 2 & 3 against the activity with 1 & 3 alone	106
Figure 5.7	Streptavidin specifically bound biotinylated DNA substrates	108
Figure 5.8	PCNA stimulated <i>Sso</i> XPF nuclease activity on biotin-streptavidin end blocked DNA	109
Figure 5.9	Cartoon of Fen1 nuclease activity & SDS-PAGE of purified recombinant <i>Sso</i> Fen1	110
Figure 5.10	<i>Sso</i> Fen1 stimulation by PCNA on biotin-streptavidin end blocked double flap substrate	111
Figure 5.11	PCNA stimulation of splayed duplex cleavage by <i>Sso</i> XPF	113
Figure 5.12	<i>Sso</i> XPF cleaved 3' flap DNA in the absence of PCNA	115
Figure 5.13	PCNA stimulation of 3' flap cleavage by <i>Sso</i> XPF	115
Figure 5.14	PCNA stimulation of 5' double flap and 5' flap cleavage by <i>Sso</i> Fen1	117

TABLES & FIGURES

Figure 5.15	Quantification of DNA cleavage by <i>Sso</i> XPF and <i>Sso</i> Fen1 in the absence of PCNA	118
Figure 5.16	Quantifying PCNA stimulation of <i>Sso</i> XPF and <i>Sso</i> Fen1 nuclease activity	119
Figure 6.1	Effect of SSB on <i>Sso</i> XPF cleavage of splayed duplex DNA	124
Figure 6.2	Effects of SSB and Alba on <i>Sso</i> XPF cleavage of 3' flap DNA	125
Figure 6.3	Effect of SSB on <i>Sso</i> XPF cleavage of nicked duplex DNA	126
Figure 6.4	Identification of proteins interacting with PCNA subunit 2	127
Figure 6.5	<i>Sso</i> XPB1 interacted with PCNA	130
Figure 6.6	<i>Sso</i> XPB1 is monomeric	131
Figure 6.7	<i>Sso</i> XPB1 domain organisation	132
Figure 6.8	Electrophoretic mobility shift analysis of <i>Sso</i> XPB1 binding with different DNA structures	134
Figure 6.9	Quantification of <i>Sso</i> XPB1 binding splayed duplex DNA with and without PCNA	135
Figure 6.10	<i>Sso</i> XPB1 helicase activity	136
Figure 7.1	Evolutionary scheme for XPF/Mus81 proteins	140

ABBREVIATIONS

3'	3 prime end (DNA)
5'	5 prime end (DNA)
3D PSSM	3 dimensional position-specific score matrix
A ₂₈₀	absorbance at 280 nm
A ₆₀₀	absorbance at 600 nm
AEBSF	4-(2-aminoethyl)-benzenesulphonylfluoride.HCl
<i>Ape</i> XPF	<i>Aeropyrum pernix</i> XPF
BER	base excision repair
BIR	break induced replication
bp	base pair
CPD	cyclobutane pyrimidine dimers
CS	Cockayne syndrome
CSA/B	Cockayne syndrome complementation group A/B
D loop	displacement loop
DDB	damage DNA binding factor
DSB	double strand (DNA) break
DSBR	double strand (DNA) break repair
dsDNA	double stranded deoxyribonucleic acid
Eme1	essential meiosis endonuclease 1
ERCC1	excision repair cross complementing 1
Fen1	flap endonuclease 1
GG-NER	global genome nucleotide excision repair
GST	glutathione-S-transferase

ABBREVIATIONS

Hef	helicase-associated endonuclease for fork structures
HhH	helix-hairpin-helix
HJRE	Holliday junction resolving enzyme
ICL	interstrand cross-link
ITC	isothermal titration calorimetry
K_D	dissociation constant
kDa	kilo Dalton
MALDI ToF	matrix-assisted laser desorption/ionization time of flight
Mfd	mutation frequency decline protein
MMR	mismatch repair
MMS4	methylmethane sulfate sensitivity factor 4
Mus81*	Mus81 dimer with partner protein (Mms4 or Emel)
NaBH_4	sodium borohydride
NER	nucleotide excision repair
nt	nucleotide
PCNA	proliferating cell nuclear antigen
RFC	replication factor C
RPA	replication protein A
SDSA	synthesis dependent strand annealing
SSB	single stranded binding protein
ssDNA	single stranded deoxyribonucleic acid
<i>SsoFen1</i>	<i>Sulfolobus solfataricus</i> Fen1
<i>SsoXPB1</i>	<i>Sulfolobus solfataricus</i> XPB1
<i>SsoXPF</i>	<i>Sulfolobus solfataricus</i> XPF
TC-NER	transcription coupled nucleotide excision repair

ABBREVIATIONS

TCA-DOC	trichloroacetic acid - deoxycholate
TFIIH	transcription factor II H
UV	ultra violet
XP	xeroderma pigmentosum
XPA/B/C/ D/E/F/G	xeroderma pigmentosum complementation group A/B/C/D/E/F/G

DECLARATION

I, Jennifer Roberts, hereby certify that this thesis, which is approximately 27,500 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date ..13/1/05..... Signature of candidate

I was admitted as a research student in October 2001 and as a candidate for the degree of PhD in October 2002; the higher study for which this is a record was carried out in the University of St Andrews between 2001 and 2004.

Date ..13/1/05..... Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate of the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ..13/1/05..... Signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date ..13/1/05..... Signature of candidate

ABSTRACT

DNA is constantly under attack from a variety of endogenous and environmental factors and all organisms have evolved a variety of pathways to repair this damage and maintain the integrity of their genome. Archaea share many similarities with eukarya in their DNA information processing pathways but their DNA repair pathways are not well understood. This thesis presents the characterisation of the archaeon *Sulfolobus solfataricus* XPF protein (*SsoXPF*), identified as a homologue of the eukaryal 5' nucleotide excision repair (NER) endonuclease XPF from sequence analyses.

Structural and functional studies revealed *SsoXPF* had properties similar to, but broader than either eukaryal XPF or the related endonuclease Mus81 involved in rescuing stalled DNA replication and recombination. This suggests fundamental conservation of this nuclease through billions of years of evolution and divergence into the specialised enzymes seen in eukarya today.

SsoXPF also displayed unique properties including an interaction with the sliding clamp protein PCNA necessary for nuclease activity, and the mechanism of this PCNA stimulation was investigated. The heterotrimeric nature of *S. solfataricus* PCNA and subunit specific interaction with *SsoXPF* opened the possibility of a second protein interacting simultaneously to create a ternary *SsoXPF*-PCNA-? complex and define the role of *SsoXPF* in vivo. Two putative helicases were identified as possible candidates and initial characterisation of one, a homologue of the eukaryal NER 3' to 5' helicase XPB showed tight binding to splayed duplex DNA consistent with a role in archaeal NER.

ACKNOWLEDGEMENTS

I would especially like to thank my supervisors Professor Malcolm White for all his teaching, guidance and encouragement throughout my project, and Professor Jim Naismith for all his help with the protein crystallography. I also wish to thank my lab colleagues past and present for their help, assistance and friendship. Also thanks to Dr Catherine Botting, Dr Robin Antrobus and Alex Houston here at St Andrews for all the samples they analysed for me by mass spectrometry, and Paul Talbot for the protein sequencing.

In addition I am very grateful to Dr Steve Bell, MRC Cancer Cell Unit, Cambridge for the purified recombinant PCNA protein I used initially and the DNA constructs enabling me to subsequently express and purify the proteins myself. Also thanks to Dr Emma Warbrick for very useful discussion and help identifying PCNA interacting motifs, and the BBSRC for funding.

Lastly but by no means least I want to thank my family and friends for all their support and for making the last few years so enjoyable.

CHAPTER 1: Introduction

1.1 The 3 domains of life

In 1977 Fox and Woese proposed that cellular life could be classified into 3 domains based on ribosomal RNA sequence analysis – bacteria, eukarya and archaea (Figure 1.1). Archaea resemble bacteria in cellular morphology i.e. a single cell containing a circular chromosome and lacking eukaryal organelles (Bernander, 2000). However archaeal information processing pathways (DNA replication, DNA transcription etc.) clearly resemble those of eukarya and are distinct from the bacterial equivalent (Bell and Jackson, 1998; Keeling and Doolittle, 1995). Such archaeal pathways are simplified versions of the eukaryal pathways, stripped down to the core factors and consequently provide useful model systems.

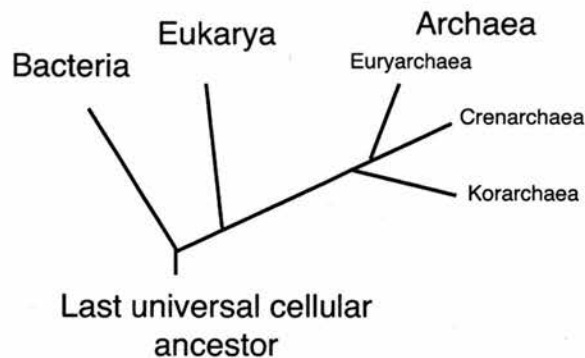


Figure 1.1. The 3 domains of life (adapted from White 2003).

Archaea are divided into 3 subdomains – euryarchaea, crenarchaea and korarchaea. Korarchaea were discovered relatively recently and are only known from environmental DNA sequence (Barns et al., 1996). Some fundamental differences exist between euryarchaea and crenarchaea. Generally euryarchaea display the greater homology to eukarya. For example most euryarchaea have histone-like proteins to stabilise and package DNA whereas crenarchaea lack true histones and

rely on small double stranded DNA binding proteins (Reeve et al., 1997). *Sulfolobus solfataricus* was used in this study. This hyperthermophilic crenarchaeon grows aerobically in acidic (pH 2-4) sulphur rich pools at approximately 80 °C. Because *S. solfataricus* is a hyperthermophile its proteins are generally stable and therefore easily expressed and highly amenable to structural and biophysical studies, often difficult in mesophilic counterparts. The genome of *S. solfataricus* strain P2 has been sequenced (She *et al* 2001) and this work looks at the putative DNA repair protein encoded by gene SSO0729.

1.2 DNA damage

DNA in all organisms is subject to constant assault from a variety of endogenous and environmental factors that modify and damage the primary structure (Lindahl, 1993). This causes genomic instability and mutations, likely to be a major factor in both ageing and carcinogenesis. Water and oxygen cause hydrolysis and oxidation of DNA respectively and together with errors in base pairing introduced during DNA replication, represent the major endogenous DNA damage. For example the intrinsic lability of the *N*-glycosyl bond in DNA makes it susceptible to hydrolysis causing loss of the base and therefore genetic information. In addition DNA base residues are themselves susceptible to hydrolytic deamination, for example the hydrolysis of cytosine to uracil, which can result in the mutagenic base pair change cytosine:guanine to thymine:adenine. During normal aerobic metabolism active oxygen species are produced, for example hydroxyl and superoxide radicals, which modify bases, in particular guanine and thymine, again potentially resulting in a point mutation. Other reactive molecules that are found in living cells can also non-

enzymatically modify DNA. Probably the most important of these is *S*-adenosylmethionine (SAM), which is a methyl donating cofactor but non-enzymatic methylation of DNA bases also occurs as a result of its weak alkylating activity. Such spontaneous chemical DNA damage described above is accelerated at elevated temperature (Lindahl, 1993), thus would pose an increased threat to hyperthermophiles such as *S. solfataricus* (Grogan, 2000).

For many organisms including *S. solfataricus* the ultraviolet (UV) component of sunlight is the most important environmental DNA damaging factor causing dipyrimidine photoproducts, most commonly cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts. These are both mutagenic and cytotoxic, distorting the DNA helix, disturbing base pairing and preventing DNA transcription and replication. Other environmental DNA damaging factors include ionising radiation that causes double strand breaks in DNA and exposure to certain chemicals, for example human exposure to tobacco smoke.

1.3 Overview of DNA repair pathways

DNA damage must be detected and repaired accurately and rapidly to maintain the integrity of the genome. Bacteria and eukarya have evolved a variety of overlapping DNA repair pathways to do this (Griffin, 1996; Lindahl and Wood, 1999). These fall into 3 mechanistic categories. (1) The damage can be directly reversed in photoreactivation or alkyl transfer pathways. Photoreactivation involves photolyase enzymes that use visible light as an energy source to convert UV lesions back to the original undamaged bases. DNA methyltransferases remove methyl groups from modified bases such as methylguanine. (2) The damage can be excised

and the undamaged strand used as a template for resynthesing the DNA. Base excision repair (BER) involves a range of DNA glycosylases that recognise and remove the modified base and an apurinic/apyrimidinic endonuclease subsequently cleaves the abasic site, allowing insertion of the correct base. Correction of most endogenous lesions is by this pathway, which shows strong evolutionary conservation. Nucleotide excision repair (NER) acts on a wide variety of mainly exogenous bulky helix-distorting lesions such as UV photoproducts and intra- and inter-molecular cross-links. Incision either side of the lesion releases a short oligonucleotide containing the damage and the gap is filled by repair replication. NER is described in more detail in the next section. Finally in the excision category, mismatch repair (MMR) primarily acts to reverse errors resulting from misincorporation of bases by DNA polymerase, missed by the polymerase's proofreading activity. This involves the highly conserved MutS and MutL families of proteins: MutS recognises the mismatched base and interaction of MutL activates the endonuclease. Subsequent exonuclease digestion up to and past the mispaired base creates a gap that is filled by repair replication. (3) The DNA lesion may be bypassed by a special class of polymerases able to synthesise past certain lesions or by homologous recombination. Recombination is discussed further later.

These repair pathways all involve groups of interacting proteins and although they repair different types of DNA damage there is some overlap between them as well as with other pathways.

1.4 Nucleotide excision repair (NER)

As mentioned above NER acts to remove a wide range of mainly exogenous, bulky, helix-distorting lesions from DNA. Eukaryal and bacterial pathways are mechanistically very similar, involving damage recognition, unwinding around the lesion, dual incision either side of the lesion and DNA re-synthesis using the undamaged strand as a template. In both bacteria and eukarya there are 2 modes of NER that differ in damage recognition. Global genome NER (GG-NER) repairs lesions spread over the entire genome whereas transcription coupled NER (TC-NER) repairs lesions in transcriptionally active strands where transcription elongation has been blocked (de Laat et al., 1999; Sancar, 1996). However bacterial and eukaryal NER machinery is unrelated.

1.4.1 NER in bacteria

In bacterial GG-NER, 3 proteins UvrA, UvrB and UvrC carry out the complete process of damage recognition and excision (Sancar, 1996), where the complex of UvrA₂B tracks along the DNA until it recognises a DNA lesion. In TC-NER the stalled RNA polymerase constitutes the damage recognition signal. The transcription coupling repair factor Mfd binds, releasing RNA polymerase and simultaneously recruiting UvrA₂B (Selby and Sancar, 1993). In both pathways a transient UvrA₂B-DNA complex is formed at the lesion, kinking and unwinding the DNA. The UvrA₂ is displaced enabling UvrB to form a stable pre-incision complex with DNA upon ATP hydrolysis. UvrC subsequently binds UvrB and cleaves the DNA 4-5 nt 3' of the lesion followed by a second cut 7-8 nt 5' of the lesion to release a 12-13 nt oligo corresponding to the footprint of UvrBC. UvrD (also called helicase II) is required for release of both UvrC and the excised oligomer containing the

damage (Orren et al., 1992), leaving the 3' OH group accessible to DNA polymerase to fill in the gap and displace UvrB.

1.4.2 NER in eukarya

Eukaryal NER is more complex involving over 30 polypeptides and reconstruction of the cut and paste reaction with purified proteins has defined the core components (Aboussekhra et al., 1995). Evidence supports a sequential stepwise assembly of repair proteins and subcomplexes rather than recruitment of a preassembled repairosome complex (Riedl et al., 2003; Volker et al., 2001).

In eukaryal GG-NER, XPC in complex with HR23B binds the site of DNA damage inducing a bend in the DNA (Janicijevic et al., 2003). For some types of lesion that are poorly recognised by XPC-HR23B, for example CPDs, the damage DNA binding (DDB) factor may facilitate identification of the lesion. XPC-HR23B recruits the multi-protein complex TFIIH. The XPB and XPD subunits have DNA dependent ATPase and helicase functions, unwinding the DNA around the lesion in 3' to 5' and 5' to 3' directions respectively. This creates an initial opening of about 10 nt around the lesion. XPA binds the damaged DNA with an affinity proportional to the helical distortion induced by XPC-HR23B. XPA is thought to verify NER lesion identification and recruit other NER factors. The single stranded DNA binding protein replication protein A (RPA) binds the ssDNA opposite the lesion in cooperation with XPA. In the presence of both XPA and RPA, HR23B facilitates displacement of XPC through modifying XPC binding activity (You et al., 2003) resulting in a stable XPA-RPA complex at the damage site.

XPC-HR23B is not involved in TC-NER. Instead the blocked elongating RNA polymerase II acts as the damage recognition signal recruiting the transcription coupled repair specific CSA and CSB proteins, TFIIH and possibly other factors

(Svejstrup, 2002). This results in displacement of the stalled RNA polymerase II from the lesion and formation of a stable XPA-RPA complex at the damage site as for GG-NER. TC-NER in eukarya is still not fully understood.

RPA has several roles in NER. It displays 2 ssDNA binding modes: an initial interaction with 8-10 nt followed by a 30 nt binding mode that is approximately 100 fold more stable. This transition between modes may facilitate formation of the open repair complex. In addition RPA binds the undamaged strand with defined polarity, positioning the nucleases XPG and XPF-ERCC1 and directing cleavage to the damaged strand (de Laat et al., 1998b). Cleavage is asymmetrical to the lesion and reflects the boundaries of the open complex. XPG cuts 2-8 nt 3' and XPF-ERCC1 15-24 nt 5' of the lesion to remove a 24-34 nt patch containing the damaged DNA, the exact size of which depends on the type of damage and the DNA sequence. TFIIH has been proposed to regulate incisions by XPG and XPF-ERCC1, preventing incision prior to full open complex formation (Winkler et al., 2001). Consistent with this is the observation that a mutation affecting the C-terminal 40 amino acids of the XPB subunit resulted in defective 5' incision (Evans et al., 1997). In addition XPG has structural, nuclease independent roles in NER promoting stability of the preincision complex (Mu et al., 1996) and the presence of XPG is a prerequisite for the 5' incision by XPF-ERCC1 (Wakasugi et al., 1997).

RPA, proliferating cell nuclear antigen (PCNA), replication factor C (RFC) and either DNA polymerase δ or ϵ are sufficient for repair DNA resynthesis *in vitro* (Aboussekhra et al., 1995). RFC binds the 3' end of dsDNA at the gap and facilitates loading PCNA. PCNA encircles DNA acting as a sliding platform for DNA polymerase δ or ϵ replication across the gap. PCNA may have additional roles linking excision and resynthesis, and promoting turnover of nucleases through an

interaction with XPG (Gary et al., 1997; Nichols and Sancar, 1992). Eukaryal NER is summarised in Figure 1.2.

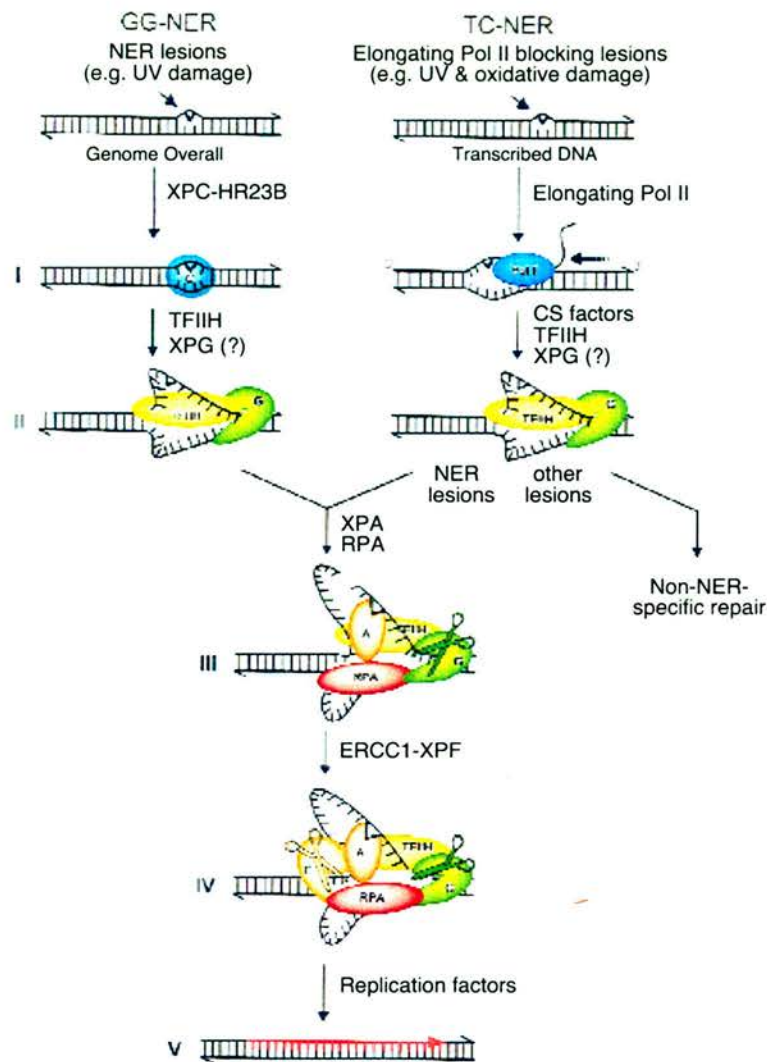


Figure 1.2. Summary of nucleotide excision repair (NER) in eukarya (taken from de Laat *et al* 1999). In the global genome repair pathway (GG-NER) XPC-HR23B (C) recognizes and binds the lesion. In transcription coupled NER (TC-NER) stalling of the elongating RNA polymerase at a lesion is the damage sensor. XPC-HR23B recruits TFIIH (and possibly XPG (G)) and TFIIH helicase subunits XPB and XPD unwind the DNA around the lesion (II, left). CSA, CSB, TFIIH, XPG and possibly other factors displace the RNA polymerase and, depending on the type of lesion, repair is completed by NER or an alternative repair pathway (II, right). XPA (A) and RPA stabilize the full open complex structure and together with TFIIH position XPG for the 3' incision (III). XPF-ERCC1 (F) is recruited to the complex and positioned by XPA and RPA to make the 5' incision (IV). The gap is filled by DNA synthesis and ligation (V).

In humans defects in NER cause several genetic disorders with very different clinical features (Lehmann, 2003). Xeroderma pigmentosum (XP) is characterised by an extreme photosensitivity and predisposition to skin cancer. Eight XP complementation groups have been identified representing the distinct repair genes XPA-E. Cockayne syndrome (CS) is a very pleiotropic condition characterised by severe neurological, developmental and premature aging in addition to photosensitivity. CS is specifically associated with a defect in TC repair and the additional symptoms indicate TC repair and/or CS proteins have roles beyond NER (Svejstrup, 2002). Some patients defective in XPB or XPD display both CS and XP whereas others suffer from the CS-like brittle hair syndrome trichothiodystrophy (TTD).

1.4.3 NER in archaea

Little is known about NER in archaea. Most information is derived from sequence analyses and lacks biochemical evidence (Grogan, 2000; White, 2003). Most archaea have homologues of eukaryal NER helicases XPB and XPD and the nucleases XPG and XPF. In addition some mesophilic methanogenic and halophilic euryarchaea have bacterial UvrABC proteins thought to be the result of lateral gene transfer between bacteria and eukarya (Grogan, 2000). In *Methanothermobacter thermautotrophicus* at least one of these pathways is functional shown by *in vitro* excision of an oligonucleotide containing a UV photoproduct (Ogrunc et al., 1998). Archaeal NER pathways are generally thought to represent simpler perhaps ancestral versions of eukaryal NER (Grogan, 2000; White, 2003).

However there remains the problem of damage recognition as archaea have no obvious homologues of eukaryal XPC and XPA GG-NER damage recognition proteins. One candidate is the archaeal single stranded binding (SSB) protein, which

has been reported to detect DNA damage, melt dsDNA and recruit repair proteins (Iftode et al., 1999). Photolyase is an alternative candidate. In the absence of light photolyase cannot catalyse the splitting reaction but remains stably bound to damaged DNA and increases CPD removal by the NER pathways in both *E. coli* and yeast (Sancar, 1996).

Interestingly plants and the simple eukarya *Plasmodium falaparum* also lack obvious XPC and XPA homologues (Gardner et al., 2002; Liu et al., 2000). In addition GG repair is less conserved than TC repair and whereas human cells exhibit substantial GG repair, levels in rodent cells are low (Svejstrup, 2002). Although no evidence for TC repair in archaea has been published it is plausible that archaeal RNA polymerase stalled at an NER lesion could act as the signal to initiate NER. However again there is no archaeal equivalent for either bacterial Mfd or eukaryal CS proteins involved in bypass/removal of stalled RNA polymerases and recruitment of the repair machinery. The interaction between RNA polymerase and SSB in *S. solfataricus* may be relevant for archaeal TCR (Richard et al., 2004).

The SSO0729 gene product was identified as the *S. solfataricus* homologue of eukaryal XPF (Figure 1.3), sharing 24 % sequence identity and 44 % sequence similarity with the C-terminal third of human XPF. After an introduction to XPF proteins, the remainder of this thesis describes the work done to characterise the *S. solfataricus* homologue (referred to as *SsoXPF* from here on) with the aim to help our understanding of archaeal DNA repair that may have implications for equivalent eukaryal pathways.

These proteins share some conserved domains but there are also distinct differences (Figure 1.4).

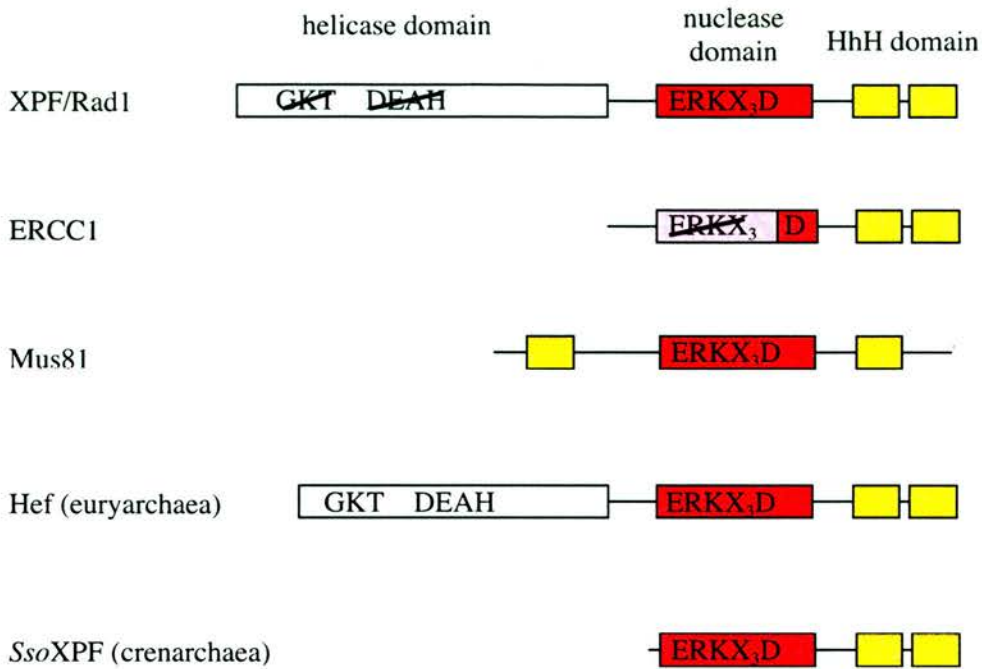


Figure 1.4. Domain organisation of XPF proteins. Conserved domains are represented by boxes: helicase domains are white, nuclease domains red and helix-hairpin-helix (HhH) domains yellow. Conserved catalytic residues are shown and scored out where these active motifs are disrupted. ERCC1 has a region homologous to the nuclease domain but lacking the active site residues (pink box). Diagram not to scale.

Eukaryal XPF/Rad1 and euryarchaeal Hef have a superfamily 2 DNA/RNA helicase domain at the N-terminus (Komori et al., 2002; Sgouros et al., 1999) that is not present in Mus81 or crenarchaeal XPF (Figure 1.4). In eukaryal XPF/Rad1 the helicase motifs necessary for activity are disrupted suggesting a structural role in polynucleotide binding since both SF2 helicases and XPF/Rad1 bind a ss/dsDNA junction with the same polarity (Gaillard and Wood, 2001; Sgouros et al., 1999). In euryarchaeal Hef the helicase motif is intact and shows DNA structure dependent

ATPase activity (although no helicase activity was detected) and a DNA remodelling role has been proposed (Komori et al., 2002).

*Sso*XPF shares the conserved metal dependent nuclease motif ERKX₃D present in XPF/Mus81 proteins (Figure 1.4). This was mapped using affinity cleavage and confirmed by mutagenesis of the conserved residues (Figure 1.3)(Enzlin and Scharer, 2002). The crystal structure of the nuclease domain of *P. furiosus* Hef revealed this formed a similar fold to the type II restriction endonucleases (Nishino et al., 2003). These are sequence specific endonucleases found ubiquitously in prokarya and are constituents of the restriction modification system functioning to protect the host genome from foreign DNA (Pingoud and Jeltsch, 2001). Other enzymes including the archaeal Holliday junction resolvase Hjc (Bond et al., 2001) and the *E. coli* GT mismatch recognition protein Vsr also share this fold. The major difference between the type II restriction enzymes signature motif PDX_n(D/E)XK and XPF/Rad1/Mus81 GDX_nERKX₃D is the conserved arginine in the latter, which interacts electrostatically with 2 conserved glutamates proposed to play a role in recognition of branched DNA structures (Nishino et al., 2003). The active site comprises 2 V shaped helical pairs forming a groove containing conserved acidic amino acids involved in metal ion coordination and catalysis (Figure 1.5).

Also similar to eukaryal XPF/Rad1 *Sso*XPF has a predicted C-terminal double helix-hairpin-helix (HhH) domain conserved in other DNA repair proteins (Aravind et al., 1999) and thought to have a role in DNA binding (Doherty et al., 1996). Indeed the bacterial NER nuclease UvrC also has this C-terminal double HhH (Singh et al., 2002), deletion of which leads to loss of the 5' incision (Moolenaar et al., 1998). The HhH domains in Mus81 are not in tandem but flanking the central nuclease domain (Figure 1.4) (Fu and Xiao, 2003; Interthal and Heyer, 2000). Differences in domain

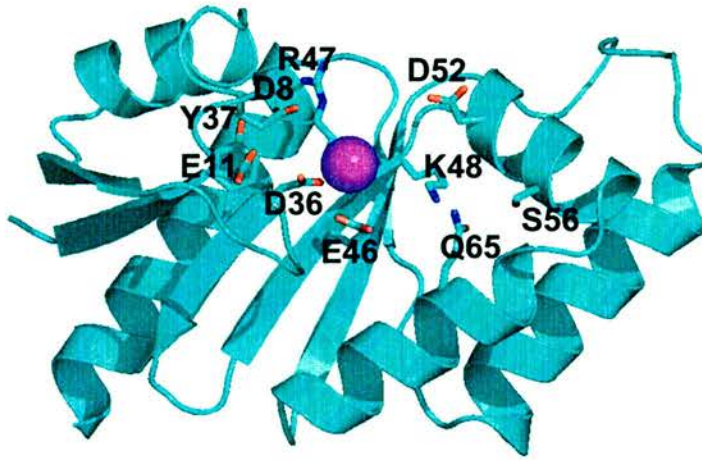


Figure 1.5. Active site model of *P. furiosus* Hef showing conserved residues numbered according to their position in *SsoXPF*. The metal II ion is shown as a magenta sphere coordinated by the main chain carbonyl oxygen of arginine 47 and the carboxyl side chains of aspartate 36 and glutamate 46. In addition the metal ion coordinates 3 water molecules held in place by aspartate 8, lysine 48 and aspartate 52. Diagram made by Professor Malcolm White.

architecture may reflect differences in recruitment of XPF and Mus81 to branched DNA (Komori et al., 2002).

All XPF/Rad1/Mus81 proteins form dimers. XPF forms a heterodimer with ERCC1 (Park et al., 1995). ERCC1 shares significant sequence homology with the nuclease and HhH domains of XPF (Figure 1.4) and is thought to have arisen from a gene duplication event (Gaillard and Wood, 2001). However ERCC1 has no nuclease activity as it lacks the catalytic site residues and is important for protein-protein interactions. Similarly yeast Rad1 forms a heterodimer with Rad10 (Bardwell et al., 1992). Mus81 forms a heterodimer with Mms4 in *S. cerevisiae* (Kaliraman et al., 2001) and Eme1 in *S. pombe* (Boddy et al., 2001). The human Mus81 partner has been named both Mms4 and Eme1 (Ciccina et al., 2003; Ogrunc and Sancar, 2003). These partner proteins have limited amino acid sequence homology but form similar predicted tertiary structures (Ogrunc and Sancar, 2003). For simplicity Mus81-

Mms4/Eme1 will be abbreviated to Mus81*. In contrast euryarchaeal Hef forms a homodimer. Two dimer interfaces were identified, one between the nuclease domain and the other between HhH domains (Nishino et al., 2003). For all XPF/Mus81 proteins dimerisation is essential for stability and coordinated substrate binding and cleavage (Bardwell et al., 1992; Boddy et al., 2001; Gaillard and Wood, 2001; Kaliraman et al., 2001). The work by Nishino *et al.* suggests dimerisation through both nuclease and HhH domains is important for Hef activity. Given XPF/Rad1/Mus81 cleave only 1 strand of the dsDNA presumably only 1 active site is required so perhaps the second nuclease domain is required in DNA binding (Yang, 2003) and must be correctly positioned by dimerisation.

Despite catalytic site and domain organisation similarities XPF/Mus81 have different substrate specificities. Where as XPF-ERCC1 and Rad1-Rad10 proteins cleave dsDNA 5' of a junction with ssDNA (Bardwell et al., 1994; Davies et al., 1995; de Laat et al., 1998a), Mus81* cleavage is directed to dsDNA 5' of a 5' DNA end (Figure 1.6) (Bastin-Shanower et al., 2003; Osman et al., 2003). This difference in substrate specificity reflects the distinct *in vivo* roles of XPF-ERCC1 and Mus81*

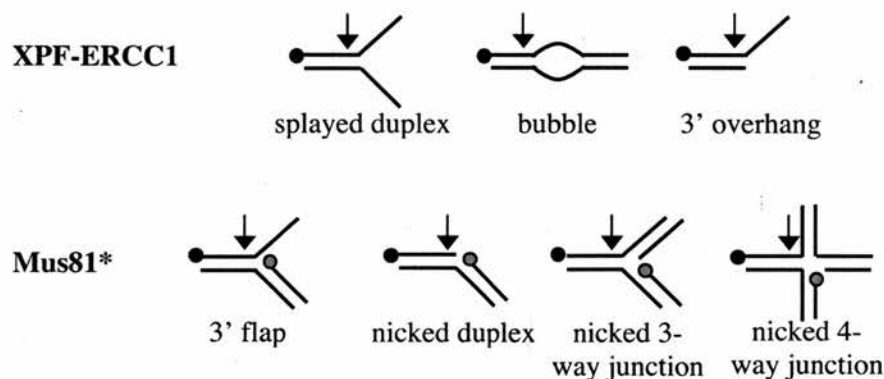


Figure 1.6. Substrate specificity of XPF-ERCC1 & Mus81*. Black circles indicate 5' ends and arrows show the site of cleavage. Red circles show the 5' end positioning and stimulating Mus81* cleavage. Such 5' DNA ends inhibit Rad1-Rad10 (and by extension XPF-ERCC1) cleavage.

that will now be introduced.

1.6 Roles of XPF-ERCC1/Rad1-Rad10

1.6.1 NER

NER was discussed in detail in section 1.4. XPF-ERCC1 diffuses freely through the nucleus but upon UV damage it is transiently immobilised (Houtsmuller et al., 1999). Interaction with XPA and RPA recruits and positions XPF-ERCC1 at the NER site to make the 5' incision (de Laat et al., 1998b; Matsunaga et al., 1996; Park et al., 1995; Park and Sancar, 1994). The physical presence of XPG is necessary for 5' incision (Mu et al., 1996; Wakasugi et al., 1997) and an interaction between the C-terminus of the XPB helicase and XPF (Evans et al., 1997) is thought to coordinate open complex formation and incision (Winkler et al., 2001). XPF-ERCC1 appears to be bound to the NER site for the duration of repair (Houtsmuller et al., 1999) before being released again.

1.6.2 XPF-ERCC1/Rad1-Rad10 & recombination

The initiating factor for recombination is a double strand break (DSB) in the DNA. DSBs arise from various events including replication of a nicked template strand, processing of stalled replication forks or exposure to γ -irradiation or certain chemicals. The 3 main recombination pathways that repair these breaks – classical double strand break repair (DSBR), synthesis dependent strand annealing (SDSA) and break induced replication (BIR) are introduced in Figure 1.7 (Paques and Haber, 1999). All involve an initial single strand annealing step: 5' to 3' degradation at the DSB produces 3' ssDNA ends that invade homologous duplex DNA. A region of non-homology in the invading 3' ssDNA end would result in a 3' ss flap. Rad1-

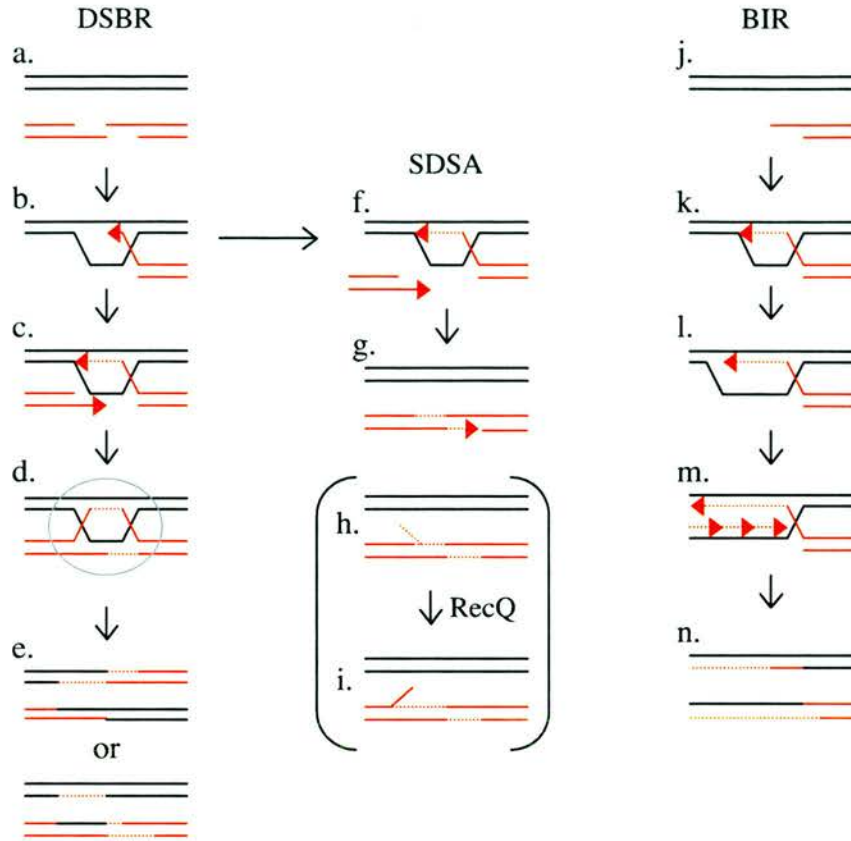


Figure 1.7. Overview of recombination mechanisms. Invading duplex DNA is red, dotted lines represent newly synthesised DNA and arrowheads represent 3' ends. In the classical double strand break repair (DSBR) model the 5' ends at the break are resected by an exonuclease to create a 3' ssDNA end (a) that invades homologous dsDNA (b). DNA synthesis is initiated to create a displacement (D) loop (c) and capture of the second end leads to formation of a double Holliday junction (d). Cleavage by a Holliday junction resolving enzyme gives crossed over (CO) or non-crossed over (NCO) product dependent on the sites of cleavage (e). Resolution of both Holliday junctions in the same direction or opposite directions result in NCO and CO respectively. Alternatively, in synthesis dependent strand annealing (SDSA) the invading strand from the D loop reanneals with the other end of the break and acts as a template to fill the gap (g). Over-synthesis of the invading strand can lead to a 3' flap upon reannealing (h) which in the presence of RecQ helicases can isomerise to a 5' flap (i). Break induced replication (BIR) is similar to DSBR but only one end of a DSB is present or homologous to the template (j). The replication bubble converts to a full replication fork that can progress to a chromosome end or until it encounters a block (k-n).

Rad10 (Fishman-Lobell and Haber, 1992) and XPF-ERCC1 (Adair et al., 2000; Sargent et al., 1997) are implicated in the removal of these tails to allow recombination to proceed. However XPF-ERCC1 is still required for gene targeting even when the ends of the targeting construct are homologous and a role in processing the heteroduplex intermediates generated as a result of stalled branch migration has been proposed (Niedernhofer et al., 2001).

In yeast the majority of DSBs are repaired by Rad52 mediated pathways (Paques and Haber, 1999). Rad52 binds ssDNA, promoting efficient annealing of homologous ssDNA but does not catalyse the initial invasion of the homologous dsDNA by the 3' ssDNA end. Interaction between Rad52 and XPF forms a stable Rad52-XPF-ERCC1 ternary complex (Motycka et al., 2004). This both stimulates XPF-ERCC1 endonuclease activity and attenuates the DNA strand annealing activity of Rad52. Thus, XPF-ERCC1 (and Rad1-Rad10 by implication) could be recruited by Rad52 for processing recombination intermediates during DSB repair and gene targeting.

1.6.3 Inter-strand cross-link repair

Chemicals including nitrogen mustards, cisplatin and psoralen cause covalent linking of the 2 DNA strands of the duplex. These inter-strand cross-links (ICLs) are cytotoxic as they prevent opening of the helix required for DNA replication and transcription. In some cases ICLs generate double strand breaks (DSBs) in dividing cells presumably from processing of the arrested replication fork (De Silva et al., 2000). In *E. coli* and *S. cerevisiae* repair of ICLs depends on a combination of both NER and recombination processes (Figure 1.8A) but less is known about mammalian mechanisms. XPF and ERCC1 defective cells are much more sensitive to DNA cross-linking agents than cells deficient in other NER factors (De Silva et al., 2000;

Hoy et al., 1985; Kaye et al., 1980) indicating a function of XPF-ERCC1 in ICL repair beyond a role in NER. Two models to explain this have been proposed. In the presence of RPA, XPF-ERCC1 acts as a 3' to 5' exonuclease to degrade DNA surrounding a psoralen-induced ICL on 1 strand to uncouple the ICL (Mu et al., 2000). A model incorporating this along with DSB formation was proposed (Figure 1.8C)(De Silva et al., 2000). Secondly XPF-ERCC1 alone can uncouple a psoralen-induced ICL by incising on both the 5' and 3' side of the cross-link (Kuraoka et al., 2000). This requires an unpaired region immediately 3' to ICL, expected to arise frequently from stalling of replication forks or elongating transcription complexes at the ICL, or by the actions of a DNA helicase or 5' to 3' exonuclease digestion of the bottom DNA strand (Figure 1.8B).

XPF-ERCC1 is required for the uncoupling step in the repair of psoralen- and nitrogen mustard-induced ICL (De Silva et al., 2000), but this does not account for the sensitivity of XPF and ERCC1 mutants to cisplatin-induced ICLs (De Silva et al., 2002). Uncoupling of cisplatin-induced ICL apparently requires a complete set of NER proteins, not just XPF-ERCC1. XPF-ERCC1 may have a dual role in both excision of the cross-link and homologous recombination (similar to that described in section 1.6.2), the relative importance of which is dependent on the type of ICL.

ERCC1 interacts with the mismatch repair MutS homologue, hMsh2 in HeLa cell extracts and both ERCC1 and Msh2 cooperate in resistance to cisplatin (Lan et al., 2004). Thus Msh2 may be a candidate for recruitment of XPF-ERCC1 to sites of ICL.

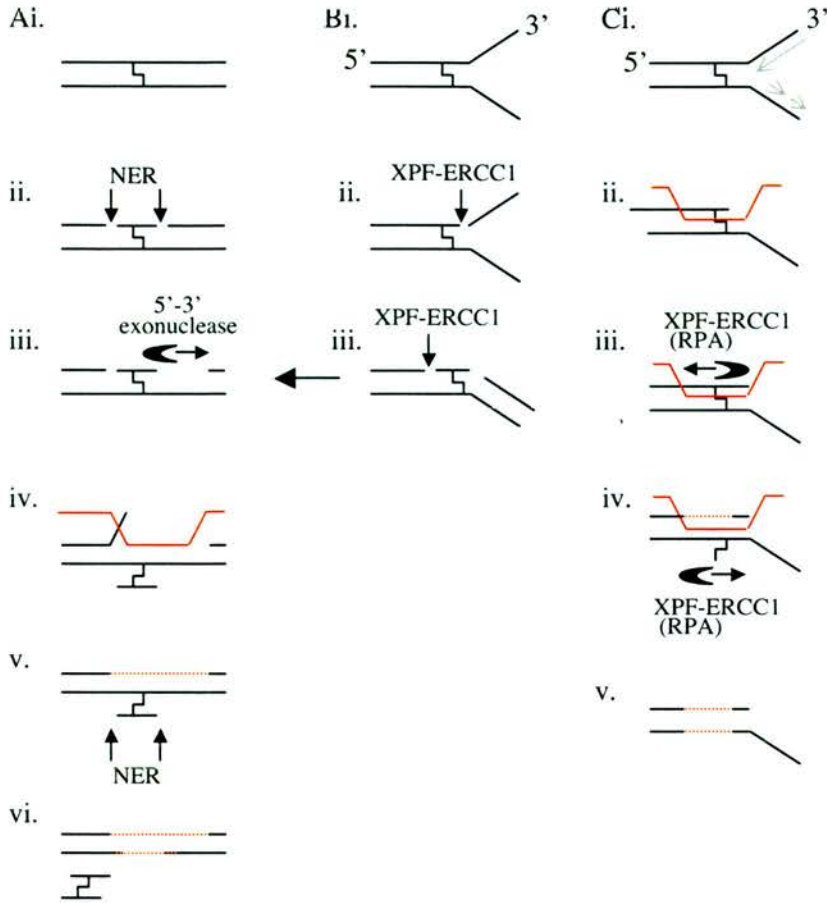


Figure 1.8. Models for repair of interstrand cross-links (ICLs). (A) NER-recombination model. NER machinery cleaves either side of the ICL on one strand (ii) and a gap 3' of the ICL created by a 5' to 3' exonuclease (iii). Strand exchange with a homologous donor (iv), resolution of heteroduplex and a second NER incision (v) results in removal of the ICL and resynthesis (vi). (B) Model proposed by Kuraoka *et al* 2000 to incorporate the unique incision properties of XPF-ERCC1. A Y structure is created near the ICL (i) and XPF-ERCC1 cleaves the 3' side (ii) followed by the 5' side (iii) of the cross-link on one strand. Exonuclease and recombination steps proceed as in (A). (C) Model proposed by De Silva *et al* 2000 to account for DSB formation and the exonuclease of XPF-ERCC1 in the presence of RPA. A DSB is formed at the site of a stalled replication fork (ii) and the DSB end is repaired by BIR (not shown). Strand invasion (ii) is followed by 3' to 5' exonuclease degradation by XPF-ERCC1 in the presence of RPA to uncouple the ICL (iii). A second excision and resynthesis event and resolution of recombination completes removal of the ICL.

1.6.4 XPF summary

All the activities of XPF-ERCC1 are dependent on its ability to cleave dsDNA 5' of a 3' ssDNA arm but the precise function is largely determined by the protein:protein interactions it makes.

1.7 Roles of Mus81*

1.7.1 Mus81 & repair of stalled replication forks

The replisome routinely encounters obstructions in the DNA it copies, leading to stalling of the replication fork. This may be a transient block by other protein complexes, for example DNA repair or transcription complexes, so it is reasonable to suspend DNA replication until the obstruction is out of the way. However stalling of the replication machinery at bulky DNA lesions would prevent access to DNA repair complexes so suspended synthesis is not an option. There are 2 avenues for rescue of stalled replication forks. Translesion synthesis by bypass polymerases enables replication past the lesion but is prone to errors (Vaisman, 2002). More favourable under normal conditions involves replication fork regression and/or recombination systems. The sensitivity of Mus81, Mms4 and Eme1 mutants to agents that stall replication (e.g. hydroxyurea, methylmethane sulfite (MMS) and camptothecin (CPT)) is consistent with a role in DNA repair (Haber and Heyer, 2001). Further, interactions of *S. pombe* Mus81 with the replication checkpoint kinase Cds1 (Boddy et al., 2000), *S. cerevisiae* Mus81 with the recombination repair Rad54 protein (Interthal and Heyer, 2000) and the recruitment of human Mus81 to sites of UV damage specifically in S phase (Fu and Xiao, 2003) are consistent with a role in rescuing stalled replication forks.

Much of our understanding of the replication fork regression and recombination comes from extensive genetic and biochemical characterisation of bacterial proteins (Cox, 2002; McGlynn and Lloyd, 2002).

A lesion in the leading strand can lead to uncoupling of the leading and lagging strand DNA polymerases and extension of the lagging strand beyond the leading strand (Figure 1.9a). *In vitro* studies show this is the favoured substrate for Mus81* (Bastin-Shanower et al., 2003; Doe et al., 2002; Kaliraman et al., 2001; Whitby et al., 2003) which cleaves the equivalent of the leading strand template to remove the leading strand arm creating a substrate for BIR. Alternatively the stalled replication fork can regress, i.e. unwind so the parental strands reanneal and the nascent strands are extruded and anneal together to form a 4-way junction or Holliday junction (Michel, 2000). In *E. coli* this regression is catalysed by RecG (McGlynn and Lloyd, 2002) but similar structures may arise from regression by positive supercoiling in the DNA occurring transiently ahead of RNA polymerase (Postow et al., 2001) or in RecA mediated strand exchange (Robu et al., 2001). This regression may clear the replication machinery away from the block giving repair machinery access to the lesion.

Replication fork regression is not all or nothing and different intermediate structures are formed depending on which template strand is blocked and whether the leading and lagging strand DNA polymerases have uncoupled (Sogo et al., 2002). *In vitro* Mus81* cleaves model replication forks partially regressed to a nicked 3-way structure (Figure 1.9i) or 3-way structures containing either a 3' or 5' tail (Figures 1.10b & 1.9j respectively), to collapse the fork (Whitby et al., 2003).

Following complete replication fork regression to Holliday junction, a number of potential outcomes exist shown in Figures 1.9 & 1.10. Initially there was much

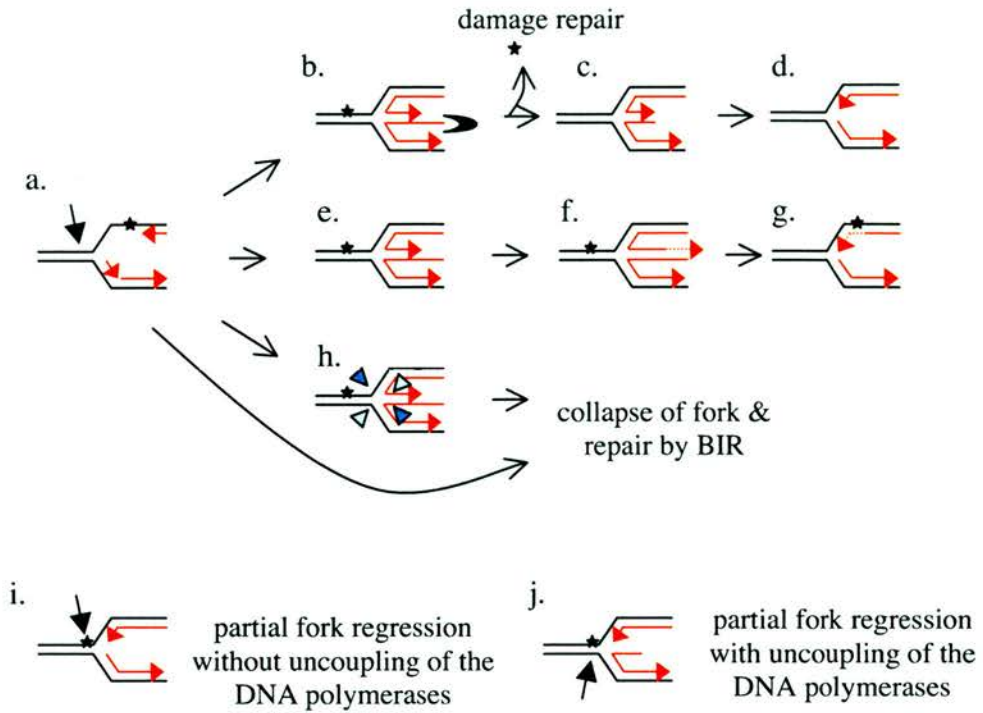


Figure 1.9. Processing of replication forks stalled at a leading strand lesion (star). Template and nascent strands are black and red respectively with arrowheads indicating the 3' end and direction of DNA synthesis. Cleavage sites for Mus81* and Holliday junction resolving enzymes are shown as black arrows and blue triangles respectively. A leading strand block can be cleaved directly by Mus81* and repaired by BIR (a). Alternatively the replication fork may partially regress with or without uncoupling of the DNA polymerases to form structures also cleaved by Mus81* (i & j). Complete regression forms a Holliday junction. Processing by a 5'-3' exonuclease digests the 5' nascent lagging strand tail (b) and repair of the lesion is followed by reverse branch migration, restoration of fork structure and replication restart (d). In template switching the nascent lagging strand tail is used as a template for extension of the 3' nascent leading strand (f) thereby bypassing the lesion on the leading strand template (g). Alternatively cleavage by a Holliday junction resolving enzyme collapses the fork and the DNA ends repaired by BIR (h).

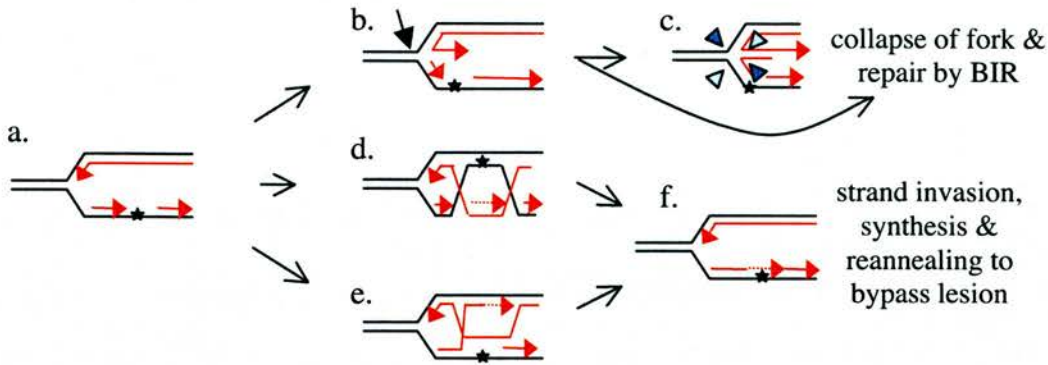


Figure 1.10. Processing of replication forks stalled at a lagging strand lesion (star). Template and nascent strands are black and red respectively with arrowheads indicating the 3' end and direction of DNA synthesis. Cleavage sites for Mus81* and Holliday junction resolving enzymes are shown as black arrows and blue triangles respectively. Lesions on the lagging strand template are generally less problematic since the discontinuous nature of lagging strand synthesis allows re-priming and continued replication downstream (a). However single strand gaps lead to activation of checkpoints that suspend replication. Similar to leading strand blockage, the replication fork can partially regress to extrude a 3' leading strand tail (b) creating a substrate for Mus81* which detaches the leading strand arm, and the resulting double strand break end is repaired by break-induced replication. The single strand gap must be filled in to prevent collapse of the fork on the subsequent round of DNA replication. In a pathway somewhat similar to template switching described previously, the single strand region of the lagging strand template containing the lesion invades the sister chromatid and the nascent leading strand acts as a template for nascent lagging strand gap filling (d). Alternatively the 3' nascent lagging strand end upstream of the gap invades the leading strand template initiating SDSA (e). In both cases displacement and reannealing reforms a fork structure in which the lagging strand lesion has been bypassed.

controversy over whether Mus81* was the eukaryal Holliday junction resolving enzyme (HJRE) (Haber and Heyer, 2001). However unlike classical HJRE such as bacterial RuvC & RusA, yeast mitochondrial Cce1 and archaeal Hjc & Hje (Lilley and White, 2001), Holliday junction cleavage by Mus81* was unsymmetrical and the ends unligatable (Boddy et al., 2001; Chen et al., 2001). In addition, Mus81* showed

a clear preference for fork and 3' flap structures over Holliday junctions *in vitro* (Ciccia et al., 2003; Doe et al., 2002; Kaliraman et al., 2001; Whitby et al., 2003).

The Holliday junction can reverse branch migrate back to a fork before or after repair/bypass of the lesion and this involves the RecQ helicase family (Wu and Hickson, 2002). Double mutations in *S. cerevisiae* and *S. pombe* Mus81 and their RecQ homologues (Sgs1 and Rqh respectively) are lethal (Boddy et al., 2000; Mullen et al., 2001) providing further evidence for Mus81* involvement in repair at stalled replication. Further, this lethality is rescued by preventing recombination suggesting Mus81* acts on recombination intermediates (Bastin-Shanower et al., 2003; Fabre et al., 2002).

To summarise Mus81* has a role in rescuing stalled replication forks before they fully regress to Holliday junctions and Mus81*-mediated collapse of forks creates substrates for BIR.

1.7.2 Mus81* & recombination

Much of the work looking at the role of Mus81* in recombination has focussed on meiotic recombination due to distinct meiotic phenotypes of *mus81* mutants in *S. cerevisiae* and *S. pombe*. 90-100 % of *S. cerevisiae mus81* mutant cells arrested in prophase as a result of unprocessed recombination intermediates triggering checkpoints (de los Santos et al., 2003; de los Santos et al., 2001). In the spores that did form viability was 40% and crossovers had decreased 2 fold. In contrast *S. pombe mus81* mutants showed over a hundred fold decrease in spore viability and 7-25 fold reduction in crossing over (Boddy et al., 2000; Osman et al., 2003). This difference in importance of *S. cerevisiae* Mus81-Mms4 and *S. pombe* Mus81-Eme1 can be explained by an additional Msh4/Msh5 dependent pathway present in *S. cerevisiae* but not in *S. pombe* so the latter has to rely more heavily on Mus81-Eme1 for meiotic

crossovers (de los Santos et al., 2003). In contrast to both yeast, *Mus81*^{-/-} mice are viable and fertile suggesting mammalian Mus81 is not essential for meiotic recombination (McPherson et al., 2004). Thus the dependence on Mus81* for meiotic recombination is species dependent.

S. solfataricus does not undergo meiosis. However, mitotic recombination mechanisms are similar, the major difference being that fewer crossovers occur (reviewed in (Paques and Haber, 1999)). The following discussion encompasses both mitotic and meiotic recombination but it is possible that the role of Mus81* in each is slightly different.

Three possible mechanisms were proposed to explain the role of Mus81* in recombination (reviewed in (Hollingsworth and Brill, 2004)). (1) As discussed earlier, Mus81* was originally proposed to resolve intact Holliday junctions, but this is no longer thought to be the case. (2) Mus81* cleaves 3' flaps generated by SDSA or strand displacement mediated crossing-over. During SDSA a fraction of 3' invading ends are over-replicated by the DNA polymerase such that following displacement and reannealing a 3' flap is created (Figure 1.7h), which Bastin-Shanower *et al* proposed to be the substrate for *S. cerevisiae* Mus81-Mms4. To explain the lethality of *S. cerevisiae mus81 sgs1* double mutants they also proposed an alternative pathway where Sgs1-Top3 displaced the downstream strand, reannealing the 3' flap and creating a 5' flap for cleavage by Rad27/Fen1 (Figure 1.7i). Consistent with this a *mus81 Rad27* double deletion is synthetically lethal (Tong et al., 2001). Similar to this *S. cerevisiae* Mus81-Mms4 was proposed to remove 3' flaps occurring in strand displacement-mediated crossing over (de los Santos et al., 2003; de los Santos et al., 2001). This is a hybrid of the DBSR and SDSA models that accounts for the presence of a meiotic specific double Holliday junction downstream

of heteroduplex DNA rather than flanking it as in the DSBR model (Allers and Lichten, 2001). (3) The final model involves Mus81* resolving D loops and half junctions before their maturation into double Holliday junctions, to give crossing over (Figure 1.11)(Osman et al., 2003). Initial cleavage in the strand complementary to invading strand would prevent SDSA from occurring and mean only crossovers could

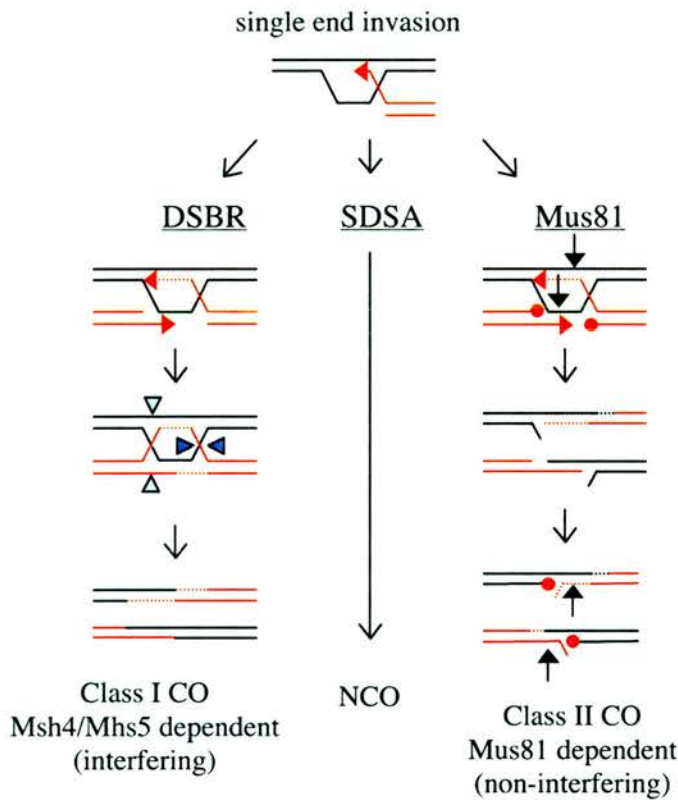


Figure 1.11. Mus81* and generation of crossovers (adapted from Hollingsworth & Brill 2004). Non-crossovers (NCO) arise from synthesis dependent strand annealing, SDSA (centre). There are 2 types of crossovers (CO). Class I show crossover interference and are dependent on Msh4 and Msh5 proteins that resolve the double Holliday junction (blue triangles) in a DSBR pathway. Class II crossovers do not show interference and are dependent on Mus81* which cleaves D loop and half junction structures (arrows) before they mature into double Holliday junctions. Cleavage results in a CO product containing 5' flaps and gaps. The flaps can be cleaved by Rad25/Fen1 or isomerize into 3' flaps for cleavage by Mus81*. Gap filling and ligation complete CO product formation. The 5' DNA ends directing Mus81* cleavage are indicated by red circles.

be generated. (However the presence of non-crossovers in *S. pombe* indicates that some D loops can escape Mus81* cleavage to follow SDSA pathway). Extension of the invading 3' strand and second end capture results in a nicked 4 way junction structure, also cleaved by Mus81* (Gaillard et al., 2003; Osman et al., 2003) which would result in linear products containing crossovers. Whether second end capture occurs before or after the first cut by Mus81 is unknown. This model is generally accepted to be the most consistent with the meiotic *in vivo* data (reviewed in (Hollingsworth and Brill, 2004)). The crossovers generated by Mus81* do not show crossover interference, a phenomenon whereby multiple crossovers on a chromosome are more widely spaced than predicted for a random distribution (de los Santos et al., 2003).

1.8 Role of euryarchaeal Hef

Unlike eukaryal XPF/Mus81 no *in vivo* activity data is available for Hef. The recombinant C-terminal nuclease domain cleaves Mus81*-like DNA structures (e.g. nicked duplex, 3' and 5' flaps, nicked Holliday junctions) but not XPF substrates such as splayed duplexes *in vitro* (Komori et al., 2002). The recombinant helicase domain showed the greatest ATPase activity with replication fork-like structures, although no helicase activity was detected. Thus helicase and endonuclease activity in the same protein suggests a very efficient mechanism for rescuing stalled replication forks: the RecQ-like helicase remodelling DNA into a substrate suitable for cleavage by the endonuclease. Given Hef is the only *P. furiosus* homologue of XPF/UvrC an additional role in NER cannot be ruled out.

CHAPTER 2: Materials & Methods

2.1 Cloning & Protein Expression

2.1.1 Cloning & vectors

The *Sulfolobus solfataricus* XPF gene was amplified from *S. solfataricus* strain P2 genomic DNA using proof-reading polymerase Pfu (Promega) and the following primers:

5' primer: 5'-CGTCGGATCCCCATGGTAATTAGAATTTATGCTGATGATAGGG-3'

3' primer: 5'-CCGGGGATCCGTCGACCTAAAGGAAATCAAATAAAGAGGTAG-3'

The PCR product was subcloned into the *NcoI/BamHI* site of vector pET19b (Novagen) for expression of the native protein, and the *BamHI/SalI* site of the N-terminal glutathione-S-transferase (GST) gene fusion vector pGEX-5X-3 (Amersham Pharmacia Biotech) for GST-XPF fusion expression.

The *Aeropyrum pernix* XPF gene was amplified from *Aeropyrum pernix* genomic DNA using Pfu polymerase (Promega) and the following primers:

5' primer: 5'-CGTCGGATCCCCATGGCTGGGGCGTGGATTTTGGAGG-3'

3' primer: 5'-CCGGGGATCCGTCGACTTATCCAGAACCAGCCTCCCC-3'

The PCR product was subcloned into the *NcoI/BamHI* site of vector pET28c (Novagen) for native protein expression.

The *S. solfataricus* Fen1 (XPG) gene was amplified from *S. solfataricus* P2 genomic DNA using KOD HiFi proofreading polymerase (Novagen) and the following primers:

5' primer: 5'-CGTCGGATCCCCATGGATTTAGCAGATTTAGTAAAAG-3'

3' primer: 5'-CCGGGGATCCGTCGACTTAAAACCATCTGTCCAATCCTGTTTGTC-3'

The PCR product was subcloned in to the *NcoI/BamHI* site of the protein expression vector pET28c.

The *S. solfataricus* XPB1 gene and XPB1 N-terminal truncated Δ 12 mutant were amplified from *S. solfataricus* strain P2 genomic DNA using Pfu polymerase (Promega) and the following primers:

*Sso*XPB1:

5' primer: 5'-CGTCGGATCCCCATGGCCTCGAGGACTTTCTATATCAAACAATGG-3'

3' primer: 5'-CCGGGGATCCGTCGACTTAATCTCTGTCGATATTTTCGGAGG-3'

N-terminal Δ 12 mutant:

5' primer:

5'-CGTCGGATCCCCATGGATGAGGATGACTTCAAGAGGTTACTATTATTCTCG-3'

3' primer: as for *Sso*XPB1

The PCR products were subcloned into the *Nco*I/*Bam*HI site of vector pET28c (Novagen) for native protein expression.

The *S. solfataricus* SSO1289 gene was amplified from *S. solfataricus* strain P2 genomic DNA using KOD polymerase (Novagen) and the following primers:

5' primer: 5'-GCTCGGATCCCCATGGGGAATTCTCTTTTGGTATTCACCTTGG-3'

3' primer: 5'-CCGGGGATCCTCGAGTTATGATTCTTTCCTATCATCTCC-3'

The PCR product was subcloned into the *Nco*I/*Bam*HI site of vector pET28c (Novagen) for native protein expression.

Dr Steve Bell (MRC Cancer Cell Unit, Cambridge) kindly donated vector constructs containing *S. solfataricus* PCNA subunits: pET33b-PCNA1, pET30a-PCNA2 and pET30a-PCNA3 for C-terminal poly-histidine tagged protein expression and pGEX-4T-3 PCNA1, 2 and 3 constructs for N-terminal GST-fusion PCNA expression (Dionne et al., 2003).

All plasmids were amplified in *E. coli* Top10 cells (Invitrogen) for miniprep (QIAprep[®] Miniprep Kit, QIAGEN) and the gene insert sequenced.

2.1.2 *Site directed mutagenesis*

The *SsoXPF* mutants (N-terminal 138 $\alpha\alpha$, active site D52A, and C-terminal truncated $\Delta 6$) were made using the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene®) with the pET19b-*SsoXPF* construct and the following primer pairs:

N-terminal 1-138 truncation (S139stop):

5'- CCAATTTGGTGAAAACAAGTAAAATCGAATTAGTTTACATAATAAGG-3'

5'- CCTTATTATGTAAACTAATTCGATTTTACTTGTTTTACCAAATTGG-3'

D52A:

5'-GGAAATCAGTAAATGCTCTAGTAAATTCAG-3'

5'- CTGAATTTACTAGAGCATTACTGATTTCC-3'

C terminal $\Delta 6$ truncation (S228stop):

5'- GTAAGAAAACCTACCTAATTATTTGATTTCC-3

5' GGAAATCAAATAATTAGGTAGTTTTCTTAC-3'

2.1.3 *Protein expression*

All proteins were expressed in BL21 Rosetta cells (Novagen). Cultures were grown in LB containing the relevant antibiotic for the plasmid (100 $\mu\text{g/ml}$ ampicillin or 35 $\mu\text{g/ml}$ kanamycin). Protein expression was induced with 0.2 mM IPTG when cultures reached an A_{600} of approximately 0.7, and grown for a further 3 hours. The cells were pelleted and frozen until required.

2.2 **Protein Purification**

2.2.1 *Purification of native recombinant SsoXPF & SsoFenI*

The bacterial pellet was thawed in ~30 ml buffer (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 100 mM NaCl, 1 mM benzamidine) and sonicated 4 x 2

minutes with cooling. The lysate was centrifuged at 48,000 x g for 20 minutes at 4 °C, and the supernatant heated to 70 °C to precipitate *E. coli* proteins before centrifugation for a further 20 minutes. The supernatant was filtered (Acrodisc® 0.45 µm syringe filter) and diluted 2-3 fold with buffer A (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT). This was applied to a 10 ml heparin column (Amersham Biosciences) equilibrated with buffer A and the bound cationic proteins eluted over a 120 ml linear gradient comprising 0-1000 mM NaCl. The fractions containing the recombinant protein, as analysed by SDS-PAGE of fractions corresponding to absorbance peaks, were pooled and concentrated to ~ 7 ml and loaded onto a HiLoad® 26/60 Superdex® 200 gel filtration column (Amersham Biosciences) equilibrated with buffer (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 150 mM NaCl). Fractions corresponding to the peak(s) were concentrated. The D52A and C-terminal Δ6 XPF mutants were purified as for wild type.

2.2.2 Purification of N-terminal 1-138 SsoXPF mutant

The bacterial pellet was thawed in ~30 ml buffer (20 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM DTT, 150 mM NaCl, 1 mM benzamidine) and sonicated, centrifuged, heated and filtered as wild type *SsoXPF*. This was applied to a 5 ml HiTrap™ Q Sepharose HP column (Amersham Biosciences) equilibrated with buffer (20 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM DTT). and the bound anionic proteins eluted over a 120 ml linear gradient comprising 0–1 M NaCl. The fractions containing N-terminal *SsoXPF* mutant were identified by SDS-PAGE, concentrated and run through gel filtration column as for wild type *SsoXPF*.

2.1.3 Purification of native recombinant ApeXPF

The bacterial pellet was lysed in ~30 ml buffer (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 500 mM NaCl, 1 mM benzamidine, 15 % glycerol, 0.1 % Triton X) and protein purification was as *SsoXPF* (2.2.1).

2.1.4 Purification of native recombinant SsoXPB1

The bacterial pellet was lysed in ~30 ml buffer (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 500 mM NaCl, 1 mM benzamidine, 0.4 mM AEBSF) by sonication. Purification was as *SsoXPF* with the following changes: 1) 1 mM benzamidine and 0.4 mM AEBSF were used in all the column buffers to minimize protein degradation; 2) a second gel filtration step through a Superose™ 6 HR 10/30 column (Amersham Pharmacia Biotech) was used to decrease the DNA in the preparation; 3) 500 mM NaCl was used in both gel filtration buffers as high concentrations of *SsoXPB1* were not stable in low salt.

2.1.5 Purification of His-tagged PCNA subunits and PCNA heterotrimer

The cell pellets were lysed in buffer (20 mM sodium phosphate pH 7.2, 500 mM NaCl), centrifuged, heated and filtered as *SsoXPF*. Cell lysates were each run through a 5 ml HiTrap™ Chelating HP column loaded with Ni²⁺ ions. The bound proteins were eluted over a 125 ml linear gradient comprising 0-500 mM imidazole. Fractions containing the PCNA subunit were identified by SDS-PAGE and pooled. Ammonium sulphate (516 g/litre) was added slowly to pooled fractions with gentle stirring at 4 °C, to precipitate proteins. This was centrifuged at 31,000 x g and the pellet re-suspended in 5-10 ml gel filtration buffer (30 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5 mM DTT, 200 mM NaCl) and loaded onto a HiLoad® 26/60 Superdex® 200 gel filtration column (Amersham Pharmacia Biotech) equilibrated with gel

filtration buffer. Fractions containing the PCNA subunit were identified by SDS-PAGE and pooled.

To form the PCNA heterotrimer approximately equal amounts of each subunit were rotated together at 4 °C for approximately 1 hour. The heterotrimer was separated from dimer and monomer by gel filtration as described above. The first elution peak (i.e. of highest molecular weight) was checked for equivalent amounts of each subunit by SDS-PAGE and concentrated.

2.1.6 Purification of GST-fusion proteins

The bacterial pellets were resuspended in ~30 ml buffer G/ 1 M NaCl (buffer G = 20 mM MES pH 6.5, 1 mM DTT, 15 % glycerol). This was sonicated for 4 x 2 minutes with cooling then centrifuged 48,000 x g for 20 minutes. The supernatant was retained and 1-2 ml glutathione agarose beads added and rotated at 4 °C for approximately 2 hours. The beads were washed in 4 x 10 ml buffer G/ 1 M NaCl with 10 minutes rotating at 4 °C and this was repeated another 3 times with buffer G/100 mM NaCl. Expression and purification of GST-fusion proteins was checked by SDS-PAGE.

2.1.7 Protein analysis & concentration determination

NuPAGE® Novex Bis-Tris-HCl gels (Invitrogen) were used according to manufacturers instructions for SDS-PAGE analysis. Protein concentrations were estimated from the absorbance of the protein solution at 280 nm (A_{280}) using the extinction coefficients obtained from ProtParam analysis of the amino acid composition (<http://ca.expasy.org/tools/protparam.html>).

2.3 Tryptic digestion of proteins

2.3.1 Tryptic digestion of SsoXPF

5 µg recombinant *SsoXPF* was incubated with doubling dilutions of trypsin from 0.2 µg to 6 ng (i.e. 1:25-1:800 w/w ratio trypsin:*SsoXPF*) and 2.5 mM CaCl₂ at 30 °C for 1½ hours and the reaction stopped by adding 50 mM AEBSF protease inhibitor. The digests were run on SDS-PAGE gel and a sample of the 1:25 digest analysed by MALDI-ToF mass spectrometry.

2.3.2 Tryptic digestion of SsoXPB1

Reactions were performed as for *SsoXPF* with 5-10 µg recombinant *SsoXPB1* and separated by SDS-PAGE. Proteins were blotted onto PVDF membrane in transfer buffer (10 mM CAPS pH 11, 10% methanol) and visualised by imidole black protein stain. Major digestion bands were cut out and N-terminal protein sequenced.

2.4 Glutaraldehyde cross-linking of proteins

A final concentration of 1% glutaraldehyde was added to 25 µg protein in 50 mM sodium phosphate buffer pH 7.6 (500µl) and incubated for 30 seconds at room temperature. 25 µl 2 M NaBH₄ was added to quench the reaction. Protein was TCA-DOC precipitated and analysed by SDS-PAGE.

2.5 GST-fusion interaction studies

2.5.1 GST-SsoXPF & GST-PCNA2 fusion affinity chromatography

1 ml glutathione agarose beads and 5 ml buffer G/ 100 mM NaCl were added to the bead-GST-SsoXPF or bead-GST-PCNA2 solution and loaded onto a disposable 5 ml polypropylene column (PIERCE). A control column with GST alone attached to beads was also set up and both columns washed with 4 x 10 ml buffer G/ 100 mM NaCl. To prepare soluble *S. solfataricus* extract, 2 g of *S. solfataricus* P2 biomass (supplied by the Centre for Extremophile Research, Porton Down, UK) per column was defrosted, suspended in buffer G/ 100 mM NaCl, sonicated for 4 x 2 minutes and centrifuged at 48,000 x g for 20 minutes. The supernatant was removed, filtered and passed through the columns. The columns were washed with 3 x 7 ml buffer G/ 100 mM NaCl to remove any unbound protein. A full column of buffer G/ 200 mM NaCl was added and the first 3 ml of the fraction collected. Columns were washed with 3 x 7 ml buffer G/ 200 mM NaCl and the fraction collection and washing steps repeated with buffer G/ 350 mM NaCl & 600 mM NaCl. Columns were stored in buffer G/ 1M NaCl. The fractions were TCA-DOC precipitated and analysed by SDS-PAGE. Bands present in GST-SsoXPF/GST-PCNA2 but absent in the GST control were cut out and identified by MALDI-ToF mass spectrometry.

2.5.2 Small scale GST fusion affinity assays

10 µl GST control beads or GST-SsoXPF/GST-PCNA subunit 1, 2 or 3 beads were rotated in 500 µl soluble *S. solfataricus* extract or 1 µg recombinant protein in buffer G/400 mM NaCl for between 2 hours and overnight. The beads were then washed with 4 x 500 µl buffer G/ 400 mM NaCl, heated in 2X LDS sample buffer (Invitrogen), and analysed by SDS-PAGE and Coomassie staining or Western

blotting. Primary antibodies raised against recombinant *SsoXPF*, recombinant *SsoXPB1* (Diagnostics Scotland) and recombinant PCNA subunits 1, 2 and 3 (Dionne et al., 2003) were used for immunological detection.

2.6 Isothermal Titration Calorimetry

2.6.1 *The VP-ITC micro calorimeter*

The VP-ITC Unit (MicroCal™) was used and is shown schematically in Figure 2.1. It consisted of two identical cells, a reference cell containing water or buffer that plays no part in the titration, and a sample cell containing the protein. These cells were enclosed in an adiabatic outer shell that eliminated heat exchange between the calorimeter and surroundings. The ligand was taken up in the syringe and precise amounts injected into the sample cell, with continuous mixing from the spinning modified end of the syringe. The temperature difference between the reference cell and sample cell was measured and calibrated to power units required to maintain temperature equilibrium. This was done electronically by administering a known quantity of power through a resistive heater element located on the sample cell. The heat released or absorbed was directly proportional to the amount of binding.

Control experiments were important to assess the background heat energy associated with dilution of protein and ligand into buffer. These should use exactly the same solutions used in the experiments and for completeness include:

1. injection of ligand into buffer
2. injection of buffer into protein solution
3. injection of buffer into buffer

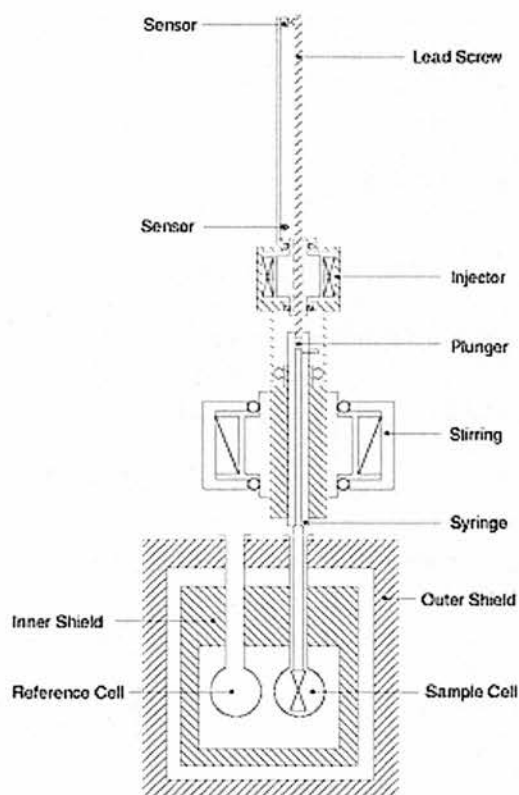


Figure 2.1. Schematic diagram of the VP-ITC Unit
(<http://www.microcalorimetry.com/?id=14>)

The heats/injection from these are removed from binding data as follows:

$$\text{heats from binding data} - (1) - (2) + (3)$$

However due to large protein requirements only control (1) was performed and subtracted from the binding data. Control (3), buffer into buffer, was not required since it is added to the binding data to prevent double subtraction of buffer-buffer dilution which is an integral part of both controls (1) and (2).

The heats of each individual injection corrected for dilutions were integrated with respect to time and plotted against the molar ratio of ligand to protein. This was fitted to an appropriate binding model enabling calculation of the enthalpy change

(ΔH), association and disassociation constants (K_A and K_D respectively) by ORIGIN software.

2.6.2 *SsoXPF-metal ion binding*

SsoXPF and D52A were dialysed overnight in buffer (50 mM Tris-HCl pH 7.4, 400 mM NaCl) and metal solutions prepared in the same buffer. Experiments were performed at either 35 °C or 55 °C and all solutions were degassed prior to each run (ThermoVac, MicroCal™). The sample cell contained 75 μ M *SsoXPF* or 69.5 μ M D52A (concentrations calculated for the monomeric proteins) or buffer (controls). The syringe contained the metal ion ligand (3 mM $MnCl_2$) that was titrated into the sample cell. An initial injection of 2 μ l was followed by 28 10 μ l injection of duration 20 seconds. Calorimetric data was analysed using MicroCal™ ORIGIN® software.

2.6.2 *SsoXPF-PCNA binding*

Proteins were dialysed overnight in buffer (50 mM Tris-HCl pH 7.4, 200 mM NaCl). Experiments were performed at 55 °C and all solutions were again degassed prior to each run. The protein in the sample cell was 5 μ M PCNA heterotrimer or 5 or 10 μ M PCNA subunit 1. The ligand was 100 μ M *SsoXPF* (concentration calculated for the dimer). *SsoXPF* was titrated into the PCNA or buffer (controls) with injections and analyses as above.

2.7 Catalytic & DNA binding assays

2.7.1 DNA substrates

The oligonucleotides used for substrates (MWG, QIAGEN or PHOENIX BioTechnologies) (Appendix 1) were 5' [³²P]-end-labelled and assembled into various structures by slow cooling from 85 °C to room temperature for 3 hours or overnight. Substrates were purified on a native 6 % acrylamide gel and electro-eluted from the gel as described in (White and Lilley, 1996) or eluted overnight into TE buffer (10 mM Tris pH 8, 1 mM EDTA) at 37 °C. DNA was ethanol precipitated from elutions, re-suspended in water and stored at -20 °C. To make the biotin-streptavidin conjugated substrates [biotin~dT] was substituted at positions 3 and 45 in the b50 oligo (QIAGEN). Substrates were formed as above and a 5x excess of streptavidin added prior to use in assays.

Size markers (G & A) were prepared from labelled substrates using standard protocols.

2.7.2 *SsoXPF* endonuclease assays

Nuclease reactions (10 µl) were performed in 30 mM HEPES pH 7.6, 5 % glycerol, 40 mM KCl, 8 mM MgCl₂, 0.1 mg/ml bovine serum albumin and 0.1 mg/ml calf thymus DNA. 80-100 nM DNA substrate was used per assay and where appropriate first incubated with 1 µM PCNA at room temperature for 2 minutes. 1 µM *SsoXPF* was added and the reaction incubated at room temperature for a further 2 minutes then at 55 °C for times depending on the substrate. Reactions were stopped by adding 90 µl stop buffer (10 mM Tris-HCl pH 8, 10 mM EDTA, 0.1 mg/ml calf thymus DNA) and the DNA was ethanol precipitated and re-suspended in formamide loading buffer. Samples were heated at 95 °C for 5 min, cooled on ice and loaded

onto a denaturing (7 M urea) 12 % polyacrylamide gel containing 1X TBE (89 mM Tris, 89 mM borate, 2 mM EDTA) and run at 50 °C and 90 watts for 1 hour. Gels were visualised by phosphorimaging (Fuji FLA5000) and/or autoradiography.

For kinetic rate measurements, reactions were set up in 30 mM HEPES pH 7.6, 5 % glycerol, 40 mM KCl, 1 mM EDTA, 0.1 mg/ml bovine serum albumin, and 0.1 mg/ml calf thymus DNA with 80 nM DNA substrate. DNA was incubated with 1 μ M PCNA at room temperature for 2 minutes. 1 μ M SsoXPF was added and reaction preheated to 55 °C for 2 minutes. Cleavage was initiated by adding MgCl₂ to a final concentration of 10 mM and mixing briefly. 5 μ l aliquots were taken at selected time points and added to chilled stop solution to terminate the reaction. DNA was ethanol precipitated and analysed as described above. Uncut and cut substrate was quantified (Image Gauge, Fuji) and the catalytic constant (k_{cat}) derived by linear regression. This value was multiplied by 60 to convert to min⁻¹.

2.7.3 SsoXPB1 electrophoretic mobility shift assays

Reactions (10 μ l) were performed in 30 mM HEPES pH 7.5, 50 mM NaCl, 5 % glycerol, 2 mM EDTA, and 0.1 mg/ml BSA. 10 nM DNA substrate was added to SsoXPB1 final dilutions of 1 μ M to 200 pM. Reactions were incubated at room temperature for 10 minutes, loading buffer (10 % Ficoll, 50 mM EDTA, 0.5 % SDS, 0.125 % bromophenol blue, 0.125 % xylene cyanol, 0.3 μ g/ml calf thymus DNA) added, and samples run on 6 % native acrylamide gel containing 1X TBE at 130 V for 3 hours. Gels were dried on a slab gel drier (Savant) and visualised by phosphorimaging (Fuji FLA5000). The unbound DNA was quantified (Image Gauge, Fuji) and the fraction bound calculated: $1 - (\text{unbound}/\text{total})$ and plotted against SsoXPB1 concentration (M). The curve was fitted with the following equation to

take into account that the concentration of DNA used (10 nM) was close to the derived dissociation constant.

$$y = \frac{((1+m1*M0+m1*1e-8)-\sqrt{((1+m1*M0+m1*1e-8)^2-(4*1e-8*m1^2*M0)})})}{(2*1e8*m1)};$$

$$m1=1e8$$

where $m1 = K_D(M)$ and $M0 = SsoXPB1$ concentration

2.7.4 *SsoXPB1* helicase assays

Reactions (10 μ l) were performed in 30 mM HEPES pH 7.6, 40 mM KCl, 1 mM DTT, 5 % glycerol, 0.1 mg/ml BSA. Dilutions of *SsoXPB1* (1 μ M to 10 nM) were incubated with 100 nM DNA with and without equimolar amounts of PCNA, and preheated to 55 °C for 2 minutes. 3 mM ATP was added and incubated for 15 minutes. Where 100 times excess of unlabelled DNA was used, this was added and incubated a further 10 minutes at 55 °C to trap the unwound DNA and prevent reannealing of the labelled oligo. Loading buffer was added and the samples run on a native acrylamide gel as described above.

**CHAPTER 3: An archaeal endonuclease dependent on
heterotrimeric PCNA**

3.1 Introduction

Eukaryal XPF (Rad1) is a structure specific endonuclease with a fundamental role in nucleotide excision repair but at the start of this project relatively little was known about its structural organisation and molecular mechanisms. The N-terminal two-thirds resembles an inactivated helicase (Gaillard and Wood, 2001; Sgouros et al., 1999). This is fused to a C-terminal domain containing the nuclease catalytic centre, as confirmed by extensive mutagenesis (Enzlin and Scharer, 2002), and a double HhH DNA binding domain (Figure 1.4). Together with partner protein ERCC1 (which appears distantly related to the C-terminal XPF nuclease domain), it cleaves double stranded DNA 5' of a junction with single stranded DNA. The role of ERCC1 is uncertain but appears to be as a bridging factor between XPF and XPA-RPA (Saijo et al., 1996). In archaea 2 XPF forms exist (Figure 1.4). Euryarchaeal XPF (Hef) is present as the helicase-nuclease fusion similar to eukarya (Komori et al., 2002) whereas crenarchaea (such as *S. solfataricus*) encode only the C-terminal nuclease domain without the N-terminal fusion. There is no homologue of ERCC1 in archaea. Studying and understanding more about this basic XPF nuclease unit as represented by *SsoXPF* may help our understanding of the more complex situation in eukarya.

In this chapter initial characterisation of *SsoXPF* is described. Firstly structural properties were identified and compared to the subsequent crystal structure of the euryarchaeal XPF homologue Hef (Nishino et al., 2003). Basic requirements for *SsoXPF* nuclease activity were also identified, including a requirement for the sliding clamp protein PCNA.

3.2 Expression & Purification of *S. solfataricus* XPF

The gene encoding *S. solfataricus* XPF (protein accession number Q9UXB7) was amplified by PCR, sequenced, and subcloned into vector pET19b for expression of native protein in *E. coli*. This recombinant *Sso*XPF protein was purified by a combination of heat treatment to precipitate *E. coli* proteins followed by 2 column steps to produce pure protein (Figure 3.1). Elution from the gel filtration column indicated dimeric (major) and tetrameric forms (Figure 3.1B).

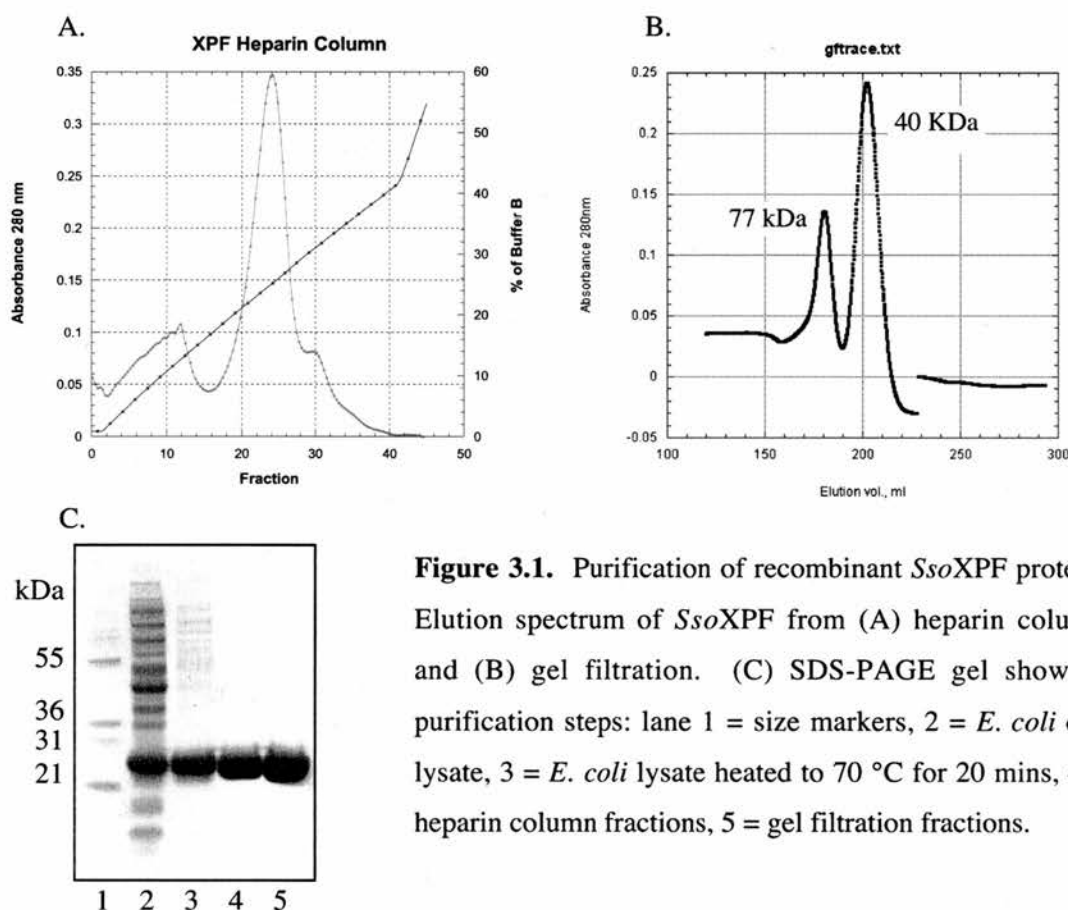


Figure 3.1. Purification of recombinant *Sso*XPF protein. Elution spectrum of *Sso*XPF from (A) heparin column and (B) gel filtration. (C) SDS-PAGE gel showing purification steps: lane 1 = size markers, 2 = *E. coli* cell lysate, 3 = *E. coli* lysate heated to 70 °C for 20 mins, 4 = heparin column fractions, 5 = gel filtration fractions.

An *Sso*XPF dimer band was sometimes seen by SDS-PAGE analysis of recombinant protein (Figure 3.2A). Further, *S. solfataricus* cell extract analysed by

Western blot using polyclonal antibodies generated against recombinant *Sso*XPF showed both *Sso*XPF monomer and dimer bands (Figure 3.2B, standard). The dimer band disappeared when the *S. solfataricus* extract was electrophoresed under reducing conditions prior to the Western blot detection of *Sso*XPF (Figure 3.2B, reducing).

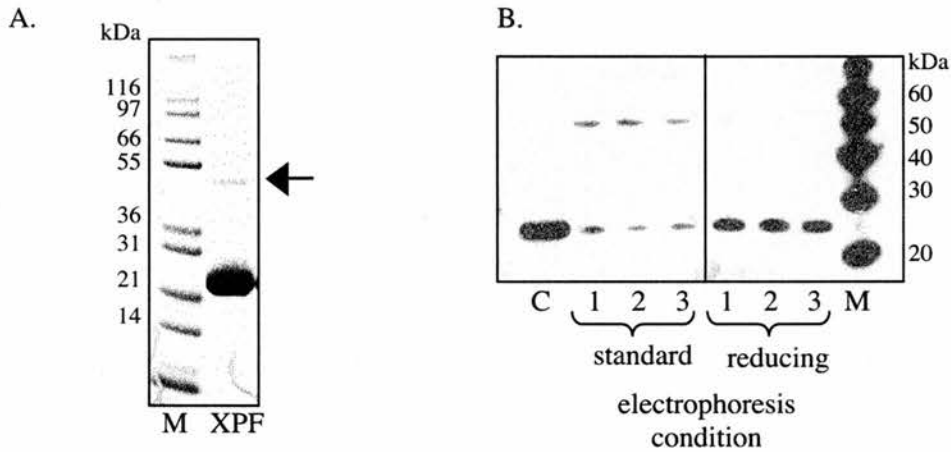


Figure 3.2. *Sso*XPF dimers. (A) SDS-PAGE gel of recombinant *Sso*XPF protein. Arrow shows the position of the band corresponding to *Sso*XPF dimer. M = size markers. (B) Western blot analysis of *S. solfataricus* cell extracts from 3 different points during the exponential growth phase ((1) beginning, (2) middle and (3) end), electrophoresed under standard and reducing conditions as described in the NuPAGE[®] instructions. Control = 64 ng recombinant *Sso*XPF. M = size markers.

3.3 Mapping *Sso*XPF domains

Recombinant *Sso*XPF was incubated with dilutions of porcine trypsin and subsequent SDS-PAGE showed two major digestion products of approximately 16 and 8 kDa (Figure 3.3A). The sites of cleavage were mapped using MALDI-ToF mass spectrometric analysis of a sample of the 1:25 trypsin:XPF digest. The spectrum shows 2 groups of peaks (Figure 3.3B), each group corresponded to a major digestion product. The molecular weight of these peptides enabled them to be mapped onto the full *Sso*XPF sequence (Figure 3.3C) given that trypsin cleaved at the

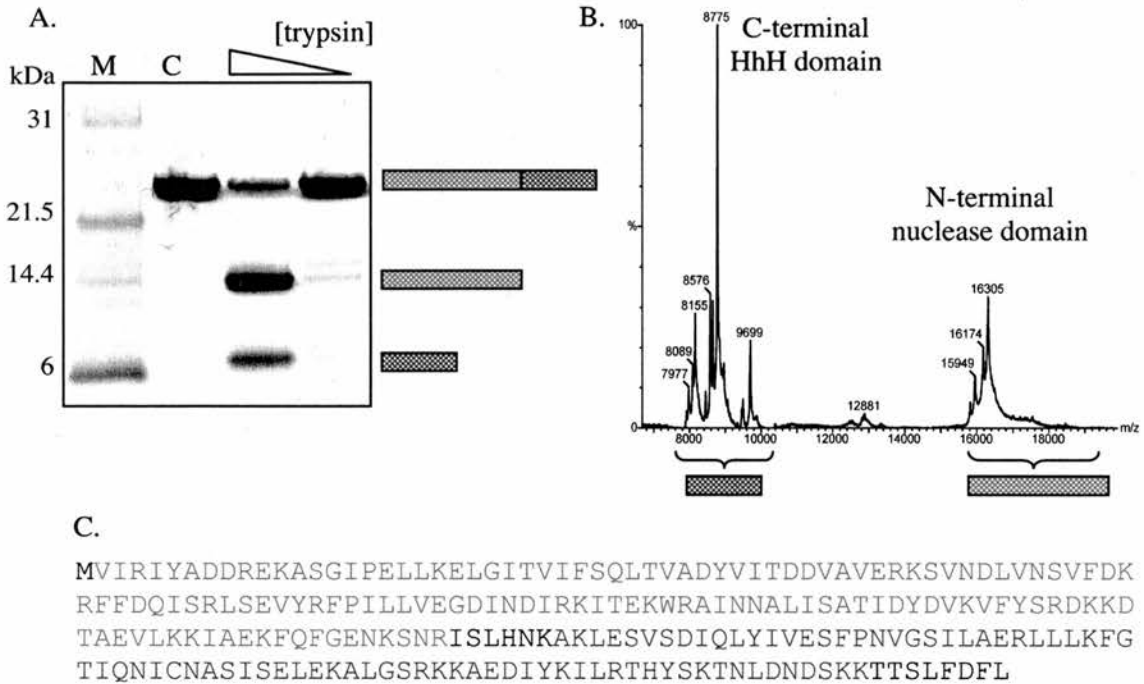


Figure 3.3. *SsoXPF* consists of 2 domains. (A) SDS-PAGE gel showing partial digestion of 2 μg *SsoXPF* at 1:25 and 1:50 ratios of porcine trypsin:*SsoXPF*. M = size markers. C = undigested *SsoXPF* control. (B) MALDI-ToF mass spectrum of a sample of the 1:25 trypsin:*SsoXPF* digest showing the 2 groups of peaks corresponding to the 2 digestion products. (C) Mapping of the peptides identified by the MALDI-ToF mass spectrometry onto the full *SsoXPF* protein sequence. Trypsin cleaved at 4 clustered lysine and arginine residues between positions 138-149, effectively cleaving *SsoXPF* into N-terminal (amino acids 1-138) and C-terminal (amino acids 150-233) domains shown in red and blue respectively.

C-terminal side of lysine and arginine residues. Cleavage was identified at a cluster of 4 basic residues between positions 138 and 149. This indicated a flexible, exposed linker region between 2 domains of *SsoXPF*. The larger N-terminal domain (amino acids 1-138) contained the nuclease motif and was stable when expressed alone (Figure 3.4B lane 2) suggesting the nuclease functionality of *SsoXPF* exists in an autonomous, folded domain. The C-terminal domain (amino acids 150-233) contains a predicted double Helix-hairpin-Helix (HhH) motif that is also present in human XPF

and ERCC1 (de Laat et al., 1998b; Interthal and Heyer, 2000). Mus81 shows a slightly different organisation of HhH domains with 1 N-terminal and 1 C-terminal of the nuclease domain (Fu and Xiao, 2003; Interthal and Heyer, 2000). This HhH motif suggests a role for non-specific double stranded DNA binding (Doherty et al., 1996). Indeed the *SsoXPF* double HhH domain binds DNA with approximately the same affinity as the full length protein (Jana Rudolf, unpublished observation), indicating this double HhH domain of *SsoXPF* confers the DNA binding specificity with no or little contribution by the C-terminal nuclease domain. A very similar domain organisation for the homologous C-terminal third of *P. furiosus* Hef has been proposed from limited proteolysis (Nishino et al., 2003).

Parallels can also be drawn with the type II restriction enzyme *FokI* which is a member of the same nuclease superfamily as XPF and Mus81. Unlike typical type II restriction enzymes that have recognition and catalytic functions integrated into a single protein domain, *FokI* has 2 independent domains similar to *SsoXPF*: an N-terminal DNA recognition domain containing 3 HhH motifs and a C-terminal endonuclease domain, separated by a flexible linker (Li et al., 1992).

3.4 *SsoXPF* forms a homodimer

Protein cross linking using glutaraldehyde (Figure 3.4 lanes 1&3) together with the elution volume of recombinant *SsoXPF* from gel filtration (Figure 3.1B) and the dimers seen on SDS-PAGE (Figure 3.2) indicated *SsoXPF* formed a homodimer. Glutaraldehyde cross-linking also showed the N-terminal nuclease domain to dimerise independently (Figure 3.4 lanes 2&4). The C-terminal HhH domain also formed a dimer under the same cross-linking conditions (Jana Rudolf, unpublished

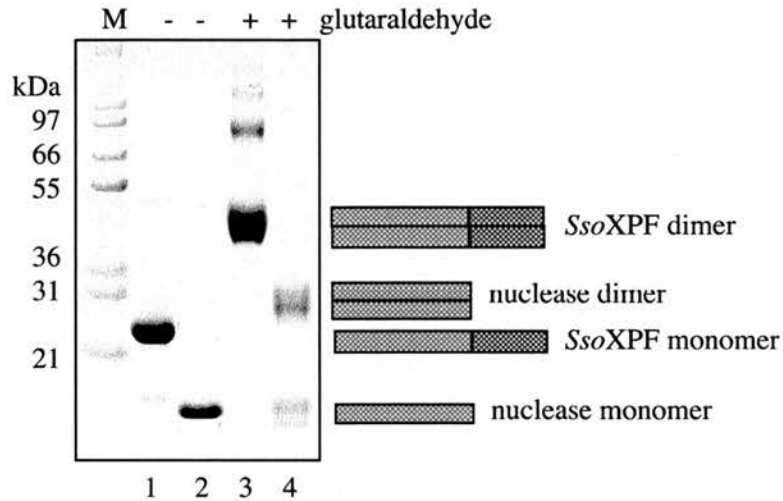


Figure 3.4. Cross-linking of full length *SsoXPF* and nuclease domain. Protein was incubated with 1 % glutaraldehyde for 30 seconds, quenched with NaBH_4 and analysed by SDS PAGE. Lanes 1 = *SsoXPF*; 2 = nuclease domain; 3 = *SsoXPF* + glutaraldehyde and 4 = nuclease domain + glutaraldehyde. M = size markers.

observation). Thus *SsoXPF* appeared to form a homodimer with 2 dimer interfaces, one between each domain. Similarly, the homologous C-terminal third of euryarchaeal Hef also formed a homodimer with 2 independent dimer interfaces and complete dimerisation was required for activity (Nishino et al., 2003). In contrast *FokI* dimerisation is mediated solely by the C-terminal nuclease domain (Wah et al., 1998). This homodimerisation in archaeal XPF orthologues probably reflects the requirement for 2 double HhH domains for DNA interaction rather than 2 active sites since only 1 DNA strand is cut therefore presumably only 1 active nuclease subunit is required (Nishino et al., 2003). Eukaryal XPF and Mus81 form heterodimers with partner proteins that lack nuclease domains but contain HhH domains, consistent with this model.

3.5 Metal ion binding

Isothermal titration calorimetry (ITC) is a versatile method for determining affinity, stoichiometry and thermodynamic parameters of ligand binding in solution and has been successfully used to study metal ion binding by several endonucleases in solution (Hadden et al., 2002; Jose et al., 1999; Kvaratskhelia et al., 1999). The VP-ITC Unit (MicroCal™) was used to investigate metal ion binding by *Sso*XPF and is described in materials and methods (section 2.6).

Manganese ions were used because magnesium and calcium ions gave little or no signals. Manganese ions show a less stringent requirement in their coordination, which often increases its affinity for metal ion binding sites (Enzlin and Scharer, 2002) explaining the stronger signal. Precise amounts of manganese ions were injected into the *Sso*XPF solution in the calorimeter sample cell. Binding was exothermic, giving a good signal at 35 °C (Figure 3.5A) which was enhanced at 55 °C (Figure 3.5B). The heat released was directly proportional to the amount of binding: as *Sso*XPF became more saturated with manganese ion ligand on subsequent injections, the heat signal diminished until all available *Sso*XPF was in complex with manganese and only the background heat of dilution remained (Figure 3.5, top panels).

These results best fit a model where 1 manganese ion binds per *Sso*XPF monomer with a similar dissociation constant, K_D , of about 85 μ M at both temperatures (Table 3.1). Other models with 2 binding sites per monomer fitted the data very poorly and when the number of binding sites, n , was floated an endpoint was not reached on fitting.

The *P. furiosus* Hef nuclease domain crystal structure showed 1 manganese or calcium ion bound the active site of each Hef monomer (Nishino et al., 2003),

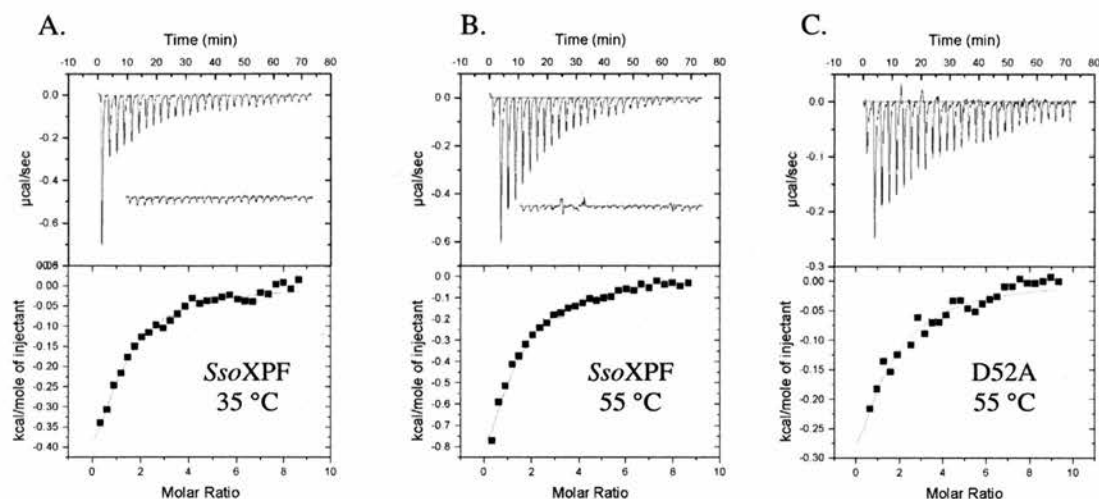


Figure 3.5. Isothermal titration calorimetry (ITC) of *SsoXPF* binding manganese ions. Upper panels show raw data for sequential 10 μl injections of manganese chloride into *SsoXPF* solutions. Lower panels show integrated heat data with a theoretical fit to 1 manganese ion binding per *SsoXPF* monomer. (A) *SsoXPF* binding manganese ions at 35 $^{\circ}\text{C}$, corrected for manganese ion dilution effects (inset). (B) *SsoXPF* binding manganese ions at 55 $^{\circ}\text{C}$ corrected for manganese ion dilution effect (inset). (C) D52A active site mutant of *SsoXPF* binding manganese ions at 55 $^{\circ}\text{C}$ corrected for dilution effect as (B). 35 $^{\circ}\text{C}$ gave a relatively weak signal so subsequent runs were performed at 55 $^{\circ}\text{C}$, in triplicate for *SsoXPF* and duplicate for D52A.

	K_A $\times 1000 \text{ M}^{-1}$	K_D μM	ΔH cal/mol	ΔG Kcal/mol	$T\Delta S$ Kcal/mol	ΔS cal/deg/mol	χ^2
XPF 35 $^{\circ}\text{C}$	1.14 ± 0.10	88 ± 7.7	-691 ± 23	-5.72	5.03	16.3	140
XPF 55 $^{\circ}\text{C}$	1.21 ± 0.12	84 ± 8.1	-1710 ± 77	-6.12 ± 0.06	4.41 ± 0.13	13.4 ± 0.4	971 ± 362
D52A 55 $^{\circ}\text{C}$	0.77 ± 0.04	131 ± 7.5	-838 ± 11	-5.83 ± 0.04	4.99 ± 0.03	15.2 ± 0.1	184 ± 20

Table 3.1 Thermodynamic parameters for *SsoXPF* binding manganese ions as calculated from the ITC data (see Appendix 2 for calculations). Standard errors are shown where available.

agreeing with the data here. However other nucleases in the same superfamily as *SsoXPF* have been shown to bind two metal ions. Whereas the yeast mitochondrial

Holliday junction resolving enzyme CceI was reported to bind 2 magnesium ions with the same affinity (Kvaratskhelia et al., 1999) T7 endonuclease I appeared to bind 2 manganese ions at significantly different affinities (Hadden et al., 2002). For the restriction enzyme *PvuII* the affinity of the 2 sites was dependent on the type of metal ion. For calcium there was a difference of about 1 order of magnitude in the binding affinities of the 2 sites whereas with manganese the affinity was about the same for each site.

Some proteins bind a second metal ion following DNA binding, for example the type II restriction enzyme *EcoRV* only generates a high affinity metal ion binding site the presence of specific DNA (Winkler et al., 1993). A 2 metal ion binding model cannot be ruled out for *SsoXPF*. No structural information is available for XPF with DNA bound. ITC is not always suitable for measuring binding especially when multiple species are involved. Alternative techniques such as electron paramagnetic resonance (EPR) could be used to look at *SsoXPF* metal ion binding in the presence of DNA. EPR detects changes in the electronic environment and rotational differences to give information about changes in structure. Determining the number of metal ions bound by *SsoXPF* could help in elucidation of the DNA cleavage mechanism.

Mutation of the *SsoXPF* active site aspartate 52 to alanine (D52A) gave a 50 % decrease in manganese ion binding affinity (Figure 3.5C; Table 3.1). Binding was not completely abolished because this conserved aspartate coordinates the manganese ion indirectly through a water molecule (Figure 1.5)(Nishino et al., 2003), and the other residues directly binding to manganese are not affected.

From the titration data, other thermodynamic parameters for *SsoXPF* manganese ion binding can be calculated (Table 3.1; Appendix 2). Manganese

binding to T7 endonuclease I has similar enthalpy and entropy changes ($\Delta H = -557.4 \pm 10.8$ cal/mol; $\Delta S = 12.93$ cal/deg/mol) (Hadden et al., 2002). In contrast the enthalpy change for Cce1 binding manganese was much greater at -4.5 kcal/mol.

3.6 Identifying *Sso*XPF protein partners

Eukaryal XPF & Mus81 form heterodimers with ERCC1 and Mms4/Eme1, respectively. Recombinant XPF-ERCC1 and Mus81-Mms4/Eme1 (Mus81*) cleave artificial DNA substrates *in vitro* (Bastin-Shanower et al., 2003; de Laat et al., 1998a; Kaliraman et al., 2001; Whitby et al., 2003). *P. furiosus* Hef forms a homodimer and is also active *in vitro* (Komori et al., 2002). However no nuclease activity was observed with recombinant *Sso*XPF homodimer under similar conditions (Figure 3.11), so perhaps another protein factor or factors were required for nuclease activity. To look for proteins interacting with *Sso*XPF, the glutathione-S-transferase (GST)-*Sso*XPF fusion protein was made and used for affinity chromatography with soluble *S. solfataricus* cell extract.

The *Sso*XPF PCR product was subcloned into the GST-fusion vector pGEX-5x-3 and expressed with N-terminal GST in *E. coli*. The GST-*Sso*XPF protein was purified from *E. coli* cell lysate by adding glutathione agarose beads, which bound the GST-*Sso*XPF. Thorough washing resulted in pure GST-*Sso*XPF immobilised on beads (Figure 3.6A). These beads were loaded into a column and *S. solfataricus* cell extract passed through. Proteins interacting with GST-*Sso*XPF bound to the column and were eluted by salt washes. After SDS-PAGE, 3 close bands of approximately 30 kDa and about equal intensity were apparent (Figure 3.6B). Each band was excised

from the gel and the proteins identified by in-gel tryptic digest and MALDI-ToF mass spectrometry as the 3 subunits of PCNA.

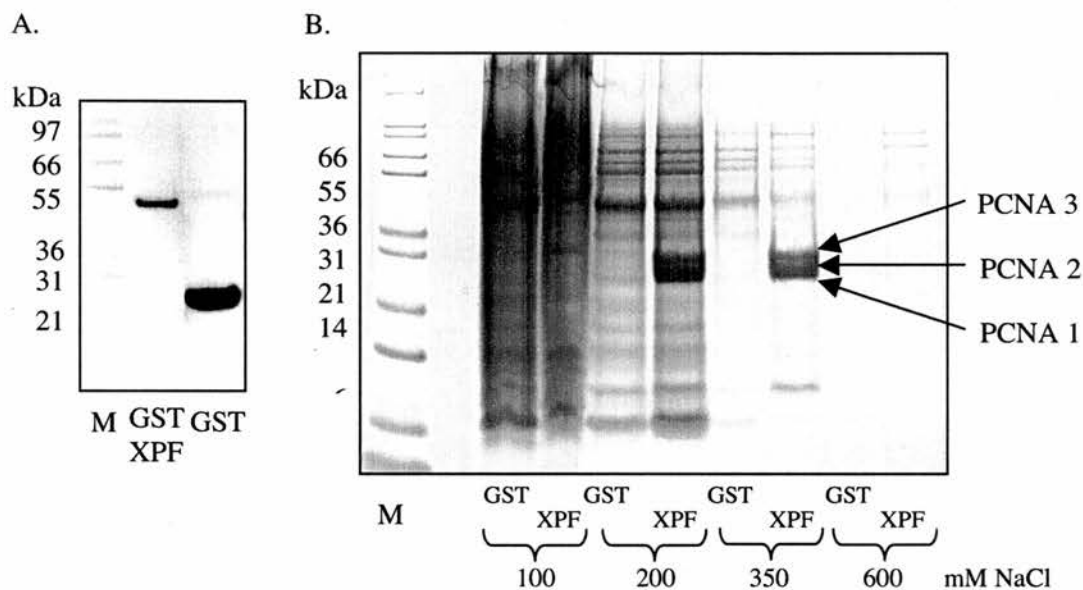


Figure 3.6. Identification of *SsoXPF* interaction with PCNA. (A) SDS-PAGE gel of purified GST-*SsoXPF* fusion and GST control protein. (B) SDS-PAGE gel of GST-*SsoXPF* affinity chromatography. *S. solfataricus* cell extract was passed over columns containing either the GST control or GST-*SsoXPF* fusion protein. Interacting proteins were eluted by a series of salt washes. The 3 prominent bands indicated were cut out and identified as the 3 subunits of PCNA by in-gel tryptic digest and MALDI-ToF mass spectrometry.

3.7 PCNA background

Proliferating cell nuclear antigen (PCNA) is a ring shaped protein that has no endogenous enzymatic activity itself, rather it encircles DNA forming a sliding clamp that mediates the interaction of other proteins with DNA in a non-specific way. There are PCNA homologues in eukarya, archaea, bacteriophage and some viruses. However, whereas eukarya and euryarchaea have homotrimeric PCNA (Krishna et al., 1994; Matsumiya et al., 2001; Schurtenberger et al., 1998) there is strong evidence for a heterotrimeric form in crenarchaea (Dionne et al., 2003). The beta-subunit of *E. coli*

DNA polymerase has a similar three-dimensional structure but is a homodimer of an unrelated protein (Kong et al., 1992).

PCNA-interacting proteins are involved in a variety of cellular processes including DNA replication, DNA repair, chromatin remodelling and cell cycle (Maga and Hubscher, 2003; Tsurimoto, 1999; Warbrick, 2000). Given the variety of proteins interacting with PCNA, it has been proposed that their coordinated binding provides a regulatory mechanism to coordinate aspects of DNA metabolism. PCNA interacts with several DNA polymerases in eukarya (e.g. DNA Pol δ (Zhou et al., 1997)) and archaea (*P. furiosus* DNA Pol I and II (Cann et al., 1999)) to increase processivity and fidelity of DNA replication. The β subunit of *E. coli* DNA polymerase has a similar processivity function in *E. coli* replication (Stukenberg et al., 1991). PCNA is required in the resynthesis step of nucleotide excision repair (NER) (Aboussekhra et al., 1995). In addition PCNA appears to have roles in DNA repair pathways before resynthesis. For examples, its orientation discriminates parental and newly synthesised DNA during mismatch repair (Umar et al., 1996). Interaction with the 3' NER endonuclease XPG in eukarya is proposed to promote turnover of nuclease activity and link excision and resynthesis (Gary et al., 1997; Nichols and Sancar, 1992). Flap endonuclease 1 (Fen1) is a homologue of XPG involved in lagging strand synthesis (Bambara et al., 1997) and base excision repair (Kim et al., 1998), and interaction with PCNA enhances Fen1 DNA binding stability allowing increased cleavage efficiency (Chapados et al., 2004; Tom et al., 2000).

3.8 Confirming *Sso*XPF interaction with PCNA

To confirm *Sso*XPF interacted with PCNA smaller scale affinity chromatography experiments were done. The GST-*Sso*XPF beads were mixed with *S. solfataricus* cell extract, washed with buffer to remove non-interacting proteins and the remaining components separated by SDS-PAGE. Polyclonal antibodies specific for each PCNA subunit were used for Western blot analysis. *Sso*XPF pulled out all 3 PCNA subunits from the *S. solfataricus* cell extract (Figure 3.7). These findings are consistent with *Sso*PCNA forming a heterotrimer rather than 3 different homotrimers. In *Aeropyrum pernix* (Daimon et al., 2002) and other crenarchaea (White, 2003) 3 PCNA genes have also been identified further indicating a heterotrimeric PCNA in crenarchaea is plausible.

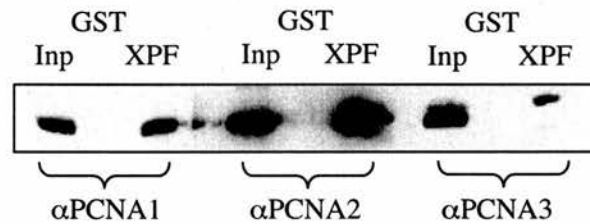


Figure 3.7. Confirmation of the *Sso*XPF interaction with PCNA. GST-*Sso*XPF (XPF) or GST control beads were mixed with *S. solfataricus* cell extract, washed with buffer to remove non-interacting proteins. The beads were analysed by SDS-PAGE and Western blot using antibodies specific for the individual PCNA subunits. Inp = 5 % *S. solfataricus* cell extract used.

Given *Sso*PCNA appears to be heterotrimeric, this permits the possibility of subunit-specific interactions. GST-PCNA subunit fusions were made and immobilised onto glutathione agarose beads (Figure 3.8A) (DNA constructs kindly donated by Steve Bell, MRC Cancer Cell Unit, Cambridge). The beads were mixed with *S. solfataricus* cell lysate as above, but polyclonal antibodies specific for

SsoXPF were used in Western blot analysis. *SsoXPF* appears to interact with PCNA subunits 1 and 3 but not appreciably with 2 (Figure 3.8B). The PCNA-*SsoXPF* interaction is investigated further in chapter 5.

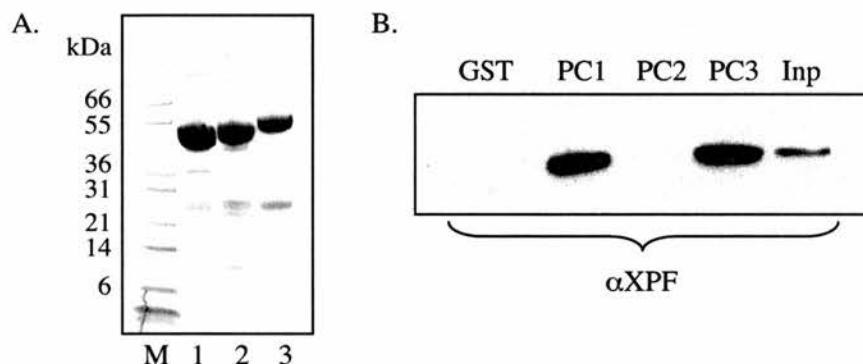


Figure 3.8. PCNA subunit specific interaction with XPF. (A) SDS-PAGE gel of purified GST-PCNA subunit fusion protein immobilised on glutathione agarose beads. Lanes 1-3 are GST-PCNA subunit 1, GST-PCNA subunit 2 and GST-PCNA subunit 3 respectively. (B) Western blot for XPF showed GST-PCNA1 and GST-PCNA3 pull-down XPF from *S. solfataricus* extract. GST-PCNA2 did not appear to interact with *SsoXPF*. Inp = 5% of *S. solfataricus* cell extract used.

3.9 *SsoXPF* has a conserved PCNA interaction motif

In eukarya PCNA interacts with a wide range of proteins involved in DNA replication, DNA repair, chromatin remodelling and cell cycle regulation (Maga and Hubscher, 2003; Tsurimoto, 1999; Warbrick, 2000). The majority of these proteins have a conserved binding motif usually located at the extreme N- or C-terminus and consisting of Qxx(h)xx(a)(a) where (h) is a moderately hydrophobic residue such as leucine, isoleucine or methionine and (a) is an aromatic residue such as phenylalanine or tyrosine (Warbrick, 1998). This was termed the PCNA interacting peptide or PIP. A similar consensus β subunit interaction motif of QL(S/D)LF exists in bacterial

DNA polymerase and the MutS1 mismatch repair protein family (Dalrymple et al., 2001). A second PCNA interacting motif termed the KA box was identified in eukarya from a random peptide library screened against PCNA (Xu et al., 2001). The consensus sequence is KA(A/L/I)(A/L/Q)xx(L/V) and was identified in DNA polymerase δ in addition to the PIP and shown to functionally interact with PCNA.

The C-terminal 6 amino acids of *Sso*XPF resembles a potential PIP PCNA-interacting motif which appears conserved in other archaeal DNA replication and repair proteins either shown to interact with or could potentially interact with PCNA given their homology to eukaryal PCNA-interacting proteins (Figure 3.9). To confirm this was the PCNA interaction motif a mutant of *Sso*XPF lacking these C-terminal 6 amino acids ($\Delta 6$) was made by site directed mutagenesis of pET19b-XPF template to insert a stop codon after serine 228 (Figure 3.10A). The GST-PCNA subunit beads were mixed with either full length recombinant *Sso*XPF or the $\Delta 6$ mutant, washed to remove unbound protein and Western blotted for *Sso*XPF. The $\Delta 6$ mutant did not interact with PCNA (Figure 3.10B).

3.10 *Sso*XPF is a PCNA-dependent, structure specific nuclease

As mentioned earlier, initial attempts to detect *Sso*XPF nuclease activity were unsuccessful. However addition of *Sso*PCNA heterotrimer to nuclease assays appeared to activate *Sso*XPF nuclease activity (Figure 3.11A). The *Sso*XPF mutant lacking the C-terminal 6 amino acids that interact with PCNA ($\Delta 6$) was not active, confirming the dependence of *Sso*XPF nuclease activity on an interaction with PCNA. Cleavage products were observed for dsDNA adjacent to 3' single strand flaps or bubble. No cleavage was detected for ds- or ssDNA (Figure 3.11C), nor for flap

substrates of opposite polarity, indicating structure specificity for dsDNA 5' of a ssDNA arm.

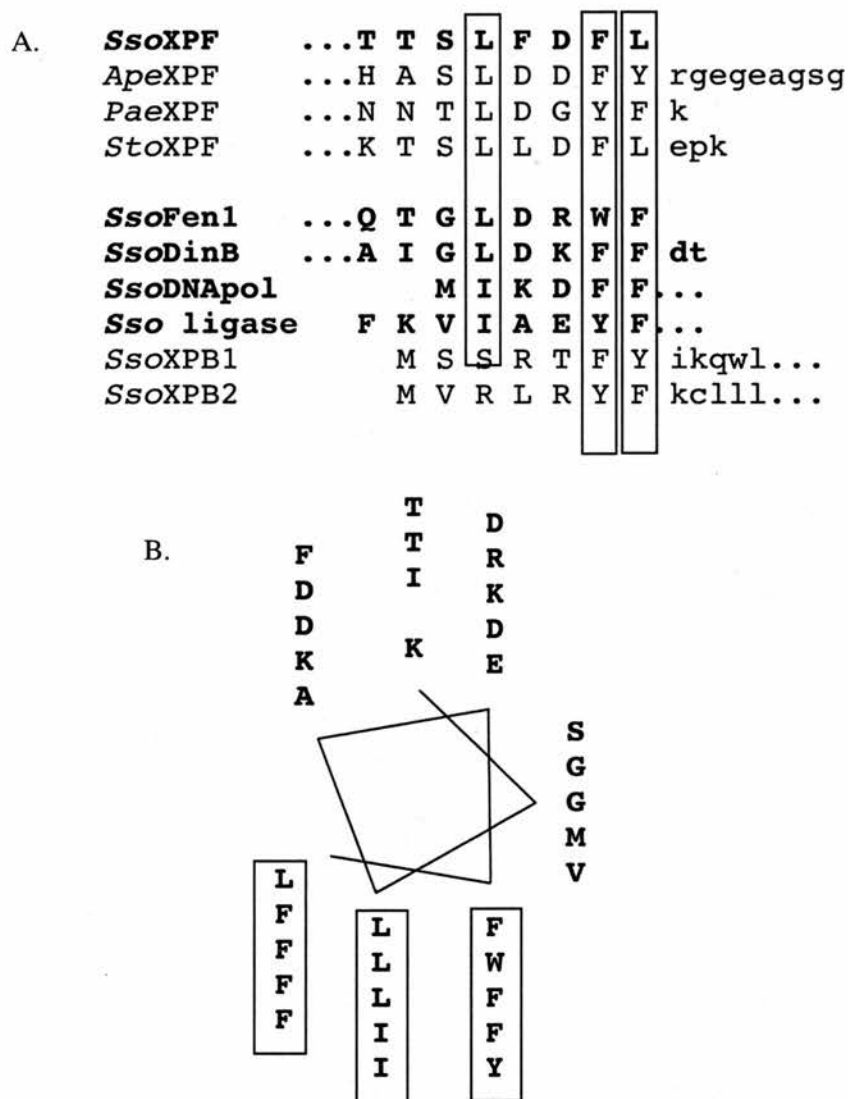


Figure 3.9. Identifying the PCNA interaction motif in *SsoXPF*. (A) Alignment of putative PCNA interaction motifs in crenarchaeal XPF homologues and other *S. solfataricus* replication/repair proteins. Conserved hydrophobic residues are highlighted in boxes. Sequences in bold are plotted on the helical wheel. (B) Helical wheel showing the conserved hydrophobic residues fall on the same face of the helix, which would facilitate interaction with the hydrophobic pocket on PCNA. *Sso* = *Sulfolobus solfataricus*; *ApeXPF* = *Aeropyrum pernix* XPF; *PaeXPF* = *Pyrobaculum aerophilum* XPF; *StoXPF* = *Sulfolobus tokadaii* XPF.

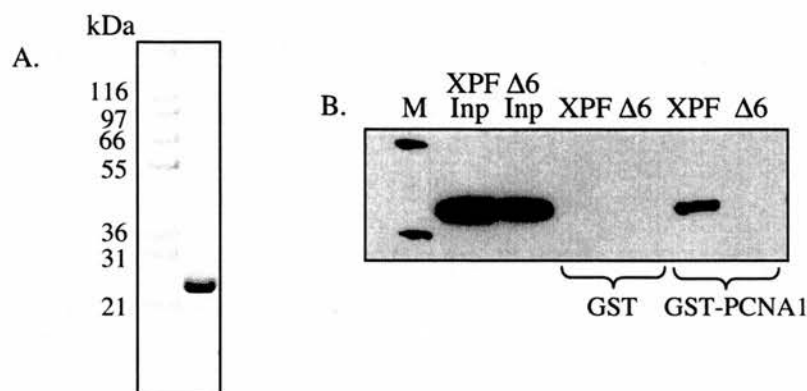


Figure 3.10. Confirming the PCNA interaction motif. (A) SDS-PAGE gel showing recombinant *SsoXPF* $\Delta 6$ mutant lacking the C-terminal 6 amino acids. (B) GST-PCNA1 fusion protein immobilised on beads was mixed with 1 μ g of either recombinant full length *SsoXPF* (XPF) or the C-terminal deletion mutant ($\Delta 6$) and Western blotted with antibodies specific for *SsoXPF*. The C-terminal deletion mutant did not interact with PCNA confirming the C-terminal 6 amino acids constitute the PCNA interacting motif. Inp = 25 % of the recombinant protein used.

Mapping of the cleavage sites showed multiple cuts were made in the 5' double strand region 3-12 bases from the single strand junction i.e. distributed over a helical turn of DNA. This apparent plasticity in DNA:protein complex may be explained by the thermal melting of DNA at 55 °C. This would create a variety of branched species with different lengths of single and double stranded DNA, subsequently locked down by protein as the PCNA-*SsoXPF* complex assembled. The DNA substrates shown in Figure 3.11 were designed with a common labelled strand (b50 strand) and the same sequence was cleaved in bubble and splayed duplex structures even though these sequences were at different positions relative to the double strand single strand junction (Figure 3.11B).

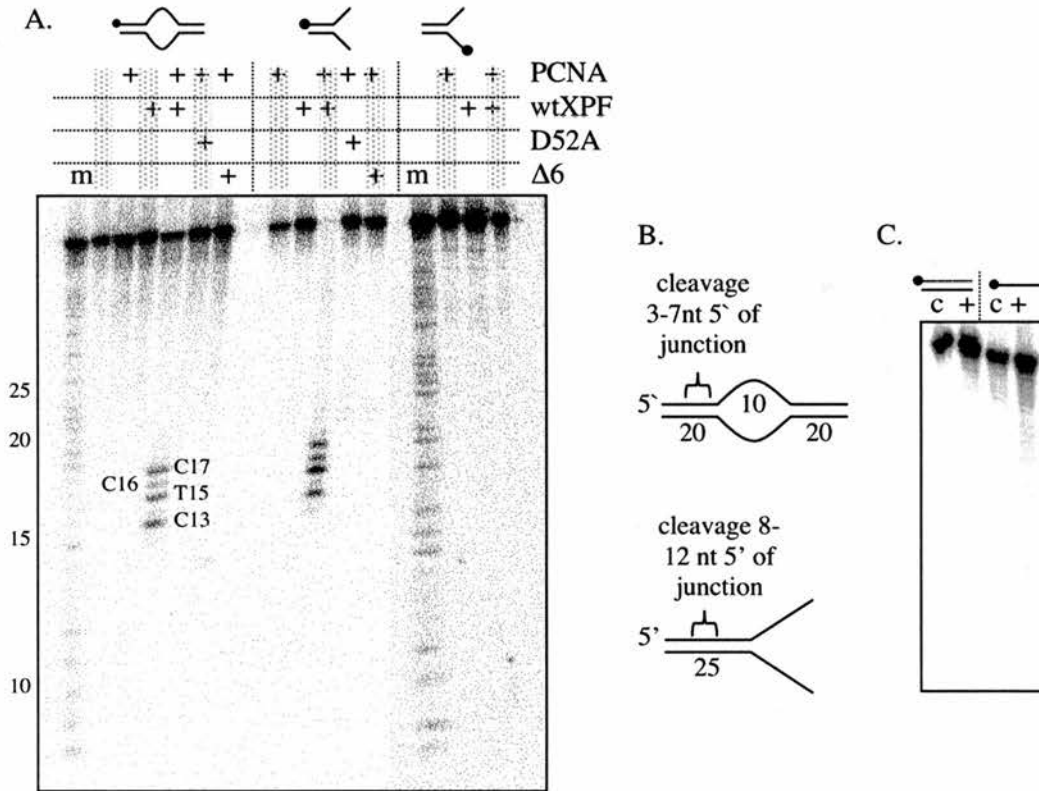


Figure 3.11 *SsoXPF* is a structure specific, PCNA-dependent endonuclease. *SsoXPF* and mutant proteins were incubated at 55 °C for 10 mins with DNA substrate in the presence and absence of PCNA heterotrimer. Reaction products were analysed by denaturing polyacrylamide gel electrophoresis. Circles indicate the position of the 5' ³²P label on the DNA substrate. (A) Denaturing polyacrylamide gel showing cleavage of bubble and splayed duplex structures. m = A&G Maxam Gilbert size markers of the labelled strand. (B) Schematic representation of bubble and splayed duplex DNA structures showing *SsoXPF*-PCNA cleavage sites. (C) Denaturing polyacrylamide gel showing *SsoXPF* in the presence of PCNA (+) does not cleave ds- or ssDNA. c = control, *SsoXPF* only.

In addition the 4 major cleavage sites were immediately 3' of a pyrimidine base. A pyrimidine preference (although not absolute) was also seen with a splayed duplex structure composed of a different sequence (r50 strand) (Figure 3.12) and a different structure, a four-way junction, with the same b50 strand labelled (Figure 3.14). This indicated a strong element of sequence specificity in addition to structural

specificity by *Sso*XPF. These are both properties of human XPF-ERCC1 (de Laat et al., 1998a). In contrast Mus81* shows no sequence specificity (Bastin-Shanower et al., 2003).

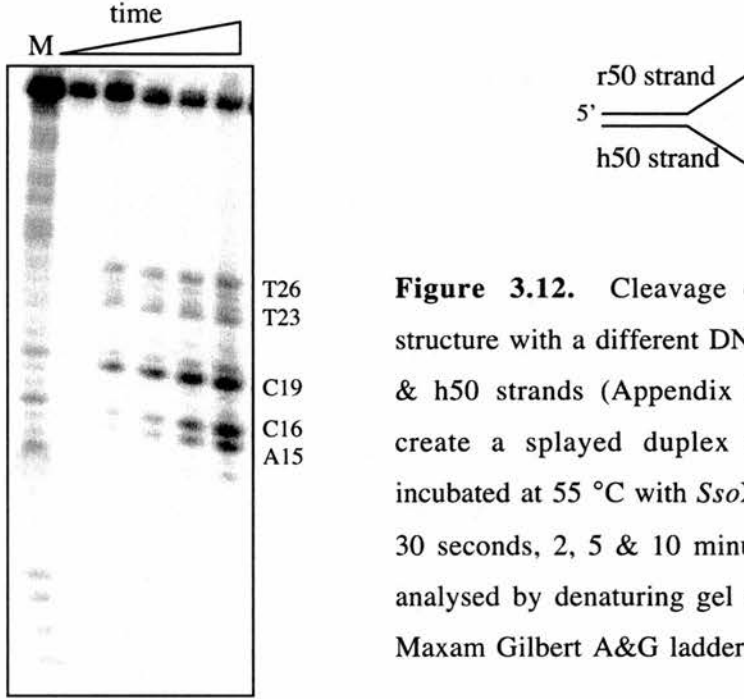


Figure 3.12. Cleavage of a splayed duplex structure with a different DNA sequence. The r50 & h50 strands (Appendix 1) were annealed to create a splayed duplex structure. This was incubated at 55 °C with *Sso*XPF and PCNA for 0, 30 seconds, 2, 5 & 10 minutes and the reactions analysed by denaturing gel electrophoresis. M = Maxam Gilbert A&G ladders of the labelled (r50)

Mutation of the *Sso*XPF presumed active site aspartate 52 to alanine (D52A) led to complete loss of detectable activity (Figure 3.11A). This was not due to disruption of the PCNA interaction since the D52A *Sso*XPF mutant interacted with GST-PCNA subunits (Figure 3.13). Mutation of the equivalent aspartate in human

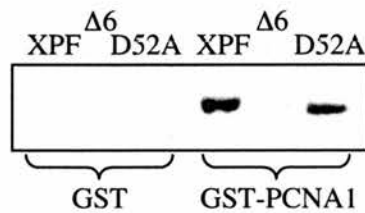


Figure 3.13. The D52A *Sso*XPF mutant interacts with PCNA. Recombinant *Sso*XPF, C-terminal $\Delta 6$ truncation and active site D52A mutant proteins were mixed with GST control or GST-PCNA1 then washed with buffer to remove non-interacting protein. The beads were analysed by SDS-PAGE and Western blot using antibodies specific for *Sso*XPF.

XPF and Mus81 also leads to loss of nuclease activity (Boddy et al., 2001; Enzlin and Scharer, 2002). This is probably due to disrupted co-ordination and/or activation of the water in the active site (Figure 1.5).

3.11 *Sso*XPF cleaved four way DNA junctions

Intermediate DNA structures arising in cells during processes such as homologous recombination, repair of double strand breaks and restart at stalled replication forks include 4 way DNA junctions, or Holliday junctions. Both the yeast XPF homologue Rad1 and more recently Mus81, were proposed as Holliday junction resolvases (Boddy et al., 2001; Chen et al., 2001; Habraken et al., 1994) before their preference for flap and fork substrates, respectively, became apparent. Similarly *Sso*XPF was also active against a 4 way DNA junction (Figure 3.14A), which it cut in all 4 strands 5' of the branch point and almost exclusively after cytosine residues (Figure 3.14B). However, unlike the Holliday junction resolving enzymes that cut symmetrically at the branch point (Lilley and White, 2001), *Sso*XPF cuts are asymmetric. Therefore *Sso*XPF does not fulfill the same criteria as the Holliday junction resolving enzymes.

3.12 *Sso*XPF was also active with other metal ions

Many nucleases show some degree of flexibility regarding the metal II ion they can use for cleavage. Figure 3.15A shows that *Sso*XPF is active with magnesium and manganese ions. The catalytic constants k_{cat} s were measured (as described in section 4.2). Activity of *Sso*XPF was 1.5 times greater in the presence of manganese ions compared to magnesium (Figure 3.15B). A similar magnitude

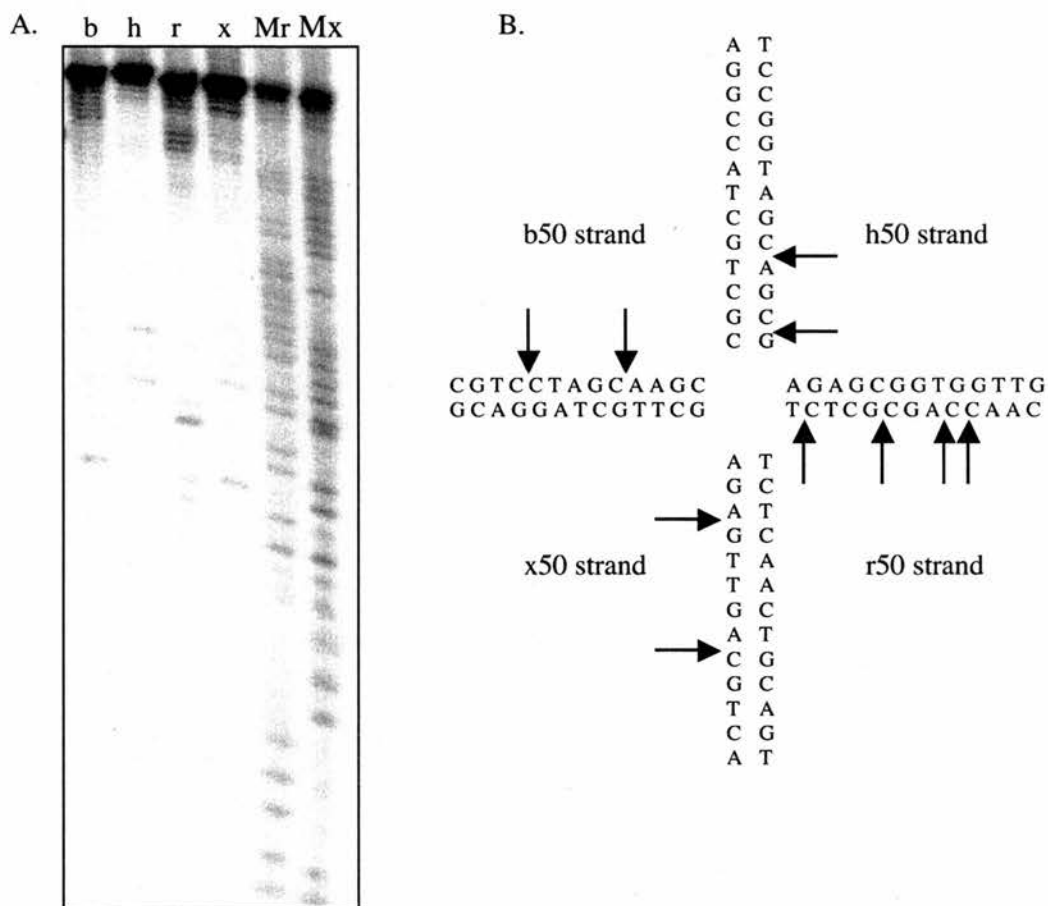


Figure 3.14. *SsoXPF* cleaved a 4-way DNA junction. (A) Denaturing polyacrylamide gel electrophoretic analysis of cleavage products resulting from incubation of *SsoXPF* and PCNA with a 4-way DNA junction. Mr & Mx = r50 and x50 strand A & G ladders, respectively. (B) Schematic representation of the central 13 base pair core of the 4 way junction. Arrows show the position of *SsoXPF* cleavage of each strand.

increase in cleavage rate with manganese is seen with human XPF (de Laat et al., 1998a). Very slight *SsoXPF* activity was seen in the presence of cobalt and no activity with calcium (Figure 3.15A). It is uncertain why calcium ions inhibit nuclease activity since they bind the same amino acid sites as magnesium and manganese ions (Nishino et al., 2003), but this inactivity is seen with other structurally related nucleases including type II restriction enzymes (Pingoud and Jeltsch, 2001).

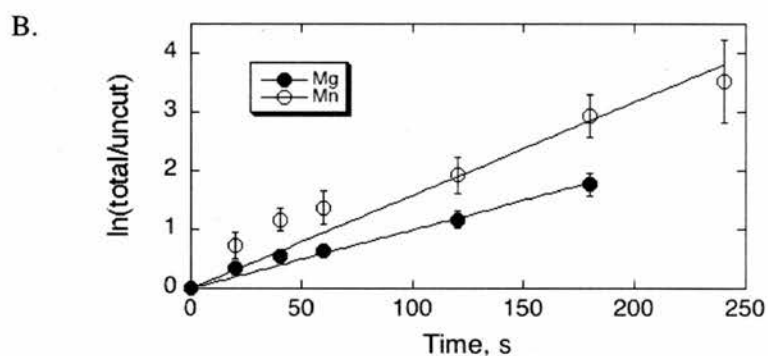
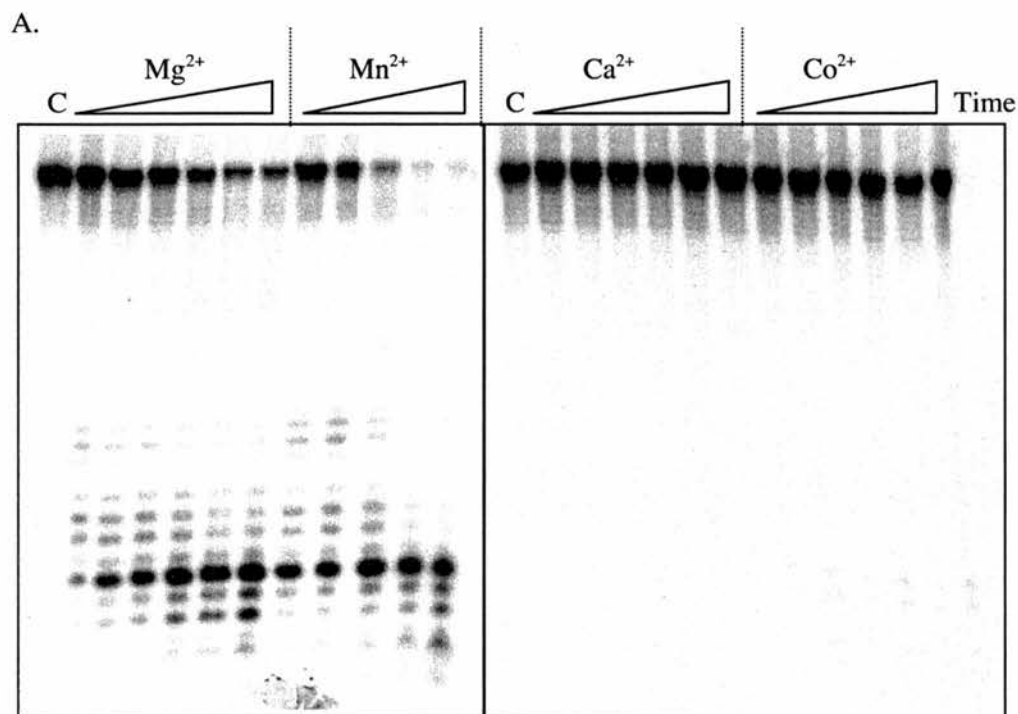


Figure 3.15. *SsoXPF* cleaved splayed duplex substrate in the presence of different metal (II) ions. (A) Denaturing polyacrylamide gel showing cleavage with 10 mM various metal ions. Time points were 20, 40 seconds, 1, 2, 3 & 4 minutes with the exception of Mn^{2+} which lacked a 1 minute time point. C = at time 0 before adding metal ions. (B) Quantification of *SsoXPF* rate with magnesium ions (filled circles) and manganese ions (open circles) (see section 4.2). The calculated k_{cat} values were 0.60 ± 0.02 and $0.95 \pm 0.06 \text{ min}^{-1}$ respectively.

3.13 Summary & conclusions

*Sso*XPF is limited to the highly conserved nuclease core of eukaryal XPF and Mus81 proteins. In this chapter we have shown *Sso*XPF comprises 2 independent domains that form a homodimer, with each monomer binding 1 manganese ion. Figure 3.16 shows a schematic representation of *Sso*XPF.

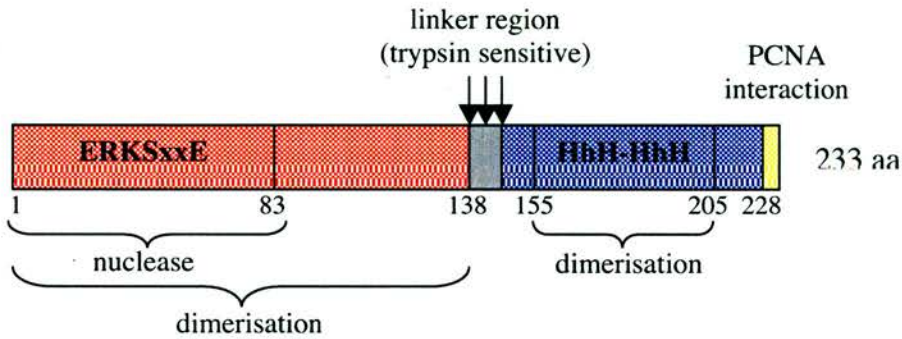


Figure 3.16. Schematic diagram of *Sso*XPF domain organisation. The nuclease domain (red) is separated from the double HhH domain (blue) by a trypsin sensitive linker region (grey). The PCNA interaction motif comprises 6 amino acids at the extreme C-terminus (yellow). Dimerisation occurs through both the nuclease and double HhH domains independently.

*Sso*XPF is a functional nuclease with a similar structure and sequence specificity to eukaryal XPF and euryarchaeal Hef, suggesting fundamental conservation of this enzyme through evolution. This is expanded on in the next chapter. However *Sso*XPF shows an apparent absolute requirement for the sliding clamp PCNA, which it interacts with through a characteristic C-terminal motif. The role of PCNA may be targeting the NER nucleases to DNA repair sites given the 3' NER nuclease XPG/Fen1 also interacts with PCNA in both archaea (Dionne et al., 2003) and eukaryotes (Gary et al., 1997). Alternatively, or perhaps additionally, PCNA may activate *Sso*XPF. In both archaea and eukarya Fen1 is active without

PCNA, in contrast to *Sso*XPF that appears dependent on the presence of PCNA for nuclease activity. The role of the PCNA-*Sso*XPF interaction is investigated further in chapter 5.

Dionne *et al.* 2003 have recently shown *S. solfataricus* to encode a heterotrimeric PCNA, in contrast to eukarya and euryarchaea, which have a homotrimer. The increased complexity of PCNA allows *S. solfataricus* to utilize subunit-specific interactions with PCNA. There is good evidence that *S. solfataricus* Fen1, DNA ligase and DNA polymerase can all bind simultaneously to a single PCNA heterotrimer to form a functional 5' flap removal and repair complex (Dionne *et al.*, 2003). An interesting observation is that both *S. solfataricus* homologues of the helicase XPB (Rad25) have potential PCNA interaction motifs (Figure 3.9). Eukaryal XPB is a component of TFIIH with a dual role in transcription and NER (Svejstrup *et al.*, 1996), and there is evidence for an interaction between XPB and XPF in eukaryal NER (Evans *et al.*, 1997). Chapter 6 looks at one *S. solfataricus* XPB homologue in more detail.

CHAPTER 4: *Sso*XPF nuclease activity

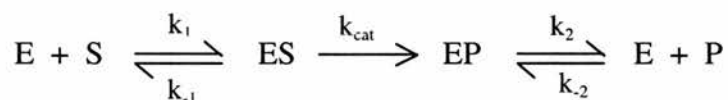
4.1 Introduction

*Sso*XPF shares the same metal-dependent nuclease motif as eukaryal XPF and Mus81 which has been mapped using an affinity cleavage assay and confirmed by mutagenesis of the conserved residues (Enzlin and Scharer, 2002). Both XPF and Mus81* are structure specific but slight differences in their substrate specificity reflect their *in vivo* functions. The XPF-ERCC1/yeast Rad1-Rad10 heterodimer cuts a variety of flap, overhang, bubble and splayed duplex structures on the 5' side of the junction between double stranded and single stranded DNA, reflecting a major *in vivo* function – the 5' incision during nucleotide excision repair (NER) (Park et al., 1995). Similarly it has additional functions including repair of interstrand cross-links (De Silva et al., 2000; De Silva et al., 2002; Kuraoka et al., 2000) and removal of 3' non-homologous single strand tails from recombination intermediates (Adair et al., 2000; Fishman-Lobell and Haber, 1992; Niedernhofer et al., 2001; Sargent et al., 1997), dependent on this ability to cleave 3' flaps and splayed duplex structures. In contrast Mus81* apparently prefers branched structures that occur at stalled replication forks and nicked 4-way junctions during replication restart and recombination (Hollingsworth and Brill, 2004; Osman et al., 2003; Whitby et al., 2003). Euryarchaeal Hef has a similar substrate preference to Mus81* (Komori et al., 2002).

This chapter focuses on the *in vitro* nuclease activity and potential *in vivo* roles of *Sso*XPF, highlighting similarities to both eukaryal XPF and Mus81. This is consistent with a scheme in which an ancestral nuclease resembling *Sso*XPF had roles in both NER and replication restart, and diverged into the specialised forms found in eukarya today.

4.2 Quantification of *Sso*XPF nuclease activity

In chapter 3 we showed *Sso*XPF was a structure specific endonuclease. Here we expand this to define *Sso*XPF specificity for a variety of DNA structures analogous to those arising *in vivo* during DNA damage and repair pathways to elucidate the function of this nuclease. The relative nuclease activity was measured under standard single turnover conditions using 80 nM ³²P-end labelled DNA, 1 μM PCNA and 1 μM *Sso*XPF. Cleavage was initiated at 55 °C by adding MgCl₂ to a final concentration of 10 mM and aliquots taken at set time points. Under these conditions the excess of enzyme (E) to DNA substrate (S) favoured the enzyme-substrate (ES) complex so the conversion of enzyme-substrate complex to enzyme and product (P) was measured in these assays i.e. the catalytic constant k_{cat} . The dissociation of the enzyme-product (EP) complex could be ignored because denaturing conditions were used.



This scheme is an oversimplification since E must encompass both PCNA and *Sso*XPF which interact. For the purposes of this analysis, where we consider relative rates of cleavage for different DNA substrates under defined concentrations of *Sso*XPF and PCNA, we assume k_1 and k_{-1} are constant.

The cleavage products were separated from the uncut substrate on a denaturing polyacrylamide gel and visualised by phosphoimaging. The uncut and cut substrate for each time point was quantified using Image Gauge software (Figure 4.1A). Unless otherwise stated this was repeated in at least triplicate for each substrate. The mean ratio of total DNA counts (uncut + cut bands) to uncut counts was used to determine a catalytic constant (k_{cat}) by linear regression (Figure 4.1B). Non-linear

regression gave a very similar k_{cat} (Figure 4.1C). Linear regression was used because it allowed easier manual interpretation of the data quality.

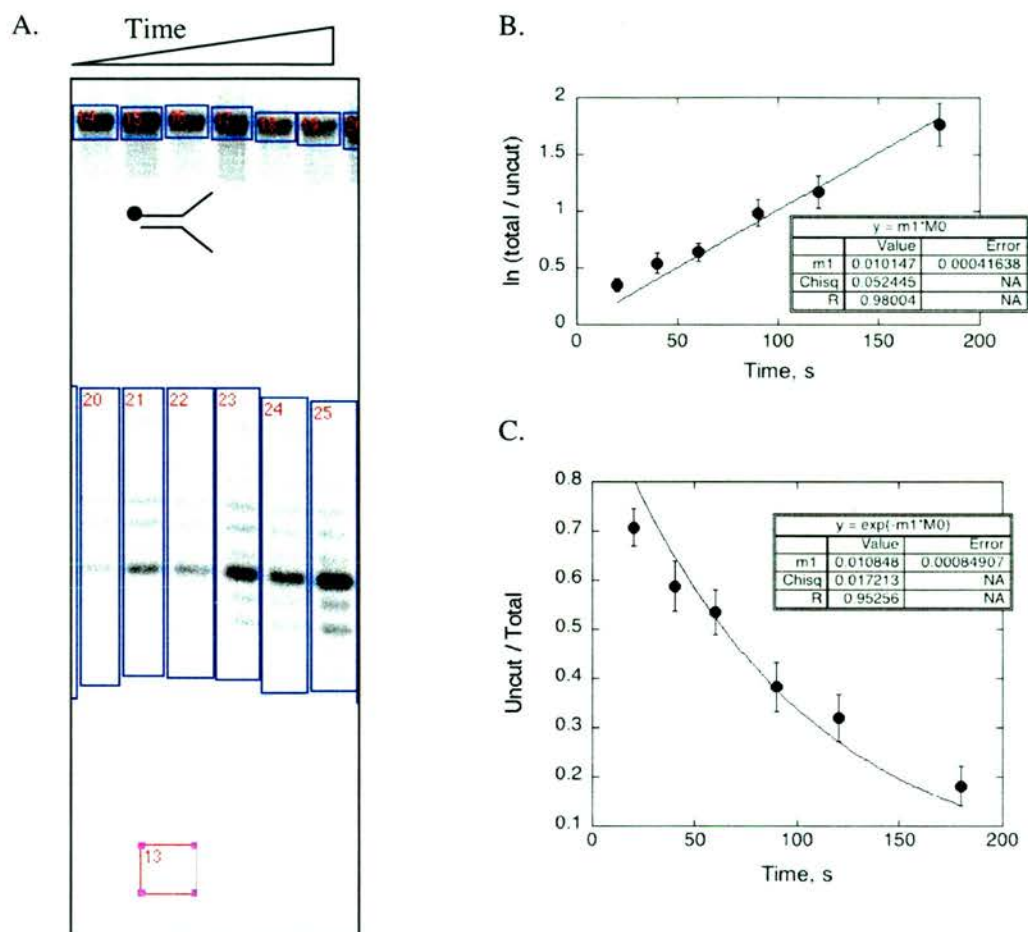


Figure 4.1. Quantifying *Sso*XPF nuclease activity for the splayed duplex DNA substrate. (A) Denaturing polyacrylamide gel showing accumulation of splayed duplex cleavage product with time points 20, 40, 60, 90, 120 & 180 seconds. The intensity of the bands corresponding to the uncut substrate and cut products were quantified using Image Gauge software and drawing boxes round the relevant areas. The background intensity (red box) was subtracted from all these values. This was repeated in quadruplicate for each time point. In addition, later rate experiments took a measurement before addition of magnesium (i.e. at time 0) and other time points were corrected accordingly. Mean values were used to calculate the k_{cat} for the *Sso*XPF cleavage reaction by linear regression (B), where $m1 = k_{\text{cat}}$ (sec^{-1}), $M0 = \text{time in seconds}$, the error is standard error and $R = \text{Pearson's R value}$. (C) k_{cat} calculated by non-linear regression.

4.3 *Sso*XPF substrate specificity

Table 4.1 and Figure 4.2 show the relative rates of *Sso*XPF activity for a variety of DNA structures.












SUBSTRATE	OLIGOS	STRUCTURE	RATE (min ⁻¹) ±S.E.
3' flap	b50, r26-50, x50		6.8
nicked duplex	b25, r26-50, x50		6.2
nicked 3-way	b50, h1-25, r26-50, x50		3.6 ± 0.2
nicked 4-way	b50, h50, r1-25, r26-50, x50		2.5 ± 0.2
splayed duplex	b50, x50		0.61 ± 0.02
replication fork	b50, h1-25, x50		0.42 ± 0.02
5' overhang	b25, x50		0.074 ± 0.003
4-way junction	b50, h50, r50, x50		0.057 ± 0.005
10 nt bubble	b50, 10nt bubble		0.051 ± 0.004
3-way junction	b50, y50, x50		0.049 ± 0.001
3' overhang	b50, x26-50		0.027 ± 0.001

Table 4.1. Relative *Sso*XPF cleavage rates for different DNA structures. Circles show the ³²P labelled 5' end and arrows indicate the approximate site of cleavage. Sequences of the oligos used are given in Appendix 1.

4.3.1 3' flaps & nicked structures – importance of a 5' DNA end

Clearly the best substrates for *Sso*XPF were the 3' flap and nicked duplex, followed by the nicked 3-way and nicked 4-way junctions. These are the preferred

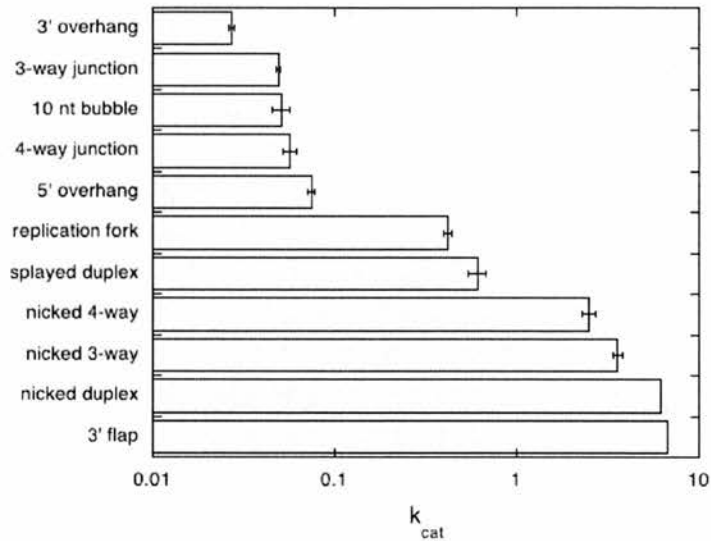


Figure 4.2. Relative *Sso*XPF cleavage rates for different DNA structures. Catalytic constants (k_{cat} s) are shown on a log scale. Standard errors are shown where available.

substrates of Mus81* (Ciccia et al., 2003; Doe et al., 2002; Kaliraman et al., 2001; Whitby et al., 2003). Efficient cleavage of nicked substrates demonstrates that a single strand flap is not an absolute requirement, rather a discontinuity in duplex DNA is sufficient to promote *Sso*XPF nuclease activity. Mus81* cleaves nicked duplex DNA but less efficiently than 3' flap (Bastin-Shanower et al., 2003). Thermal fraying of the nicked duplex under the increased temperature used in *Sso*XPF assays could result in transient formation of 3' flap structures and may explain the more efficient cleavage of the nicked duplex.

All the DNA substrates favoured by *Sso*XPF have a 5' DNA end near the branch point or point of discontinuity. The splayed duplex substrate that did not have a 5' DNA end was cut approximately 10 fold slower (Table 4.1). In addition a substrate analogue of a replication fork with a 3' DNA end near the branch point was cut relatively slowly. This requirement for a 5' end was not a requirement for adjacent duplex DNA since the 3-way and 4-way junctions that contained adjacent

duplex DNA but no 5' end were cut poorly. So *Sso*XPF showed a clear preference for a 5' DNA end near the cleavage site, similar to Mus81* but unlike Rad1-Rad10/XPF-ERCC1 whose activity is inhibited by such a 5' end (Bastin-Shanower et al., 2003; de Laat et al., 1998). *In vivo* most substrates of this type probably have a phosphate group at this 5' position. To test whether this phosphorylation was important for *Sso*XPF activity, the k_{cat} for nicked duplex with and without a phosphorylated 5' end was compared (Figure 4.3). There was a small but significant increase in k_{cat} in the presence of phosphate at the 5' end.

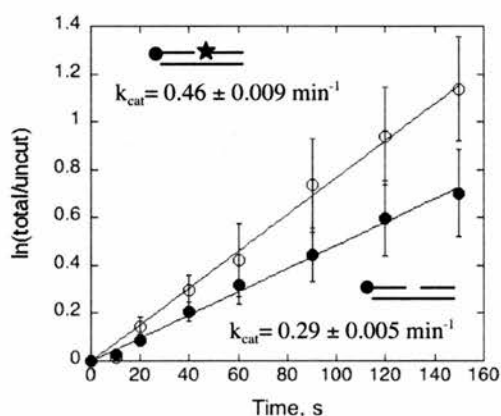


Figure 4.3. Quantifying *Sso*XPF activity for nicked duplex DNA with and without a 5' phosphorylated end (open and closed circles respectively). Protein concentrations were decreased to 0.1 μM to obtain a linear initial rate over a longer period of time for more accurate comparison. Standard errors are shown. Circles represent the ^{32}P -labelled strand and the star shows the phosphorylated 5' DNA end.

What then is the role of this 5' end in *Sso*XPF cleavage? In the case of Mus81* the 5' end directs the position of cleavage (Bastin-Shanower et al., 2003; Osman et al., 2003). *Sso*XPF cleaved the 3' flap between 4 and 12 nt 5' from the flap/5' end with the most predominant positions 4 and 8 nt (Figure 4.4). If the position of the 5' end was important in directing *Sso*XPF cleavage, truncations of the oligo downstream of the 3' flap (labelled with an asterisk in Figure 4.4B) would cause a

corresponding change in the major cleavage positions (Bastin-Shanower et al., 2003). This could be investigated.

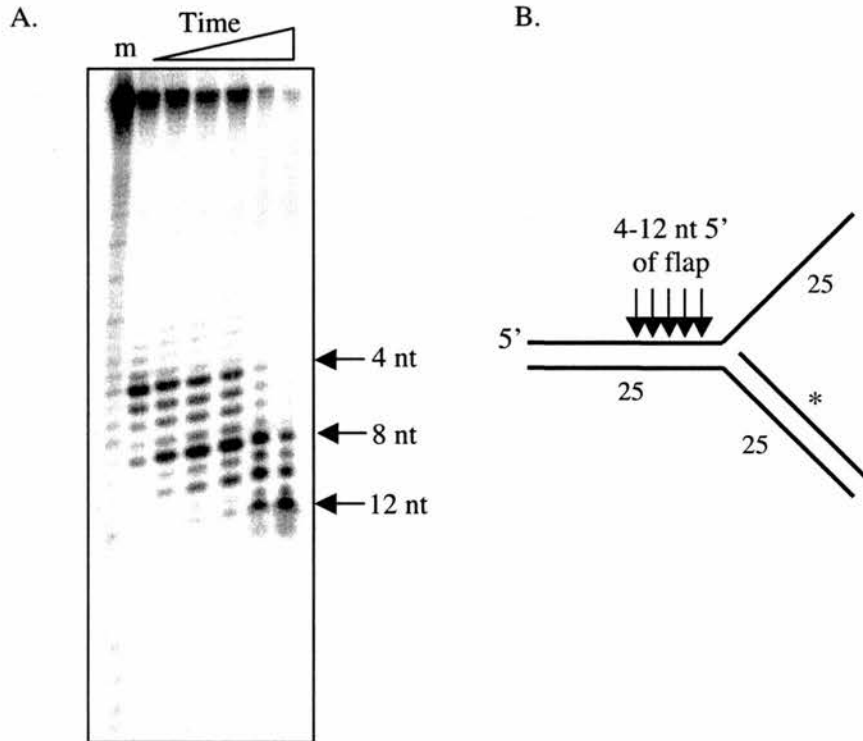


Figure 4.4. *Sso*XPF cleaved a 3' flap. (A) Denaturing polyacrylamide gel showing cleavage with increasing time (5, 10, 20, 40 seconds, 1 minute 30 & 4 minutes). Arrows and numbers show the distance of cleavage sites from the 3' flap. m = A & G size markers. (B) Schematic representation of the 3' flap substrate showing the position of *Sso*XPF cleavage (arrows). Major cleavage sites were 4 & 8 nt 5' of the flap. The asterisk marks the strand downstream of the 3' flap as referred to in the text.

4.3.2 *Sso*XPF cleaved flaps & overhangs

In the absence of a 5' DNA end a single stranded flap or flaps appear important for *Sso*XPF nuclease activity (Table 4.1; Figure 4.2). The splayed duplex substrate was cut approximately 10 fold faster than a 5' overhang and 20 fold faster than a 3' overhang, indicating single strand flaps on both the cut and uncut strand were important for *Sso*XPF activity. The 3-way and 4-way junctions that have no flap, only discontinuities in the duplex DNA, were cut poorly. This is consistent with

a role as 5' endonuclease in DNA excision repair pathways where the DNA around a lesion is unwound by helicases to form a bubble.

Decreasing the length of the 3' overhang did not change the position of cleavage (Figure 4.5A) but decreasing the overhang length from 25 nt to 12 nt decreased *Sso*XPF activity by 2 fold and further reduction of flap length to 6 nt decreased activity a further 2 fold (Figure 4.5B). So overhang length in the absence of a 5' DNA end was important. This is similar to eukaryal XPF-ERCC1 activity which also decreases with decreasing ssDNA (de Laat et al., 1998). This dependence on overhang length may explain why the 10 nt bubble was cut relatively poorly compared to the splayed duplex – the 2 single stranded ‘flaps’ are short and less flexible with respect to an *in vivo* repair bubble, for example in eukaryal NER this is approximately 28 to 32 nt long (de Laat et al., 1999).

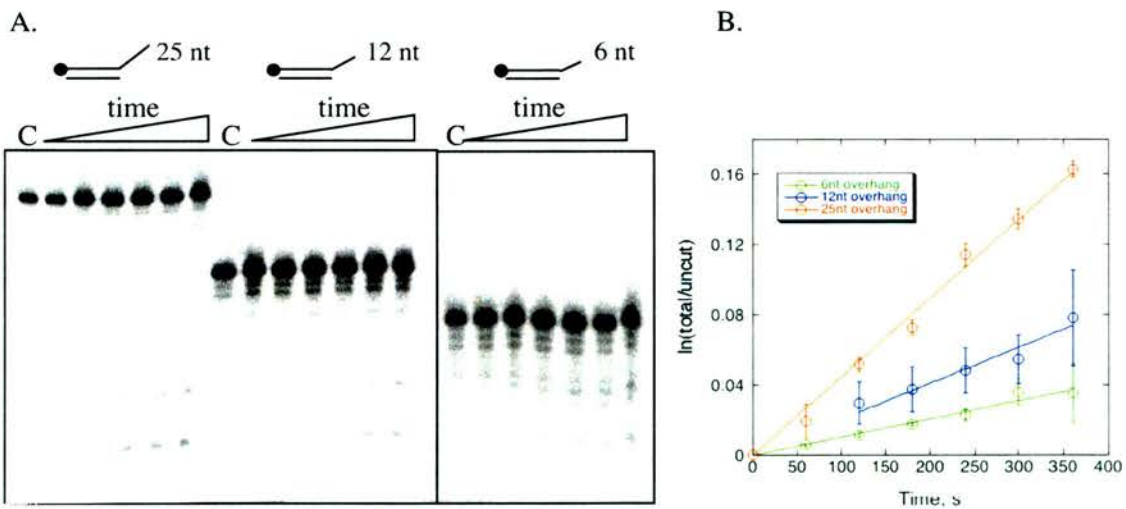


Figure 4.5. Effect of 3' overhang length on *Sso*XPF nuclease activity. (A) Denaturing polyacrylamide gel showing cleavage of duplex DNA with 25, 12 & 6 nt long 3' overhangs with time points 0, 1, 2, 3, 4, 5 & 6 minutes. (B) Quantification of 3' overhang length and *Sso*XPF nuclease activity. k_{cat} values \pm standard error for 25, 12 and 6 nt overhangs were 0.027 ± 0.0001 , 0.012 ± 0.0005 and $0.0062 \pm 0.0002 \text{ min}^{-1}$ respectively.

4.3.3 *Sso*XPF cleaved duplex DNA containing lesions weakly

Could *Sso*XPF recognise and cleave specific lesions in DNA? Mismatches in DNA base pairing and deleterious base modifications (for examples damage by UV light, hydrolytic deamination and oxidation) lead to bulges in the DNA and distortion of the double helix. *Sso*XPF cleaved a 3-way junction containing a 3 nt bulge approximately twice as fast a 3-way junction but still relatively slowly (Figure 4.6A). Similarly duplex DNA structures containing a 3 nt bulge, an abasic site or a thymine dimer were cleaved very weakly (Figure 4.6B&C; Figure 4.7). Cleavage in each case was specific for the DNA strand containing the lesion and occurred 5' of the damage suggesting cleavage was not random.

4.3.4 Summary

Together these results suggest that *Sso*XPF does not immediately function in recognition of damaged DNA but as rather part of a DNA repair pathway(s) and functions either (1) after other proteins have recognised the damage and unwound the surrounding DNA to create a suitable flap substrate for *Sso*XPF or (2) the damage has not been recognised by repair machinery, DNA replication has proceeded and become stalled at the point of damage. The next sections in this chapter apply the *in vitro* activities of *Sso*XPF to possible DNA repair situations that could arise *in vivo*.

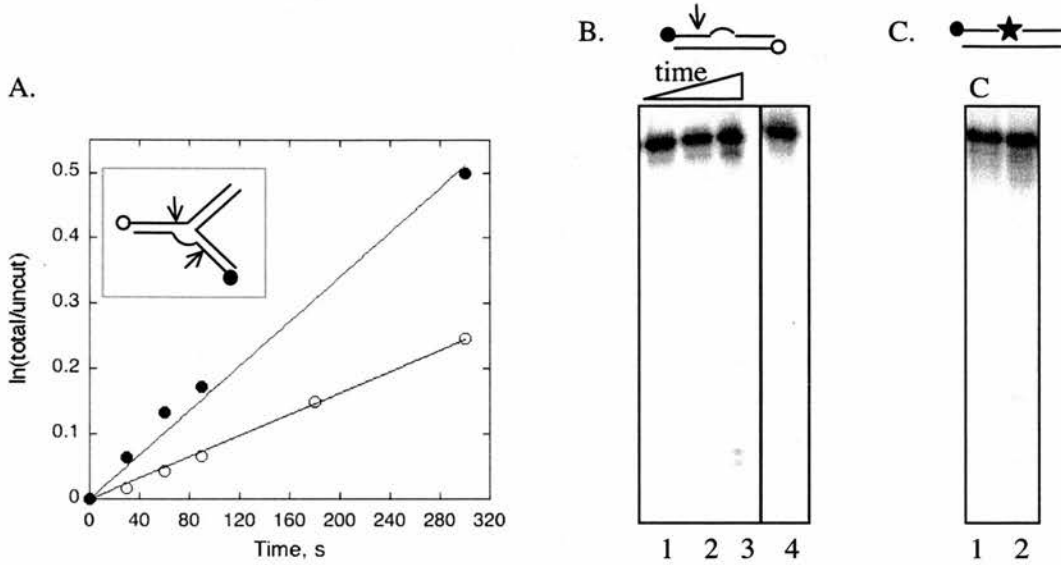


Figure 4.6. *SsoXPF* cut weakly at bulges and abasic sites in DNA. (A) Quantification of *SsoXPF* cleavage of a 3-way junction containing a 3 nt bulge at the branch point. This substrate was made by annealing YA3 (filled circle), b50 (open circle) and h50 oligos (Appendix 1). The b50 strand ($k_{\text{cat}} = 0.040 \pm 0.002 \text{ min}^{-1}$) was cut at a similar rate to the 3-way junction lacking a bulge (Table 4.1). The YA3 strand containing the bulge was cleaved over 2 fold faster ($k_{\text{cat}} = 0.10 \pm 0.004 \text{ min}^{-1}$). (B) Denaturing polyacrylamide gel showing *SsoXPF* cleaved duplex DNA containing a 3 nt bulge. This substrate was made by annealing the A3 bulge oligo (filled circle) and b50 oligo (open circle)(Appendix 1). Lanes 1-3 = DNA incubated with *SsoXPF* & PCNA for 1, 3 & 5 minutes respectively; 4 = bulge substrate labelled on opposite strand to lesion incubated with *SsoXPF* & PCNA for 5 minutes. *SsoXPF* cleaved the strand containing the bulge weakly but no cleavage of the opposite strand was detected. (C) Denaturing polyacrylamide gel showing *SsoXPF* cleaved duplex DNA containing an abasic site at position 25 very weakly. This substrate was made from b50 abasic and b50 complementary oligos (Appendix 1). Lane 1 = DNA incubated with *SsoXPF* for 15 minutes; 2 = DNA incubated with *SsoXPF* & PCNA for 15 minutes.

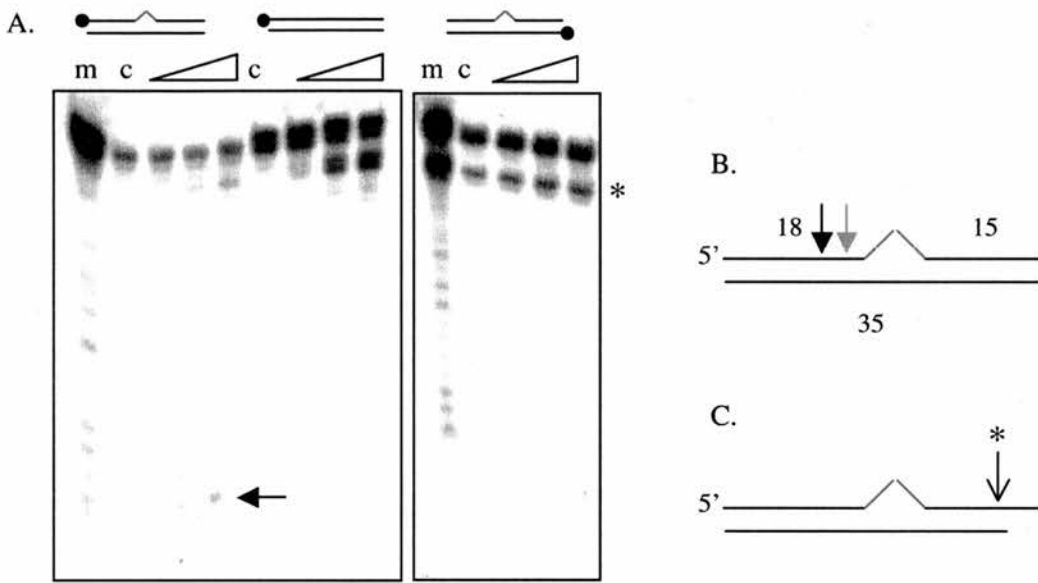


Figure 4.7. *Sso*XPF cut dsDNA containing a thymine dimer. (A) Denaturing polyacrylamide gel showing *Sso*XPF cleaved 5' of thymine dimer (red inverted V) but did not cleave duplex DNA of the same sequence nor the strand opposite the TT lesion. The ^{32}P labelled strand is indicated by circles. The black arrow shows the major cleavage site at C13, 5 nt 5' from the TT dimer. Time points were 1, 5 & 10 minutes; m = A & G size markers; c = DNA control. The asterisk highlights a problem with the quality of the complementary oligo used to make the TT dimer. This smaller band indicated the presence of a truncated form of this oligo. If the truncation was at the 5' end, annealing to the TT dimer oligo would create a 3' overhang (see C). *Sso*XPF would cleave of the TT dimer strand to remove the resulting 3' overhang on this strand, explaining the appearance of a large product with time. (B) Schematic representation of the TT dimer substrate. The TT dimer was at positions 20 & 21 of the 35 nt long oligo and arrows represent *Sso*XPF cleavage. (C) Schematic representation of the TT dimer substrate if the complementing oligo was truncated at the 5' end, creating a 3' overhang which would be cleaved by *Sso*XPF

4.4 *Sso*XPF & rescue of stalled replication forks

Blocked replication forks form a number of different structures depending on which template strand is blocked, whether the leading and lagging strand DNA polymerases can uncouple, and whether the fork reverses (Sogo et al., 2002). Assuming the *in vivo* substrate preference is similar to that observed *in vitro* the favoured substrate of *Sso*XPF, the 3' flap, would occur when the leading strand was blocked and uncoupling of the polymerases led to continued DNA replication by the lagging strand polymerase only (Figure 4.8Biib). *Sso*XPF cleavage would result in the detachment of the leading strand arm and collapse of the replication fork creating a double strand break (DSB) end (Figure 4.8Biiib), for break induced replication (BIR). Similarly regression of the stalled fork without uncoupling of the polymerases could lead to a nicked 3-way like structure that we have shown to be a good substrate for *Sso*XPF (Figure 4.8Biiia). A role for eukaryal Mus81* in the break induced repair (BIR) pathway has also been hypothesised (Constantinou et al., 2002; Doe et al., 2002; Kaliraman et al., 2001; Whitby et al., 2003).

In addition Mus81* cleaves blocked forks which have partially regressed to give either 5' nascent lagging strand tail (Figure 4.8Bivb) or 3' nascent leading strand tail (Figure 4.8Cii), prior to full Holliday junction formation (Whitby et al., 2003). Again this cleavage provides DSB ends, substrates for BIR. There are 3 potential pathways for replication fork regression in bacteria: RecA mediated (Robu et al., 2001), RecG mediated (McGlynn and Lloyd, 2002), and spontaneous regression by positive supercoiling from upstream replication processes (Postow et al., 2001). *S. solfataricus* has a RecA homologue, RadA, but appears to lack obvious RecG homologues (reviewed in (Bolt and Allers, 2004)). Given *Sso*XPF's preference for a 5' DNA end we would predict cleavage of a fork with a lagging strand block that had

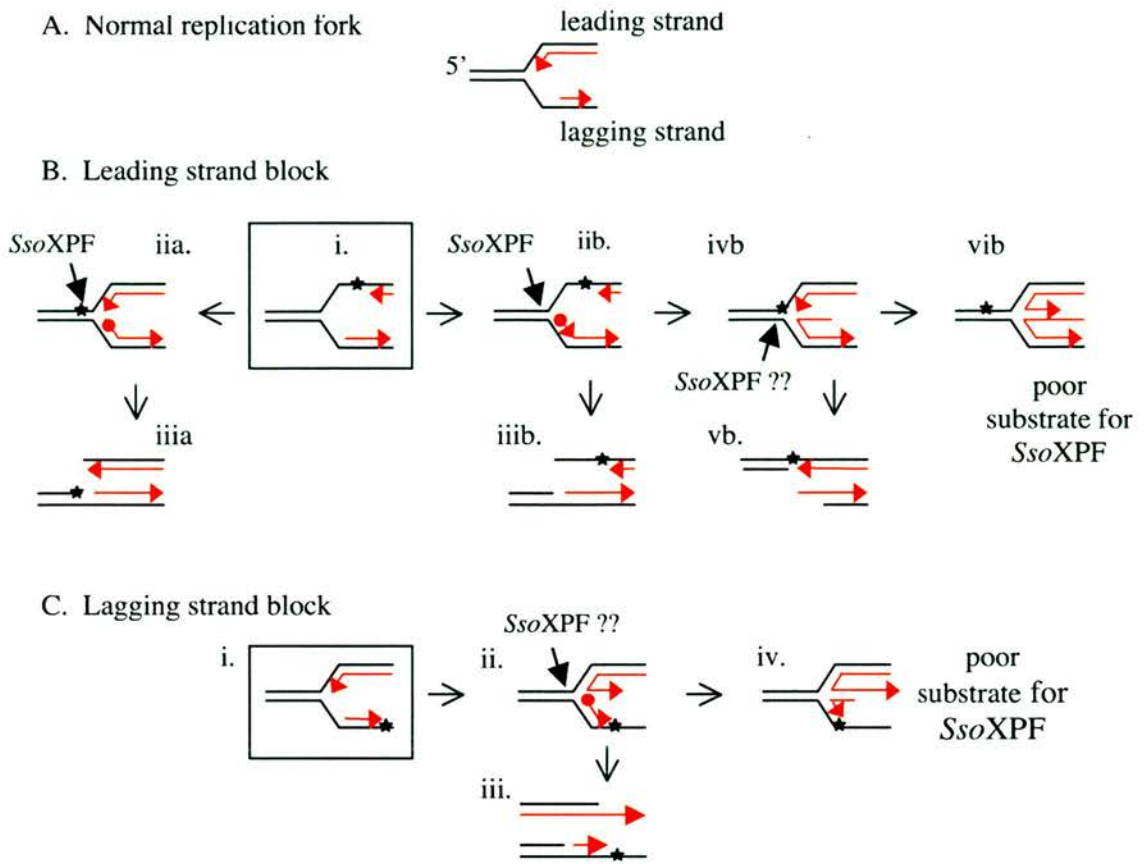


Figure 4.8. Model for *Sso*XPF cleaving stalled and reversed replication forks. Nascent strands are shown in red with an arrowhead at the 3' end. The star indicates the position of the replication fork block. *Sso*XPF cleavage sites and potential cleavage sites (??) are shown. Red circles show the 5' DNA ends stimulating *Sso*XPF activity. (A) A normal replication fork is cleaved relatively poorly by *Sso*XPF. (B) Leading strand block (i). Regression of fork to form a nicked 3-way like junction (iia) and cleavage of leading strand template by *Sso*XPF leads to collapse of the fork (iiiia). Alternatively uncoupling of DNA polymerases leads to continued DNA replication of lagging strand to create the 3' flap structure (iib). *Sso*XPF cleavage of the leading strand template leads to collapse of the fork (iiiib). The fork may regress further exposing a 5' nascent strand tail (ivb). By analogy to Mus81* *Sso*XPF may also cleave the lagging strand template resulting in a collapsed fork (vb), but there is no experimental evidence for this. Further regression of the fork forms a 4-way junction which is cleaved poorly by *Sso*XPF (vib). (C) Lagging strand block (i) and fork regression to create a 3' nascent leading strand tail (ii). By analogy to Mus81* but also given *Sso*XPF preference for the 5' DNA end and cleavage of a similar nicked 3-way structure (Biia), we predict *Sso*XPF would cleave the leading strand template leading to collapse of the fork (iii). On further regression to a 4-way junction *Sso*XPF cleaves poorly (iv).

regressed to form a 3' leading strand tail (Figure 4.8Cii), similar to Mus81*. Cleavage of a fork regressed to give a 5' lagging strand tail, which has no 5' end would probably be cut less efficiently.

In chapter 3 we showed *Sso*XPF interacted with PCNA and this was necessary for nuclease activity. This interaction may also be relevant for targeting *Sso*XPF to sites of stalled replication *in vivo*. PCNA is required for DNA replication, for the processivity of DNA polymerase, and therefore would be present at sites where replication has stalled.

4.5 *Sso*XPF & homologous recombination

The 3 main recombination pathways, classic double strand break repair (DSBR), synthesis dependent strand annealing (SDSA), and break induced replication are shown in Figure 1.7. All involve resection of a 5' end(s) by an endonuclease to create a 3' ssDNA end that invades homologous duplex DNA. Initiation of DNA synthesis and elongation of the invading 3' end creates a D loop. In SDSA, the elongated 3' end reanneals to the other end of the break and acts as a template to fill the gap. However, occasional over-synthesis of invading 3' ends leads to 3' flaps upon displacement and reannealing, and these must be removed to enable ligation of DNA ends and gap filling. Given *Sso*XPFs preference for 3' flaps, this may be an *in vivo* function of this enzyme (Figure 4.9), similar to that proposed for Mus81* (Bastin-Shanower et al., 2003).

However, *Sso*XPF cleaved nicked duplex DNA with a similar efficiency to 3' flap structures suggesting the 5' DNA end rather than a 3' single strand flap was important (although as discussed previously further work is necessary to confirm

this). The D loop recombination intermediate has such a 5' DNA end (Figure 4.9iii), and was recently shown to be cut by Mus81* (Osman et al., 2003). We investigated whether *Sso*XPF could also resolve D loop structures.

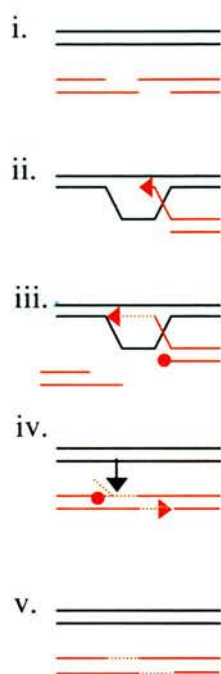


Figure 4.9. Role of *Sso*XPF in recombination - Model 1: removal of 3' flap intermediates during SDSA. Invading strands are shown in red. Dotted lines and arrowheads show the direction of DNA synthesis. The *Sso*XPF cleavage site is represented by an arrow and red circles show the 5' DNA end directing/stimulating cleavage. Following a DSB the 5' DNA ends are resected to generate 3' ssDNA ends (i) that invade homologous dsDNA (ii). Over-replication of the invading 3' DNA end (iii) and subsequent dissociation and reannealing to the other end of the break creates a 3' ssDNA flap (iv) which is removed by *Sso*XPF (iii).

First we made a 25 nt bubble, lacking an invading strand, by annealing b75 and z75 oligos (Figure 4.10A; Appendix 1). *Sso*XPF cut both strands weakly in the duplex region 5' of the bubble: the b75 strand was cut 4 nt 5' of the bubble and the z75 strand 2, 5, 8 and 11 nt 5' of the bubble (Figure 4.10). The catalytic constants were 0.024 ± 0.001 and $0.054 \pm 0.001 \text{ min}^{-1}$ for b75 and z75 strands respectively (Figure 4.12; Table 4.2), similar to the 10 nt bubble (Table 4.1). The difference in cleavage site and rate between the b75 and z75 strands presumably reflects the sequence dependence of *Sso*XPF.

In the presence of an invading strand (i.e. the complete D loop structure, Figure 4.11A), additional *Sso*XPF cleavage products and increased rates of cutting were apparent. Additional major cleavage sites were mapped to the b75 strand, 8 and

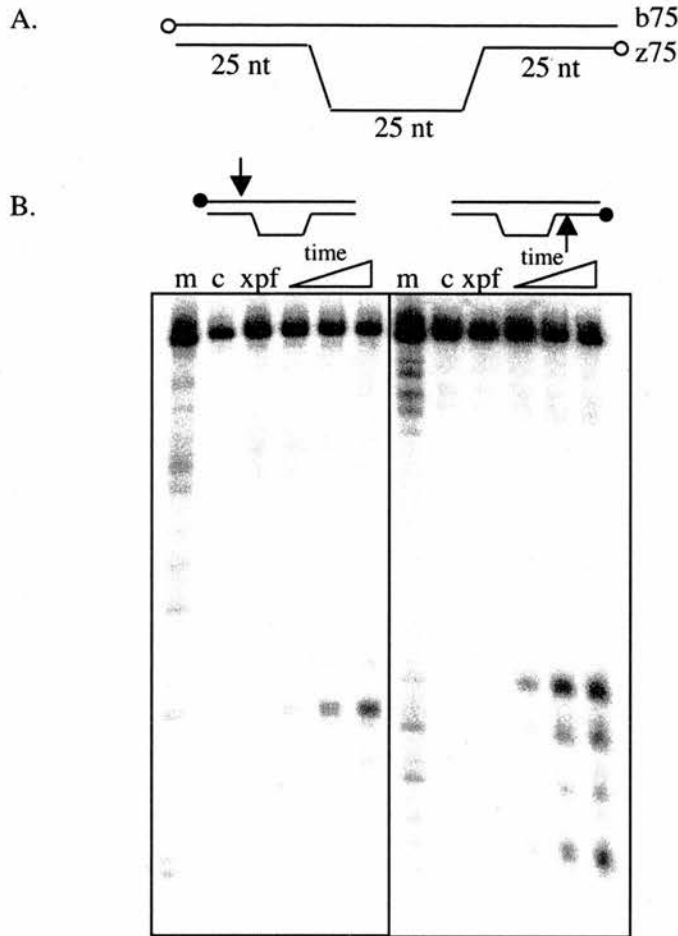


Figure 4.10. *Sso*XPF cleaved bubble DNA lacking an invading strand. (A) b75 and z75 strands (Appendix 1) were annealed to make a 25 nt bubble DNA structure. (B) This was incubated at 55 °C with *Sso*XPF and PCNA, and the uncut substrate separated from the cleavage product by denaturing polyacrylamide gel electrophoresis. Cartoons above the gel indicate the labelled strand (circle) and the position of *Sso*XPF cleavage (arrow). Time points were 1, 5 & 10 minutes; m = A & G Maxam Gilbert size markers of the labelled strand; c = DNA alone; xpf = DNA + *Sso*XPF

12 nt 5' from the 4-way junction (Figure 4.11B&C). Therefore *Sso*XPF specifically cleaved D loop structures in the strand complementary to the invading strand 8 and 12 nt 5' of the 4 way junction. This was very similar to Mus81* which cleaved the equivalent strand 3-8 nt 5' of the 4-way junction (Osman et al., 2003).

In contrast, *Sso*XPF cleaved the z75 strand at the same position in both bubble and D loop structures (Figure 4.11B&C). However, a comparison of relative rates

revealed cleavage was 10 fold faster in the D loop structure compared to the bubble (Figure 4.12; Table 4.2). Therefore the presence of the invading strand stimulated z75 cleavage. The 5' end of the r26-50 strand was relatively close to the z75 cleavage sites (Figure 4.11A) and probably accounted for this stimulation of *Sso*XPF activity. In addition this 5' end may direct and/or stimulate *Sso*XPF cleavage at the b75 D loop major site. Further investigations using D loop structures lacking r26-50 and containing an extended r26-50 strand are necessary to confirm this.

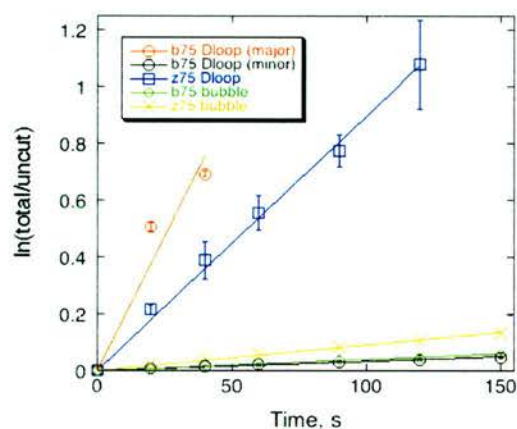


Figure 4.12. Quantifying *Sso*XPF activity for D loop & 25 nt bubble DNA. Standard errors are shown.

Table 4.2. *Sso*XPF rates for D loop & 25 nt bubble DNA. Arrows show the position of cleavage.

SUBSTRATE	STRUCTURE	RATE (min^{-1}) \pm S.E.
D loop b75 (major)		1.1
D loop z75		0.54 ± 0.01
D loop b75 (minor)		0.020 ± 0.001
bubble b75		0.024 ± 0.001
bubble z75		0.054 ± 0.001

In context with the other *Sso*XPF substrates the catalytic constant obtained for cleavage at the major b75 D loop site was between that of the nicked 4-way and the splayed duplex (Tables 4.1 & 4.2). However it must be emphasised that this was an estimated value from few initial time points that did not follow a linear relationship extending through zero (Figure 4.13). A problem in terminating the reaction (i.e. if cleavage went on for longer than anticipated) would explain this. The faster cleavage of this site compared to others would lead to a greater effect on the measurements under the same conditions.

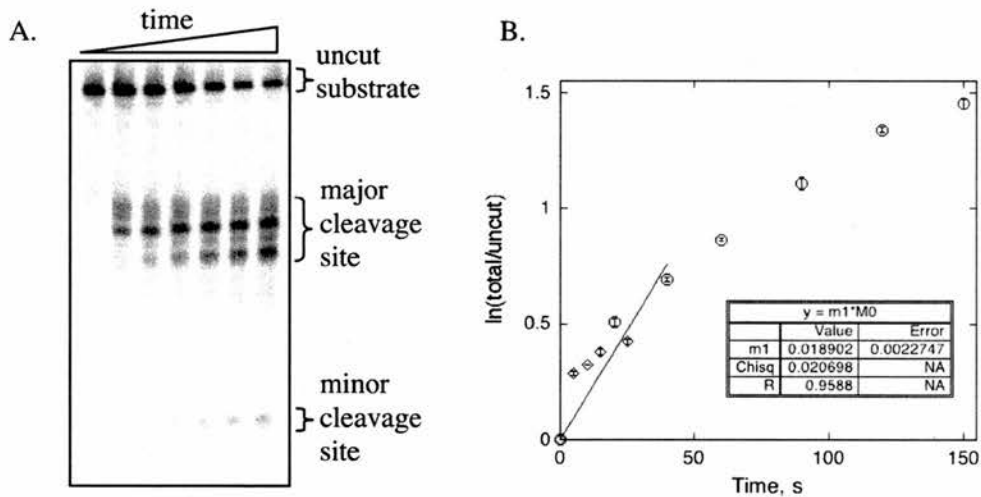


Figure 4.13. Quantifying *Sso*XPF cleavage of D loop b75 strand. (A) Denaturing polyacrylamide gel showing *Sso*XPF cleavage products at time points 0, 20, 40 seconds, 1 minute, 1 minute 30, 2 minutes & 2 minutes 30. Cleavage products at the major and minor sites 1 and 2 were quantified individually in Image Gauge. To determine the catalytic constant (k_{cat}) for cleavage at the minor site, the uncut substrate was measured as uncut substrate + major cleavage site since the major cleavage products were still substrates for cleavage at this site. (B) Calculating k_{cat} for *Sso*XPF cleavage at the major site by linear regression. The relationship is clearly not linear extending through zero. This was probably not due to non-initial rate measurements since a second set of experiments at time points 5-25 seconds was similar (diamonds). The k_{cat} was measured using initial time points only and gave a better than expected Pearson's R value. However, if this failure to extend through zero reflected a problem in terminating the reaction, this value would be an over estimation of k_{cat} .

The minor D loop cleavage sites (b75 strand, 4 nt 5' of the invading strand or bubble, and cleavage of the x50 strand) can be attributed to *Sso*XPFs ability to cleave at distortions and discontinuities in dsDNA (Figure 4.2; Table 4.1). Therefore these sites are unlikely to be relevant *in vivo*.

In summary, *Sso*XPF can cleave D loop structures at 3 sites (Figure 4.14); (a) the strand complementary to the invading strand, 8 and 12 nt 5' of the 4-way junction, (b) in the 5' dsDNA region of the displaced strand, and (c) after second end capture and DNA synthesis. We propose a second model for the role of *Sso*XPF in recombination: resolution of D loops to create cross over products, similar to that proposed for Mus81* (Osman et al., 2003). Again *Sso*XPFs requirement for PCNA for activity may be relevant in determining its D loop target site. Given this recombination scheme requires DNA replication at the 3' invading strand and at the second 3' end captured, PCNA is well placed for *Sso*XPF cleavage at sites a and c. This was proposed to be the case for Mus81*. *In vivo* *Sso*XPF cleavage at site b

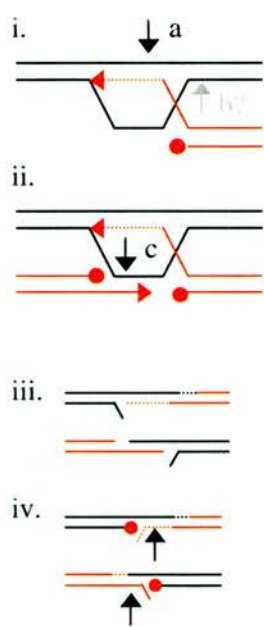


Figure 4.14. Role of *Sso*XPF in recombination - Model 2: resolution of D loop structures to create cross-over recombinants. Invading strands are shown in red. Dotted lines and arrowheads show the direction of DNA synthesis. Arrows represent *Sso*XPF cleavage sites and red circles show the 5' DNA end directing/stimulating cleavage. Following processing of a DSB, strand invasion and synthesis, a D loop is formed (i). This is efficiently cleaved by *Sso*XPF at site a. (Cleavage at site b is discussed in the text). Second end capture and DNA synthesis creates a second substrate for *Sso*XPF, a 3' flap or nicked Holliday junction depending on the extent of invading strand synthesis (ii). *Sso*XPF cleavage at a and c coupled with strand exchange creates gapped and 5' flap intermediates (iii). Gaps are filled in and ligated, and 5' flaps cleaved by Fen1 or isomerise to 3' flaps for a second round of cleavage by *Sso*XPF (iv).

would be disfavoured since PCNA is not loaded on this stretch of duplex DNA. Cleavage at sites a and b would result in breakage of the DNA arm which would be unproductive for D loop resolution.

The model of D loop cleavage by *Sso*XPF shown in Figure 4.14 results in crossover recombination products and genetic rearrangements. An alternative model where *Sso*XPF cleaves only at site a, creating a nick and allowing replication restart in a BIR pathway (without second end capture) is also plausible. This may be more favourable in the context of DNA repair as potentially deleterious genetic rearrangements would be avoided. Unfortunately very little is published regarding recombination and cross over events in *S. solfataricus* and archaea in general.

4.6 *Sso*XPF can cleave DNA processively *in vitro*

Comparing cleavage patterns over time for the 3' flap, splayed duplex and the 5' and 3' overhangs (Figure 4.15), which all have the same cut and uncut strand sequences, it appeared *Sso*XPF could cut some substrates processively. Substrates comprising a 5' flap, double or single stranded, on the uncut strand were cleaved once and the product became a substrate for subsequent cleavage and so on. This was evident from the disappearance of higher sized products and appearance of smaller products with time. The 3' overhang lacking this flap on the uncut strand was cleaved once at 1 of 3 major cleavage sites, shown by the gradual accumulation of these bands with time (Figure 4.16).

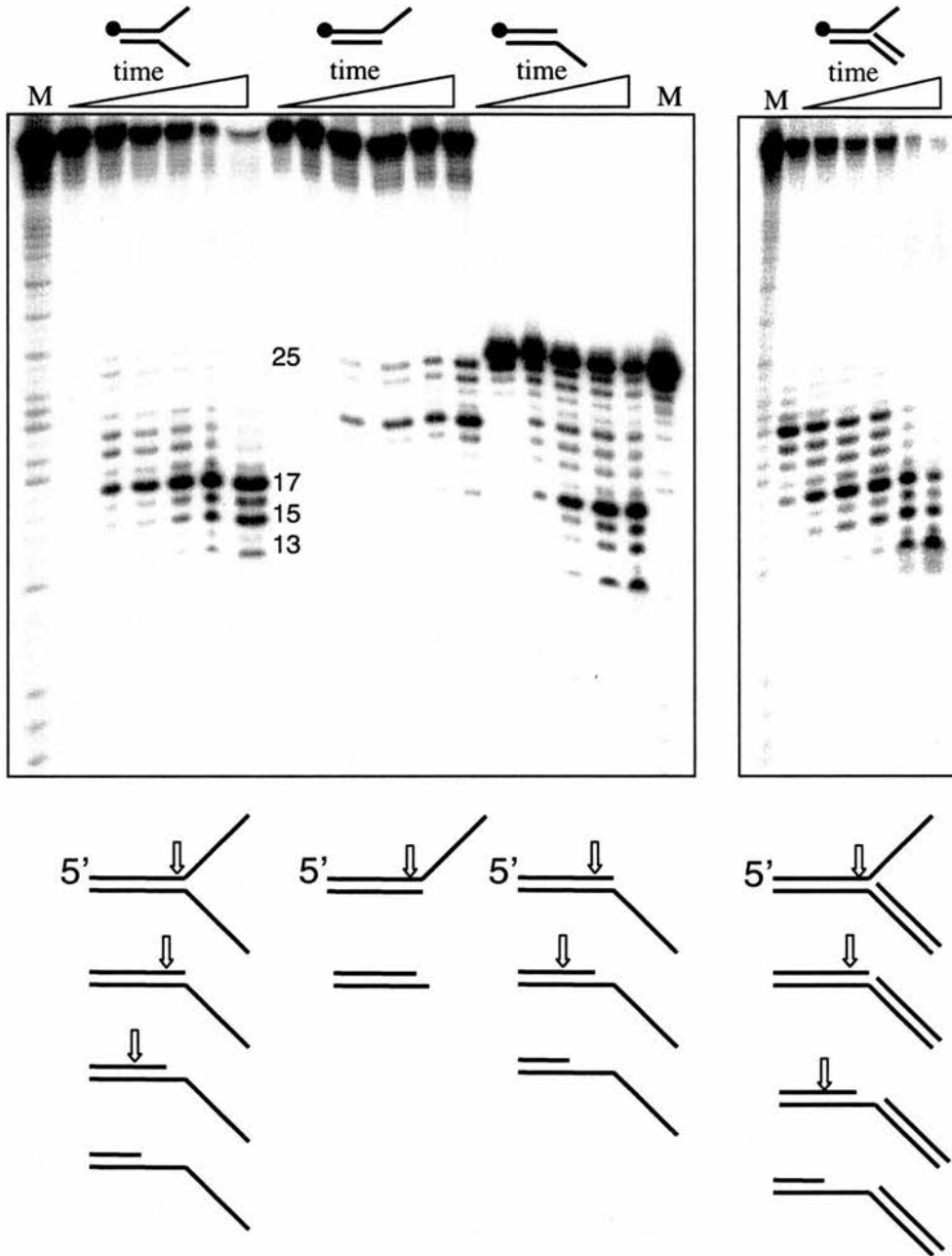


Figure 4.15. Processive cleavage of DNA flap substrates by *Sso*XPF. Time points were as follows: splayed duplex 0, 30 seconds, 1, 2, 5 & 10 minutes; 3' overhang 0, 1, 5, 10, 20 & 40 minutes; 5' overhang 0, 1, 5, 10, 20 minutes; 3' flap 5, 10, 20, 40 seconds, 1 minute 30 & 4 minutes. m = A & G size markers. Circles show the ^{32}P -labelled strand. A schematic representation of *Sso*XPF cleavage is shown below.

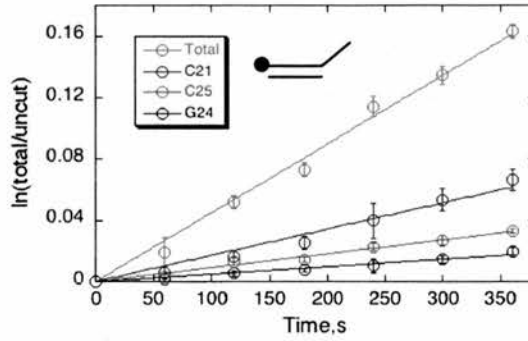


Figure 4.16. Rate of *Sso*XPF cleavage at individual sites in the 3' overhang DNA substrate. k_{cat} values for cutting at bases, \pm standard errors are: C21 = 0.010 ± 0.0004 , C25 = 0.005 ± 0.0002 and G24 = $0.003 \pm 0.0001 \text{ min}^{-1}$. The total rate $0.027 \pm 0.001 \text{ min}^{-1}$ includes other minor sites therefore is not the sum of these 3 values.

Where there was processive cutting, this stopped when the region of duplex DNA reached about 12 nt. This corresponded to the approximate footprint of PCNA and is consistent with a model whereby the PCNA-*Sso*XPF complex slides along the duplex region of DNA away from the flap, with the *Sso*XPF cleaving pyrimidines as it went. When less than 12 bp DNA remained the PCNA-*Sso*XPF complex 'falls off' the DNA end so no further processing was seen. However this appears contrary to the likely *in vivo* position of PCNA at, for example, a 3' flap structure at a stalled replication fork as mentioned in section 4.4

DNA footprinting studies comparing the protection from, for examples, DNaseI or hydroxyl radicals in the presence and absence of *Sso*XPF-PCNA could yield information about the positioning of *Sso*XPF-PCNA on DNA. However given the low binding affinity of both *Sso*XPF and PCNA for DNA, problems with footprinting techniques could be envisaged. An alternative crystallographic approach is underway in collaboration with Professor Jim Naismith, University of St Andrews, to solve the structure of *Sso*XPF and PCNA bound to a DNA substrate.

To investigate the extent of processivity *in vitro*, nicked duplex and 5' overhang substrates with extended regions (75 nt) of 5' duplex DNA were made. Both appeared to be cut processively to approximately 12 or 13 bp from the end (Figure 4.16). Also note the faster initial cleavage of nicked duplex compared to 5' overhang, consistent with the 5' DNA end stimulation described earlier.

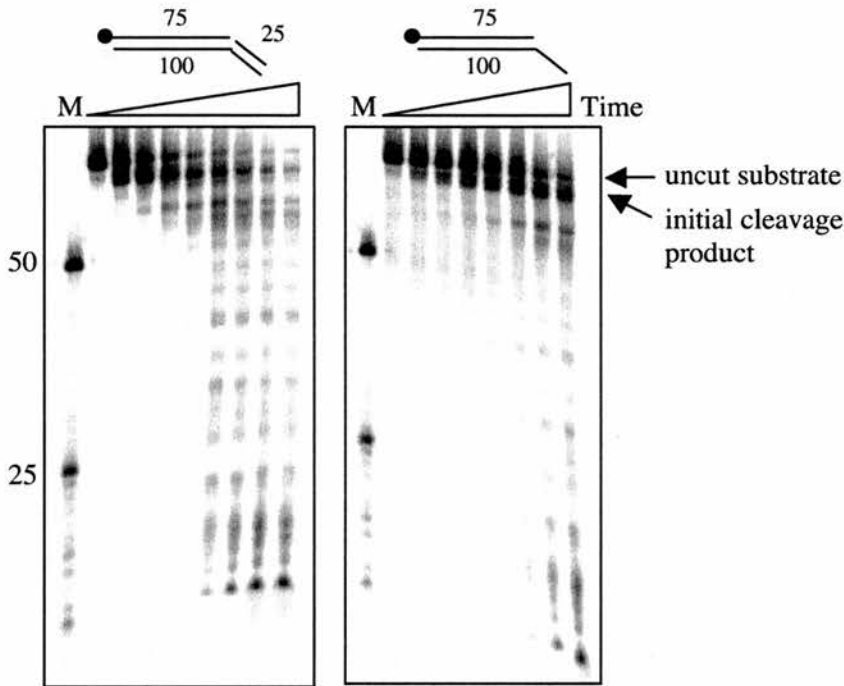


Figure 4.16. Processive cleavage of extended double strand regions of nicked duplex and 5' overhang DNA by *Sso*XPF. Time points were 0, 10, 30 seconds, 1, 2, 5, 10, 20 & 30 minutes with the exception of no 20 minute time point for the 5' overhang. M = 50 & 25 nt DNA markers (there was some degradation of the 25 nt marker).

This processive cutting may be an *in vitro* artefact. *In vivo* there could be a factor or factors that limits processive cutting by *Sso*XPF-PCNA. Alba and SSB, that coat ds- and ssDNA respectively, could potentially affect PCNA-*Sso*XPF procession (see Chapter 6) or other proteins involved in, for example, recognition or unwinding damaged DNA. Alternatively this processivity could represent a very basic form of

DNA repair. The DNA lesion may stall transcription or replication to create a splayed duplex or 3' flap structure. Alternatively another protein factor could recognise the lesion and insert a nick on the 3' side. *Sso*XPF in complex with PCNA recognises this and sequentially cleaves away from the flap or nick, removing the damaged section and allowing resynthesis. This is an attractive prospect given crenarchaea lack obvious homologues of damage recognition and mismatch repair proteins, but obviously further work is necessary to characterise this processivity and its *in vivo* relevance. An interesting experiment would use a duplex DNA substrate containing, for example, a bp mismatch or thymine dimer and a nick or ssDNA arms 3' of this to see if *Sso*XPF could processively cleave past the lesion, thereby removing it. Similar to this XPF-ERCC1 (in the presence of RPA) acts as a 3' to 5' exonuclease during repair of ICLs, digesting past the ICL leading to its uncoupling (Figure 1.8C)(De Silva et al., 2000; Mu et al., 2000).

4.7 Summary & conclusions

Kinetic analysis of the substrate preference of *Sso*XPF clearly shows the best substrates resemble stalled replication forks, D loops and nicked 4-way junctions, which all have a 5' DNA end near the branch point or discontinuity. As a consequence of its high growth temperature *S. solfataricus* suffers a higher frequency of lesions such as hydrolytic deamination, oxidation and single strand breaks (Grogan, 2000), known to stall or collapse replication forks. The substrate specificity for *Sso*XPF was consistent with a role in cleaving stalled replication forks and/or recombination intermediates.

The significant activity of *Sso*XPF against splayed duplex structures together with a domain organisation and sequence specificity similar to eukaryal XPF described in chapter 3 suggests a potential function in archaeal NER. NER pathways in archaea are unknown but most lack bacterial UvrABC homologues so is suggested that eukaryal-type NER systems may exist in archaea (Grogan, 2000; White, 2003).

The processivity of *Sso*XPF cutting is a further intriguing feature but its *in vivo* significance is unknown. Obviously *in vitro* studies of enzyme activity are not ideal as additional factors can affect activity *in vivo*. Phenotypic analysis of *Sso*XPF knockout strains would greatly help in our understanding of the *in vivo* role(s) of this protein. In eukarya XPF functions in several different repair pathways and is in part dependent on interaction with other proteins. Similarly, the specificity of *Sso*XPF *in vivo* may in part be defined by the interactions it makes with other proteins and its interaction with PCNA could provide a mechanism for further protein interactions. This is addressed in chapter 6 but first we continue with the theme of *Sso*XPF activity and look at the mechanism of *Sso*XPF stimulation by PCNA.

CHAPTER 5: *Sso*XPF interaction with PCNA

5.1 Introduction

In chapter 3 we showed a physical interaction between *Sso*XPF and PCNA that was necessary for *Sso*XPF nuclease activity. Further characterisation of *Sso*XPF-PCNA substrate specificity described in chapter 4 highlighted the preference for stalled replication and recombination intermediates, consistent with the expected location of PCNA *in vivo*. In this chapter the interaction between *Sso*XPF and PCNA was investigated further with the aim of defining the role of PCNA in *Sso*XPF nuclease stimulation. Does PCNA simply target and/or hold the otherwise weakly DNA binding *Sso*XPF on DNA enabling cleavage or does it play a role more akin to a cofactor activating *Sso*XPF DNA binding and/or nuclease activity?

5.2 Recombinant PCNA proteins

Initial characterisation of *Sso*XPF described in Chapter 3 used recombinant PCNA heterotrimer protein supplied by Dr Steve Bell (MRC Cancer Cell Unit, Cambridge). Subsequently the pET33b-PCNA subunit 1 and pET30a-PCNA subunits 2 and 3 constructs were donated (also from Dr Steve Bell) for expression and purification of the poly-histidine tagged subunits and assembly of the PCNA heterotrimer (Figure 5.1). Elution of the individual PCNA subunits from gel filtration occurred as a single peak of approximately 30 kDa indicating monomers rather than homotrimers. In contrast a mix of all 3 PCNA subunits eluted at approximately 80 kDa which, together with SDS-PAGE analysis of the concentrated fractions showing stoichiometric amounts of the 3 subunits, indicated the heterotrimer had successfully formed (Figure 5.1B).

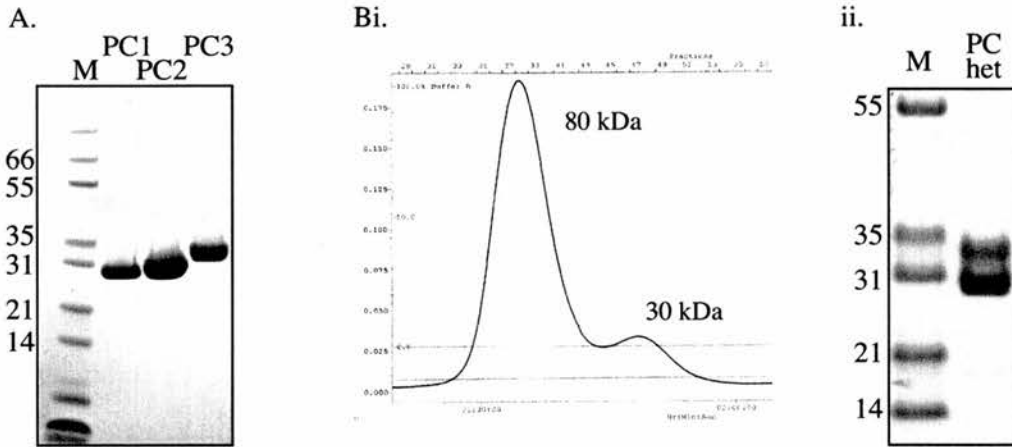


Figure 5.1. SDS-PAGE analysis of purified his-tagged PCNA subunits and assembly of the PCNA heterotrimer. PCNA subunits 1, 2 and 3 were expressed and purified separately by initial heat treatment at 70 °C for 20 minutes followed by a nickel column. Where the individual PCNA subunits were used in assays, an additional gel filtration step was used to give pure proteins (A). PC1, 2 & 3 = PCNA subunits 1, 2 & 3 respectively. To assemble the heterotrimer approximately equal amounts of the 3 PCNA subunits were mixed and trimer (major peak) separated from monomer and dimers (minor peak) by gel filtration (Bi). (Bii) SDS-PAGE gel showing pure PCNA heterotrimer (PC het). M = size markers.

5.3 *Aeropyrum pernix* XPF is active with *S. solfataricus* PCNA

Aeropyrum pernix is a hyperthermophilic crenarchaeon growing optimally at 90-95 °C and pH 7.0. Like *S. solfataricus*, it expresses a small XPF consisting of the basic nuclease core seen in the eukaryal proteins (Figure 1.4). The gene encoding this *Ape*XPF (protein accession number Q9YC15) was amplified by PCR from *A. pernix* genomic DNA, subcloned into the *Nco*I and *Bam*HI sites of vector pET28c and sequenced. Partial *Nco*I digestion of the PCR product was necessary to avoid cleavage at the internal *Nco*I site located at position 146 of the gene. Recombinant *Ape*XPF protein was expressed in *E. coli* and purified by a combination of heating and 2 column steps as described in the methods to produce pure protein (Figure 5.2A).

Under the standard nuclease assay conditions described for *Sso*XPF in Chapter 4, *Ape*XPF alone showed no nuclease activity with the splayed duplex DNA substrate but some activity when *Sso*PCNA heterotrimer was also present (Figure 5.2.B). This further supported a functional interaction between PCNA and XPF in crenarchaea. The low activity of *Ape*XPF was probably due to several factors. Firstly splayed duplex DNA is not the preferred substrate. Subsequent studies using the 3' flap showed stronger cleavage by *Ape*XPF (Newman et al., EMBO, In Press). The assay temperature (55 °C) was low compared to the optimal conditions for this organism. Higher temperatures could not be used due to instability of the DNA substrates. Lastly, subtle variation(s) in the PCNA interaction between species may mean sub-optimal association and decreased stimulation of *Ape*XPF nuclease activity by *Sso*PCNA.

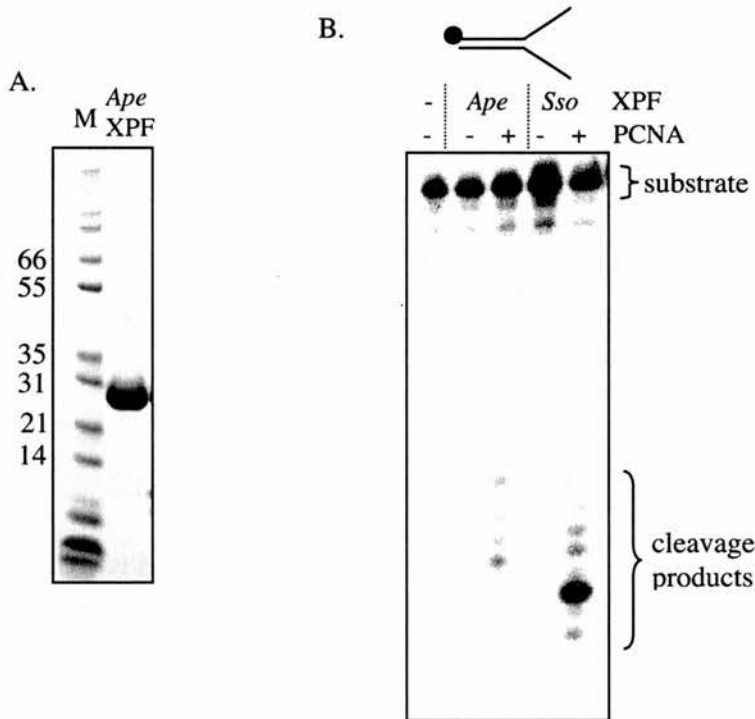


Figure 5.2. *Aeropyrum pernix* XPF. (A) SDS-PAGE gel of recombinant *Ape*XPF protein. M = size markers. (B) Denaturing polyacrylamide gel analysis showing cleavage products resulting from incubation of *Ape*XPF or *Sso*XPF with *Sso*PCNA and splayed duplex DNA at 55 °C for 10 minutes.

Such complementation of PCNA between species has been reported in eukarya where human PCNA can functionally complement *S. pombe* PCNA (Waseem et al., 1992) and also between eukarya and euryarchaea where *P. furiosus* PCNA was shown to interact functionally with calf thymus DNA polymerase delta and human replication factor C (Ishino et al., 2001).

5.4 Stoichiometry & affinity of *Sso*XPF binding to PCNA

The *Sso*XPF dimer has 2 PCNA binding motifs and the PCNA heterotrimer has 2 *Sso*XPF binding sites (subunits 1 and 3), therefore a number of PCNA-*Sso*XPF binding models are possible. One *Sso*XPF dimer could interact with 1 PCNA heterotrimer, either through 1 or both of its PCNA interaction motifs. Alternatively 1 *Sso*XPF dimer could interact with 2 PCNA heterotrimers or 2 *Sso*XPF dimers with the same PCNA heterotrimer.

To attempt to determine how *Sso*XPF interacted with PCNA isothermal titration calorimetry (ITC) was used (see section 2.6.1 for ITC background). Here the ligand was *Sso*XPF and was injected into the sample cell containing PCNA heterotrimer. Heat changes associated with binding were measured (Figure 5.3 top panels). The heats for each individual injection were corrected for dilutions (which in the absence of a complete set of control experiments was assumed to be the endpoint reached with titrations), integrated with respect to time and plotted against molar ratio of PCNA heterotrimer to *Sso*XPF dimer (Figure 5.3 lower panels). The stoichiometry (i.e. the number of *Sso*XPF dimers binding to each PCNA heterotrimer, n) was limited to possible values of 0.5, 1 or 2 for curve fitting (Figure 5.3). The dissociation constants

(K_D) and chi square values derived from models fitting the data are shown in Table 5.1.

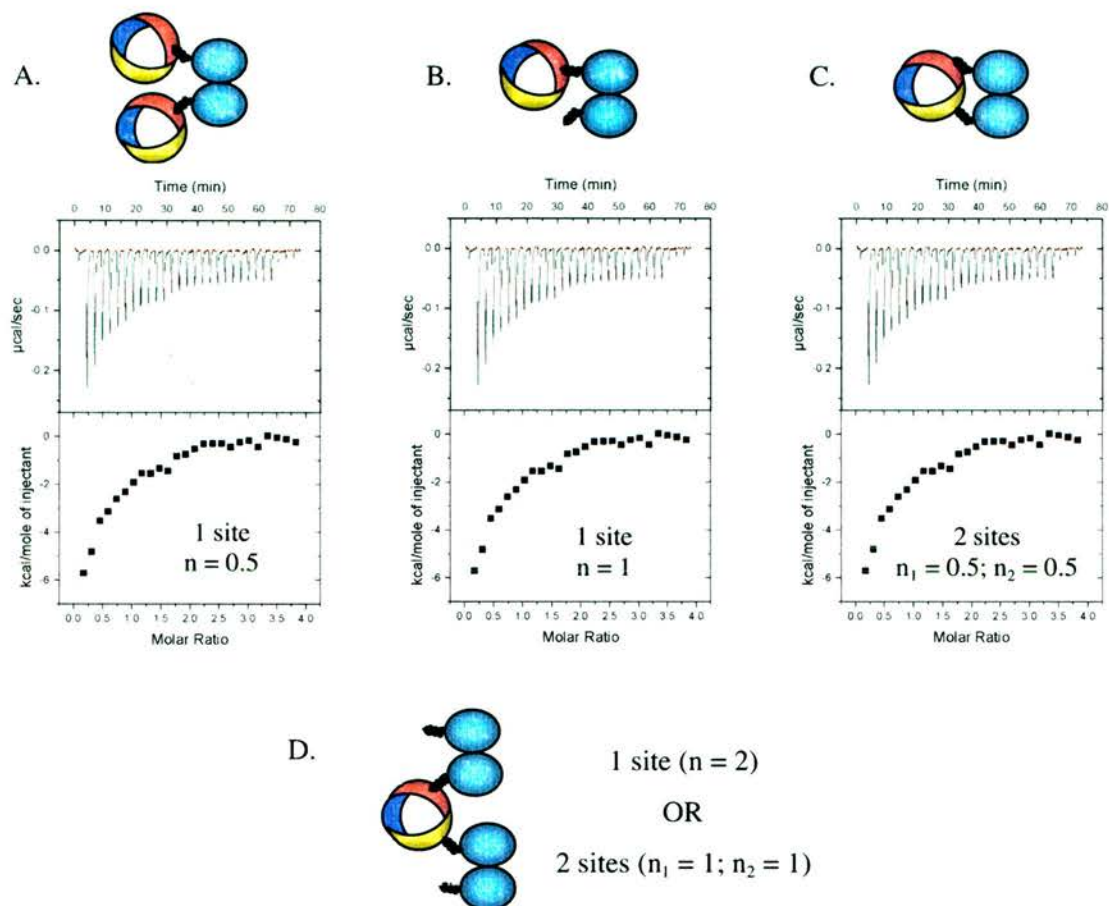
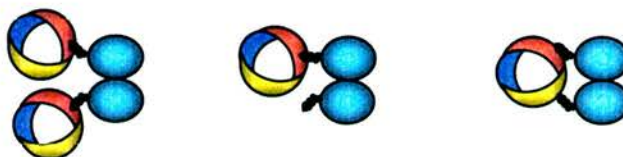


Figure 5.3. Isothermal titration calorimetry of *Sso*XPF binding to PCNA fitted to different binding models. A cartoon of the binding model is shown above each plot. The PCNA subunits 1, 2 and 3 are represented as red, blue and yellow components of the ring respectively. Blue ovals represent *Sso*XPF monomers and show the C-terminal PCNA interacting tail. Upper panels of each plot show the raw data for sequential 10 µl injections of *Sso*XPF into 5 µM PCNA heterotrimer (A inset = *Sso*XPF injections into buffer control). Lower panels show integrated heat data with a theoretical fit to (A) 1 type of binding site per PCNA heterotrimer with stoichiometry (n) of 0.5 *Sso*XPF dimers to each PCNA heterotrimer; (B) 1 type of binding site per PCNA heterotrimer with $n = 1$; (C) 2 types of PCNA binding sites with $n_1 = 0.5$ & $n_2 = 0.5$. The final 3 injections were of smaller volume and ignored in the curve fitting. (D) The data did not fit to an alternative model where 2 *Sso*XPF dimers bound each PCNA.



	1 SITE ($n = 0.5$)	1 SITE ($n = 1$)	2 SITES ($n_1 \& n_2 = 0.5$)
K_D (1) μ M	2.7 ± 0.16	1.2 ± 0.1	3.3 ± 1.3 mM
K_D (2) μ M	-	-	2.0 ± 0.1
Chi square	3.8×10^4	6.9×10^5	3.2×10^4

Table 5.1. Isothermal titration calorimetry data for *Sso*XPF binding to PCNA. Cartoons of the binding models are shown above each column as in the previous figure. K_D s are in μ M unless otherwise stated, and standard errors are shown.

The data did not fit with binding models where 2 *Sso*XPF dimers bound each PCNA heterotrimer with either the same affinity (1 site; $n = 2$) or different affinities (2 sites; $n_1 = 1 \& n_2 = 1$). Thus model D shown in Figure 5.3 could be ruled out. The simplest model of 1 *Sso*XPF dimer binding a PCNA heterotrimer through 1 of its PCNA interaction motifs was clearly not the best fit to the data (Figure 5.3B; Table 5.1). A better fit was obtained from 1 *Sso*XPF binding 2 PCNA heterotrimers with different affinities (2 sites; $n_1 = 0.5 \& n_2 = 0.5$)(Figure 5.3A; Table 5.1). However, such close proximity of 2 PCNA heterotrimers would be an unlikely situation *in vivo*. The best fit was obtained if both *Sso*XPF monomers interacted with 1 PCNA heterotrimer (Figure 5.3C; Table 5.1). Here, one affinity was very weak (K_D in the mM range) compared to that for the other site and other binding models (K_D 1-3 μ M) (Table 5.1). Thus perhaps at high concentrations *Sso*XPF favoured binding the second PCNA site given each *Sso*XPF has 2 potential PCNA interacting motifs.

Alternatively the better fit may result from the increased degrees of freedom for fitting with 2 types of sites as opposed to 1.

The stoichiometry was not clear from this ITC data, but it does not appear to be 2 *Sso*XPF dimers interacting with a single heterotrimer. An alternative model is that 1 *Sso*XPF dimer interacts with 2 sites of a PCNA heterotrimer but $n_1 \neq n_2$ i.e. a hybrid of the models shown in Figures 5.3B&C, where the second PCNA interaction motif may or may not have bound PCNA. This requires further investigation. Increased protein concentrations may help to determine the stoichiometry by ITC, along with a complete set of controls (namely injection of buffer into PCNA heterotrimer) to increase the reliability of the corrected heat data. Given the PCNA heterotrimer \leftrightarrow PCNA1-2 + PCNA3 equilibrium the heat effects of the constant association and dissociation of PCNA 3 may contribute significantly to the heats of reaction.

The dissociation constant (K_D) for the interaction appeared to be 1-3 μ M. In other words at a PCNA concentration of between 1 and 3 μ M half the *Sso*XPF would be bound to PCNA. This was much weaker than the PCNA binding affinities of between 10-60 nM reported for a variety of eukaryal proteins (reviewed (Maga and Hubscher, 2003)). The sub-optimum temperature used in our ITC experiments, 55 $^{\circ}$ C, compared to *in vivo* temperatures of about 80 $^{\circ}$ C may have contributed to this.

ITC was also used to assess binding of *Sso*XPF to PCNA subunit 1 alone. However no binding was obvious from the data (Figure 5.4) suggesting weaker if any binding and the need for increased protein concentrations that were not attainable.

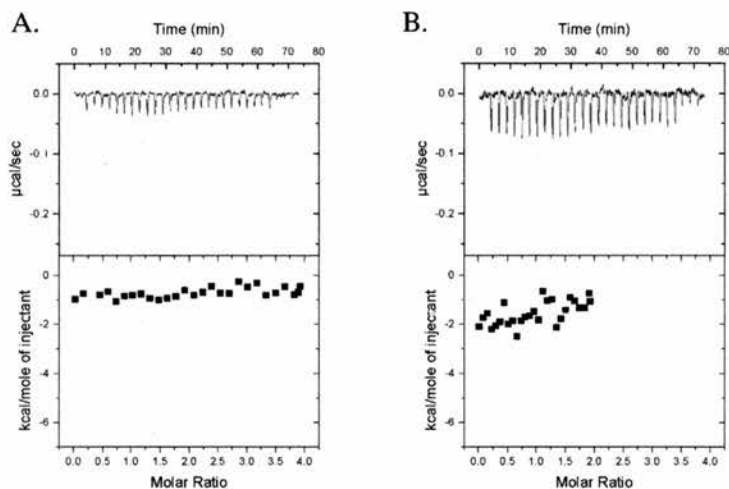


Figure 5.4. *Sso*XPF binding to PCNA subunit 1. Isothermal titration calorimetry for sequential 10 μ l injections of *Sso*XPF into (A) 5 μ M PCNA subunit 1 or (B) 10 μ M PCNA subunit 1. No binding is obvious from this data. Again the final 3 injections were of smaller volume hence smaller signal.

5.5 *Sso*XPF requires PCNA subunits 1 & 3 for activity

Next the minimal PCNA requirement for stimulation of *Sso*XPF nuclease activity was investigated by comparing *Sso*XPF activity with different combinations of PCNA subunits (Figure 5.5). A similar *Sso*XPF activity was apparent in the presence of assembled PCNA heterotrimer and a mixture of the 3 PCNA subunits. No *Sso*XPF activity was detected with each PCNA subunit alone, a mixture of PCNA subunits 1&2 or a mixture of PCNA subunits 2&3. However some activity was seen in the presence of PCNA subunits 1&3, consistent with earlier interaction studies showing *Sso*XPF bound to PCNA subunits 1 and 3 but not significantly to subunit 2 (Figure 3.8). The rates for *Sso*XPF cleavage in the presence of all 3 PCNA subunits was 60 times greater than with subunits 1&3 only (Figure 5.6).

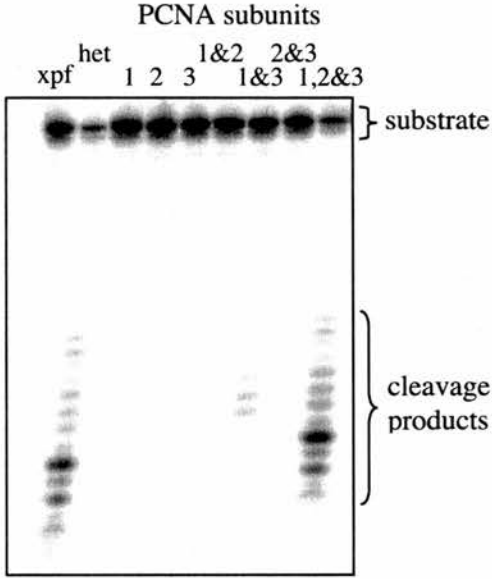


Figure 5.5. *Sso*XPF activity with combinations of PCNA subunits. 80 nM splayed duplex DNA substrate was incubated with 1 μ M *Sso*XPF and 1 μ M PCNA heterotrimer or a combination of PCNA subunits at 55 $^{\circ}$ C for 10 minutes. Cleavage products were separated from uncut substrate by denaturing gel electrophoresis and visualised by phosphoimaging. xpf = DNA + *Sso*XPF; het = DNA + PCNA heterotrimer.

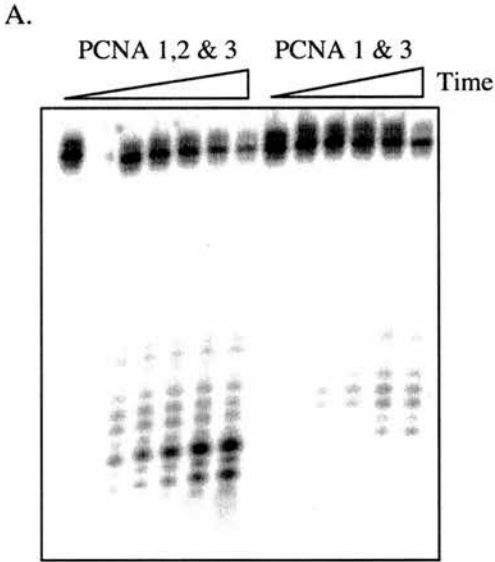
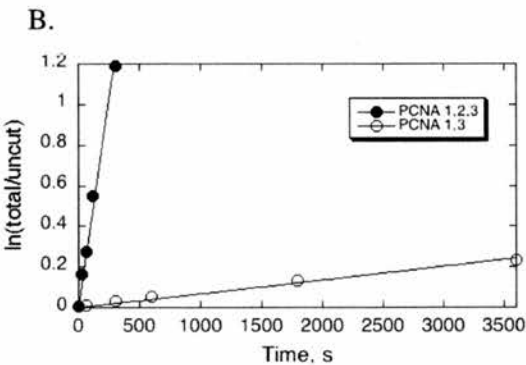


Figure 5.6. Comparing *Sso*XPF activity with PCNA subunits 1, 2 & 3 against activity with 1&3 alone. (A) Denaturing polyacrylamide gel showing cleavage of splayed duplex DNA with time in the presence of all 3 PCNA subunits or subunits 1 & 3 only. 80 nM DNA was incubated at 55 $^{\circ}$ C with 1 μ M *Sso*XPF and 1 μ M of each PCNA subunit used. Time points were 0, 1, 5, 10, 20 & 30 minutes with all 3 PCNA subunits and 0, 5, 10, 20, 30 & 60 minutes with PCNA subunits 1&3. (B) Quantification of *Sso*XPF cleavage with PCNA subunits 1, 2 & 3 (filled circles) and subunits 1&3 (open circles) by linear regression. k_{cat} values were 0.24 & 0.004 min^{-1} respectively.



These results are consistent with a model whereby each monomer of *Sso*XPF interacted with PCNA, one with subunit 1 the other with subunit 3, causing a conformational change and activation of *Sso*XPF. However we cannot rule out the possibility that PCNA trimer formed (i.e. PCNA 1₂3 or PCNA 13₂) and the presence of this PCNA trimer ring accounted for the *Sso*XPF activity observed. Given no significant *Sso*XPF cleavage in the presence of PCNA subunits 1&2 or 2&3 this was unlikely. A similar set of experiments using *Sso*Fen1, which binds to PCNA 1 only (Dionne et al., 2003) would provide a good comparison.

5.6 Loading of PCNA onto DNA

There are 2 possible mechanisms for loading of *Sso*PCNA onto DNA. PCNA could slide onto a DNA end. Alternatively the PCNA heterotrimer may assemble on the DNA given that subunits 1 and 2 associate very tightly and subunit 3 relatively weakly and so in solution there is equilibrium between PCNA heterotrimer and PCNA 1-2 heterodimer with PCNA 3 monomer. To determine the loading mechanism 3' overhang substrates containing a biotin-streptavidin block at both the 5' and 3' end was constructed. Although not the preferred substrate of *Sso*XPF, the 3' overhang was used because PCNA has been shown to slide onto short single stranded DNA (Tom et al., 2000) so if the splayed duplex or 3' flap substrates were used more than 2 biotinylation sites would be necessary to block all possible routes for PCNA sliding onto the DNA.

To confirm that the streptavidin was binding to the biotinylated DNA, electrophoretic mobility shift assays were done under the same conditions. In the presence of streptavidin all biotinylated DNA showed a decrease in mobility

indicating the biotin-streptavidin complex had formed (Figure 5.7). DNA lacking biotin was unaffected.

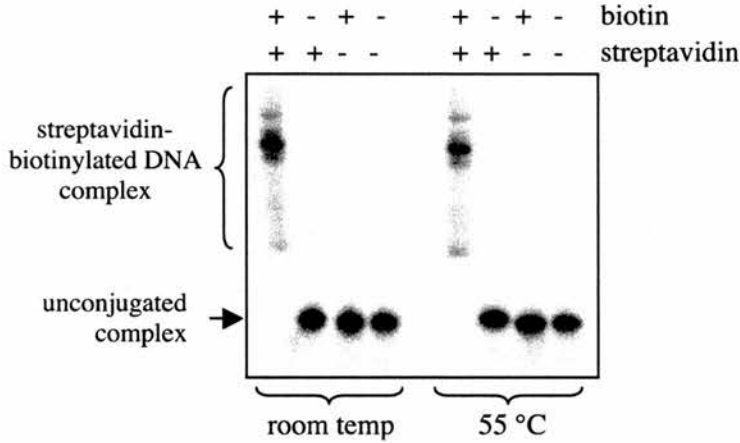


Figure 5.7. Streptavidin specifically bound biotinylated DNA substrates. 80 nM ^{32}P -labelled biotinylated DNA substrate or unbiotinylated control was mixed with a 5 times excess of streptavidin and incubated for 10 minutes at room temperature or 55 °C. Samples were run on a 6 % native acrylamide gel and visualised by phosphoimaging. All the biotinylated DNA in the presence of streptavidin showed decreased electrophoretic mobility consistent with formation of a streptavidin-biotin-DNA complex. Streptavidin did not bind the unbiotinylated DNA.

*Sso*XPF showed no activity in the absence of PCNA but retained nuclease activity in the presence of PCNA regardless of whether both DNA ends were blocked (Figure 5.8). This activity was dependent on an interaction between *Sso*XPF and PCNA since the *Sso*XPF C-terminal $\Delta 6$ mutant, unable to interact with PCNA, showed no activity in the presence of PCNA. This was consistent with a model for PCNA assembling on the DNA.

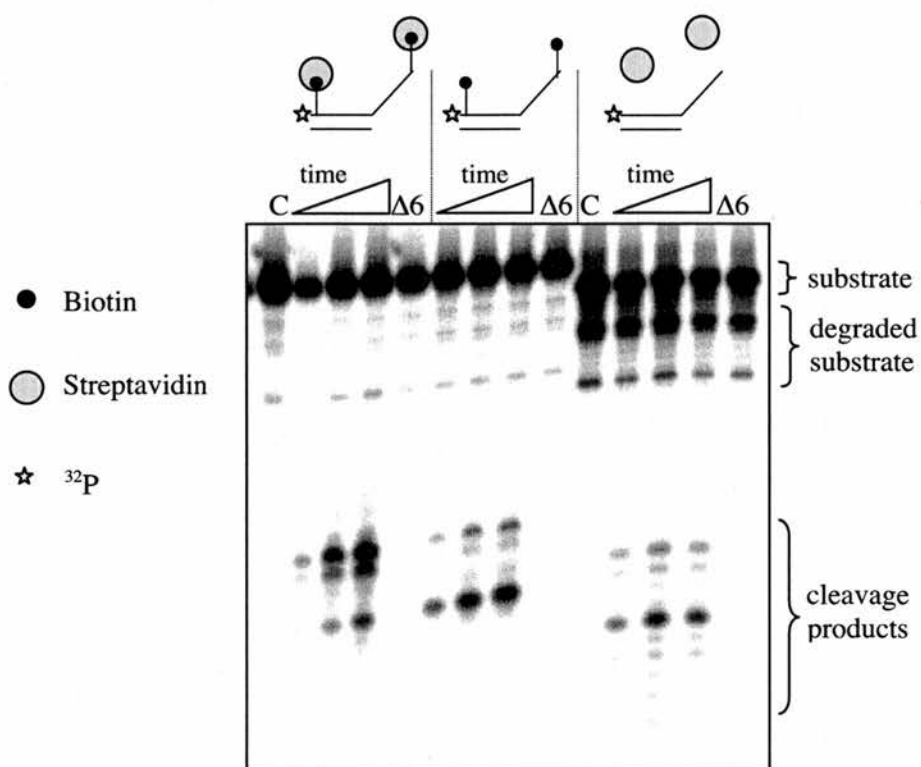


Figure 5.8. PCNA stimulated *Sso*XPF nuclease activity on biotin-streptavidin end blocked DNA substrate. A modified b50 oligo with biotin at positions 3 & 45 was annealed to the x(26-50) oligo (Appendix 1) to make a 3' overhang substrate. Addition of streptavidin resulted in a biotin-streptavidin conjugate at both ends that would prevent PCNA sliding onto the DNA. There was some degradation of DNA substrate that was independent of protein. The C-terminal deletion mutant of *Sso*XPF ($\Delta 6$) that cannot interact with PCNA showed no cleavage. *Sso*XPF cleaved the biotin-streptavidin end blocked DNA in the presence of PCNA showing loading of PCNA onto DNA was not affected when the ends of the substrate were blocked. Time points were 5, 10 & 15 minutes. C = DNA alone.

Fen1 is a highly conserved structure specific endonuclease with roles in lagging strand synthesis (Bambara et al., 1997) and base excision repair (Kim et al., 1998), cleaving double flap and 5' flap structures (Kaiser et al., 1999). Fen1 also interacts with PCNA, which stimulates nuclease activity 5-50 fold (Dionne et al., 2003; Li et al., 1995; Tom et al., 2000; Wu et al., 1996) therefore it was a good positive control for these experiments. The scheme for Fen1 cleavage is shown in Figure 5.9A.

The *S. solfataricus* Fen1 gene was amplified from genomic DNA and subcloned into vector pET28c for native protein expression in *E. coli*. The recombinant *Sso*Fen1 protein was purified to homogeneity (Figure 5.9B) as described in the materials and methods. Blocking both ends of the duplex DNA with biotin streptavidin did not affect stimulation of Fen1 nuclease activity by PCNA (Figure 5.10) consistent with the PCNA heterotrimer assembling on the DNA.

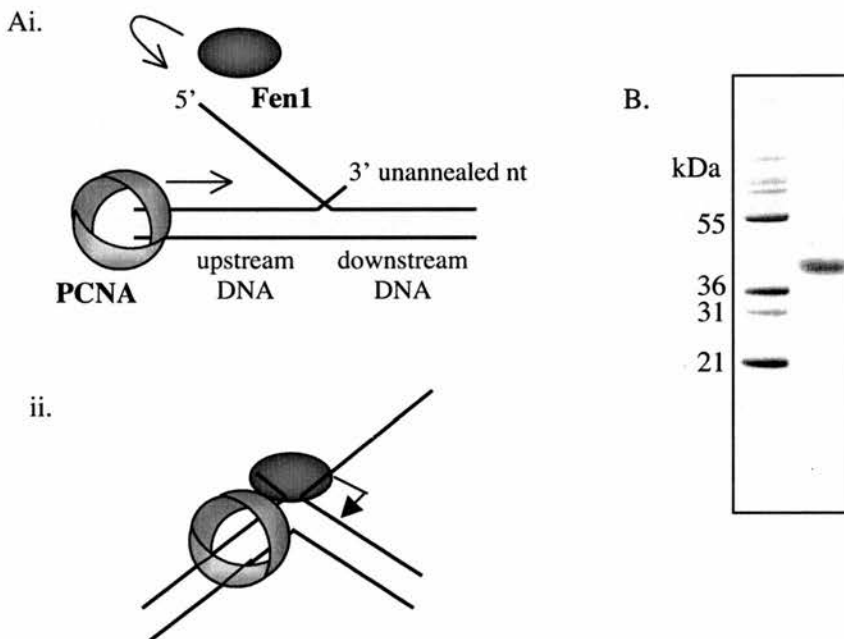


Figure 5.9. Cartoon of Fen1 nuclease activity & SDS-PAGE of purified recombinant *Sso*Fen1. (A) Fen1 cleaves double flap structures occurring when DNA polymerase displaces damaged DNA or RNA to form a 5' ssDNA flap and the newly synthesized and displaced DNA compete for base pairing with template. 5' ssDNA flaps lacking the 3' nt are also cut but more weakly. Fen1 recognises and tracks along the 5' ssDNA flap (i). Upon reaching the branch point Fen1 binds the unpaired 3' nt which positions cleavage to the first base pair preceding the 5' flap, removing it to leave nicked duplex DNA (ii). PCNA loads onto the upstream duplex DNA and stimulates Fen1 cleavage. Fen1 stimulation by PCNA is described in section 5.8. (B) SDS-PAGE gel of purified recombinant *Sso*Fen1.

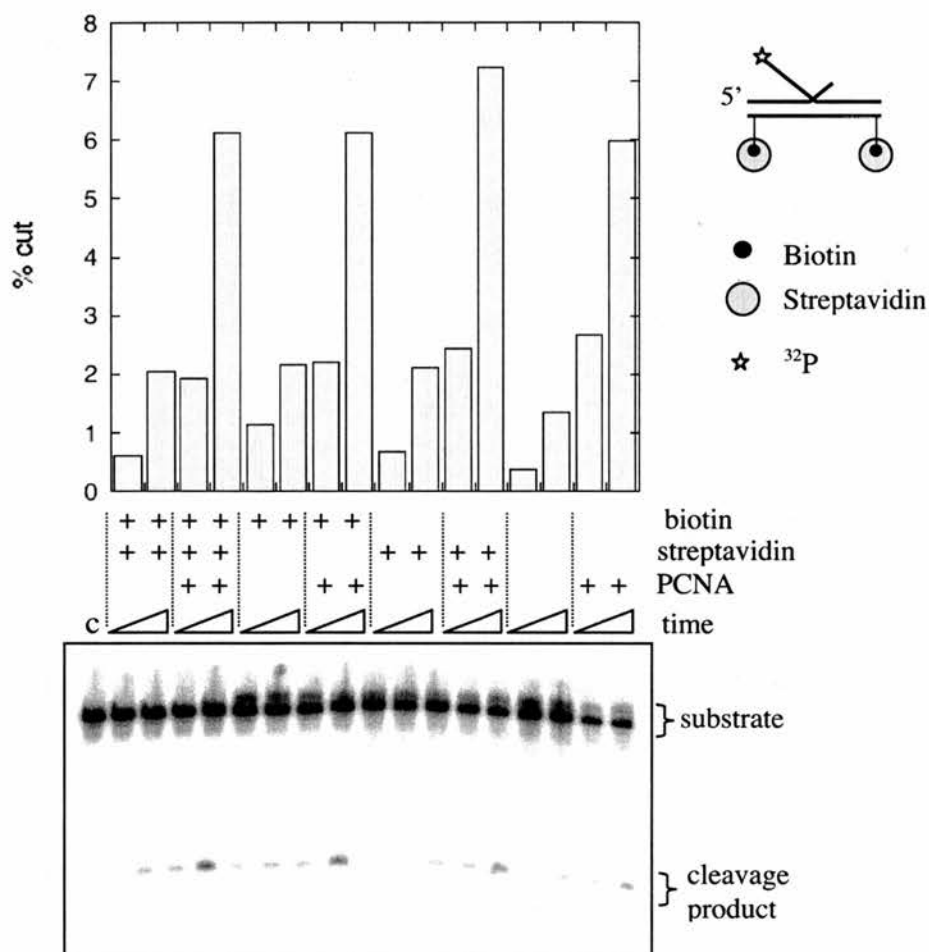


Figure 5.10. *Sso*Fen1 stimulation by PCNA on biotin-streptavidin end blocked double flap substrate. A double flap DNA structure with an unpaired 3' nt overlapping a 25 nt 5' ssDNA flap was made by annealing h(1-26), x50 and b50 biotin oligos (Appendix 1). The star shows the ^{32}P -labelled strand. Addition of streptavidin resulted in a biotin streptavidin block at each end of the dsDNA as indicated. Fen1 was incubated with biotinylated or non-biotinylated DNA in the presence and absence of streptavidin and PCNA for 1 & 3 minutes. Cut and uncut DNA was separated on a denaturing polyacrylamide gel and the percentage cut for each condition quantified. Fen1 stimulation by PCNA was similar regardless of whether the streptavidin block was present.

The ability of *Sso*PCNA to self-assemble on DNA is in contrast to eukaryal PCNA that requires the replication factor C (RFC) protein to open and close the

PCNA ring in an ATP dependent reaction to load it onto and unload it from DNA (reviewed in (Ellison and Stillman, 2001)). Although archaea have a homologue of RFC that interacts with PCNA (Cann et al., 1999; Henneke et al., 2002) both *P. furiosus* and *A. pernix* PCNAs are able to stimulate DNA polymerase synthetic activity on closed circular DNA in the absence of RFC (Cann et al., 1999; Daimon et al., 2002). From the crystal structure of *P. furiosus* PCNA, weakened intermolecular interactions between the PCNA subunits, in contrast to the stable hydrophobic and hydrogen bond interactions at eukaryal PCNA monomer interfaces, was proposed to explain this self-loading of archaeal PCNA (Matsumiya et al., 2001).

5.7 Stimulation of *Sso*XPF nuclease activity by PCNA

Initial characterisation of *Sso*XPF focussed on the splayed duplex as a model substrate and no nuclease activity was detected in the absence of PCNA (Figure 3.11). The dissociation constant of *Sso*XPF binding splayed duplex DNA is about 16 μM (Jana Rudolf, unpublished observation), and initial assays contained only 1 μM , well below this. However even at 50 μM *Sso*XPF, no cleavage of splayed duplex DNA was detected in the absence of PCNA (Figure 5.11Ai). In contrast cleavage products were detected for the same substrate at just 0.1 μM *Sso*XPF when 0.05 μM PCNA was present (Figure 5.11Aii). One hypothesis for *Sso*XPF stimulation by PCNA is that the *Sso*XPF-PCNA complex is the active nuclease form i.e. PCNA has a role akin to a cofactor. In this case, activity would directly reflect the association of *Sso*XPF with PCNA, in turn dependent on protein concentrations.



To investigate this 0.1 μM *Sso*XPF was assayed with increasing concentrations of PCNA to achieve a saturating level of PCNA stimulation that could be quantified (Figure 5.11B). PCNA stimulation of *Sso*XPF nuclease activity was apparent at 0.1 μM PCNA (i.e. a 1:1 ratio of PCNA to *Sso*XPF) and maximal at about 5 μM (a 50 fold excess of PCNA over *Sso*XPF). This data fitted a simple binding curve to give a dissociation constant of $0.60 \pm 0.9 \mu\text{M}$. This was similar to the dissociation constant of 1-3 μM for *Sso*XPF binding to PCNA determined from the ITC experiments earlier and is therefore consistent with a model for PCNA binding to and stimulating or activating *Sso*XPF nuclease activity.

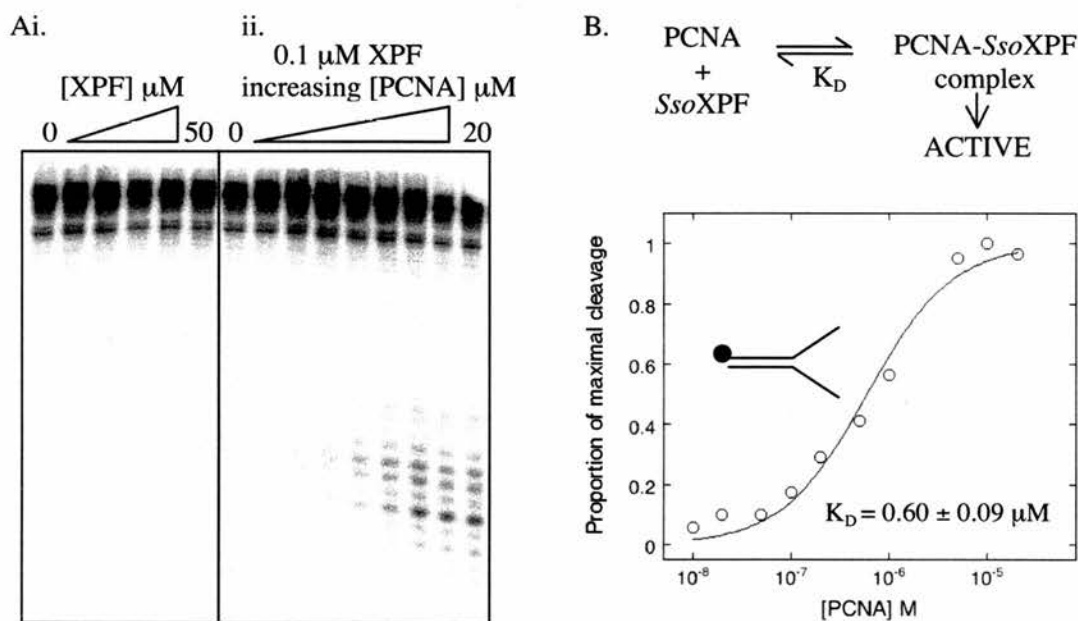


Figure 5.11. PCNA stimulation of splayed duplex cleavage by *Sso*XPF. (A) Denaturing polyacrylamide analysis of cleavage products after incubating DNA with (i) increasing concentrations of *Sso*XPF (0, 1, 10, 20, 35 & 50 μM) and (ii) 0.1 μM *Sso*XPF with increasing concentrations of PCNA (0, 0.01, 0.05, 0.1, 0.5, 1, 5, 10 & 20 μM) at 55 $^{\circ}\text{C}$ for 10 minutes. (B) Quantification of splayed duplex cleaved by *Sso*XPF with increasing PCNA. The proportion of maximal *Sso*XPF cleavage was calculated by dividing the % cut at each PCNA concentration by the maximum % cut. Each point was measured in triplicate and fitted to the binding curve $y = (M_0 * m_1) / (1 + m_1 * M_0)$ where M_0 is the PCNA concentration and m_1 is the association constant (K_A). The dissociation constant (K_D) was calculated: $1/K_A$, \pm standard error.

Further characterisation of *Sso*XPF substrate specificity in chapter 4 revealed a 3' flap structure was the favoured substrate, being cut about 10 fold faster than the splayed duplex in the presence of PCNA (Table 4.1; Figure 4.2). In contrast to the splayed duplex, μM amounts of *Sso*XPF cleaved the 3' flap in the absence of PCNA (Figure 5.12A) and the initial rate of cleavage was linear with respect to *Sso*XPF concentration (Figure 5.12B). This started to tail off at higher *Sso*XPF concentrations as the DNA substrate became saturated with *Sso*XPF, approaching the maximal level of *Sso*XPF activity. Ideally higher *Sso*XPF should have been used to achieve complete saturation, enabling estimation of the maximum *Sso*XPF cleavage rate in the absence of PCNA. The k_{cat} at 20 μM *Sso*XPF was estimated at 0.011 min^{-1} from the 10 minute time point shown in Figure 5.12. Since 20 μM is very close to the K_D of 16 μM for *Sso*XPF-DNA binding (Jana Rudolf, unpublished observation), we'd expect the maximum rate, in the absence of PCNA, to be about twice this i.e. 0.022 min^{-1} . In contrast the k_{cat} for just 1 μM *Sso*XPF in the presence of 1 μM PCNA was over 300 times faster (Table 4.1). Such a large difference is consistent with PCNA directly increasing the activity of *Sso*XPF. If PCNA primarily increased *Sso*XPF binding to DNA, the maximum *Sso*XPF cleavage rate (V_{max}) would not change so significantly in the presence of PCNA. This is explained in more detail in the next section.

The stimulatory effect of PCNA for 3' flap cleavage was measured with 0.01 μM *Sso*XPF and increasing amounts of PCNA to achieve a saturating level of PCNA stimulation that could be quantified (Figure 5.13). PCNA stimulation was again apparent at 0.1 μM PCNA but this time represented a 10:1 ratio of PCNA to *Sso*XPF. Fitting the data to a simple binding curve as before gave a dissociation constant of $0.88 \pm 0.01 \mu\text{M}$, further consistent with PCNA-dependent activation of *Sso*XPF cleavage.

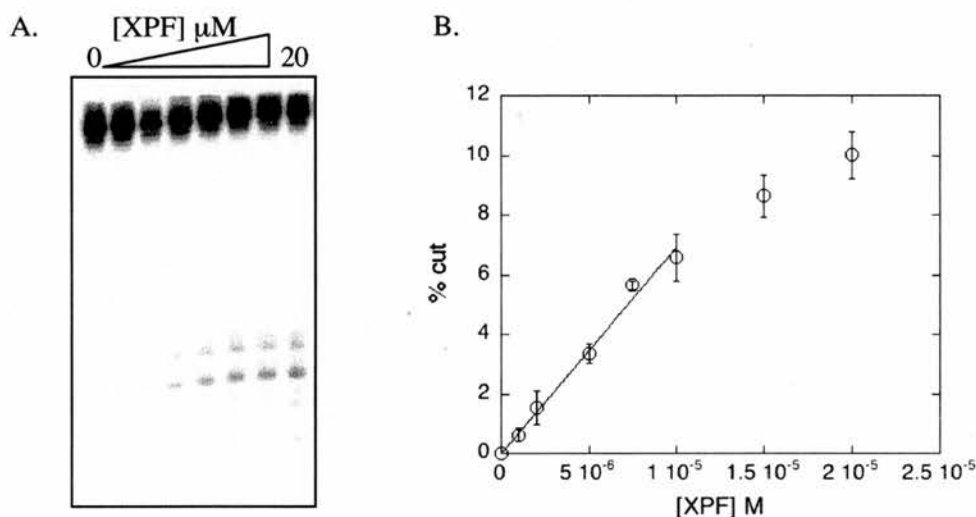


Figure 5.12. *Sso*XPF cleaved 3' flap DNA in the absence of PCNA. (A) Denaturing polyacrylamide analysis of cleavage products after incubating DNA with increasing concentrations of *Sso*XPF (0, 1, 2, 5, 7.5, 10, 15 & 20 μM) at 55 °C for 10 minutes. (B) Quantification of cleavage with increasing *Sso*XPF concentration. Initial time points were fitted to $y = m1 * M0$ where $M0$ is *Sso*XPF concentration and $m1$ is a constant reflecting the dependence of activity on *Sso*XPF concentration. Standard error bars are shown. The k_{cat} at 20 μM *Sso*XPF was estimated from the mean 10 minute time point: $(\ln(\text{total}/\text{uncut}))/10 = 0.011 \text{ min}^{-1}$.

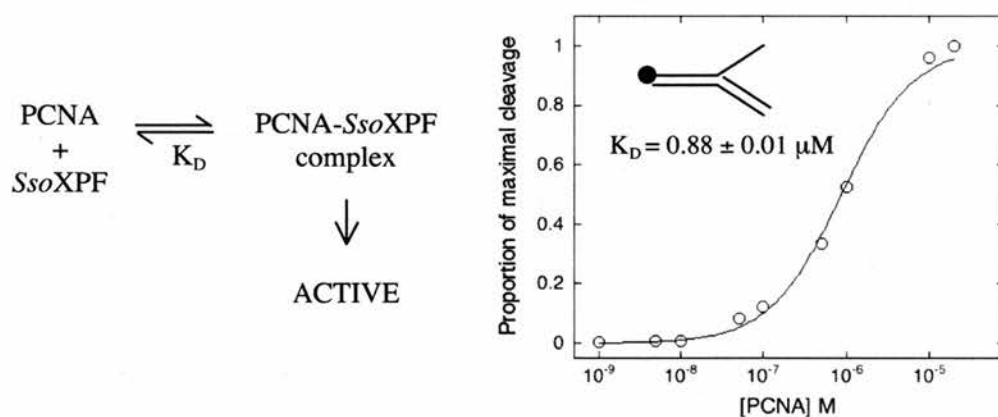


Figure 5.13. PCNA stimulation of 3' flap DNA cleavage by *Sso*XPF. DNA was incubated with 0.01 μM *Sso*XPF and increasing PCNA (0, 0.001, 0.01, 0.1, 0.2, 0.5, 1, 2, 5, 7.5 μM) at 55 °C for 10 minutes. This was done in duplicate. The proportion of the maximal cleavage was calculated as before and plotted against PCNA concentration. Data was fitted to the binding curve $(M0 * m1) / (1 + m1 * M0)$ where $M0$ was the PCNA concentration and $m1$ was the association constant (K_A). The dissociation constant (K_D) was calculated: $1/K_A$, ± standard error.

5.8 Comparing PCNA stimulation of *Sso*XPF & *Sso*Fen1

There are 2 alternatives: (1) PCNA could improve binding of *Sso*XPF to its DNA substrate leading to an increase in activity at sub-saturating concentrations of *Sso*XPF; or (2) PCNA directly increases *Sso*XPF activity. If PCNA improved DNA binding, the maximum reaction velocity (V_{\max}) would not change, but the Michaelis constant (K_m) would decrease. In contrast a large change in V_{\max} would indicate that PCNA increased catalytic activity, which, from crude analyses described in the last section, appears to be the case for *Sso*XPF.

Quantitative analyses were reported for Fen1 stimulation by PCNA and provide a useful comparison for *Sso*XPF (Tom et al., 2000). As mentioned earlier Fen1 cleaves DNA structures with an unpaired 3' nt overlapping a 5' ssDNA flap (Kaiser et al., 1999) and cleavage is stimulated 5-50 fold by PCNA (Tom et al., 2000; Wu et al., 1996). Tom *et al.* showed that PCNA decreased the K_m of Fen1 11-12 fold. Thus the primary effect of PCNA was to increase Fen1 binding to the DNA substrate. Consistent with this they also showed that increasing Fen1 concentration decreased the level of PCNA stimulation. In addition a small, 2 fold increase in V_{\max} suggested PCNA altered the conformation or orientation of Fen1 to give a small increase in catalytic rate. The subsequent crystal structure of Fen1 revealed an interaction between the unstructured C-terminus of PCNA and an unstructured region of Fen1 linking a DNA binding helix and the characteristic PCNA interacting motif, to form an ordered intermolecular β -zipper (Chapados et al., 2004). They proposed that this conformation change enhanced and orientated Fen1 binding, increasing the likelihood of cleavage and accounting for the decreased K_m and slight increase in V_{\max} observed by Tom *et al.* We compared the PCNA stimulation of *Sso*Fen1 to *Sso*XPF to see if it would give us further insight to the mechanism of *Sso*XPF stimulation by PCNA.

Increasing amounts of *Sso*Fen1 were incubated with double flap DNA (Figure 5.14Ai) or the less favoured 5' flap DNA (Figure 5.14Bi) in the absence of PCNA.

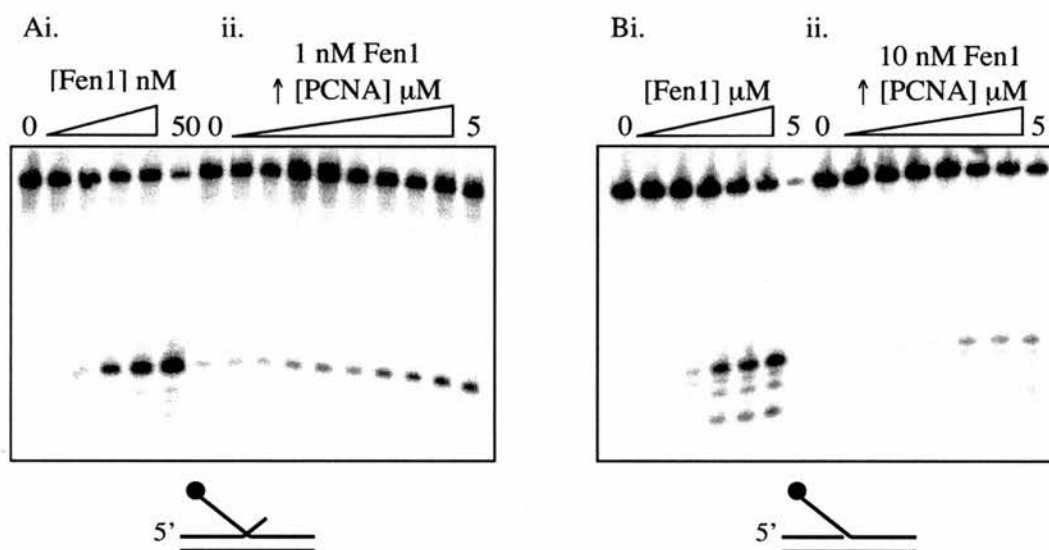


Figure 5.14. PCNA stimulation of double flap and 5' flap cleavage by Fen1. (A) Denaturing polyacrylamide analysis of cleavage products after incubating double flap DNA with 1 nt overlap with (i) increasing concentrations of *Sso*Fen1 (0, 0.1, 1, 5, 10 & 50 nM) and (ii) 1 nM Fen1 with increasing concentrations of PCNA (0, 0.0001, 0.001, 0.01, 0.1, 0.2, 0.5, 1, 2 & 5 μM) at 55 °C for 10 minutes. (B) Denaturing polyacrylamide analysis of cleavage products after incubating 5' flap DNA with (i) increasing concentrations of *Sso*Fen1 (0, 0.001, 0.01, 0.1, 0.5, 1, 5 μM) and (ii) 0.01 μM Fen1 with increasing concentrations of PCNA (0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1 & 5 μM) at 55 °C for 10 minutes.

Fen1 activity with the double flap was approximately 100 fold greater than the 5' flap (Figure 5.15). In addition cleavage was limited to a single site for the double flap DNA compared to 3 sites in the 5' flap. The crystal structure of *Archaeoglobus fulgidus* Fen1 bound to DNA identified a 3' flap binding site (Chapados et al., 2004), which, with the 3' DNA end bound results in conformational ordering of the Fen1 helical clamp around DNA. This both anchors the DNA in the defined orientation for cleavage and extrudes bulk solvent from the active site, explaining the higher Fen1

activity and the cleavage site specificity seen for the double flap substrate (Chapados et al., 2004).

Compared to *Sso*XPF, *Sso*Fen1 showed a much greater nuclease activity at a lower protein concentration in the absence of PCNA i.e. *Sso*Fen1 was intrinsically a more active nuclease (Figure 5.15).

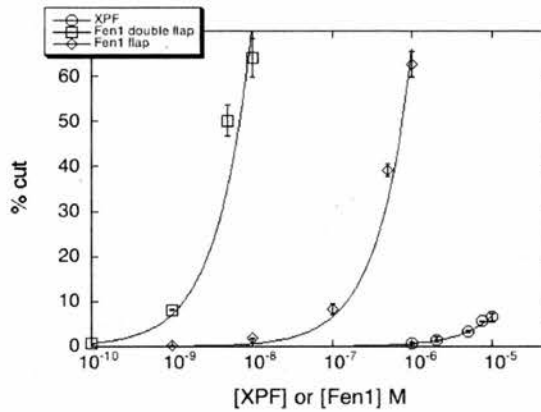


Figure 5.15. Quantification of DNA cleavage by *Sso*XPF and *Sso*Fen1 in the absence of PCNA. Increasing amounts of *Sso*XPF or *Sso*Fen1 were incubated with 100 nM DNA substrate (*Sso*XPF 3' flap; *Sso*Fen1 double flap and 5' flap) for 10 minutes at 55 °C. Cleavage products were separated from uncut substrate by denaturing polyacrylamide gel electrophoresis as shown in Figures 5.12A and 5.14. The % cut was calculated, plotted against protein concentration (log scale) and points fitted to $y = m1 * M0$ where $M0$ is *Sso*XPF or *Sso*Fen1 concentration and $m1$ is a constant reflecting the dependence of activity on protein concentration. This was done in triplicate for each and standard error bars are shown.

*Sso*Fen1 cleavage of its DNA substrates increased with increasing PCNA concentration (Figures 5.14Aii & Bii). The percentage cut was quantified and plotted against PCNA concentration (Figure 5.16A). Similar to *Sso*XPF, *Sso*Fen1 activity was stimulated by at least 0.1 μ M PCNA (i.e. 100 and 10 fold excess PCNA to *Sso*Fen1 for double flap and 5' flap DNA respectively), and was maximal at about 10 μ M PCNA. In the presence of 1 μ M PCNA upwards the % of double flap and 5' flap

DNA cut by *Sso*Fen1 were very similar, reaching a maximum of about 25 % after 10 minutes. An explanation is that PCNA compensated for the lack of a 3' nt in the 5' flap substrate by increasing binding of *Sso*Fen1 to the DNA. Consistent with this, the estimated fold PCNA stimulation of 5' flap cleavage by *Sso*Fen1 was greater than for the double flap DNA (Table 5.2).

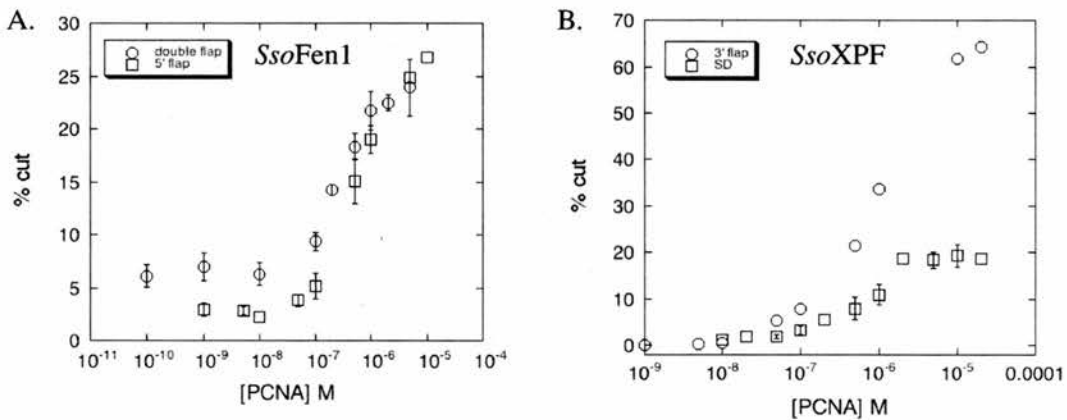


Figure 5.16. Quantifying PCNA stimulation of *Sso*XPF and *Sso*Fen1 activity. (A) PCNA stimulation of *Sso*Fen1 cleavage of double flap (circles) and 5' flap DNA (squares). 100 nM double flap or 5' flap DNA was incubated with 0.001 and 0.01 μ M Fen1 respectively with increasing PCNA for 10 minutes at 55 °C. Cleavage products were separated from uncut substrate by denaturing polyacrylamide gel electrophoresis as shown in Figure 5.14. The % cut was calculated and plotted against PCNA concentration. Standard error bars are shown. (B) Quantification of PCNA stimulation of *Sso*XPF cleavage of 3' flap (circles) and splayed duplex DNA (squares). 100 nM 3' flap or splayed duplex DNA was incubated with 0.01 & 0.1 μ M *Sso*XPF respectively for 10 minutes at 55 °C. Cleavage products were separated from uncut substrate by denaturing polyacrylamide gel electrophoresis as shown in Figure 5.11Aii for the splayed duplex. The % cut was calculated and plotted against PCNA concentration. Standard error bars are shown.

In contrast PCNA could not substitute for 5' DNA end stimulation of *Sso*XPF (Figure 5.16B); 64 % of 3' flap and 19 % of splayed duplex was cut after 10 minutes

in the presence of 20 μ M PCNA. In addition the maximum fold stimulation of *Sso*XPF cleavage of the 3' flap DNA was much greater than that for *Sso*Fen1 (Table 5.2). Such results would be expected if PCNA activated *Sso*XPF nuclease activity.

PROTEIN	[PROTEIN] μ M	DNA SUBSTRATE	k_{cat} min^{-1}	k_{cat} +PCNA min^{-1}	FOLD PCNA STIMULATION
<i>Sso</i> Fen1	0.001	double flap	7.9×10^{-3}	0.030	4
	0.01	5' flap	6.6×10^{-4}	0.029	43
<i>Sso</i> XPF	0.01	3' flap	6.9×10^{-6}	0.10	14,500
	0.1	splayed duplex	-	0.019	-

Table 5.2. Estimated fold stimulation of *Sso*XPF & *Sso*Fen1 nuclease activity in the presence of PCNA. The rate (k_{cat}) for *Sso*Fen1 cutting double flap DNA without PCNA was estimated from the 10 minute control time points, shown in Figure 5.14Aii. The k_{cat} values for *Sso*Fen1 5' flap cleavage and *Sso*XPF 3' flap cleavage were extrapolated from the data shown in Figure 5.15 since no cutting was detected under these conditions in the absence of PCNA. No activity was detected for *Sso*XPF with splayed duplex DNA in the absence of PCNA therefore this k_{cat} could not be calculated. The maximum rate of activity reached in the presence of PCNA was estimated from the 10 minute time points, shown in Figure 5.16: $(\ln(\text{total}/\text{uncut}))/10$, and the fold PCNA stimulation calculated.

However there are problems with these crude kinetic analyses. Firstly different protein concentrations were used. This was necessary to achieve curves approaching saturating levels of PCNA stimulation given 1) the different intrinsic activities of *Sso*XPF and *Sso*Fen1 and 2) the different activities of these proteins for different DNA structures. This would have affected the amount of *Sso*XPF or *Sso*Fen1 bound at a given PCNA concentration. Also, the PCNA stimulation may depend on the nuclease concentration. For instance increasing the Fen1 concentration has the same

effect as adding PCNA because PCNA does not change the basic Fen1 reaction mechanism (Tom et al., 2000). This seems less likely in the case of *Sso*XPF given its much lower intrinsic nuclease activity and much larger fold stimulation by PCNA compared to *Sso*Fen1 (Figure 5.17A; Table 5.2) but further experiments investigating PCNA stimulation at different *Sso*XPF concentrations are necessary.

5.9 Summary & Conclusions

S. solfataricus PCNA is heterotrimeric and monomers can self-assemble on DNA *in vitro*, in contrast to eukaryal homotrimeric PCNA that requires RFC to assemble monomers on DNA. *Sso*PCNA subunits 1 and 3 alone were sufficient for stimulation of *Sso*XPF nuclease activity, consistent with previous interaction studies in chapter 3. A model whereby *Sso*XPF interacts with both PCNA subunits 1 and 3, leading to a conformational change and activation of the nuclease is attractive. Recently Fen1 was shown to undergo a conformation change upon interacting with PCNA, increasing DNA binding (Chapados et al., 2004; Tom et al., 2000). Although a role for PCNA in increasing the stability of *Sso*XPF binding to DNA cannot be ruled out from the data here, there were clear differences between PCNA stimulation of *Sso*XPF and *Sso*Fen1. The presence of PCNA does not change the DNA binding affinity of *Sso*XPF as measured by electrophoretic mobility shift assays (Jana Rudolf, unpublished observation), further supporting PCNA activation of *Sso*XPF nuclease activity. However, further investigation is necessary to confirm the precise role of PCNA in *Sso*XPF nuclease stimulation.

CHAPTER 6: *Sso*XPF interaction with other proteins

6.1 Introduction

In vivo there are many additional protein factors that could potentially affect the activity of *Sso*XPF. All domains of life have proteins that bind single and double stranded DNA. In eukarya the single stranded DNA binding protein RPA has multiple roles in NER (reviewed in (de Laat et al., 1999)). It binds DNA opposite the lesion, aiding unwinding and stabilisation of the open complex, and positions XPG and XPF-ERCC1 for cleavage of the damaged strand (de Laat et al., 1998). *S. solfataricus* encodes a homologue of RPA, SSB, that has the characteristic OB fold for DNA binding and a flexible C-terminal tail for protein:protein interactions (Wadsworth and White, 2001). A role for SSB binding the single strand regions of the repair bubble during NER has been proposed (White, 2003) and we investigated its effect on *Sso*XPF activity in this chapter. The effect of the double stranded DNA binding protein Alba, reported to coat duplex DNA giving protection from nuclease digestion (reviewed in (White and Bell, 2002)), was also tested.

*Sso*XPF specifically interacted with PCNA subunits 1 and 3 (Figure 3.8) therefore it is possible that another factor(s) interact with PCNA 2 simultaneously to form a ternary complex. In chapter 4 we showed *Sso*XPF appeared to have the activities of both eukaryal XPF-ERCC1 in NER and Mus81* in rescue of stalled replication and recombination. An intriguing possibility is that the role of *Sso*XPF is dictated by which protein is bound simultaneously to PCNA 2. The major part of this chapter attempts to identify and characterise such a potential PCNA 2 interacting protein.

6.2 Effects of *S. solfataricus* SSB & Alba on *Sso*XPF activity

Recombinant *S. solfataricus* SSB protein was kindly provided by Dr Liza Cubeddu (Centre for Biomolecular Sciences, University of St Andrews). Figure 6.1 shows complete inhibition of splayed duplex cleavage by *Sso*XPF in the presence of SSB. A small amount of cleaved product was detected at 5 minutes when SSB lacking the C-terminal 29 amino acids proposed to form a protein interacting tail (Wadsworth and White, 2001) was used.

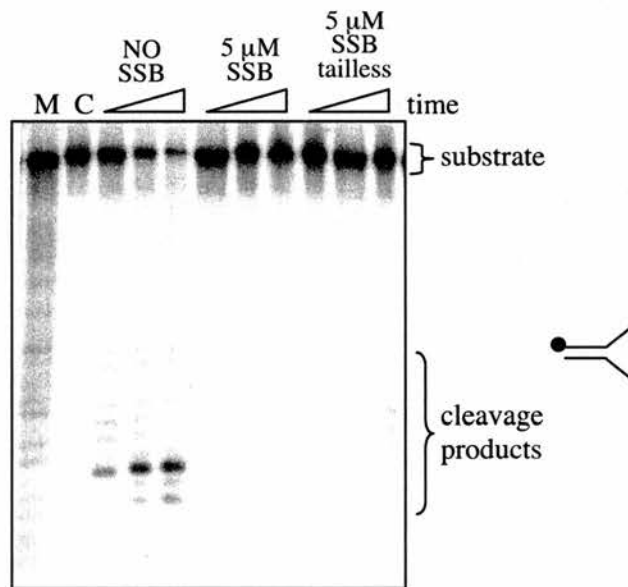


Figure 6.1. Effect of SSB on *Sso*XPF cleavage of splayed duplex DNA. Splayed duplex DNA (80 nM) was incubated with 1 μ M *Sso*XPF and PCNA in the presence of 5 μ M SSB or SSB lacking the C-terminal 29 amino acids (SSB tailless) at 55 $^{\circ}$ C for 1, 2.5 and 5 minutes. Cleaved products were separated from uncut substrate by denaturing polyacrylamide gel electrophoresis. M = A & G size marker. C = DNA + *Sso*XPF.

SSB also inhibited 3' flap cleavage by *Sso*XPF (Figure 6.2A). Initial cleavage products appeared at later time points in the presence of SSB. The cleavage of 3' flap but not splayed duplex DNA in the presence of SSB probably reflects the greater activity of *Sso*XPF for the 3' flap (Table 4.1). Later time points for 3' flap cleavage

in the presence of SSB are necessary to ascertain if SSB limited processivity of cleavage as described in section 4.6.

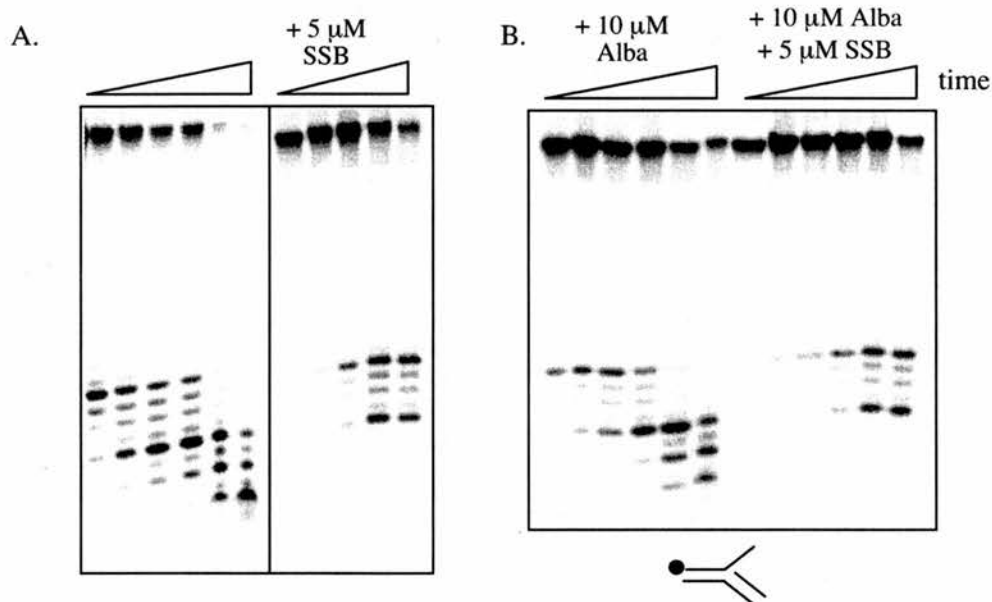


Figure 6.2. Effect of SSB and Alba on *Sso*XPF cleavage of 3' flap DNA. DNA (80 nM) was incubated at 55 °C with 1 μM *Sso*XPF and PCNA in the presence of 5 μM SSB (A), or 10 μM Alba with and without SSB (B). Cleaved products were separated from uncut substrate by denaturing polyacrylamide gel electrophoresis. Time points were 5, 10, 20, 40, 90 & 240 seconds except where 5 μM SSB was used and there was no 90 second time point.

Inclusion of the double stranded DNA binding protein Alba (provided by Dr Clare Jelinska, Centre for Biomolecular Sciences, University of St Andrews) in assays along with *Sso*XPF and PCNA had little effect on 3' flap cleavage (Figure 6.2, compare amounts of uncut substrate). In addition Alba had little effect on the inhibition of that cleavage by SSB (Figure 6.2B).

SSB also inhibited cleavage of nicked duplex DNA (Figure 6.3A) that lacked single stranded arms. This result appeared contrary to a model for SSB binding single stranded flaps and interacting with *Sso*XPF to inhibit cleavage or limit the processive cleavage observed. There are 2 possibilities. The initial cut by *Sso*XPF was

unaffected by SSB but subsequent cuts were inhibited in the presence of SSB, which bound the single stranded DNA exposed after the first cut, interacting with *Sso*XPF to inhibit further cleavage. Alternatively breathing of the duplex DNA could create transient single strand flaps or bubbles that are maintained in the presence of SSB and interfere with PCNA-*Sso*XPF assembly on the duplex DNA (Figure 6.3B). Similarly DNA breathing and SSB binding could change the position of the 5' DNA end so it could not stimulate *Sso*XPF cleavage.

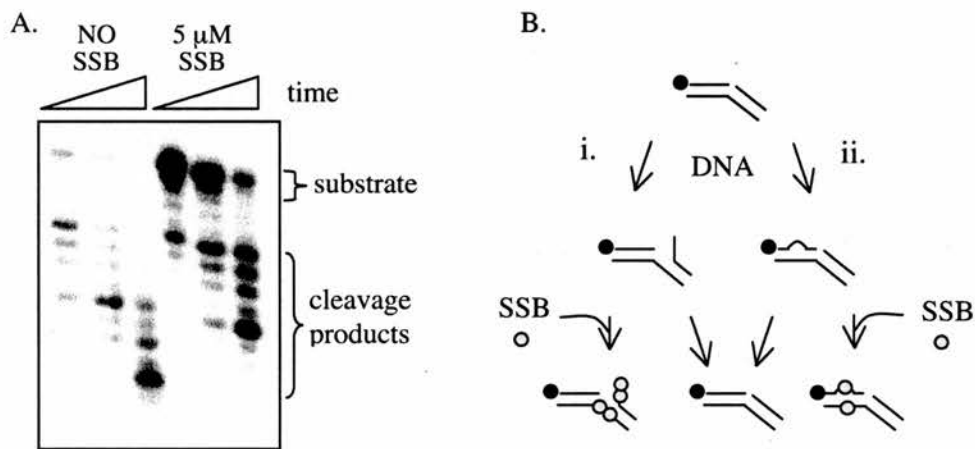


Figure 6.3. Effect of SSB on *Sso*XPF cleavage of nicked duplex DNA. (A) DNA (80 nM) was incubated at with 1 μM *Sso*XPF and PCNA, with and without 5 μM SSB at 55 °C for 10, 40 & 240 seconds. Cleaved products were separated from uncut substrate by denaturing polyacrylamide gel electrophoresis. (B) Model for *Sso*SSB inhibition of *Sso*XPF cleavage of nicked duplex DNA. (i) Breathing of DNA creates a transient 5' flap which is bound by SSB preventing reannealing, thus removing the 5' end that stimulates *Sso*XPF nuclease activity. Similarly 3' flap structures could be affected (not shown). (ii) Breathing and SSB binding could also create bulges that inhibits the assembly of PCNA-*Sso*XPF on the DNA.

6.3 PCNA interacts with SSO1289 & XPB1

*Sso*XPF specifically interacts with PCNA subunits 1 and 3 therefore it is possible that another factor(s) interact with PCNA simultaneously to form a ternary

complex (Figure 6.4A). GST-PCNA subunit 2 fusion protein was used for affinity chromatography with soluble *S. solfataricus* cell extract as described in section 3.6. After SDS-PAGE a band of approximately 55 kDa was apparent (Figure 6.4B arrow). This was excised from the gel and in-gel tryptic digest and MALDI-ToF mass spectrometry identified SSO1289 and *Sso*XPB1 as high scoring hits on 2 and 3 separate runs, respectively.

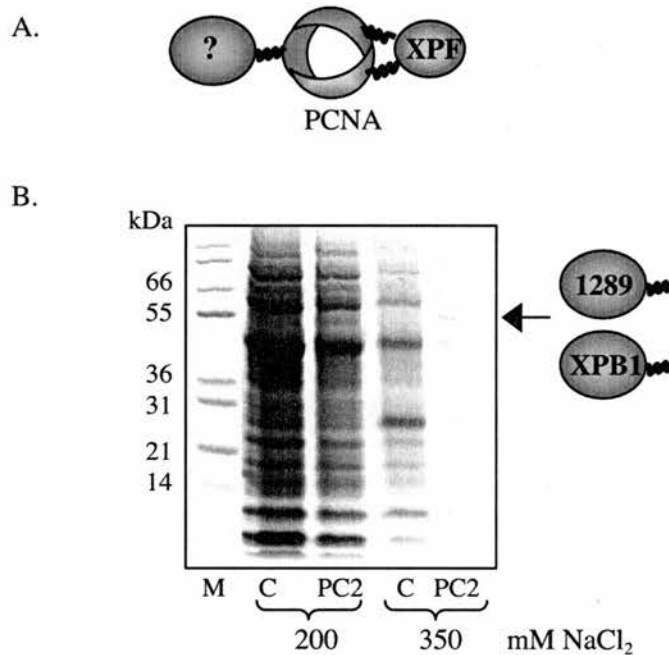


Figure 6.4. Identification of proteins interacting with PCNA subunit 2. (A) Model for a *Sso*XPF-PCNA ternary complex. PCNA subunits 1, 2 & 3 are represented as red, blue and yellow components of the ring respectively. The *Sso*XPF dimer (blue oval) interacts with PCNA subunits 1&3 via 2 C-terminal PCNA interaction motifs, one from each *Sso*XPF monomer. Another protein (green oval) could bind PCNA2 simultaneously, creating a ternary complex. (B) GST-PCNA2 affinity chromatography to identify possible PCNA2 interacting proteins. *S. solfataricus* cell extract was passed over columns containing either the GST control (C) or GST-PCNA2 fusion (PC2). Interacting proteins were eluted by a series of salt washes and separated by SDS-PAGE. The faint band indicated by the arrow was cut out and in-gel tryptic digest and MALDI-ToF mass spectrometry revealed high scoring hits for *Sso*XPB1 and SSO1289.

6.4 SSO1289

SSO1289 is a conserved hypothetical protein of molecular weight 53 kDa. Bioinformatic analyses searching against amino acid sequence only identified other conserved hypothetical proteins. However comparing predicted protein folds by 3D-PSSM (<http://www.sbg.bio.ic.ac.uk/~3dpssm>) identified the bacterial RuvB and UvrB helicases with 14 % and 12 % identity respectively. RuvB is part of the RuvABC bacterial complex that processes recombination intermediates (West, 1997). RuvB is a helicase and together with RuvA promotes ATP-dependent branch migration to form a Holliday junction that is subsequently cleaved by RuvC (Dunderdale et al., 1991). UvrB is involved in bacterial NER as was described in section 1.4.1.

A potential hydrophobic PCNA interacting motif was identified at the N-terminus of the sequence (LLVFTL).

The SSO1289 gene was amplified from *S. solfataricus* P2 genomic DNA, cloned into vector pET28c for native protein expression in *E. coli* and sequenced fully. However problems with solubility in expression trials resulted in the focus being put upon *Sso*XPB1.

6.5 *Sso*XPB1

S. solfataricus XPB1 is a 64 kDa protein, slightly larger than indicated by the SDS-PAGE analysis of GST-PCNA 2 affinity chromatography (Figure 6.4). It is similar to eukaryal XPB with 26 % identity and 47 % similarity and the 7 conserved helicase motifs (Figure 6.7).

In eukarya, XPB (also called Rad25) is a subunit of the TFIIH complex involved in transcription and NER (Svejstrup et al., 1996). Specifically, XPB was shown to

melt 8-10 nt around the promoter during initiation of RNA polymerase II transcription (Douziech et al., 2000) and is thought to be involved in unwinding DNA downstream of the lesion in NER (Evans et al., 1997). However although eukaryal XPB does have weak 3'-5' helicase activity, this is not detected in the TFIIH complex (Coin et al., 1998). In addition the extreme C-terminus of XPB interacts with XPF-ERCC1, promoting 5' incision by recruiting XPF-ERCC1 to the NER lesion (Evans et al., 1997) and helps coordinate open complex formation with incision (Winkler et al., 2001). Mutations in XPB can give rise to 3 inherited disorders in humans, xeroderma pigmentosum, Cockayne Syndrome and trichothiodystrophy (Lehmann, 2003) with a range of transcriptional and NER defects.

Most archaea sequenced to date have 1 or 2 XPB homologues. *S. solfataricus* has 2 that share 58 % amino acid sequence identity and long stretches of nucleotide identity. *Sso*XPB2 is constitutively expressed whereas *Sso*XPB1 mRNA is barely detectable in cells under normal conditions, but shows a 10 fold increase following UV exposure, supportive of a role in DNA repair (Salerno et al., 2003). So whereas *Sso*XPB2 may function in transcription, *Sso*XPB1 may be specifically involved in repair. The initial characterisation of *Sso*XPB1 is described in the remainder of this chapter.

6.6 Expression & purification of *Sso*XPB1

The gene encoding *S. solfataricus* XPB1 (protein accession number Q97ZF9) was amplified by PCR, sequenced fully and subcloned into vector pET28c for expression of the native protein in *E. coli*. This recombinant *Sso*XPB1 protein was purified to homogeneity as described in the methods (Figure 6.6; Figure 6.7A). The

recombinant protein ran at different sizes according with the size marker used but was consistent with the band observed in the GST-PCNA subunit 2 affinity chromatography.

6.7 Confirming the *Sso*XPB1 interaction with PCNA

To confirm the interaction between *Sso*XPB1 and PCNA, small scale affinity chromatography experiments were done. GST-PCNA subunit beads were mixed with recombinant *Sso*XPB1 and washed with buffer to remove non-interacting proteins. Recombinant *Sso*XPB1 was used rather than *S. solfataricus* cell extract because *Sso*XPB1 expression *in vivo* was low. The washed GST-PCNA beads were analysed by SDS-PAGE and Western blot using polyclonal antibodies specific for *Sso*XPB1. Substantial *Sso*XPB1 was detected with the GST-PCNA2 beads compared to only background levels with GST-PCNA 1&3 (Figure 6.5). This confirmed a specific

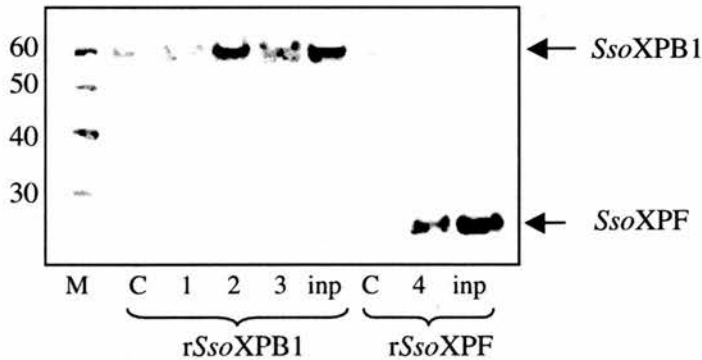


Figure 6.5. *Sso*XPB1 interacted with PCNA. GST control (C) or GST-PCNA subunit 1, 2 or 3 beads (lanes 1, 2 & 3 respectively) were mixed with 1 μ g recombinant *Sso*XPB1 (r*Sso*XPB1) and washed to remove unbound protein. GST-PCNA1 beads mixed with recombinant *Sso*XPF (r*Sso*XPF) was used as a positive control (lane 4). The beads were analysed by SDS-PAGE and Western blot using antibodies specific for *Sso*XPB1 and *Sso*XPF. Inp = 5 % of the recombinant protein used.

interaction between *Sso*XPB1 and PCNA subunit 2. A mutant lacking the N-terminal 12 amino acids comprising the predicted PCNA interaction motif ($\Delta 12$) was made by site directed mutagenesis and tested for an interaction with PCNA as before. This was attempted multiple times but high background binding to GST control (or no signal at all) indicated the procedure needed further fine-tuning.

6.8 Mapping *Sso*XPB1 domains

Protein cross-linking using glutaraldehyde showed *Sso*XPB1 was monomeric (Figure 6.6). Partial digestion of recombinant protein by trypsin and subsequent SDS-PAGE analysis revealed 3 major digestion products of approximately 25 kDa, 40 kDa, and a slightly truncated version of the full length (Figure 6.7A). The site of cleavage was confirmed as lysine 329 using a combination of in gel tryptic digest and MALDI-ToF and N-terminal protein sequencing. The truncation observed occurred at the C-terminus, given the intact N-terminal sequence, but the exact position of cleavage could not be determined.

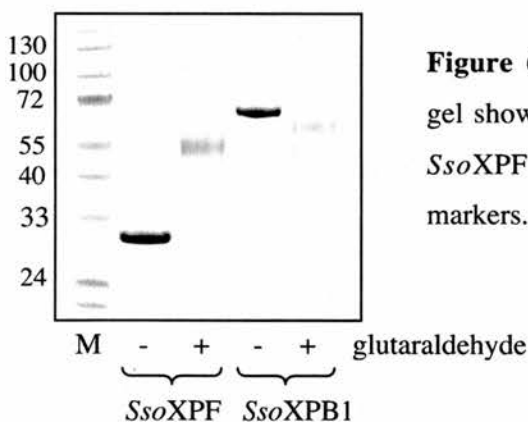


Figure 6.6. *Sso*XPB1 is monomeric. SDS-PAGE gel showing proteins cross linked by glutaraldehyde. *Sso*XPF was used as a positive control. M = size markers.

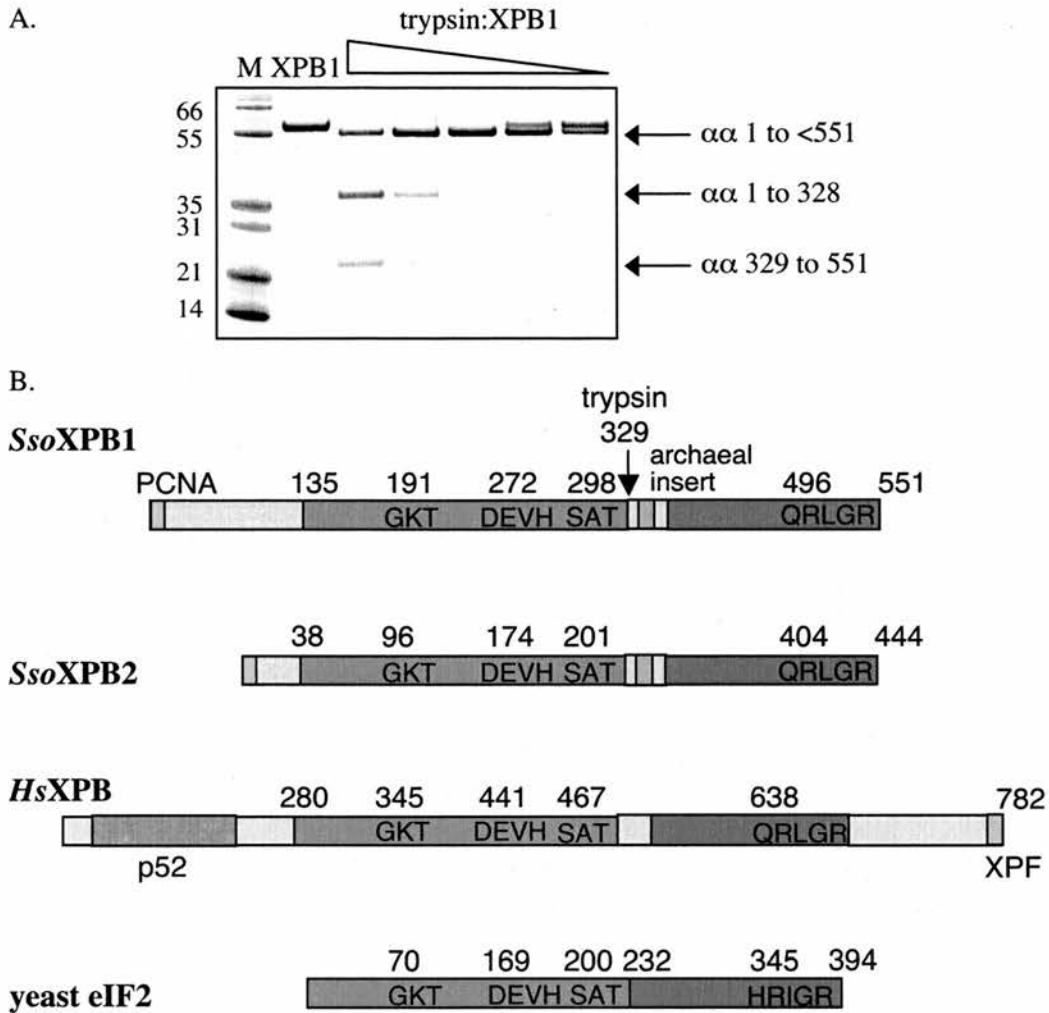


Figure 6.7. *SsoXPB1* domain organization. (A) SDS-PAGE gel showing partial digestion of 5 μ g *SsoXPB1* at 1:25, 1:50, 1:100, 1:200 & 1:400 ratio porcine trypsin to *SsoXPB1*. The 3 products were mapped using in-gel tryptic digest and MALDI-ToF mass spectrometry, and N-terminal protein sequencing. The amino acids ($\alpha\alpha$) represented by each band is shown. M = size markers; XPB1 = undigested *SsoXPB1* control. (B) Comparison between domain structures of *SsoXPBs* 1&2, human XPB (*HsXPB*) and the minimal SFII helicase eIF4A. Position of conserved helicase motifs and predicted protein interaction domains are shown.

Sequence analysis of the N- and C-terminal digestion products revealed *SsoXPB1* comprised the core superfamily II (SFII) helicase motor – 2 RecA-like domains in tandem containing the helicase motifs – with an additional N-terminal

extension (Figure 6.7B). The limited trypsinolysis was consistent with the presence of a flexible linker between the 2 RecA-like domains as exemplified by the minimum SFII helicase yeast eIF4A (Caruthers et al., 2000). The N-terminal extension may confer substrate recognition and/or provide a protein interaction interface (Singleton and Wigley, 2002). These N-terminal extensions vary in length between *Sso*XPB1 and 2 but both sequences have a potential PCNA interaction motif at the extreme N-terminus as discussed previously (Figure 3.10Ai). In eukarya XPB interacts with the TFIIH subunit p52 through the N-terminal extension domain, which anchors it to the TFIIH complex (Jawhari et al., 2002)(Figure 6.7B). Archaea do not have a homologue of p52.

6.9 *Sso*XPB1 specifically bound splayed duplex DNA

Electrophoretic mobility shift assays were used to look at *Sso*XPB1 binding to different DNA structures (Figure 6.8). There was clear binding to a splayed duplex and slight binding to a 10 nt bubble. Although there was disappearance of unbound ss- and dsDNA at high *Sso*XPB1 concentrations there was no obvious band corresponding to bound DNA, only hold up in the wells, so this was considered non-specific. *Sso*XPB1 did not bind either 3' or 5' overhang DNA showing that both the 3' and 5' flaps were necessary for binding. However key structures – 3' flap, replication fork and nicked 3-way junction – were not tested. Thus it is unknown whether *Sso*XPB1 specifically required single strand arms (as would occur during NER) or if 1 or both could be double stranded (as at stalled replication forks). Addition of equimolar amounts of PCNA heterotrimer to *Sso*XPB1 binding assays

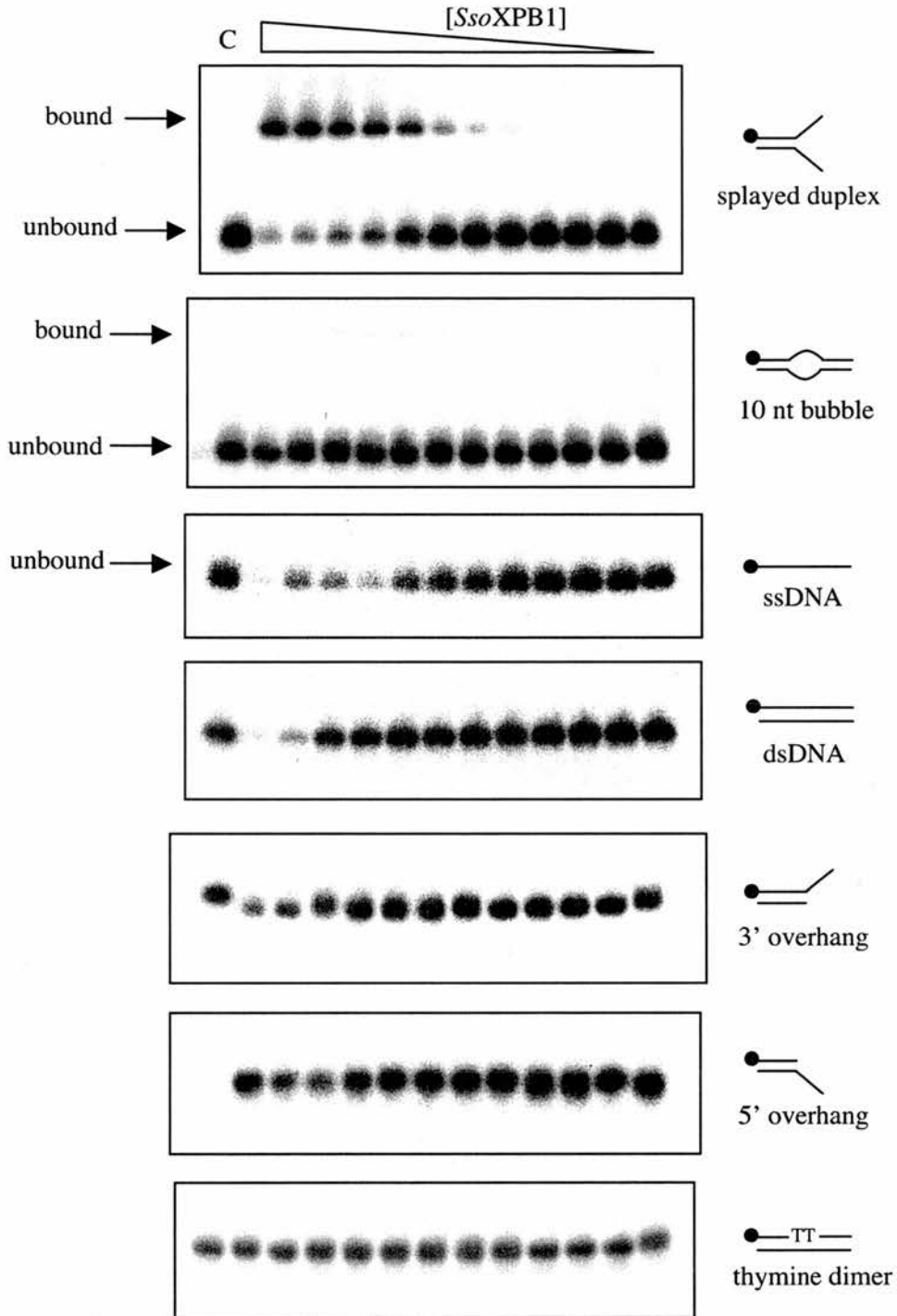


Figure 6.8. Electrophoretic mobility shift analysis of *Sso*XPB1 interaction with different DNA structures. 10 nM ^{32}P -end labelled DNA was incubated with dilutions of *Sso*XPB1 (1000, 500, 200, 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2 nM) for 10 minutes at room temperature. The bound DNA was separated from unbound by native gel electrophoresis and visualized by phosphoimaging. C = DNA alone.

resulted in a small but significant increase in splayed duplex binding affinity (Figure 6.9). Whether this is significant *in vivo* is uncertain.

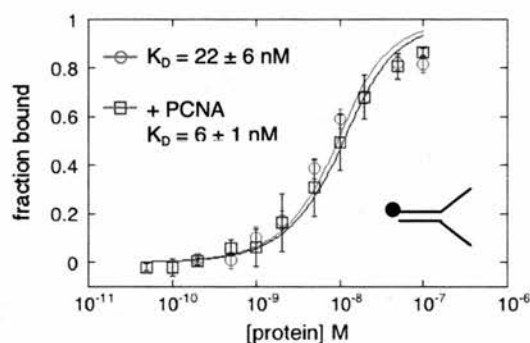


Figure 6.9. Quantification of *Sso*XPB1 binding splayed duplex DNA with and without PCNA. Equimolar concentrations of *Sso*XPB1 and PCNA heterotrimer were used. The fraction of DNA bound was calculated as 1-(unbound/total). This was plotted against the molar protein concentration and the curve fitted to a binding equation as described in the methods.

6.10 *Sso*XPB1 helicase activity

To look for helicase activity, *Sso*XPB1 was incubated with ³²P-labelled splayed duplex DNA in the presence of ATP, with and without PCNA. On 2 occasions a small amount of unwound substrate was seen in the presence of PCNA (Figure 6.10) but this was not consistently reproducible. A major problem was DNA binding by *Sso*XPB1 that remained despite using calf thymus DNA and adding a large excess of unlabelled splayed duplex DNA prior to loading the samples on the gel. In addition the positive DNA control did not show complete separation of DNA strands, suggesting strands were re-annealing despite the 100 fold excess of unlabelled b50 DNA used to stop this. A variety of assay buffers and conditions were used but all showed similar results. However we do not necessarily expect *Sso*XPB1 to show

processive helicase activity, given eukaryal XPB shows only very weak helicase activity.

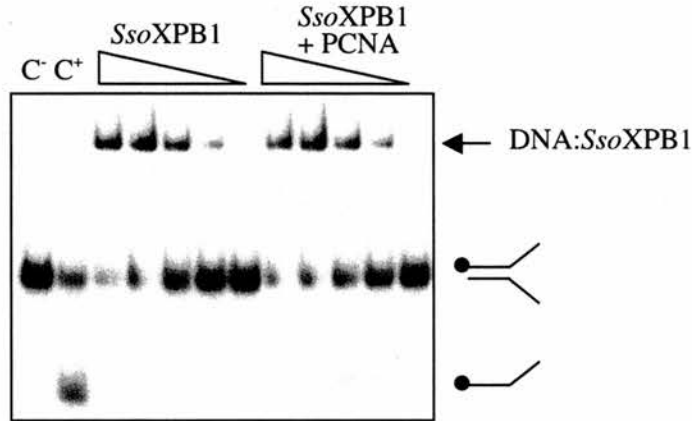


Figure 6.10. *Sso*XPB1 helicase assay. Splayed duplex DNA was incubated with *Sso*XPB1 (1 μ M, 500, 100 50 & 10 nM) with or without equimolar amounts of PCNA, in the presence of ATP at 55 $^{\circ}$ C for 15 minutes. A 100 fold excess of unlabelled J150b oligo was added and reaction incubated for a further 10 minutes to trap unwound DNA. A 100 fold excess of unlabelled splayed duplex DNA was added and substrate and product separated by native acrylamide gel electrophoresis and visualised by phosphoimaging. C⁻ = DNA negative control; C⁺ = boiled DNA, positive control.

6.11 Summary & conclusions

SSB inhibited *Sso*XPF nuclease activity. A large part of this inhibition may have arisen from *Sso*SSB trapping transient ssDNA regions arising during breathing of the DNA, disrupting *Sso*XPF cutting sites. If SSB did indeed inhibit *Sso*XPF nuclease activity, this may be important in the regulation of *Sso*XPF activity during archaeal NER. Alternatively SSB may limit processive cleavage by *Sso*XPF rather than have a general inhibitory effect on nuclease activity. However this would be in contrast to the eukaryal homologue RPA that stimulates XPF-ERCC1 exonuclease activity (Mu et al., 2000). Whether SSB inhibits *Sso*XPF activity and/or processivity requires further study.

Two proteins that interacted with PCNA subunit 2 were identified - the putative helicase SSO1289 and *Sso*XPB1. Initial characterisation of *Sso*XPB1 identified a SFII helicase member that bound specifically to splayed duplex DNA consistent with a role in archaeal NER. *Sso*XPB1 could bind simultaneously to PCNA-*Sso*XPF, targeting the ternary complex to NER sites through its high affinity for splayed duplex DNA structures. Alternatively, if another factor with a high binding affinity for replication fork DNA structures bound PCNA-*Sso*XPF, nuclease activity would be directed to these sites.

This chapter highlights additional avenues for investigation. Further elucidation of both the effect of SSB on *Sso*XPF nuclease activity, and the possibility of functional *Sso*XPF ternary protein complexes via PCNA may help piece together archaeal DNA repair pathways and the role(s) played by *Sso*XPF.

CHAPTER 7: Conclusions & future work

SsoXPF shares structural and functional properties with both eukaryal XPF and Mus81 proteins but also has unique differences. Structurally *SsoXPF* appears more akin to eukaryal XPF containing the core nuclease domain with an N-terminal double HhH domain, in contrast to Mus81 where the 2 HhH domains flank the nuclease domain. However in contrast to both XPF and Mus81 that form active homodimers, *SsoXPF* forms a homodimer and nuclease activity is largely dependent on the sliding clamp protein PCNA with which *SsoXPF* interacts through a conserved C-terminal PCNA interaction motif. *In vitro* the substrate preference of *SsoXPF*-PCNA more closely resembles Mus81*, cleaving DNA structures including 3' flaps, nicked duplex DNA, nicked 3- and 4-way junctions, and D loops. As for Mus81*, a 5' DNA near the branch point or flap is important for cleavage but whether this positions *SsoXPF* cleavage in a similar way was not established. *SsoXPF* also showed significant cleavage of splayed duplex DNA, the favoured substrate of eukaryal XPF, and displayed a similar sequence dependence, cleaving at pyrimidine bases. Thus functionally *SsoXPF* has a range of activities similar to but broader than either eukaryal enzyme suggesting *SsoXPF* may represent an ancestral form that diverged into more specialised enzymes in eukarya (Figure 7.1).

Based on the *in vitro* activity of *SsoXPF* and its similarity to eukaryal XPF and Mus81 proteins, roles in archaeal NER and processing of stalled replication forks and recombination intermediates seem likely. The 3' to 5' processive cleavage by *SsoXPF* raises an interesting possibility: a replication fork stalled at a lesion creates a 3' flap which is processively cleaved by *SsoXPF* past the lesion, thereby removing it. A similar mechanism was proposed for XPF-ERCC1 removal of interstrand cross-links (Figure 1.8C) (De Silva et al., 2000) and perhaps represents the remnants of such a basic DNA repair mechanism. A nick 3' of a lesion could also initiate *SsoXPF*

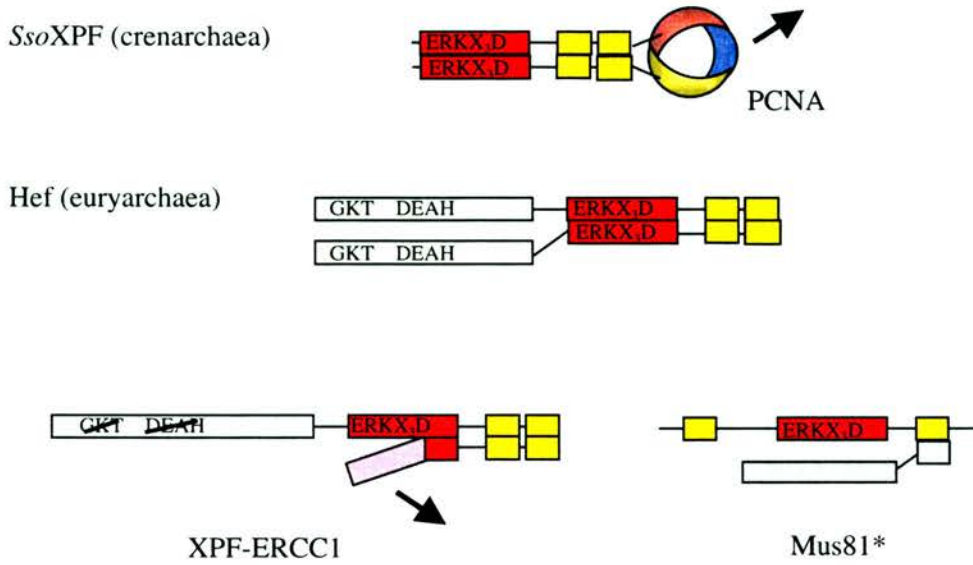


Figure 7.1 Evolutionary scheme for XPF/Mus81 proteins. An ancestral protein resembling crenarchaeal XPF consisted of a nuclease domain (red) and a double HhH domain (yellow), which formed a homodimer and nuclease activity was largely dependent on an interaction with PCNA. This protein had roles in NER, rescuing stalled DNA replication and resolving recombination intermediates. Gene fusion with an SFII helicase (white) dispensed with the PCNA requirement, as represented by euryarchaeal Hef. Gene duplication probably gave rise to the eukaryal XPF partner protein ERCC1, specialized for protein:protein interaction (arrow) necessary for coordinating the more complex eukaryal NER pathway. Gene duplication of the nuclease domain and rearrangement of the 2 HhH domains gave rise to Mus81 that forms a heterodimer, specialized for rescue of stalled DNA replication and resolving recombination intermediates.

processive digestion and removal of the lesion, since *SsoXPF* activity does not depend on the presence of a flap. These are intriguing possibilities since archaea have no obvious NER damage recognition or MMR protein homologues. Whether *SsoXPF* can cleave past DNA lesions is a key experiment lacking here. Obviously *in vitro* studies have their limitations and *in vivo* data would provide greater support for the cellular role(s) of *SsoXPF*. Knocking out the *SsoXPF* gene and comparing the

sensitivity of these mutants to DNA damaging agents such as UV light and chemicals such as hydroxyurea that stall DNA replication, would be one way to address this.

The requirement for PCNA is unique to crenarchaeal XPF proteins and there are several possible roles that PCNA may play.

- 1) PCNA may target or recruit *SsoXPF* to the required sites on the DNA. Indeed, PCNA's involvement in DNA replication means it would be located at sites of stalled DNA replication and at D loop recombination intermediates where the invading 3' DNA end is extended, consistent with *SsoXPF* cleavage sites. Such a role is less certain in archaeal NER since DNA resynthesis must occur after the dual incision.
- 2) PCNA may increase the intrinsically low DNA binding ability of *SsoXPF*, increasing the time that *SsoXPF* is bound to the DNA thus increasing the likelihood of cleavage. This is the mechanism of Fen1 stimulation by PCNA (Chapados et al., 2004; Tom et al., 2000). However in contrast to *SsoFen1*, *SsoXPF* shows substantially lower intrinsic nuclease activity and greater stimulation by PCNA. Also, addition of PCNA to electrophoretic mobility shift assays did not affect DNA binding by *SsoXPF* (Jana Rudolf, unpublished observations).
- 3) Binding to PCNA may activate *SsoXPF* nuclease activity i.e. PCNA may function more akin to a cofactor. Indeed half maximal activity was observed at a PCNA concentration consistent with half the *SsoXPF* binding PCNA. *SsoXPF* interacts with PCNA subunits 1 and 3 and alone these 2 subunits can stimulate *SsoXPF* nuclease activity. Possibly the 2 PCNA interaction motifs in the *SsoXPF* dimer interact with different PCNA subunits, changing conformation of the *SsoXPF* monomers and activating the nuclease.
- 4) PCNA may provide an interface for protein:protein interaction. In crenarchaea PCNA is a heterotrimer therefore subunit specific interaction is possible.

Indeed a functional PCNA-DNA polymerase-DNA ligase1-Fen1 complex was identified in *S. solfataricus* (Dionne et al., 2003). The protein bound simultaneously to PCNA subunit 2 in the PCNA-*Sso*XPF complex could define the function of *Sso*XPF i.e. NER, rescuing stalled DNA replication or resolving recombination intermediates. Similarly, the role of eukaryal XPF-ERCC1 appears dependent on the protein interactions it makes. Two putative helicases were found to interact with PCNA subunit 2. *Sso*XPB1 is a homologue of eukaryal XPB and showed high binding affinity for splayed duplex DNA. Therefore it is possible that *Sso*XPB1 forms a ternary complex with PCNA and *Sso*XPF, directing it to NER sites via its high binding affinity for splayed duplex DNA. Obviously further work is required to confirm existence and functionality of such a complex, with either *Sso*XPB1 or the other putative helicase identified, SSO1289.

REFERENCES

- Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protic, M., Hubscher, U., Egly, J.M. and Wood, R.D. (1995) Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell*, **80**, 859-868.
- Adair, G.M., Rolig, R.L., Moore-Faver, D., Zabelshansky, M., Wilson, J.H. and Nairn, R.S. (2000) Role of ERCC1 in removal of long non-homologous tails during targeted homologous recombination. *Embo J*, **19**, 5552-5561.
- Allers, T. and Lichten, M. (2001) Intermediates of yeast meiotic recombination contain heteroduplex DNA. *Mol Cell*, **8**, 225-231.
- Aravind, L., Walker, D.R. and Koonin, E.V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res*, **27**, 1223-1242.
- Bambara, R.A., Murante, R.S. and Henricksen, L.A. (1997) Enzymes and reactions at the eukaryotic DNA replication fork. *J Biol Chem*, **272**, 4647-4650.
- Bardwell, A.J., Bardwell, L., Tomkinson, A.E. and Friedberg, E.C. (1994) Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. *Science*, **265**, 2082-2085.
- Bardwell, L., Cooper, A.J. and Friedberg, E.C. (1992) Stable and specific association between the yeast recombination and DNA repair proteins RAD1 and RAD10 in vitro. *Mol Cell Biol*, **12**, 3041-3049.
- Barns, S.M., Delwiche, C.F., Palmer, J.D. and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A*, **93**, 9188-9193.

- Bastin-Shanower, S.A., Fricke, W.M., Mullen, J.R. and Brill, S.J. (2003) The mechanism of Mus81-Mms4 cleavage site selection distinguishes it from the homologous endonuclease Rad1-Rad10. *Mol Cell Biol*, **23**, 3487-3496.
- Bell, S.D. and Jackson, S.P. (1998) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol*, **6**, 222-228.
- Bernander, R. (2000) Chromosome replication, nucleoid segregation and cell division in archaea. *Trends Microbiol*, **8**, 278-283.
- Boddy, M.N., Gaillard, P.H., McDonald, W.H., Shanahan, P., Yates, J.R., 3rd and Russell, P. (2001) Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell*, **107**, 537-548.
- Boddy, M.N., Lopez-Girona, A., Shanahan, P., Interthal, H., Heyer, W.D. and Russell, P. (2000) Damage tolerance protein Mus81 associates with the FHA1 domain of checkpoint kinase Cds1. *Mol Cell Biol*, **20**, 8758-8766.
- Bolt, E. and Allers, T. (2004) New enzymes, new mechanisms? DNA repair by recombination in the archaea. *The Biochemist*, **26**, 19-21.
- Bond, C.S., Kvaratskhelia, M., Richard, D., White, M.F. and Hunter, W.N. (2001) Structure of Hjc, a Holliday junction resolvase, from *Sulfolobus solfataricus*. *Proc Natl Acad Sci U S A*, **98**, 5509-5514.
- Cann, I.K., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K. and Ishino, Y. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol*, **181**, 6591-6599.
- Caruthers, J.M., Johnson, E.R. and McKay, D.B. (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A*, **97**, 13080-13085.

- Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B. and Tainer, J.A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell*, **116**, 39-50.
- Chen, X.B., Melchionna, R., Denis, C.M., Gaillard, P.H., Blasina, A., Van de Weyer, I., Boddy, M.N., Russell, P., Vialard, J. and McGowan, C.H. (2001) Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Mol Cell*, **8**, 1117-1127.
- Ciccia, A., Constantinou, A. and West, S.C. (2003) Identification and characterization of the human mus81-eme1 endonuclease. *J Biol Chem*, **278**, 25172-25178.
- Coin, F., Marinoni, J.C., Rodolfo, C., Fribourg, S., Pedrini, A.M. and Egly, J.M. (1998) Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat Genet*, **20**, 184-188.
- Constantinou, A., Chen, X.B., McGowan, C.H. and West, S.C. (2002) Holliday junction resolution in human cells: two junction endonucleases with distinct substrate specificities. *Embo J*, **21**, 5577-5585.
- Cox, M.M. (2002) The nonmutagenic repair of broken replication forks via recombination. *Mutat Res*, **510**, 107-120.
- Daimon, K., Kawarabayasi, Y., Kikuchi, H., Sako, Y. and Ishino, Y. (2002) Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon *Aeropyrum pernix*: interactions with the two DNA polymerases. *J Bacteriol*, **184**, 687-694.
- Dalrymple, B.P., Kongsuwan, K., Wijffels, G., Dixon, N.E. and Jennings, P.A. (2001) A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc Natl Acad Sci U S A*, **98**, 11627-11632.

- Davies, A.A., Friedberg, E.C., Tomkinson, A.E., Wood, R.D. and West, S.C. (1995) Role of the Rad1 and Rad10 proteins in nucleotide excision repair and recombination. *J Biol Chem*, **270**, 24638-24641.
- de Laat, W.L., Appeldoorn, E., Jaspers, N.G. and Hoeijmakers, J.H. (1998a) DNA structural elements required for ERCC1-XPF endonuclease activity. *J Biol Chem*, **273**, 7835-7842.
- de Laat, W.L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N.G. and Hoeijmakers, J.H. (1998b) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev*, **12**, 2598-2609.
- de Laat, W.L., Jaspers, N.G. and Hoeijmakers, J.H. (1999) Molecular mechanism of nucleotide excision repair. *Genes Dev*, **13**, 768-785.
- de Laat, W.L., Sijbers, A.M., Odijk, H., Jaspers, N.G. and Hoeijmakers, J.H. (1998c) Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res*, **26**, 4146-4152.
- de los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J. and Hollingsworth, N.M. (2003) The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics*, **164**, 81-94.
- de los Santos, T., Loidl, J., Larkin, B. and Hollingsworth, N.M. (2001) A role for MMS4 in the processing of recombination intermediates during meiosis in *Saccharomyces cerevisiae*. *Genetics*, **159**, 1511-1525.
- De Silva, I.U., McHugh, P.J., Clingen, P.H. and Hartley, J.A. (2000) Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Mol Cell Biol*, **20**, 7980-7990.

- De Silva, I.U., McHugh, P.J., Clingen, P.H. and Hartley, J.A. (2002) Defects in interstrand cross-link uncoupling do not account for the extreme sensitivity of ERCC1 and XPF cells to cisplatin. *Nucleic Acids Res*, **30**, 3848-3856.
- Dionne, I., Nookala, R.K., Jackson, S.P., Doherty, A.J. and Bell, S.D. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell*, **11**, 275-282.
- Doe, C.L., Ahn, J.S., Dixon, J. and Whitby, M.C. (2002) Mus81-Eme1 and Rqh1 involvement in processing stalled and collapsed replication forks. *J Biol Chem*, **277**, 32753-32759.
- Doherty, A.J., Serpell, L.C. and Ponting, C.P. (1996) The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res*, **24**, 2488-2497.
- Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M. and Coulombe, B. (2000) Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIIH revealed by protein-DNA photo-cross-linking. *Mol Cell Biol*, **20**, 8168-8177.
- Dunderdale, H.J., Benson, F.E., Parsons, C.A., Sharples, G.J., Lloyd, R.G. and West, S.C. (1991) Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. *Nature*, **354**, 506-510.
- Ellison, V. and Stillman, B. (2001) Opening of the clamp: an intimate view of an ATP-driven biological machine. *Cell*, **106**, 655-660.
- Enzlin, J.H. and Scharer, O.D. (2002) The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *Embo J*, **21**, 2045-2053.

- Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M. and Wood, R.D. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *Embo J*, **16**, 6559-6573.
- Fabre, F., Chan, A., Heyer, W.D. and Gangloff, S. (2002) Alternate pathways involving Sgs1/Top3, Mus81/ Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication. *Proc Natl Acad Sci U S A*, **99**, 16887-16892.
- Fishman-Lobell, J. and Haber, J.E. (1992) Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. *Science*, **258**, 480-484.
- Fu, Y. and Xiao, W. (2003) Functional domains required for the *Saccharomyces cerevisiae* Mus81-Mms4 endonuclease complex formation and nuclear localization. *DNA Repair (Amst)*, **2**, 1435-1447.
- Gaillard, P.H., Noguchi, E., Shanahan, P. and Russell, P. (2003) The endogenous Mus81-Eme1 complex resolves Holliday junctions by a nick and counternick mechanism. *Mol Cell*, **12**, 747-759.
- Gaillard, P.H. and Wood, R.D. (2001) Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res*, **29**, 872-879.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman,

- S.L., Newbold, C., Davis, R.W., Fraser, C.M. and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498-511.
- Gary, R., Ludwig, D.L., Cornelius, H.L., MacInnes, M.A. and Park, M.S. (1997) The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J Biol Chem*, **272**, 24522-24529.
- Griffin, S. (1996) DNA damage, DNA repair and disease. *Curr Biol*, **6**, 497-499.
- Grogan, D.W. (2000) The question of DNA repair in hyperthermophilic archaea. *Trends Microbiol*, **8**, 180-185.
- Haber, J.E. and Heyer, W.D. (2001) The fuss about Mus81. *Cell*, **107**, 551-554.
- Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1994) Holliday junction cleavage by yeast Rad1 protein. *Nature*, **371**, 531-534.
- Hadden, J.M., Declais, A.C., Phillips, S.E. and Lilley, D.M. (2002) Metal ions bound at the active site of the junction-resolving enzyme T7 endonuclease I. *Embo J*, **21**, 3505-3515.
- Henneke, G., Gueguen, Y., Flament, D., Azam, P., Querellou, J., Dietrich, J., Hubscher, U. and Raffin, J.P. (2002) Replication factor C from the hyperthermophilic archaeon *Pyrococcus abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RFC PCNA-binding domain. *J Mol Biol*, **323**, 795-810.
- Hollingsworth, N.M. and Brill, S.J. (2004) The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. *Genes Dev*, **18**, 117-125.

- Houtsmuller, A.B., Rademakers, S., Nigg, A.L., Hoogstraten, D., Hoeijmakers, J.H. and Vermeulen, W. (1999) Action of DNA repair endonuclease ERCC1/XPF in living cells. *Science*, **284**, 958-961.
- Hoy, C.A., Thompson, L.H., Mooney, C.L. and Salazar, E.P. (1985) Defective DNA cross-link removal in Chinese hamster cell mutants hypersensitive to bifunctional alkylating agents. *Cancer Res*, **45**, 1737-1743.
- Iftode, C., Daniely, Y. and Borowiec, J.A. (1999) Replication protein A (RPA): the eukaryotic SSB. *Crit Rev Biochem Mol Biol*, **34**, 141-180.
- Interthal, H. and Heyer, W.D. (2000) MUS81 encodes a novel helix-hairpin-helix protein involved in the response to UV- and methylation-induced DNA damage in *Saccharomyces cerevisiae*. *Mol Gen Genet*, **263**, 812-827.
- Ishino, Y., Tsurimoto, T., Ishino, S. and Cann, I.K. (2001) Functional interactions of an archaeal sliding clamp with mammalian clamp loader and DNA polymerase delta. *Genes Cells*, **6**, 699-706.
- Janicijevic, A., Sugasawa, K., Shimizu, Y., Hanaoka, F., Wijgers, N., Djurica, M., Hoeijmakers, J.H. and Wyman, C. (2003) DNA bending by the human damage recognition complex XPC-HR23B. *DNA Repair (Amst)*, **2**, 325-336.
- Jawhari, A., Laine, J.P., Dubaele, S., Lamour, V., Poterszman, A., Coin, F., Moras, D. and Egly, J.M. (2002) p52 Mediates XPB function within the transcription/repair factor TFIIH. *J Biol Chem*, **277**, 31761-31767.
- Jose, T.J., Conlan, L.H. and Dupureur, C.M. (1999) Quantitative evaluation of metal ion binding to PvuII restriction endonuclease. *J Biol Inorg Chem*, **4**, 814-823.
- Kaiser, M.W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L. and Lyamichev, V.I. (1999) A comparison of eubacterial and archaeal structure-specific 5'-exonucleases. *J Biol Chem*, **274**, 21387-21394.

- Kaliraman, V., Mullen, J.R., Fricke, W.M., Bastin-Shanower, S.A. and Brill, S.J. (2001) Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes Dev*, **15**, 2730-2740.
- Kaye, J., Smith, C.A. and Hanawalt, P.C. (1980) DNA repair in human cells containing photoadducts of 8-methoxypsoralen or angelicin. *Cancer Res*, **40**, 696-702.
- Keeling, P.J. and Doolittle, W.F. (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc Natl Acad Sci U S A*, **92**, 5761-5764.
- Kim, K., Biade, S. and Matsumoto, Y. (1998) Involvement of flap endonuclease 1 in base excision DNA repair. *J Biol Chem*, **273**, 8842-8848.
- Komori, K., Fujikane, R., Shinagawa, H. and Ishino, Y. (2002) Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst*, **77**, 227-241.
- Kong, X.P., Onrust, R., O'Donnell, M. and Kuriyan, J. (1992) Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding clamp. *Cell*, **69**, 425-437.
- Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233-1243.
- Kuraoka, I., Kobertz, W.R., Ariza, R.R., Biggerstaff, M., Essigmann, J.M. and Wood, R.D. (2000) Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. *J Biol Chem*, **275**, 26632-26636.
- Kvaratskhelia, M., George, S.J., Cooper, A. and White, M.F. (1999) Quantitation of metal ion and DNA junction binding to the Holliday junction endonuclease Cce1. *Biochemistry*, **38**, 16613-16619.

- Lan, L., Hayashi, T., Rabeya, R.M., Nakajima, S., Kanno, S., Takao, M., Matsunaga, T., Yoshino, M., Ichikawa, M., Riele, H., Tsuchiya, S., Tanaka, K. and Yasui, A. (2004) Functional and physical interactions between ERCC1 and MSH2 complexes for resistance to cis-diamminedichloroplatinum(II) in mammalian cells. *DNA Repair (Amst)*, **3**, 135-143.
- Lehmann, A.R. (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie*, **85**, 1101-1111.
- Li, L., Wu, L.P. and Chandrasegaran, S. (1992) Functional Domains in Fok I Restriction Endonuclease. *Proc Natl Acad Sci U S A*, **89**, 4275-4279.
- Li, X., Li, J., Harrington, J., Lieber, M.R. and Burgers, P.M. (1995) Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *J Biol Chem*, **270**, 22109-22112.
- Lilley, D.M. and White, M.F. (2001) The junction-resolving enzymes. *Nat Rev Mol Cell Biol*, **2**, 433-443.
- Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709-715.
- Lindahl, T. and Wood, R.D. (1999) Quality control by DNA repair. *Science*, **286**, 1897-1905.
- Liu, Z., Hossain, G.S., Islas-Osuna, M.A., Mitchell, D.L. and Mount, D.W. (2000) Repair of UV damage in plants by nucleotide excision repair: Arabidopsis UVH1 DNA repair gene is a homolog of *Saccharomyces cerevisiae* Rad1. *Plant J*, **21**, 519-528.
- Maga, G. and Hubscher, U. (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*, **116**, 3051-3060.

-
- Matsumiya, S., Ishino, Y. and Morikawa, K. (2001) Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci*, **10**, 17-23.
- Matsunaga, T., Park, C.H., Bessho, T., Mu, D. and Sancar, A. (1996) Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. *J Biol Chem*, **271**, 11047-11050.
- McGlynn, P. and Lloyd, R.G. (2002) Genome stability and the processing of damaged replication forks by RecG. *Trends Genet*, **18**, 413-419.
- McPherson, J.P., Lemmers, B., Chahwan, R., Pamidi, A., Migon, E., Matysiak-Zablocki, E., Moynahan, M.E., Essers, J., Hanada, K., Poonepalli, A., Sanchez-Sweatman, O., Khokha, R., Kanaar, R., Jasin, M., Hande, M.P. and Hakem, R. (2004) Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science*, **304**, 1822-1826.
- Michel, B. (2000) Replication fork arrest and DNA recombination. *Trends Biochem Sci*, **25**, 173-178.
- Moolenaar, G.F., Uiterkamp, R.S., Zwijnenburg, D.A. and Goosen, N. (1998) The C-terminal region of the *Escherichia coli* UvrC protein, which is homologous to the C-terminal region of the human ERCC1 protein, is involved in DNA binding and 5'-incision. *Nucleic Acids Res*, **26**, 462-468.
- Motycka, T.A., Bessho, T., Post, S.M., Sung, P. and Tomkinson, A.E. (2004) Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J Biol Chem*, **279**, 13634-13639.

- Mu, D., Bessho, T., Nechev, L.V., Chen, D.J., Harris, T.M., Hearst, J.E. and Sancar, A. (2000) DNA interstrand cross-links induce futile repair synthesis in mammalian cell extracts. *Mol Cell Biol*, **20**, 2446-2454.
- Mu, D., Hsu, D.S. and Sancar, A. (1996) Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem*, **271**, 8285-8294.
- Mullen, J.R., Kaliraman, V., Ibrahim, S.S. and Brill, S.J. (2001) Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in *Saccharomyces cerevisiae*. *Genetics*, **157**, 103-118.
- Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P., White, M.F. and McDonald, N.Q. (EMBO, In Press) Crystal structure of a xeroderma pigmentosum group F endonuclease with and without DNA suggests a model for recognition of branched DNA substrates.
- Nichols, A.F. and Sancar, A. (1992) Purification of PCNA as a nucleotide excision repair protein. *Nucleic Acids Res*, **20**, 2441-2446.
- Niedernhofer, L.J., Essers, J., Weeda, G., Beverloo, B., de Wit, J., Muijtjens, M., Odijk, H., Hoeijmakers, J.H. and Kanaar, R. (2001) The structure-specific endonuclease Ercc1-Xpf is required for targeted gene replacement in embryonic stem cells. *Embo J*, **20**, 6540-6549.
- Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2003) X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. *Structure (Camb)*, **11**, 445-457.
- Ogrunc, M., Becker, D.F., Ragsdale, S.W. and Sancar, A. (1998) Nucleotide excision repair in the third kingdom. *J Bacteriol*, **180**, 5796-5798.

- Ogrunc, M. and Sancar, A. (2003) Identification and characterization of human MUS81-MMS4 structure-specific endonuclease. *J Biol Chem*, **278**, 21715-21720.
- Orren, D.K., Selby, C.P., Hearst, J.E. and Sancar, A. (1992) Post-incision steps of nucleotide excision repair in *Escherichia coli*. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *J Biol Chem*, **267**, 780-788.
- Osman, F., Dixon, J., Doe, C.L. and Whitby, M.C. (2003) Generating crossovers by resolution of nicked Holliday junctions: a role for Mus81-Eme1 in meiosis. *Mol Cell*, **12**, 761-774.
- Paques, F. and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev*, **63**, 349-404.
- Park, C.H., Bessho, T., Matsunaga, T. and Sancar, A. (1995) Purification and characterization of the XPF-ERCC1 complex of human DNA repair excision nuclease. *J Biol Chem*, **270**, 22657-22660.
- Park, C.H. and Sancar, A. (1994) Formation of a ternary complex by human XPA, ERCC1, and ERCC4(XPF) excision repair proteins. *Proc Natl Acad Sci U S A*, **91**, 5017-5021.
- Pingoud, A. and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res*, **29**, 3705-3727.
- Postow, L., Ullsperger, C., Keller, R.W., Bustamante, C., Vologodskii, A.V. and Cozzarelli, N.R. (2001) Positive torsional strain causes the formation of a four-way junction at replication forks. *J Biol Chem*, **276**, 2790-2796.

- Reeve, J.N., Sandman, K. and Daniels, C.J. (1997) Archaeal histones, nucleosomes, and transcription initiation. *Cell*, **89**, 999-1002.
- Richard, D.J., Bell, S.D. and White, M.F. (2004) Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase. *Nucleic Acids Res*, **32**, 1065-1074.
- Riedl, T., Hanaoka, F. and Egly, J.M. (2003) The comings and goings of nucleotide excision repair factors on damaged DNA. *Embo J*, **22**, 5293-5303.
- Robu, M.E., Inman, R.B. and Cox, M.M. (2001) RecA protein promotes the regression of stalled replication forks in vitro. *Proc Natl Acad Sci U S A*, **98**, 8211-8218.
- Saijo, M., Kuraoka, I., Masutani, C., Hanaoka, F. and Tanaka, K. (1996) Sequential binding of DNA repair proteins RPA and ERCC1 to XPA in vitro. *Nucleic Acids Res*, **24**, 4719-4724.
- Salerno, V., Napoli, A., White, M.F., Rossi, M. and Ciaramella, M. (2003) Transcriptional response to DNA damage in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res*, **31**, 6127-6138.
- Sancar, A. (1996) DNA excision repair. *Annu Rev Biochem*, **65**, 43-81.
- Sargent, R.G., Rolig, R.L., Kilburn, A.E., Adair, G.M., Wilson, J.H. and Nairn, R.S. (1997) Recombination-dependent deletion formation in mammalian cells deficient in the nucleotide excision repair gene ERCC1. *Proc Natl Acad Sci U S A*, **94**, 13122-13127.
- Schurtenberger, P., Egelhaaf, S.U., Hindges, R., Maga, G., Jonsson, Z.O., May, R.P., Glatter, O. and Hubscher, U. (1998) The solution structure of functionally active human proliferating cell nuclear antigen determined by small-angle neutron scattering. *J Mol Biol*, **275**, 123-132.

- Selby, C.P. and Sancar, A. (1993) Molecular mechanism of transcription-repair coupling. *Science*, **260**, 53-58.
- Sgouros, J., Gaillard, P.H. and Wood, R.D. (1999) A relationship between a DNA-repair/recombination nuclease family and archaeal helicases. *Trends Biochem Sci*, **24**, 95-97.
- Singh, S., Folkers, G.E., Bonvin, A.M., Boelens, R., Wechselberger, R., Niztayev, A. and Kaptein, R. (2002) Solution structure and DNA-binding properties of the C-terminal domain of UvrC from E.coli. *Embo J*, **21**, 6257-6266.
- Singleton, M.R. and Wigley, D.B. (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol*, **184**, 1819-1826.
- Sogo, J.M., Lopes, M. and Foiani, M. (2002) Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science*, **297**, 599-602.
- Stukenberg, P.T., Studwell-Vaughan, P.S. and O'Donnell, M. (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J Biol Chem*, **266**, 11328-11334.
- Svejstrup, J.Q. (2002) Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol*, **3**, 21-29.
- Svejstrup, J.Q., Vichi, P. and Egly, J.M. (1996) The multiple roles of transcription/repair factor TFIIH. *Trends Biochem Sci*, **21**, 346-350.
- Tom, S., Henricksen, L.A. and Bambara, R.A. (2000) Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *J Biol Chem*, **275**, 10498-10505.
- Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghbizadeh, S., Hogue, C.W., Bussey, H., Andrews, B., Tyers, M. and

- Boone, C. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2364-2368.
- Tsurimoto, T. (1999) PCNA binding proteins. *Front Biosci*, **4**, D849-858.
- Umar, A., Buermeyer, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M. and Kunkel, T.A. (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell*, **87**, 65-73.
- Vaisman, A. (2002) The bypass of DNA lesions by DNA and RNA polymerases. *Mutation Research*, **510**, 1-7.
- Volker, M., Mone, M.J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H., van Driel, R., van Zeeland, A.A. and Mullenders, L.H. (2001) Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell*, **8**, 213-224.
- Wadsworth, R.I. and White, M.F. (2001) Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. *Nucleic Acids Res*, **29**, 914-920.
- Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10564-10569.
- Wakasugi, M., Reardon, J.T. and Sancar, A. (1997) The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J Biol Chem*, **272**, 16030-16034.
- Warbrick, E. (1998) PCNA binding through a conserved motif. *Bioessays*, **20**, 195-199.
- Warbrick, E. (2000) The puzzle of PCNA's many partners. *Bioessays*, **22**, 997-1006.

- Waseem, N.H., Labib, K., Nurse, P. and Lane, D.P. (1992) Isolation and analysis of the fission yeast gene encoding polymerase delta accessory protein PCNA. *Embo J*, **11**, 5111-5120.
- West, S.C. (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet*, **31**, 213-244.
- Whitby, M.C., Osman, F. and Dixon, J. (2003) Cleavage of model replication forks by fission yeast Mus81-Eme1 and budding yeast Mus81-Mms4. *J Biol Chem*, **278**, 6928-6935.
- White, M.F. (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans*, **31**, 690-693.
- White, M.F. and Bell, S.D. (2002) Holding it together: chromatin in the Archaea. *Trends Genet*, **18**, 621-626.
- White, M.F. and Lilley, D.M. (1996) The structure-selectivity and sequence-preference of the junction-resolving enzyme CCE1 of *Saccharomyces cerevisiae*. *J Mol Biol*, **257**, 330-341.
- Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J*, **12**, 1781-1795.
- Winkler, G.S., Sugawara, K., Eker, A.P., de Laat, W.L. and Hoeijmakers, J.H. (2001) Novel functional interactions between nucleotide excision DNA repair proteins influencing the enzymatic activities of TFIIH, XPG, and ERCC1-XPF. *Biochemistry*, **40**, 160-165.
- Wu, L. and Hickson, I.D. (2002) RecQ helicases and cellular responses to DNA damage. *Mutat Res*, **509**, 35-47.

-
- Wu, X., Li, J., Li, X., Hsieh, C.L., Burgers, P.M. and Lieber, M.R. (1996) Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. *Nucleic Acids Res*, **24**, 2036-2043.
- Xu, H., Zhang, P., Liu, L. and Lee, M.Y. (2001) A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry*, **40**, 4512-4520.
- Yang, W. (2003) Pruning DNA: structure-specific endonucleases (XPF/Rad1/Mus81). *Structure (Camb)*, **11**, 365-366.
- You, J.S., Wang, M. and Lee, S.H. (2003) Biochemical analysis of the damage recognition process in nucleotide excision repair. *J Biol Chem*, **278**, 7476-7485.
- Zhou, J.Q., He, H., Tan, C.K., Downey, K.M. and So, A.G. (1997) The small subunit is required for functional interaction of DNA polymerase delta with the proliferating cell nuclear antigen. *Nucleic Acids Res*, **25**, 1094-1099.

Appendix 1

Oligonucleotides used for DNA substrates

NAME	SEQUENCE (5' to 3')
b50	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTTCTGGACC
b50 biotin	CC[Bio~dT]CGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTTC [Bio~dT]GGACC
b(1-37)	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGG
b(1-31)	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGC
b(1-25)	CCTCGAGGGATCCGTCCTAGCAAGC
b75	GGAGCGGTGGTTGAATTCCTCGACGCCTCGAGGGATCCGTCCTAGCAAGC CGCTGCTACCGGAAGCTTCTGGACC
b50 abasic	CCTCGAGGGATCCGTCCTAGCAAGXCCTGCTACCGGAAGCTTCTGGACC
b complement	GGTCCAGAAGCTTCCGGTAGCAGCGGCTTGCTAGGACGGATCCCTCGAGG
b complement 7 nt bubble	GGTCCAGAAGCTTCCGGTAGCATAACCGCAGCTAGGACGGATCCCTCGAGG
b complement 10 nt bubble	GGTCCAGAAGCTTCCGGTAGGCTACCGCACCTAGGACGGATCCCTCGAGG
x50	GCTCGAGTCTAGACTGCAGTTGAGAGCTTGCTAGGACGGATCCCTCGAGG
x(26-50)	GCTTGCTAGGACGGATCCCTCGAGG
h50	GGTCCAGAAGCTTCCGGTAGCAGCGAGAGCGGTGGTTGAATTCCTCGACG
h(1-25)	GGTCCAGAAGCTTCCGGTAGCAGCG
h(1-26)	GGTCCAGAAGCTTCCGGTAGCAGCGA
h(26-50)	AGAGCGGTGGTTGAATTCCTCGACG
r50	CGTCGAGGAATTCAACCACCGCTCTTCTCAACTGCAGTCTAGACTCGAGC
r(1-25)	CGTCGAGGAATTCAACCACCGCTCT
r(26-50)	TCTCAACTGCAGTCTAGACTCGAGC
y50	CGTCGAGGAATTCAACCACCGCTCTGCTTGCTAGGACGGATCCCTCGAGG
y(A3 bulge)	CGTCGAGGAATTCAACCACCGCTCTAAAGCTTGCTAGGACGGATCCCTCG AGG
z75	GGTCCAGAAGCTTCCGGTAGCAGCGTTTTTTTTTTTTTTTTTTTTTTTTTTTCG TCGAGGAATTCAACCACCGCTCC
TT dimer	CGGGATCGAGCACCAGAAT [*] TCACGAGTACCTGCGG
TT control	CGGGATCGAGCACCAGAATTCACGAGTACCTGCGG
TT complement	CCGCAGGTA [*] CTCGTGAATTCTGGTGCTCGATCCCG

Appendix 2

Calculating thermodynamic parameters for *Sso*XPF interaction with manganese ions from isothermal titration calorimetry data

The interaction of *Sso*XPF and manganese ions at 35 °C is used as the example.

The temperature was calculated in degrees Kelvin (°K)

$$T (\text{°K}) = T (\text{°C}) + 273 = 35 + 273 = 308$$

The curve fit gave the apparent association constant K_{app} from which the dissociation constant (K_{D}) was calculated:

$$K_{\text{D}} = 1/K_{\text{App}} = 1/11410 = 8.7642 \times 10^{-5} = 87.6 \mu\text{M}$$

The free energy change (ΔG) was calculated

$$\Delta G = -RT \ln(K_{\text{A}}) = -1.987 \times 308 \times \ln(11410) = -5717 \text{ calmol}^{-1}$$

where R was the gas constant ($1.987 \text{ calmol}^{-1}\text{deg}^{-1}$) and T was the temperature (°K)

The curve fit also gave the enthalpy change (ΔH) that together with ΔG calculated above was used to calculate the entropy changes ($T\Delta S$ and ΔS)

$$\Delta G = \Delta H - T\Delta S \text{ was rearranged to } T\Delta S = \Delta H - \Delta G$$

$$T\Delta S = -691.2 - (-5717) = 5026 \text{ calmol}^{-1}$$

$$\Delta S = 5026/T = 5026/308 = 16.3 \text{ calmol}^{-1}\text{°K}^{-1}$$

The results are shown in Table 3.1.

An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA

J. A. Roberts,¹ S. D. Bell,² and M. F. White^{1*}

¹Centre for Biomolecular Science, St Andrews University, North Haugh, St Andrews, Fife KY16 9ST, UK.

²MRC Cancer Cell Unit, Hutchison MRC Centre, Hills Road, Cambridge, CB2 2XZ, UK.

Summary

Archaea share many similarities with eukarya in their information processing pathways and have proven to be a useful model for studies of DNA replication and transcription, but DNA repair pathways are not well understood in archaea. Nucleotide Excision Repair (NER) deals with many bulky DNA lesions and involves over 30 proteins in eukarya. Archaeal NER has not been characterized biochemically, but homologues of the human repair nucleases XPF and XPG have been identified by homology searches. Crenarchaeal XPF proteins have a simplified domain structure, consisting of the C-terminal nuclease domain conserved in XPF and Mus81 but lacking the N-terminal 'helicase' domain that is found in eukaryal and euryarchaeal sequences. Unexpectedly, *Sulfolobus* XPF is only active in the presence of the sliding clamp PCNA, which is a heterotrimer in this organism. Interactions with two of the three subunits of PCNA are mediated via a C-terminal interaction motif. The PCNA-XPF complex acts as a structure-specific nuclease on a similar range of DNA flap, bubble and junction substrates as the human protein, suggesting a fundamental conservation through billions of years of evolution.

Introduction

Over the last few years genome sequencing projects have produced overwhelming evidence that the archaea share strong similarities in information processing pathways with the eukarya. Biochemical and structural studies, particularly of DNA replication and transcriptional machineries,

have confirmed both the relationship between the archaeal and eukaryal domains and the utility of the former as a model system in this context.

One area where we still have only a very limited understanding is the question of DNA repair in the archaea. All organisms devote a considerable proportion of their gene content, at least 100 genes, to the repair of DNA damage and possess a selection of overlapping repair pathways designed to cope with every conceivable insult to the genetic material. Base excision repair (BER) pathways are conserved in their main characteristics across all three domains of life, consisting of a large number of specialized glycosylases built around a handful of protein superfamilies (Aravind *et al.*, 1999). Archaeal double-strand break repair (homologous recombination) appears similar to its eukaryal equivalent, with homologues of Rad51 (Seitz and Kowalczykowski, 2000) and Rad50/Mre11 (Constantinesco *et al.*, 2002). However archaea encode a unique Holliday junction resolving enzyme, Hjc (Komori *et al.*, 1999; Kvaratskhelia *et al.*, 2000) and have no recognizable branch migration motor for the Holliday junction. The archaeal machinery for mismatch repair (MMR) is not obvious from sequence comparisons, despite the fact that MMR is conserved in its essentials from *E. coli* to humans (Grogan, 2000). Lastly, the nucleotide excision repair (NER) pathway, which deals with a wide variety of bulky lesions on DNA including 6–4 photoproducts and cyclobutane pyrimidine dimers generated by UV irradiation, remains an enigma in archaea.

There are very few published studies of NER activity in archaea, and this may be due, in part, to the erroneous impression that archaeal NER is essentially equivalent to the bacterial UvrABC pathway. Initial comparisons of archaeal NER activity focussed on the organism *Methanothermobacterium thermoautotrophicum* and demonstrated patch repair size diagnostic of bacterial NER (Ogrunc *et al.*, 1998). It subsequently became clear that this organism does possess a set of UvrABC homologues, however, these are not found in the majority of archaea, and are likely to have resulted from a lateral gene transfer event (Grogan, 2000). Most archaea, including *M. thermoautotrophicum*, have genes encoding homologues of the eukaryal NER nucleases XPF (Rad1) and XPG/Fen1 (Rad2), and helicases XPB (Rad25) and XPD (Rad3).

Accepted 18 December, 2002. *For correspondence. E-mail mfw2@st-andrews.ac.uk; Tel. (+44) 1334 463432; Fax (+44) 1330 462595.

However, with the exception of XPG, little biochemical evidence is available for the functionality of these proteins. Furthermore, there are no recognizable homologues of the eukaryal damage recognition proteins XPA and XPC in any of the archaea (Aravind *et al.*, 1999). This also seems to be the case in the primitive eukaryote *Plasmodium falciparum*, suggested that these damage recognition proteins may have been later additions to the NER pathway (Gardner *et al.*, 2002). Nucleotide excision repair in the archaea thus presents as a puzzle where many of the pieces are missing and we don't understand how the ones we have can fit together.

In this paper we present biochemical data on the role of the archaeal XPF homologue in nucleotide excision repair. Human XPF is a heterodimer with ERCC1 (the equivalent proteins in yeast are Rad1 and Rad10). The heterodimer recognizes junctions between double strand and single stranded regions of DNA where the single strand leaves the junction with a 5'-3' polarity. Model structures such as flaps, Y and bubble structures are cleaved 2-8 bp from the junction point (de Laat *et al.*, 1998a). *In vivo*, XPF-ERCC1 is known to assemble at the site of DNA damage as part of a complex repair machine that recognizes, unwinds and removes the damaged DNA strand (reviewed in Araujo and Wood, 1999; de Laat *et al.*, 1999). XPF-ERCC1 catalyses the strand scission on the 5' side of the lesion, and has also been shown to remove interstrand psoralen cross-links in DNA (Kuraoka *et al.*, 2000).

The domain organization of XPF is intriguing, with an N-terminal helicase domain fused to a C-terminal nuclease (Fig. 2A). The partner protein ERCC1 appears distantly related to the C-terminal nuclease domain of XPF (Sgouros *et al.*, 1999; Gaillard and Wood, 2001). A homologue of XPF, Mus81, has recently been shown to function as a Holliday junction resolving enzyme with a possible role in the rescue of stalled replication forks (Boddy *et al.*, 2001; Chen *et al.*, 2001). Mus81 forms heterodimers with a variety of partners, such as Cds1 and Eme1/Mms4 (Boddy *et al.*, 2000; 2001; Kaliraman *et al.*, 2001).

The archaeal XPF homologues come in two distinct forms. In the euryarchaea, XPF is present as a helicase:nuclease fusion as in eukarya, with a polypeptide length of 700-800 amino acids. By contrast, in the crenarchaea the XPF gene encodes only the C-terminal nuclease domain of approximately 240 amino acids without the N-terminal fusion (Figs 1A and 2A). Here we report the expression, purification and characterization of the XPF homologue from *Sulfolobus solfataricus*, and demonstrate that it possesses flap-endonuclease activity that is diagnostic of eukaryal XPF proteins, but displays an unexpected requirement for a functional interaction with the archaeal sliding clamp PCNA.

Results

Expression and purification of *Sulfolobus* XPF

The gene encoding *Sulfolobus* XPF (protein accession number Q9UXB7) was amplified by PCR, sequenced and subcloned into vector pET19b for expression of the native protein, and vector pGEX-5X for expression as a GST-fusion. Three mutant forms of the protein were also designed: mutant S139stop for expression of the N-terminal nuclease domain; mutant S228stop to delete the presumed PCNA interaction motif (also denoted ' $\Delta 6'$ '), and mutant D52A to abrogate the nuclease activity. These mutants were constructed in vector pET19b and sequenced fully. The native wild-type and mutant proteins were purified by a combination of heat treatment followed by two column chromatography steps, resulting in essentially pure, homogeneous protein (Fig. 1C), as described in the methods. The GST-XPF fusion protein was purified directly on glutathione-agarose beads as described in the *Experimental procedures*.

Mapping XPF domains

Limited trypsinolysis of XPF results in cleavage at a cluster of basic residues between positions 138-149, producing an N-terminal domain (amino acids 1-138) and a C-terminal portion (amino acids 150-233) (Fig. 1D). The larger N-terminal domain is stable when expressed in the form of a (1-138) truncation mutant in *E. coli* (Fig. 1C), suggesting that the nuclease functionality of XPF exists in an autonomous, folded domain. The C-terminal domain contains a predicted Helix-hairpin-Helix (HhH) that is also present in the human XPF, Mus81 and ERCC1 proteins (de Laat *et al.*, 1998a; Interthal and Heyer, 2000), suggesting a role for non-specific dsDNA binding for this region (Doherty *et al.*, 1996).

Sulfolobus XPF interacts with PCNA

The GST-XPF fusion protein was immobilized on glutathione-agarose beads and used for affinity chromatography using extracts of lysed *Sulfolobus* cells. Figure 3A shows the result on one such experiment. After SDS-PAGE three closely spaced proteins of approximate mass 30 kDa were apparent. Each band was excised from the gel and the protein identified by in-gel tryptic digest and MALDI-TOF mass spectrometry. The three proteins were identified as the three subunits of *Sulfolobus* PCNA, which has recently been shown to exist as a heterotrimer (Dionne *et al.*, 2003).

To confirm the interaction with PCNA, antibodies raised against each subunit of PCNA were used to demonstrate that all three subunits of PCNA are indeed present in the immobilized fraction on the GST-XPF column (Fig. 3B). In

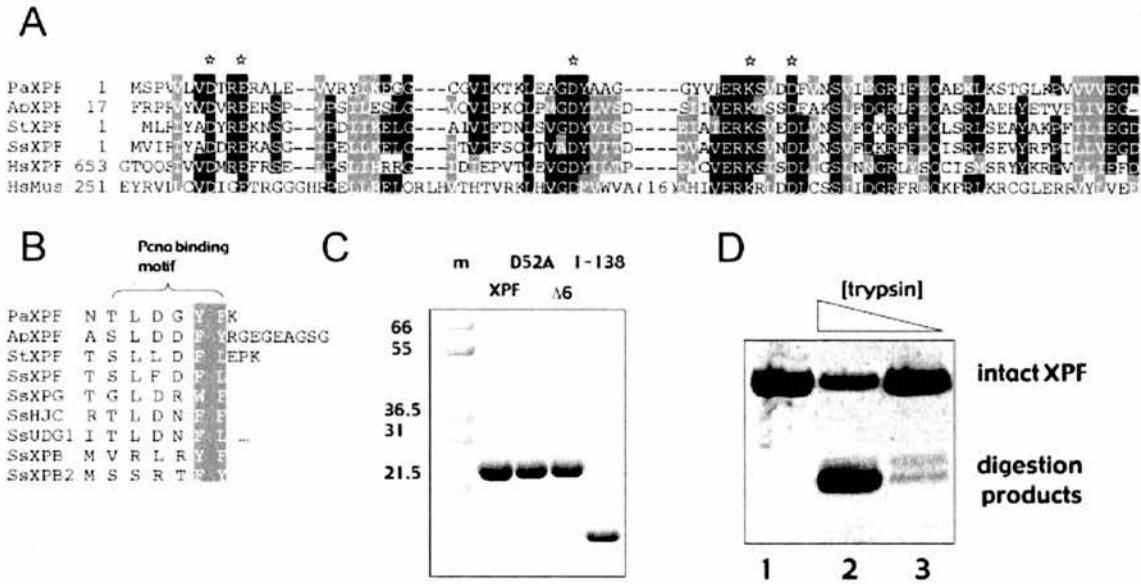


Fig. 1. A. Alignment of the nuclease domain conserved between crenarchaeal XPF sequences and human XPF and Mus81. Residues marked with an asterisk have been mutated in human XPF and shown to be essential for nuclease activity.

B. Alignment of putative PCNA interaction motifs in the C-termini of crenarchaeal XPF orthologues and other proteins from *Sulfolobus solfataricus*. (SsXPF *Sulfolobus solfataricus* XPF, Q9UXB7; PaXPF *Pyrobaculum aerophilum* XPF, Q8ZWLO; ApXPF, *Aeropyrum pernix* XPF, Q9YC15; StXPF, *Sulfolobus tokadaii* XPF, Q975H3; HsXPF, *Homo sapiens* XPF, Q92889; HsMus81, *Homo sapiens* Mus81, Q96NY9; SsHJC, *Sulfolobus solfataricus* Holliday junction resolving enzyme Hjc; SsUDG1, *Sulfolobus solfataricus* uracil glycosylase; SsXPB, *Sulfolobus solfataricus* XPB homologue; SsXPB2, *Sulfolobus solfataricus* XPB homologue 2).

C. Lanes 2–5 show purified recombinant wild-type XPF (XPF), active site mutant D52A, C-terminal 6 amino acid deletion mutant ($\Delta 6$) and the N-terminal truncation mutant (1–138), respectively, analysed on a Coomassie blue stained gel.

D. Tryptic mapping of domain boundaries in XPF. Recombinant wt XPF was incubated with serial dilutions of porcine trypsin and the resultant degradation products were analysed by SDS-PAGE and MALDI-TOF. Trypsin introduced cuts at four clustered lysine and arginine residues between positions 138–149, effectively cleaving XPF into an N-terminal domain (amino acids 1–138) and a C-terminal portion (amino acids 150–233). The N-terminal domain was refractory to further digestion, and is stable when expressed in the form of a (1–138) truncation mutant (Fig. 1C).

addition, GST-fusions of each of the three subunits of PCNA were used in pull-down experiments with extracts of *Sulfolobus* cell lysate as described above. This resulted in the observation that XPF interacts strongly with PCNA1 and PCNA3, but not appreciably with PCNA2 (Fig. 3C).

Sulfolobus XPF has a conserved PCNA interaction motif that is essential for PCNA binding

Investigation of the C-terminus of *Sulfolobus* XPF revealed a sequence resembling the PCNA interaction motif found in a variety of DNA replication and repair proteins in the archaea (Matsumiya *et al.*, 2001; Dionne *et al.*, 2003) (Fig. 1B). This consists of a short sequence ending with two consecutive hydrophobic (often aromatic) residues, similar to the consensus sequence defined for PCNA interaction motifs in eukaryal proteins (reviewed in Warbrick, 2000). The motif is present in orthologues of the short form of XPF from other crenarchaeal sequences, as well as other DNA repair enzymes such as XPG (Fen1)

where an interaction has been confirmed biochemically in eukaryal systems (Gary *et al.*, 1997). A *Sulfolobus* XPF mutant lacking the last six C-terminal residues was constructed and tested for interaction with PCNA (Fig. 3D). The XPF:PCNA interaction was abrogated in this mutant, confirming that the C-terminal motif mediates the interface between the two proteins.

Sulfolobus XPF is a PCNA-dependent, structure specific nuclease

Eukaryal XPF-ERCC1 and Rad1-Rad10 complexes are active as nucleases on many substrates possessing a ssDNA/dsDNA junction (Davies *et al.*, 1995; de Laat *et al.*, 1998a). However, initial attempts to detect nuclease activity associated with *Sulfolobus* XPF were unsuccessful. Addition of *Sulfolobus* PCNA heterotrimer to nuclease assays resulted in the activation of XPF, allowing cleavage of a variety of flap and bubble structures. Multiple strand scissions were observed 5' of the ss/ds junction, in a region of duplex DNA 3–12 nt from the junction (Fig. 4A).

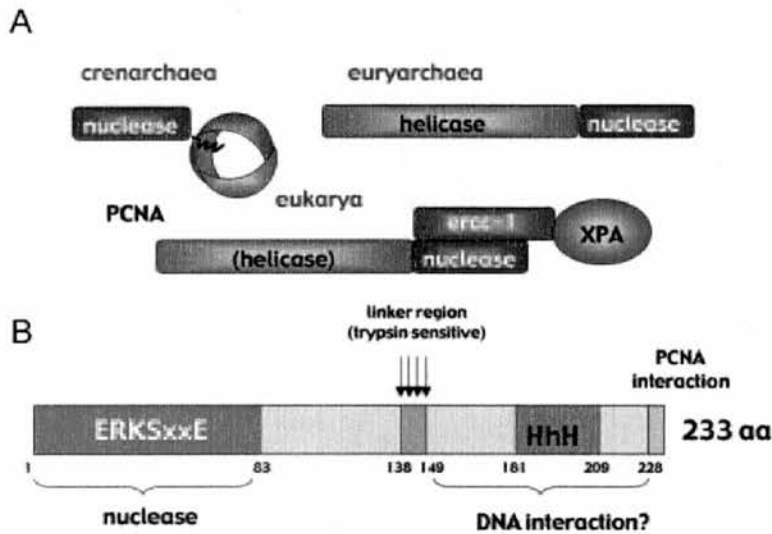


Fig. 2. Defining the domain organization of XPF.

A. Gross domain organization of XPF homologues. The crenarchaeal XPF orthologues lack the helicase domain present in the eukaryotic and euryarchaeal XPF proteins, but compensate for this via an interaction with the sliding clamp PCNA. Eukaryal XPF forms a heterodimer with ERCC1, which in turn interacts with the damage recognition protein XPA during excision repair.

B. Summary of domain organization of *Sulfolobus* XPF. The nuclease domain homologous to eukaryotic XPF and Mus81 is located in the N-terminal 85 amino acids. The N-terminal 138 amino acids form a stable domain separated from the C-terminal domain by a trypsin sensitive linker region. The C-terminal domain consists of a PCNA interaction sequence at the extreme terminus, and has a consensus Helix-hairpin-Helix (HhH) motif. This motif is also found in eukaryotic XPF and many other DNA binding and repair proteins, where it is known to constitute a sequence non-specific DNA binding module.

As observed previously for XPF and Mus81 (Boddy *et al.*, 2001; Enzlin and Schärer, 2002), mutation of the presumed active site residue Asp-52 to alanine (D52A mutant) results in complete loss of detectable nuclease activity. This is not due to a disruption of the interaction with PCNA, which is unaffected by the mutation (data not shown). Similarly, a mutant XPF lacking the C-terminal six residues that constitute the PCNA interaction motif is also deficient in nuclease activity (Fig. 4A).

The DNA substrates were designed with a common labelled DNA strand (the b strand). Mapping of the cleavage products produced by *Sulfolobus* XPF shows that the same sequences are cleaved in both the bubble and played duplex structures, even though these sequences are set at different positions with respect to the ss/ds junction (Fig. 4A). Further, all four major cleavage sites are introduced 3' of pyrimidine residues. This sequence preference is also observed in cleavage of four-way DNA junctions (Fig. 5 below). Together, these observations suggest there is a strong element of sequence specificity overlying the structural specificity of XPF. These properties are identical to the human XPF-ERCC1 enzyme, which also prefers to cut after pyrimidines (de Laat *et al.*, 1998a), suggesting a fundamental property of XPF that has been conserved through billions of years of evolution.

Sequential cleavage of substrates by XPF

Analysis of the cleavage products generated by XPF using a 3'-flap substrate shows clearly that the initial point of cleavage is at a cytosine residue 4 nt 5' of the branch site (Fig. 4B). This is followed by further processing of the

cleaved strand at subsequent pyrimidine residues up to 12 nt 5' of the branch site. The sequential appearance of strand scissions progressively further away from the initial discontinuity in the DNA helix suggests that, *in vitro*, XPF trims back the DNA strand by acting on DNA products that have been initially cleaved near the branch point.

Sulfolobus SSB modulates XPF activity

In eukaryotes, the single stranded DNA binding protein RPA is intimately involved in NER. In addition to a role in binding the ssDNA generated during the NER process, RPA influences the repair process directly, partly through an interaction with XPA that in turn contacts ERCC1. RPA also stimulates the nuclease activity of XPF-ERCC1 for substrates with 5' protruding single stranded arms whereas *E. coli* SSB does not (de Laat *et al.*, 1998b). *Sulfolobus* encodes a homologue of RPA, SsoSSB, that has a single OB fold for DNA binding and a flexible C-terminal tail for protein:protein interactions (Wadsworth and White, 2001). Clearly SsoSSB would bind the single stranded regions of the substrates, and is likely to play a role in NER in the archaea. Inclusion of SsoSSB in the cleavage reactions along with PCNA and XPF results in an inhibition of substrate cleavage by XPF (Fig. 4B). Initial cleavage products appear at later time points when SSB is present, and the substrate is not fully processed to the mature form seen in the absence of SSB. It is interesting to note that the presence of a similar concentration of the double-strand DNA binding protein Alba has no effect either on the rate of cleavage of branched substrates by XPF or on the inhibition of that cleavage by SSB.

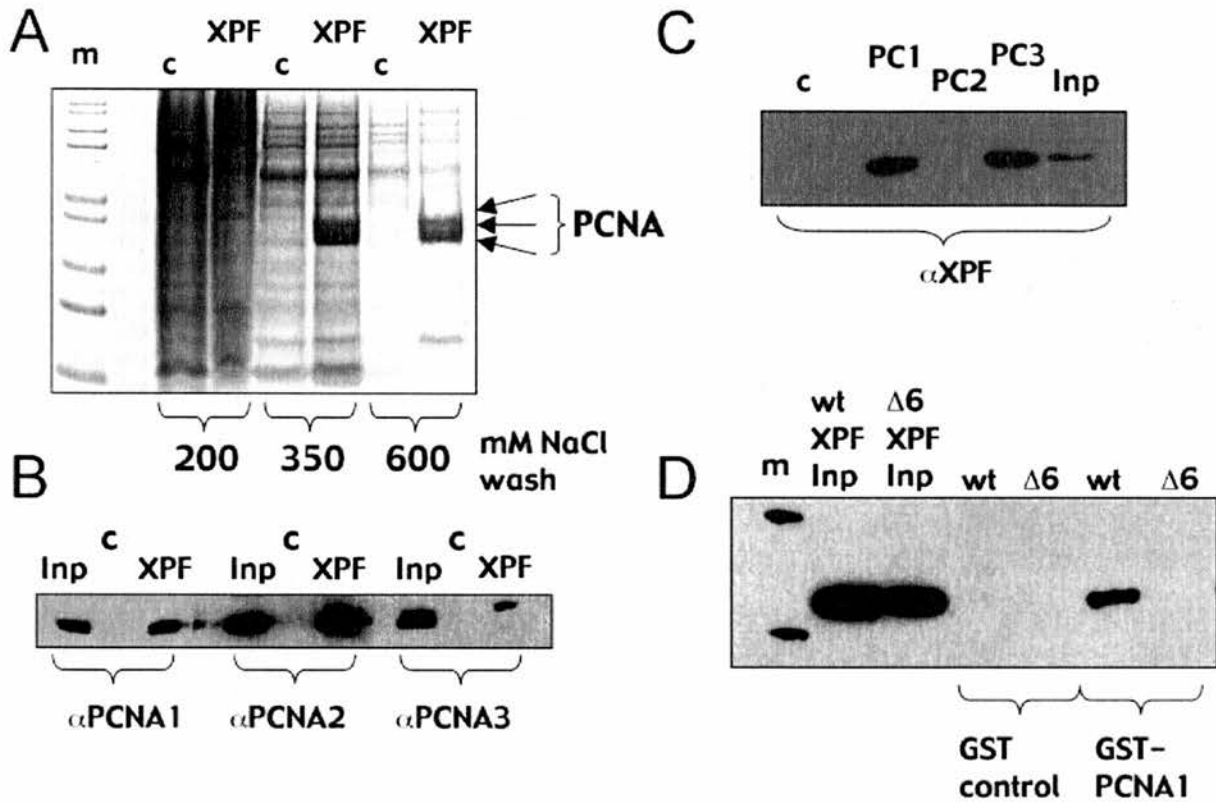


Fig. 3. *Sulfolobus* XPF interacts with heterotrimeric PCNA

A. SDS-PAGE analysis of GST-XPF affinity chromatography fractions. *S. solfataricus* lysate was passed through a column containing GST-XPF fusion protein (XPF) immobilized on glutathione agarose beads. Proteins with affinity for XPF bound to the column and were eluted in 100 mM, 200 mM, and 350 mM NaCl washes. A GST column (c) was used as a control. The three bands shown were identified as the three subunits of PCNA by MALDI-TOF mass spectrometry. Lane m = molecular weight markers (Mark 12™, Invitrogen).

B. GST-XPF pull-down of all 3 PCNA subunits from *S. solfataricus* lysate as shown by Western blot with antibodies specific for individual PCNA subunits. Input = 5% of the *S. solfataricus* lysate used. c = GST control protein pull-down.

C. XPF interacts specifically with PCNA subunits 1 and 3. GST-PCNA1 (PC1) and GST-PCNA3 (PC3) can pull-down XPF from *S. solfataricus* cell lysate as shown by a Western blot with antibodies specific for XPF. GST-PCNA2 did not appear to interact with XPF. c = GST control. input = 5% of the *S. solfataricus* lysate used.

D. The XPF-PCNA interaction is dependent on the C-terminal 6 residues of XPF. GST-PCNA pull-downs with 1 μ g of either full length recombinant XPF (wt) or the C-terminal deletion mutant ($\Delta 6$) lacking the proposed PCNA interaction motif. Input = 25% of the recombinant protein used. m = MagicMark™ Western blot standard markers (Invitrogen).

Sulfolobus XPF cuts Holliday junctions

The yeast XPF homologue Rad1 cuts mobile Holliday junctions and was originally proposed to be a bona fide Holliday junction resolving enzyme (Habraken *et al.*, 1994), before its higher specificity for splayed substrates became apparent. More recently, Mus81 has been shown to cleave fixed and mobile Holliday junction substrates *in vitro* (Boddy *et al.*, 2001; Chen *et al.*, 2001) and has been proposed to play a role in the rescue of stalled replication forks via Holliday junction cleavage. *Sulfolobus* XPF is also active against immobile four-way DNA junctions (Fig. 5), demonstrating that this activity is intrinsic to this whole class of flap endonucleases. Nicks in all four strands are introduced 5' of the branch point, almost

exclusively after cytosine residues. A Holliday junction containing a mobile core capable of 13 steps of branch migration was also cleaved by *Sulfolobus* XPF (data not shown).

Structure preference of *Sulfolobus* XPF

To identify the preferred substrates for XPF in more detail, we used a single-turnover kinetic assay to estimate the relative rates of cleavage of three different branched DNA structures by the enzyme (Fig. 6). The best substrate for XPF was a 3' flap structure, which was cut too quickly to allow accurate determination of the reaction rate using manual sampling methods. This rate was estimated as 6.8 min⁻¹ from two initial time points, but this probably

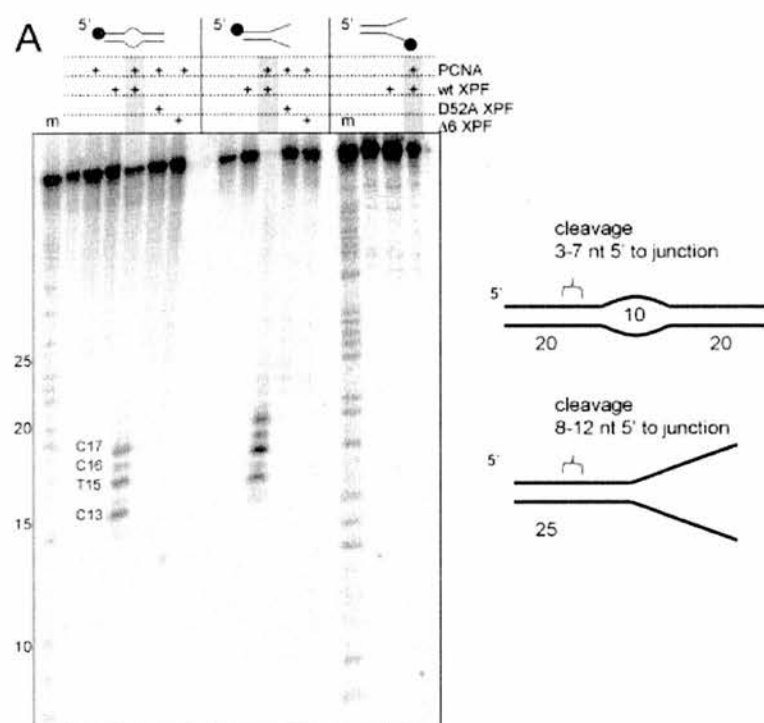


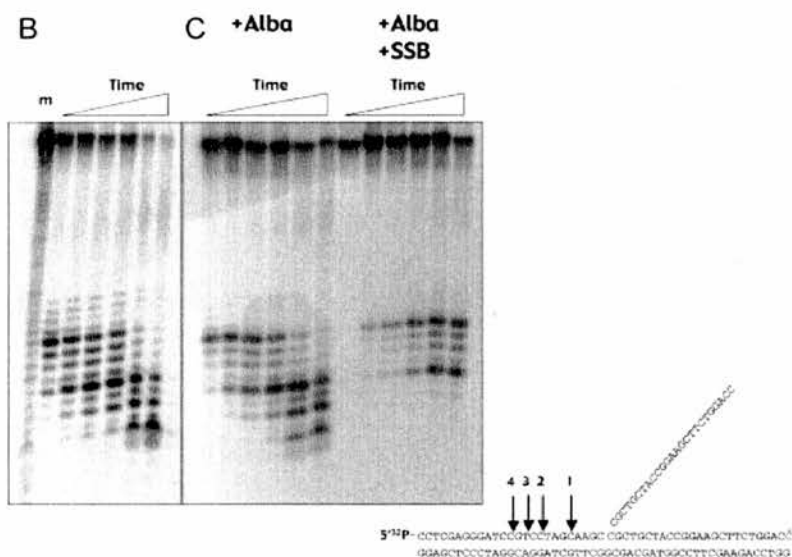
Fig. 4. XPF is a structure-specific, PCNA-dependent endonuclease.

A. Denaturing polyacrylamide gel electrophoretic analysis of reaction products resulting from incubation of wild-type and mutant XPF proteins with substrate DNA species in the presence and absence of PCNA. XPF activity is dependent on: an active nuclease site, presence of PCNA and presence of PCNA interaction motif at C-terminus. Under these conditions cleavage products are observed for both substrates with a labelled strand leaving a ds/ss junction with a 5'-3' polarity, but not with the opposite polarity. In both substrates XPF cleavage sites map to 5 pyrimidine residues that vary in their distance from the junction position, as seen for human XPF-ERCC1. m-A + G Maxam Gilbert size markers of the labelled strand. Black circles indicate the position of the ³²P-label on the cartoons of the substrates.

B. Sequential cleavage of a branched substrate by XPF

Denaturing polyacrylamide gel electrophoretic analysis of reaction products resulting from incubation of wild-type XPF + PCNA proteins with a 3'-flap substrate over time. The cleavage reaction was initiated by the addition of magnesium, and time points were taken at 5, 10, 20, 40, 90 and 240 s respectively. Initial products cleaved near the branch point are processed further to yield a mixture of mature products 8-12 nt 5' of the ds/ss DNA junction.

C. XPF activity is modulated by *Sulfolobus* SSB. Time-course of XPF cleavage of a 5'-branched substrate in the presence of *Sulfolobus* DNA binding proteins. The cleavage reaction was carried out as for Fig. 4B, but in the presence of either the dsDNA binding protein Alba (10 μM) or the presence of Alba (10 μM) plus the single-stranded DNA binding protein SSB (5 μM). Inhibition of XPF activity by SSB was observed, whilst Alba had no discernible effect on activity. SSB inhibits XPF activity to an identical extent in the absence of Alba (data not shown). Reaction time points are 5, 10, 20, 40, 90 and 240 s respectively.



represents an underestimate of its true activity. The played duplex was cleaved at a rate of $0.60 \pm 0.02 \text{ min}^{-1}$, at least 10-fold more slowly than the 3' flap substrate, and the fixed four-way junction was cut with a rate of $0.057 \pm 0.005 \text{ min}^{-1}$, 10-fold more slowly than the played duplex and 100-fold more slowly than the flap substrate. As the four-way junction is cut in all four arms,

the true rate of cleavage of this substrate is probably fourfold faster, but this doesn't change the overall order of preference of XPF. These studies confirm that, *Sulfolobus* XPF probably functions a 3'-flap repair endonuclease, rather than a Holliday junction resolving enzyme and is therefore more similar to eukaryal XPF than to the homologue Mus81 protein.

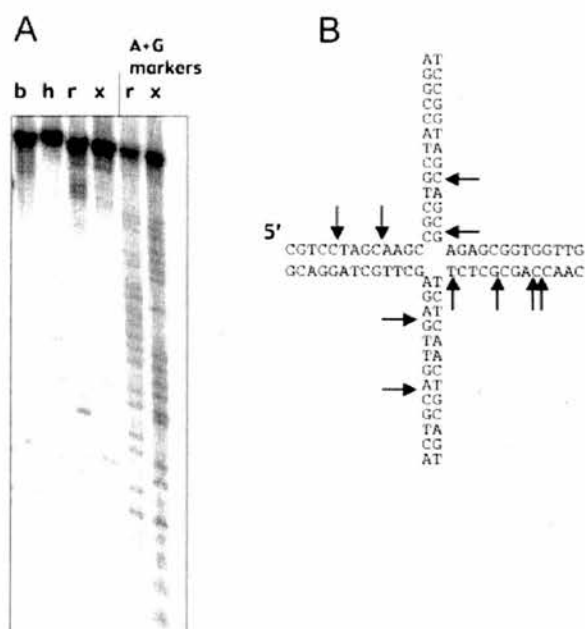


Fig. 5. Holliday junctions are cleaved by XPF.

A. Denaturing polyacrylamide gel electrophoretic analysis of reaction products resulting from incubation of wild-type XPF + PCNA proteins with a fixed four-way DNA junction. All four strands of the junction are cut at positions 5' of the branch point, and all but one of the cleavage sites is immediately 3' of a cytosine residue.

B. The central 13 base pairs of the junction are shown, with the positions of the strand scissions indicated by arrows. Cleavage sites are introduced up to 9 nt 5' of the branch point of the junction.

Discussion

The structure-specific endonuclease XPF (Rad1) plays a fundamental role in the eukaryotic Nucleotide Excision Repair pathway, but we know relatively little about the structural organization and molecular mechanism. The C-terminal third of human XPF contains the catalytic domain, as confirmed by extensive mutagenesis studies (Enzlin and Schärer, 2002). The N-terminal two-thirds of XPF resembles a helicase, though it has diverged from the canonical sequence (Sgouros *et al.*, 1999). There is little clear evidence for its function, though it is probably not a functional helicase or ATPase. It may be important for conferring specificity for branched substrates, or have a general DNA binding activity. The role of the partner protein ERCC1 is also unclear, though it performs a bridging interaction with XPA-RPA (Saijo *et al.*, 1996). XPF is also conserved in the archaea. As is often the case when comparing archaeal and eukaryal proteins, the archaeal XPF homologues present both some clues and some further puzzles. There is a clear split between the two fundamental subdomains euryarchaeota and crenarchaeota. The former all possess a full length XPF homologue

with an N-terminal helicase domain and a C-terminal structure-specific nuclease domain (Komori *et al.*, 2002). Whereas no helicase activity has yet been demonstrated, the helicase motifs appear intact and have an ATPase activity that is highly stimulated by the presence of branched DNA (Komori *et al.*, 2002). The crenarchaeal XPF is limited to the highly conserved nuclease core of about 240 amino acids (Fig. 1A). As we have shown, this is a fully functional nuclease with a similar structural specificity to eukaryal XPF. However it displays an absolute requirement for the sliding clamp PCNA, with which it forms a stable interaction through a characteristic C-terminal motif.

How can we piece together this evolutionary puzzle? It seems reasonable to suggest that the ancestral form of the XPF protein resembled the nuclease domain found in present day crenarchaea, which has the required specificity and nuclease activity but relies on an interaction with PCNA for efficient targeting to repair sites. An ancient origin for PCNA in this role is suggested by its interaction with the nuclease XPG, which is conserved throughout archaea and eukarya. There is a pleasing symmetry in the possibility that both the nucleases involved in NER might originally have been targeted to damaged DNA via PCNA. At some point the XPF nuclease domain fused with a DEXH box helicase and dispensed with a requirement for PCNA. The helicase may still have a DNA remodelling function in euryarchaea, though no biochemical evidence is currently available. In the eukarya, the helicase motifs are degenerate, suggesting that this part of the protein may play a generalized role in DNA binding or provide a platform for protein:protein interactions. The partner protein ERCC1 is also a eukaryal invention, as no homologues exist in archaea. Its known interactions with XPA

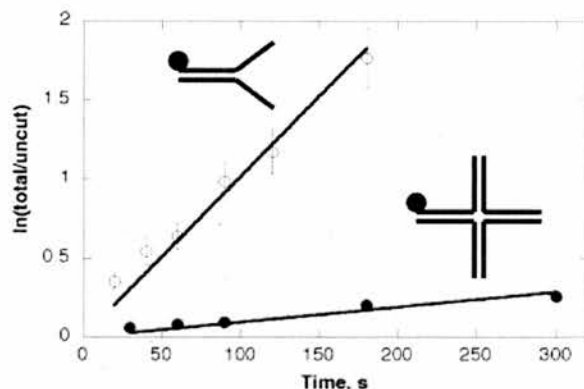


Fig. 6. Relative reaction rates for cleavage of a splayed duplex and a fixed four-way DNA junction by XPF. Assays were carried out in triplicate and standard deviations are shown. *Sulfolobus* XPF displayed a clear preference for cleavage of a 3'-flap structure (reaction rate estimated as 7 min^{-1}) over a splayed duplex (rate 0.6 min^{-1}) and four-way junction (rate 0.06 min^{-1} for the B arm).

suggest a bridging role in connecting the complex machinery of NER as it assembles around the DNA lesion. It appears distantly related to the C-terminal third of XPF and may have arisen by gene duplication of the latter (Gaillard and Wood, 2001), followed by divergence towards a new function. The alternative hypothesis that the ancestral form of XPF consisted of the helicase plus nuclease domains cannot be ruled out, but seems less likely as proteins tend to become more complex during evolution, partly by domain fusion to recruit new activities.

Sulfolobus solfataricus has recently been shown to encode a heterotrimeric PCNA (Dionne *et al.*, 2003). By contrast, the euryarchaea and eukarya encode only one PCNA gene and therefore utilize a homotrimeric PCNA protein. The increased complexity of PCNA allows *Sulfolobus* to utilize subunit-specific interactions with PCNA. Dionne *et al.* (2003) have presented convincing evidence that *Sulfolobus* XPG/Fen1 displays an absolute dependence on the heterotrimeric PCNA ring for enzymatic activity, and that XPG, DNA ligase and DNA polymerase can all bind simultaneously to a single PCNA heterotrimer to create a functional 5'-flap removal and repair complex. Our data suggest that XPF interacts with PCNA subunits 1 and 3 but not significantly with subunit 2. *Sulfolobus* XPG also interacts strongly with subunit 1 (Dionne *et al.*, 2003), suggesting the two proteins would compete for PCNA and could not be bound simultaneously to one heterotrimer. This seems sensible biologically, as it is hard to envisage a situation where the two endonucleases should be held such a close association. It is more likely that individual PCNA heterotrimers at either side of a DNA lesion target XPF and XPG separately to catalyse strand scission. There remains the possibility that a second protein interacts with PCNA simultaneously with XPF, via a contact with the PCNA2 subunit. Several *Sulfolobus* proteins have candidate PCNA interaction motifs (Fig. 1B). Most are proteins that have been shown in other systems to form complexes with PCNA, including uracil DNA glycosylase UNG1 (Yang *et al.*, 2002), the XPG nuclease, DNA ligase and DNA polymerase (Dionne *et al.*, 2003). A more provoking observation is that both of the *Sulfolobus* homologues of the helicase XPB (Rad25), which is a component of TFIIH in eukarya and plays a role in the NER pathway, also have a potential PCNA interaction motif at their N-termini. The possibility of a ternary XPB-PCNA-XPF interaction during archaeal NER is currently under investigation. It is worth noting that there is evidence for a functional interaction between XPB and XPF in eukaryotic NER (Evans *et al.*, 1997).

We have shown that the substrate specificity of *Sulfolobus* XPF, both in terms of DNA structures cleaved and the base preference underlying cleavage, is remarkably similar to the human XPF protein, underlining the fundamental conservation of this enzyme across evolutionary time.

We can conclude that archaeal NER follows the eukaryotic pattern and not, as was suggested originally, the bacterial UvrABC pathway (Ogrunc *et al.*, 1998). Further study of the proteins and complexes involved in archaeal NER promise to shed light on the equivalent eukaryal proteins and the evolutionary origins of the pathway.

Experimental procedures

Cloning and expression of the *S. solfataricus* XPF gene

The *S. solfataricus* XPF gene was amplified from *S. solfataricus* strain P2 genomic DNA using proof-reading polymerase Pfu (Promega) and the following primers: 5' primer: 5'-CGTCGGATCCCCATGGTAAATTAGAAATTTATGCTGATGATA GGG-3'

3' primer: 5'-CCGGGGATCCGTCGACCTAAAGGAAATCA AATAAAGAGGTAG-3'

The PCR product was subcloned into the *NcoI/BamHI* site of vector pET19b (Novagen) for expression of the native protein, and the *BamHI/SalI* site of the GST gene fusion vector pGEX-5X-3 (Amersham Pharmacia Biotech) for GST-XPF fusion expression. Protein expression was carried out in BL21 Rosetta cells induced by adding 0.2 mM IPTG when cultures reached A_{600} 0.7, grown for further 3 h and the cells pelleted and frozen until required.

The XPF mutants (N-terminal 138 $\alpha\alpha$, active site D52A, and C terminal truncated $\Delta 6$) were made using the Quick-Change™ site-directed mutagenesis kit (Stratagene®) with pET19bXPF construct and the following oligonucleotide primer pairs:

N-terminal 1–138 truncation (S139stop):

5'-CCAAATTTGGTGAAAACAAGTAAAA TCGAATTAGTTTAC ATAATAAGG-3'

5'-CCTTATTATGTAAACTAATTCGATTTTACTTGTTTTCACC AAATTGG-3'

D52a

5'-GGAAATCAGTAAATGCTCTAGTAAATTCAG-3'

5'-CTGAATTTACTAGAGCATTACTGATTTCC-3'

C terminal $\Delta 6$ truncation (S228stop):

5'-GTAAGAAAACCTACCTAATTATTTGATTTCC-3'

5'-GGAAATCAAATAATTAGGTAGTTTTCTTAC-3'

Purification of native recombinant XPF

The bacterial pellet was thawed in ~35 ml buffer (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 100 mM NaCl, 1 mM benzamidine) and sonicated 4 × 2 min with cooling. The lysate was centrifuged at 48 000 *g* for 20 min, 4°C and the supernatant heated to 70°C to precipitate *E. coli* proteins before centrifugation for a further 20 min. The supernatant was filtered (Acrodisc® 0.2 μ m syringe filter) and diluted twofold to threefold with buffer A (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT). This was applied to a 10 ml heparin column (Amersham Pharmacia Biotech) equilibrated with buffer A and the bound cationic proteins eluted over a 120 ml linear gradient comprising 0–1000 mM NaCl. The fractions containing XPF as analysed by SDS-PAGE of fractions corresponding to absorbance peaks were pooled and concentrated to ~7 ml and loaded onto a HiLoad® 26/60 Superdex® 200 gel

filtration column (Amersham Pharmacia Biotech) equilibrated with buffer (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 150 mM NaCl). Fractions corresponding to the peak(s) were concentrated as before and the protein concentration calculated from the A_{280} . Protein was stored at 4°C until required. Mutant proteins were purified as for the wild type.

Purification of XPF-GST fusion

The bacterial pellets were resuspended in ~35 ml buffer G/1 M NaCl (buffer G = 20 mM MES pH 6.5, 1 mM DTT, 15% glycerol). This was sonicated for 4 × 2 min with cooling then centrifuged 48 000 g for 20 min. The supernatant was retained and 1–2 ml glutathione agarose beads added and rotated at 4°C for 2 h or overnight. The beads were washed in 4 × 10 ml buffer G/1 M NaCl with 10 min rotating at 4°C and this was repeated another three times with buffer G/100 mM NaCl. Expression and purification of GST-XPF was checked by SDS-PAGE.

Trypsin digestion of XPF

Recombinant XPF (5 µg) was incubated with doubling dilutions of trypsin from 0.2 µg to 6 ng (i.e. 1:25–1:800 w/w ratio trypsin:XPF) and 2.5 mM CaCl_2 at 30°C for 1.5 h and the reaction stopped by adding 50 mM AEBSF protease inhibitor. The digests were run on SDS-PAGE gel and a sample of the 1:25 digest analysed by MALDI-ToF mass spectrometry.

GST-XPF fusion affinity chromatography

Glutathione agarose beads (1 ml) and 5 ml buffer G/100 mM NaCl were added to the bead-GST-XPF solution and loaded onto a disposable 5 ml polypropylene column (Pierce). A control column with GST alone attached to beads was also set up and both washed with 4 × 10 ml buffer G/100 mM NaCl. To prepare the soluble *S. solfataricus* extract 2 g of *S. solfataricus* P2 biomass (supplied by the Centre for Extremophile Research, Porton Down, UK) per column was defrosted, suspended in buffer G/100 mM NaCl, sonicated for 4 × 2 min with cooling and centrifuged at 48 000 g for 20 min. The supernatant was removed, filtered and passed through the columns. The columns were washed with 3 × 7 ml buffer G/100 mM NaCl to remove any unbound protein. A full column of buffer G/200 mM NaCl was added and the first 3 ml of the fraction collected. Columns were washed with 3 × 7 ml buffer G/200 mM NaCl and the fraction collection and washing steps repeated with buffer G/350 mM NaCl and 600 mM NaCl. Columns were stored in buffer G/1 M NaCl. The fractions were TCA precipitated and analysed by SDS-PAGE. Bands present in GST-XPF but absent in GST control were cut out and identified by MALDI-ToF mass spectrometry.

GST pull-downs

Ten microlitres of GST/GST-XPF or GST-PCNA beads were rotated in 500 µl soluble *S. solfataricus* extract for between 2 h and overnight. The beads were then washed with

4 × 200 µl buffer G/400 mM NaCl. Beads were heated in 2 × LDS sample buffer (Invitrogen), and samples analysed by SDS-PAGE and Coomassie staining or Western blotting. The same procedure was utilized with GST-PCNA subunit 1, 2 and 3 beads (purified as described in Dionne *et al.*, 2003). Primary antibodies raised against recombinant XPF (Diagnostics Scotland) and recombinant PCNA1, 2 and 3 (Dionne *et al.*, 2003) were used for immunological detection.

DNA substrates

The following oligos were used for catalytic assays:

b-strand
5'-CCTCGAGGGA TCCGTCCTAGCAAGCCGCTGCTACCG
GAAGCTTCTGGACC-3'

h-strand
5'-GGTCCAGAAGCT TCCGGTAGCAGCGAGAGCGGTGG
TTGAATTCCTCGAGC-3'

r-strand
5'-CGTCGAGGAAT TCAACCACCGCTCT TCTCAACTGCA
GTCTAGACTCGAGC-3'

x-strand
5'-GCTCGAGTCTAGACTGCAGT TGAGAGCTTGCTAGGA
CGGATCCCTCGAGG-3'

b complement with 10 base bubble
5'-GGTCCAGAAGCTTCCGGTAGGCTACCGCACCTAGGA
CGGATCCCTCGAGG-3'

R25-strand
5'-TCTCAACTGCAGTCTAGACTCGAGC-3'

The b, h, r and x strands form a fixed four way junction (Junction 1) with 25 bp arms when hybridized together. Hybridization of the b- and x- strands results in formation of a splayed duplex as used in Fig. 4A. Hybridization of b- and b-comp bubble strands forms a 10 nt bubble flanked on either side by 20 bp duplex DNA, as used in Fig. 4A. Hybridization of the b and x strands together with the r25 strand formed the 3' flap structure used in kinetic assays (Fig. 4B). Oligonucleotides were 5' [^{32}P]-end-labelled and assembled into various structures by slow cooling from 85°C to room temperature overnight, purified on a native 6% acrylamide gel and electroeluted from the gel as described in (White and Lilley, 1996). Size markers (G + A) were prepared from labelled oligonucleotides using standard protocols.

Endonuclease assays

Nuclease reactions (10 µl) were performed in 30 mM Hepes pH 7.6, 5% glycerol, 40 mM KCl, 8 mM MgCl_2 , 0.1 mg ml⁻¹ bovine serum albumin and 0.1 mg ml⁻¹ calf thymus DNA. DNA substrate (50–100 nM) was used per assay and where appropriate first incubated with 2.5 µM PCNA at room temperature for 5 min, 2 µM XPF was added and the reaction incubated at room temperature for a further 5 min then at 55°C for 5–15 min depending on the substrate. Reactions were stopped by adding 90 µl TE buffer (10 mM Tris pH 8, 1 mM EDTA) and the DNA was ethanol precipitated and

resuspended in formamide loading buffer. Samples were heated at 95°C for 5 min, cooled on ice and loaded onto a denaturing (7 M urea) 12% polyacrylamide gel containing 1 × TBE and run at 50°C and 90 watts for 1 h. Gels were visualized by phosphorimaging (Fuji FLA5000) and/or autoradiography.

For single turnover kinetic rate measurements, reactions (20 µl final volume) were set up in 30 mM Hepes pH 7.6, 5% glycerol, 40 mM KCl, 1 mM EDTA, 0.1 mg ml⁻¹ bovine serum albumin, and 0.1 mg ml⁻¹ calf thymus DNA with 80 nM DNA substrate. DNA was incubated with 2.5 µM PCNA at room temperature for 5 min. 1 µM fresh XPF was added and reaction preheated to 55°C for 2 min. Cleavage was initiated by adding MgCl₂ to a final concentration of 10 mM and mixing briefly. Aliquots (5 µl) were taken at selected time points and added to chilled stop solution (10 mM Tris pH 8, 5 mM EDTA, 0.1 mg ml⁻¹ calf thymus DNA) to terminate the reaction. The DNA was ethanol precipitated and processed as described above. In experiments where SSB or Alba proteins were included in the assay, they were added to the DNA after the addition of PCNA and before the addition of XPF. Addition of SSB prior to addition of XPF gave the same effect as addition after XPF.

Acknowledgements

Thanks to Ross Wadsworth for initial subcloning of the XPF gene, Catherine Botting for expert MALDI-TOF analyses and to Derek Richard for helpful discussion. Thanks to Emma Warbrick for very helpful advice on PCNA interaction motifs. M.F.W. is a Royal Society University Research Fellow. This work was funded by the BBSRC.

References

- Araujo, S.J., and Wood, R.D. (1999) Protein complexes in nucleotide excision repair. *Mutat Res* **435**: 23–33.
- Aravind, L., Walker, D.R., and Koonin, E.V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* **27**: 1223–1242.
- Boddy, M.N., Lopez-Girona, A., Shanahan, P., Interthal, H., Heyer, W.D., and Russell, P. (2000) Damage tolerance protein Mus81 associates with the FHA1 domain of checkpoint kinase Cds1. *Mol Cell Biol* **20**: 8758–8766.
- Boddy, M.N., Gaillard, P.H., McDonald, W.H., Shanahan, P., Yates, J.R., III, and Russell, P. (2001) Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell* **107**: 537–548.
- Chen, X.B., Melchionna, R., Denis, C.M., Gaillard, P.H., Blasina, A., Van de Weyer, I., et al. (2001) Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Mol Cell* **8**: 1117–1127.
- Constantinesco, F., Forterre, P., and Elie, C. (2002) NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. *EMBO Rep* **3**: 537–542.
- Davies, A.A., Friedberg, E.C., Tomkinson, A.E., Wood, R.D., and West, S.C. (1995) Role of the rad1 and rad10 proteins in nucleotide excision-repair and recombination. *J Biol Chem* **270**: 24638–24641.
- Dionne, I., Nookala, R.K., Jackson, S.P., Doherty, A.J., and Bell, S.D. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell* **11**: 275–282.
- Doherty, A.J., Serpell, L.C., and Ponting, C.P. (1996) The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res* **24**: 2488–2497.
- Enzlin, J.H., and Scharer, O.D. (2002) The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J* **21**: 2045–2053.
- Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J* **16**: 6559–6573.
- Gaillard, P.H., and Wood, R.D. (2001) Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res* **29**: 872–879.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498–511.
- Gary, R., Ludwig, D.L., Cornelius, H.L., MacInnes, M.A., and Park, M.S. (1997) The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J Biol Chem* **272**: 24522–24529.
- Grogan, D.W. (2000) The question of DNA repair in hyperthermophilic archaea. *Trends Microbiol* **8**: 180–185.
- Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1994) Holliday junction cleavage by yeast Rad1 protein. *Nature* **371**: 531–534.
- Interthal, H., and Heyer, W.D. (2000) MUS81 encodes a novel helix-hairpin-helix protein involved in the response to UV- and methylation-induced DNA damage in *Saccharomyces cerevisiae*. *Mol Gen Genet* **263**: 812–827.
- Kaliraman, V., Mullen, J.R., Fricke, W.M., Bastin-Shanower, S.A., and Brill, S.J. (2001) Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes Dev* **15**: 2730–2740.
- Komori, K., Fujikane, R., Shinagawa, H., and Ishino, Y. (2002) Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst* **77**: 227–241.
- Komori, K., Sakae, S., Shinagawa, H., Morikawa, K., and Ishino, Y. (1999) A Holliday junction resolvase from *Pyrococcus furiosus*: functional similarity to *Escherichia coli* RuvC provides evidence for conserved mechanism of homologous recombination in bacteria, eukarya, and archaea. *Proc Natl Acad Sci USA* **96**: 8873–8878.
- Kuraoka, I., Kobertz, W.R., Ariza, R.R., Biggerstaff, M., Essigmann, J.M., and Wood, R.D. (2000) Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. *J Biol Chem* **275**: 26632–26636.
- Kvaratskhelia, M., Wardleworth, B.N., Norman, D.G., and White, M.F. (2000) A conserved nuclease domain in the archaeal holliday junction resolving enzyme Hjc. *J Biol Chem* **275**: 25540–25546.
- de Laat, W.L., Appeldoorn, E., Jaspers, N.G., and Hoeijmak-

- ers, J.H. (1998a) DNA structural elements required for ERCC1-XPF endonuclease activity. *J Biol Chem* **273**: 7835–7842.
- de Laat, W.L., Appeldoorn, E., Sugawara, K., Weterings, E., Jaspers, N.G., and Hoeijmakers, J.H. (1998b) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev* **12**: 2598–2609.
- de Laat, W.L., Jaspers, N.G., and Hoeijmakers, J.H. (1999) Molecular mechanism of nucleotide excision repair. *Genes Dev* **13**: 768–785.
- Matsumiya, S., Ishino, Y., and Morikawa, K. (2001) Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci* **10**: 17–23.
- Ogrunc, M., Becker, D.F., Ragsdale, S.W., and Sancar, A. (1998) Nucleotide excision repair in the third kingdom. *J Bacteriol* **180**: 5796–5798.
- Saijo, M., Kuraoka, I., Masutani, C., Hanaoka, F., and Tanaka, K. (1996) Sequential binding of DNA repair proteins RPA and ERCC1 to XPA in vitro. *Nucleic Acids Res* **24**: 4719–4724.
- Seitz, E.M., and Kowalczykowski, S.C. (2000) The DNA binding and pairing preferences of the archaeal RadA protein demonstrate a universal characteristic of DNA strand exchange proteins. *Mol Microbiol* **37**: 555–560.
- Sgouros, J., Gaillard, P.H., and Wood, R.D. (1999) A relationship between a DNA-repair/recombination nuclease family and archaeal helicases. *Trends Biochem Sci* **24**: 95–97.
- Wadsworth, R.I., and White, M.F. (2001) Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. *Nucleic Acids Res* **29**: 914–920.
- Warbrick, E. (2000) The puzzle of PCNA's many partners. *Bioessays* **22**: 997–1006.
- White, M.F., and Lilley, D.M.J. (1996) The structure-selectivity and sequence-preference of the junction-resolving enzyme CCE1 of *Saccharomyces cerevisiae*. *J Mol Biol* **257**: 330–341.
- Yang, H., Chiang, J.H., Fitz-Gibbon, S., Lebel, M., Sartori, A.A., Jiricny, J., Slupska, M.M., and Miller, J.H. (2002) Direct interaction between uracil-DNA glycosylase and a proliferating cell nuclear antigen homolog in the crenarchaeon *Pyrobaculum aerophilum*. *J Biol Chem* **277**: 22271–22278.

AN ARCHAEOAL ENDONUCLEASE DISPLAYS KEY PROPERTIES OF BOTH EUKARYAL XPF-ERCC1 AND MUS81

Jennifer Roberts and Malcolm F. White

From the Centre for Biomolecular Sciences, University of St Andrews, St Andrews, KY16 9ST, United Kingdom.

Running Title: Archaeal XPF endonuclease

Address correspondence to: Malcolm F. White, Centre for Biomolecular Sciences, University of St Andrews, St Andrews, KY16 9ST, United Kingdom. Tel +44-1334-463432; Fax +44-1334-462595; email mfw2@st-and.ac.uk;

Structure specific nucleases of the XPF/Mus81 family function in several DNA recombination and repair pathways in eukaryotes, cleaving a variety of flap and branched DNA substrates. Mus81 and XPF are clearly related evolutionarily, but differ markedly in their substrate specificity and protein partners. We demonstrate that the XPF endonuclease from *Sulfolobus solfataricus*, which is dependent on the sliding clamp PCNA for activity, represents an ancestral form of the XPF/Mus81 family, with key properties in common with both enzymes. The archaeal XPF has a domain organisation and sequence preference very similar to eukaryal XPF-ERCC1. However, the archaeal enzyme has a pronounced preference for Mus81-type substrates such as D loops, nicked four way junctions and 3' flaps. These all have in common a 5' DNA end next to the cleavage site. The availability of the sliding clamp PCNA may dictate the activity of *Sulfolobus* XPF *in vivo*.

Branched DNA structures are created in many circumstances, arising both as a direct consequence of DNA damage and as intermediates in cellular DNA recombination and repair pathways. These structures must be processed correctly by structure specific nucleases prior to DNA replication. In eukaryotes, the related nucleases XPF-ERCC1 and Mus81-Eme1/Mms4 (abbreviated here to Mus81*) process a variety of DNA substrates with 3' branches and bubbles (see Table 2 for a

definition of the DNA substrates). The cellular functions of these important enzymes are still being delineated, as are the molecular basis for their interaction with DNA substrates.

Most euryarchaea encode a clear homolog of eukaryal XPF, including the N-terminal helicase and C-terminal nuclease domains. The homolog from *Pyrococcus furiosus*, Hef, cleaves a variety of fork and flap structures that are substrates for eukaryal Mus81* (1). The crystal structure of the nuclease domain of *Pyrococcus* Hef unveiled the relationship of this domain with a superfamily of nucleases including the restriction enzymes, Holliday junction resolving enzymes Hjc and T7 endonuclease I and other DNA repair enzymes (2).

Crenarchaeal XPF homologs lack the N-terminal helicase domain altogether, and consist only of the nuclease and HhH₂ domains (Figure 3). In this respect they may represent the simplest and perhaps ancestral version of the XPF/Mus81 nuclease. We showed previously that the XPF homolog from *Sulfolobus solfataricus* (*SsoXPF*) cuts flap or splayed duplex structures (3), and that its activity is crucially dependent on an association with the sliding clamp PCNA, via a C-terminal PCNA interaction motif (PIP) (3). Here we report the detailed kinetic analysis of the substrate specificity of *Sulfolobus* XPF. We demonstrate key similarities with both the eukaryal XPF and Mus81* nucleases.

Materials and Methods

Recombinant proteins

Sulfolobus solfataricus XPF and PCNA heterotrimer were expressed and purified as described previously (3), (4).

DNA substrates

The oligonucleotides used for substrates (Table 1) were 5'-[³²P]-end-labelled and assembled into various structures (Table 2) by slow cooling from 85 °C to room temperature for 3 hours or overnight. Substrates were purified on a native 6 % acrylamide gel followed by electro-elution and ethanol precipitation. Size markers (G & A) were prepared from labelled substrates using standard protocols.

Endonuclease assays

Nuclease reactions (10 µl) were assembled in 30 mM HEPES pH 7.6, 5 % glycerol, 40 mM KCl, 0.1 mg/ml BSA and 0.1 mg/ml calf thymus DNA with 80 nM DNA substrate and 1 µM XPF-PCNA and equilibrated at 55 °C. Cleavage was initiated by adding MgCl₂ to a final concentration of 10 mM, mixed briefly, and incubated at 55 °C. 5 µl aliquots were taken at selected time points and added to chilled stop solution (10 mM Tris-HCl pH 8, 10 mM EDTA, 0.1 mg/ml calf thymus DNA) to terminate the reaction. DNA was ethanol precipitated and analysed on a denaturing polyacrylamide/urea/TBE gel. Uncut substrate and products were quantified (Image Gauge, Fuji) and the reaction rate obtained by linear regression. In the experiment to determine the influence of a 5' phospho-group on activity, the PCNA-XPF concentration was reduced to 0.1 µM to allow more accurate quantitation.

RESULTS

Structure specificity and requirements of *Sulfolobus* XPF

To pin down the function of structure specific nucleases, it is necessary to define their specificity for the different branched DNA structures that arise *in vivo* due to

DNA damage and repair pathways. Quantitative data for eukaryal XPF-ERCC1 and Mus81* are not yet available. We have carried out a systematic quantitative analysis of the substrate preference of *Sso*XPF using a panel of branched DNA substrates (Table 2). In each case, the sequence of the strand cleaved by XPF is the same, although its length varies in some cases.

In general terms, the order of substrate preference of *Sso*XPF matches that of the Mus81* family rather than the eukaryal XPF family. The best substrate is a 3'-flap, followed by a nicked duplex, nicked 3-way and nicked 4-way junctions. The efficient cleavage of a nicked duplex substrate demonstrates that no single stranded flap is required, rather a discontinuity in the DNA duplex is enough to promote the nuclease activity. These DNA species are all the preferred substrates of Mus81*, and share in common a 5' DNA end near the branch point or point of discontinuity that has been shown to be essential in directing the activity of Mus81* (5). The 3'-flap and nicked 3-way substrates are analogs of a stalled and reversed replication fork, respectively (6). In contrast, a substrate analog of an active replication fork with a 3' DNA end near the branch point is cut much more slowly, again in common with Mus81* (6). A splayed duplex species, the preferred substrate of eukaryotic XPF-ERCC1, is cleaved approximately 10-fold more slowly than a 3'-flap substrate. Lastly, four-way junctions are very poor substrates for *Sso*XPF, cut more than 2 orders of magnitude more slowly than the preferred substrates.

The clear preference of *Sso*XPF for a 5' DNA end near the cleavage site mirrors that reported previously for Mus81* from *S. cerevisiae* and *S. pombe* (6), (5). *In vivo*, most substrates of this type would have a phosphate group at this 5' position, though this was not present in the synthetic substrates used for Mus81*. To test whether phosphorylation of this terminus was important for the activity of *Sso*XPF,

we compared the rates of cleavage with a nicked duplex substrate with and without a 5'-phosphate at the nick (Figure 1). We found a measurable though modest increase in the cleavage rates of a phosphorylated substrate compared to a non-phosphorylated one, suggesting that the archaeal XPF (and probably also Mus81*) does not require the presence of a 5'-phosphate to recognise this DNA end.

Recent studies of Mus81* suggest it can play an important role in the generation of crossover recombination events by cleavage of D loop (Displacement loop) substrates formed as a consequence of strand invasion after a chromosome sustains a double strand break (7), (8). A D loop bubble is a poor substrate for *SsoXPF* in the absence of an invading 3' ssDNA end (Figure 2B, left). This structure is equivalent to an NER bubble, the favoured substrate of eukaryal XPF-ERCC1 (9). Weak cleavage of the b75 strand is seen at the splayed duplex junction (white arrows in figure 2B, rate 0.02 min^{-1}). The z75 strand is also cut weakly at the equivalent site on the other side of the bubble (rate 0.054 min^{-1}). Once the invading strand is present, however, *SsoXPF* is strongly activated to cut the b75 strand at a new position (black arrows in figure 2B; a) in figure 2C, 2D). The invading strand creates a good substrate for *SsoXPF*, as observed for Mus81*, probably stimulated by the 5' DNA end paired with the invading strand (5' end of oligo r26-50 in figure 2A) (7). The rate of cleavage at this new position is at least 50-fold faster than those for the substrate without the invading strand. Cleavage at site b) on the z75 strand (grey arrows in figure 2B; b in figure 2C, 2D) is also stimulated by the presence of the invading duplex, but is cut significantly more slowly than site a). Second end capture and DNA synthesis results in formation of a 3' flap substrate (site c) that is the most preferred substrate for *SsoXPF* (Table 2). This is the second site proposed by Osman et al for Mus81* (7). Cleavage at sites a) and c), coupled

with capture of the second strand, DNA synthesis and strand exchange would result in crossover recombination products (figure 2D). Any 3' or 5' flaps generated as a result of this process could be removed by XPF or Fen1. Cleavage at site a) alone would dissociate a D loop, and could be functionally relevant for resetting of collapsed replication forks (10). Every strand scission was mapped 3' to pyrimidine residues, confirming that although *SsoXPF* cuts preferentially at Mus81*-type substrates, it retains key characteristics of XPF-ERCC1.

DISCUSSION

In addition to the clear similarities in their domain organisation (Figure 3), the archaeal XPF has the same sequence preference for cleavage at pyrimidines as eukaryal XPF (3). In contrast, Mus81* has little or no sequence dependence (5). *SsoXPF* has a significant activity against splayed duplex substrates that are found at the 5' side of lesions repaired by the eukaryal XPF protein during NER. These substrates are cut very poorly by Mus81*. Taken together, these data suggest a close relationship between the eukaryal and archaeal XPF enzymes.

On the other hand, our kinetic analysis of the substrate preference of *SsoXPF* demonstrates clearly that the best substrates resemble stalled replication forks, D loops and nicked Holliday junctions, with a 5' DNA end near the branch point or strand discontinuity (Table 2 and Figure 3) – a feature in common with Mus81*. Cleavage of 3'-flap substrates by Mus81-Mms4 is strongly stimulated by the presence of a 5'-terminus within 4 nt of the branch point (5). The highly efficient cleavage of a nicked duplex substrate demonstrates that the archaeal XPF protein seeks out strand discontinuities that allow DNA deformation, flanked on either side by duplex DNA, rather than branched DNA substrates *per se*. This is also the case for

Mus81*, which cleaves nicked duplexes efficiently to generate 5 nt gapped duplexes (5). These substrates are cut very poorly by eukaryal XPF-ERCC1. D loop substrates are also cut by *SsoXPF* and Mus81* in a similar way, with strong stimulation observed when strand invasion occurs.

Therefore, the archaeal XPF clearly has the potential to act as both a Mus81-like flap endonuclease and an XPF-like repair endonuclease. Previous comparisons of the activity of Mus81-Mms4 with Rad1-Rad10 have come to the conclusion that the specificities of the two enzymes do not overlap. The presence of a 5' DNA terminus within 4 nt of the branch point results in strong stimulation of Mus81-Mms4 and reduces the activity of Rad1-Rad10 such that the two enzymes have at least a 100 fold preference for their specific substrates (5). *SsoXPF* bridges these two diverged proteins, and likely represents the ancestral form of both enzymes.

Archaeal DNA recombination and repair

Hyperthermophiles such as *Sulfolobus solfataricus* suffer a higher frequency of DNA damage such as hydrolytic deamination, oxidation and single strand breaks as a consequence of their higher growth temperatures (11). Many of these lesions are known to stall or collapse replication forks – for example a single strand break is converted to a double strand break if replicated. Stalled and collapsed forks are a frequent event in bacteria and eukarya, which consequently have efficient mechanisms to recover from these potentially catastrophic events. One mechanism for fork rescue is via the formation of a Holliday junction, which can be resolved to allow replication restart (reviewed in (12)). The substrate specificity we have described for *SsoXPF* is consistent with a role in the cleavage of stalled replication forks to generate

substrates for recombination, and/or resolution of early recombination intermediates (D loops), similar to Mus81*. In this regard the requirement of *SsoXPF* for PCNA as an essential cofactor may be relevant, as the enzyme's target sites may be determined in part by PCNA loading. This is well illustrated by consideration of the D loop substrate (Figure 2D). The recombination scheme requires extension by DNA replication of both the invading 3' strand (strand x50) and of the second DNA end after capture. This replication will result in loading of PCNA on both these duplexes. This may direct *SsoXPF* to cut at sites a) and c), thus generating recombinogenic DNA ends. In contrast, cleavage at site b) could be disfavoured *in vivo* as PCNA will not be loaded on this duplex.

The data presented here do not rule out the possibility that *SsoXPF* also functions in an archaeal NER pathway. The function of eukaryal XPF-ERCC1 is largely determined by the protein:protein interactions it makes. In Nucleotide Excision Repair, this includes specific interactions between ERCC1 and XPA (13), and between XPF and the C-terminus of the helicase XPB, a component of the multiprotein complex TFIIH (14). Thus XPF-ERCC1 is recruited into the assembling repair complex and positioned to introduce a single strand cleavage 5' of a DNA lesion. XPF-ERCC1 also has roles in the resolution of DNA inter-strand cross-links (15), (16) and in targeted gene replacement (17). These functions link XPF-ERCC1 to recombinational pathways and necessitate recruitment of the enzyme into different DNA processing complexes from those formed during NER. Similarly, interactions of *SsoXPF* mediated by PCNA may help to define its role *in vivo*.

REFERENCES

1. Komori, K., Fujikane, R., Shinagawa, H., and Ishino, Y. (2002) *Genes Genet. Syst.* **77**, 227-241
2. Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2003) *Structure (Camb)* **11**, 445-457
3. Roberts, J. A., Bell, S. D., and White, M. F. (2003) *Mol. Micro.* **48**, 361-371
4. Dionne, I., Nookala, R. K., Jackson, S. P., Doherty, A. J., and Bell, S. D. (2003) *Mol Cell Biol* **11**, 275-282
5. Bastin-Shanower, S. A., Fricke, W. M., Mullen, J. R., and Brill, S. J. (2003) *Mol Cell Biol* **23**, 3487-3496
6. Whitby, M. C., Osman, F., and Dixon, J. (2003) *J Biol Chem* **278**, 6928-6935
7. Osman, F., Dixon, J., Doe, C. L., and Whitby, M. C. (2003) *Mol Cell Biol* **12**, 761-774
8. Hollingsworth, N. M., and Brill, S. J. (2004) *Genes Dev* **18**, 117-125
9. de Laat, W. L., Appeldoorn, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998) *J Biol Chem* **273**, 7835-7842
10. Doe, C. L., Osman, F., Dixon, J., and Whitby, M. C. (2004) *Nucleic Acids Res* **32**, 5570-5581
11. Grogan, D. W. (2000) *Trends Microbiol* **8**, 180-185
12. McGlynn, P., and Lloyd, R. G. (2002) *Nat Rev Mol Cell Biol* **3**, 859-870
13. Li, L., Lu, X., Peterson, C. A., and Legerski, R. J. (1995) *Mol. Cell. Biol.* **15**, 5396-5402
14. Evans, E., Moggs, J. G., Hwang, J. R., Egly, J. M., and Wood, R. D. (1997) *Embo J* **16**, 6559-6573
15. Kuraoka, I., Kobertz, W. R., Ariza, R. R., Biggerstaff, M., Essigmann, J. M., and Wood, R. D. (2000) *J Biol Chem* **275**, 26632-26636
16. Niedernhofer, L. J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A. F., de Wit, J., Jaspers, N. G., Beverloo, H. B., Hoeijmakers, J. H., and Kanaar, R. (2004) *Mol Cell Biol* **24**, 5776-5787
17. Niedernhofer, L. J., Essers, J., Weeda, G., Beverloo, B., de Wit, J., Muijtjens, M., Odijk, H., Hoeijmakers, J. H., and Kanaar, R. (2001) *Embo J* **20**, 6540-6549

FOOTNOTES

Thanks to the BBSRC for financial support. MFW is a Royal Society URF.

FIGURE LEGENDS

Figure 1. Effect of phosphorylation on substrate cleavage rate.

The rate of XPF-PCNA cleavage for nicked duplex substrates with and without a 5' phosphorylated end adjacent to the nick (open and closed circles, respectively). The presence of a 5' phospho-group results in only a two-fold stimulation of endonuclease activity. Standard errors are shown. Black spots indicate the 5' end of the ³²P-labeled strand, and the star indicates the 5' phosphate adjacent to the nick.

Figure 2. Sulfolobus XPF cleaves a D loop substrate.

- (A) Schematic diagram showing the D loop strand nomenclature used in the text. Open circles indicate 5' DNA ends.
- (B) Denaturing polyacrylamide gel showing cleavage of D loop DNA substrate. Cartoons

above gel represent the substrate, with the labeled strand indicated by a black circle at its 5' end. Major cleavage sites are indicated with black and grey arrows. Minor cleavage of the splayed duplex junction on strand b75 is indicated by the white arrow. Time points were 1, 5 & 10 minutes; m = A & G size markers of the labeled strand; control c1 = DNA alone; c2 = DNA + *Sso*XPF; c3 = DNA + PCNA.

- (C) Schematic representation of the central 60 bp of the D loop substrate showing the invading strand in bold. Black and grey arrows indicate the major sites of cleavage by *Sso*XPF. All strand scissions are mapped 3' to pyrimidine residues.
- (D) Cartoon of D loop formation and cleavage by *Sulfolobus* XPF. Following a double strand break of the grey DNA duplex and generation of 3' tails, strand invasion generates substrates for *Sso*XPF, which cuts most efficiently at site a). Second end capture and DNA synthesis generates a second *Sso*XPF substrate (a 3'-flap or nicked Holliday junction) at c). Cleavage at sites a) and c) coupled with strand exchange, gap filling, flap removal and ligation would generate crossover recombinant products. Cleavage at site b) may be disfavoured *in vivo* due to the requirement for PCNA, which would be loaded during DNA replication near sites a) and c).





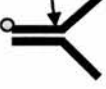
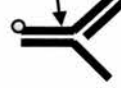





Figure 3. Structure and evolution of the XPF/Mus81 nuclease family

The conserved nuclease (nuc) domain is found in many other nucleases involved in DNA restriction and repair pathways. In the XPF/Mus81 family, the nuclease domain is linked to a helix hairpin helix (HhH₂) DNA binding domain. The archaeal XPF enzymes are homodimers whereas the eukaryotic enzyme has duplicated to form heterodimeric proteins, allowing subunit specialisation, and to form nucleases with differing substrate specificity. The archaeal and Mus81 enzymes show a distinct preference for substrates with a 5' DNA terminus (red circle), whereas eukaryal XPF-ERCC1 has specialised for cleavage of splayed duplex or bubble substrates.

Table 1. Oligonucleotides used for DNA substrates

No	NAME	SEQUENCE (5' to 3')
1	b50	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTTCTGGACC
2	b25	CCTCGAGGGATCCGTCCTAGCAAGC
3	b75	GGAGCGGTGGTTGAATTCCTCGACGCCTCGAGGGATCCGTCCTAGCAAGC CGCTGCTACCGGAAGCTTCTGGACC
4	10nt bubble	GGTCCAGAAGCTTCCGGTAGGCTACCGCACCTAGGACGGATCCCTCGAGG
5	x50	GCTCGAGTCTAGACTGCAGTTGAGAGCTTGCTAGGACGGATCCCTCGAGG
6	x26-50	GCTTGCTAGGACGGATCCCTCGAGG
7	h50	GGTCCAGAAGCTTCCGGTAGCAGCGAGAGCGGTGGTTGAATTCCTCGACG
8	h25	GGTCCAGAAGCTTCCGGTAGCAGCG
9	h26-50	AGAGCGGTGGTTGAATTCCTCGACG
10	r50	CGTCGAGGAATTCAACCACCGCTCTTCTCAACTGCAGTCTAGACTCGAGC
11	r25	CGTCGAGGAATTCAACCACCGCTCT
12	r26-50	TCTCAACTGCAGTCTAGACTCGAGC
13	Y50	CGTCGAGGAATTCAACCACCGCTCTGCTTGCTAGGACGGATCCCTCGAGG
14	z75	GGTCCAGAAGCTTCCGGTAGCAGCGTTTTTTTTTTTTTTTTTTTTTTTTTTTCG TCGAGGAATTCAACCACCGCTCC

Table 2. Substrate preference of *Sulfolobus* XPF^a

SUBSTRATE	OLIGOS	STRUCTURE	RATE (min ⁻¹) ±S.E.
3' flap	b50, r26-50, x50		6.8
nicked duplex	b25, r26-50, x50		6.2
nicked 3-way	b50, h25, r26- 50, x50		3.6 ± 0.2
nicked 4-way	b50, h50, r25, r26-50, x50		2.5 ± 0.2
splayed duplex	b50, x50		0.61 ± 0.07
replication fork	b50, h25, x50		0.42 ± 0.02
5' overhang	b25, x50		0.074 ± 0.003
4-way junction	b50, h50, r50, x50		0.057 ± 0.005
10 nt bubble	b50, 10nt bubble		0.051 ± 0.006
3-way junction	b50, y50, x50		0.049 ± 0.001
3' overhang	b50, x26-50		0.027 ± 0.001

^a DNA substrates were assembled with oligos indicated (sequences in table 1). The labelled strand is indicated with a circle at the 5' end, and cleavage sites with an arrow. All rates were measured under pseudo single turnover conditions as described in materials and methods. Rates for the two fastest-cut substrates were estimated from initial time points, and no standard error is indicated.

Figure 1

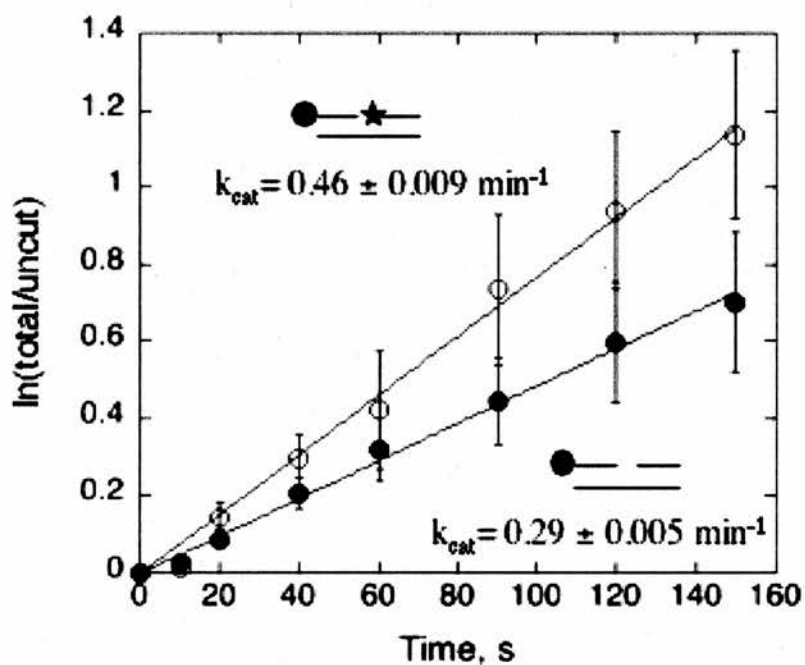
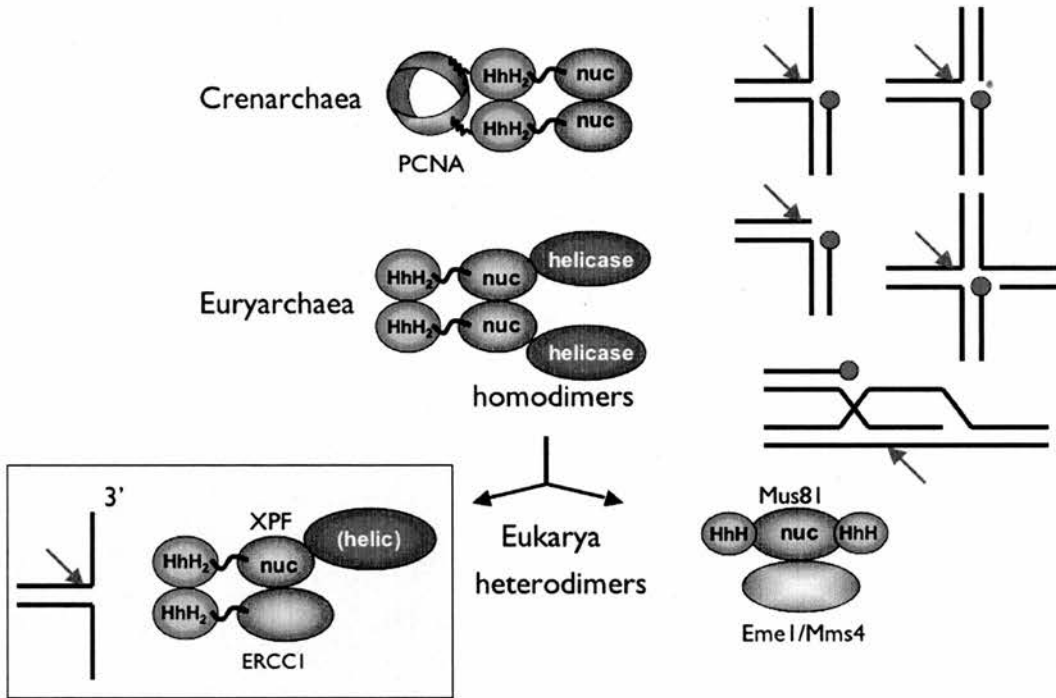


Figure 3



EXPRESSION OF DELAYED CELL DEATH AND DNA REPAIR IN HUMAN EPITHELIAL CELL LINES FOLLOWING EXPOSURE TO ULTRAVIOLET RADIATION

G. Bakirtzis, C. V. Briscoe, C. Peddie and A. C. Riches
School of Biology, Medical Sciences and Human Biology
University of St Andrews, St Andrews, Fife, KY16 9TS, UK

Abstract—The long term effects of UVA and UVB have been investigated using two human epithelial cell lines, HTori-3 (a human thyroid epithelial cell line) and 340 RPE-T53 (a human retinal pigment epithelial cell line). There was a marked difference in clonogenic survival following exposure between the two cell lines. DNA repair studies were undertaken using ara-C treatment. Ara-C administered immediately after UVB exposure, reduced survival in both cell lines indicating that DNA repair was inhibited. The plating efficiency, as an index of delayed cell death of both cell lines measured up to 20 population doublings following exposure to UV was reduced in a dose dependent manner after exposure to UVB but not to UVA.

INTRODUCTION

Over the past few decades, an increased incidence of skin cancer has been reported in fair-skinned populations around the world. There is growing evidence that the continuing depletion of the ozone layer leads to increased levels of UV irradiation which causes the increased incidence of skin cancer.

A large number of studies have used a variety of cell lines to show that exposure to irradiation or certain genotoxic agents can induce a type of delayed cell damage. This delayed damage is expressed as a reduction in plating efficiency in the progeny of irradiated cells. This reduction in plating efficiency is usually used as an endpoint to measure this effect. This effect has been given different names and in the current study it will be referred to as delayed cell death (DCD). Other endpoints used for measurement of DCD include giant cell formation, chromosome instability, micronucleus formation and increase in the mutation frequency in specific genes.

Several theories have been published to explain the mechanism of DCD. Little⁽¹⁾, suggested that a persistent increase in oxy-radical generation was produced in the progeny of irradiated cells leading to oxidative damage and therefore an increase in cell death, and chromosome aberrations. Stamato *et al*⁽²⁾, suggested that persistence of DNA lesions like the cyclobutane pyrimidine dimers increased the probability of base misrepairing errors during DNA replication leading eventually to the expression of this phenotype. He also proposed another theory, implying that DCD developed as a secondary consequence of DNA damage. This damage could create a mutator phenotype by reducing the ability of the replication machinery or DNA repair process. The nature of the initial lesion which induced this phenomenon remains unclear. It has been suggested that DNA breaks must play a very significant role in the expression of DCD. Support for this theory comes from Chang *et al*⁽³⁾, who showed that restriction endonucleases which can

cause double strand breaks, produced DCD in CHO cells whereas UVC did not.

The predominant mechanism by which DCD occurs has been shown to be apoptosis. The majority of the colonies which are produced from the survivors of irradiated cells contain apoptotic bodies. Mothersill *et al*⁽⁴⁾, showed high levels of bcl-2 expression after gamma irradiation in cells which have no expression of bcl-2 before irradiation. Since bcl-2 functions as an anti-apoptotic gene, it therefore promotes survival of a proportion of irradiated cells which may carry unrepaired damage.

The level of unrepaired damage in surviving cells may be related to DNA repair activity. Alper *et al*⁽⁵⁾ demonstrated that DCD was not observed in DNA repair deficient cell lines. DNA breaks must also play a very significant role in the expression of DCD. It has been suggested that DNA breaks on their own are unlikely to trigger this phenomenon but that DNA breaks combined with oxidative stress might be able to induce DCD⁽⁶⁾.

The induction of DCD after exposure to UV irradiation is poorly understood. CHO-K1 cells irradiated with UVC irradiation did not exhibit DCD, as expressed by the reduction in plating efficiency as an endpoint⁽³⁾. However, Stamato *et al*⁽²⁾ concluded that UVC can induce DCD expression in the CHO cell line, when mutational events on the G6PD and the HPRT locus were analysed. Recently, it has been shown that DCD can be expressed in the HACAT cell line following either UVA or UVB exposure⁽⁷⁾.

The importance which DCD might play in human UV induced carcinogenesis, suggested a more detailed study of this process should be undertaken. The primary purpose of the current study was to establish if DCD occurs following exposure to UVA or UVB. Since the majority of human cancers are of epithelial origin, it is important to understand the mechanism by which UV may induce long term alterations in human cells. The investigation of the delayed decrease in stability of the genome may

give an insight into the accumulation of long term multiple genetic changes usually observed in tumours. In addition, DNA repair studies have been conducted using ara-C as an inhibitor of DNA repair in order to elucidate any existing differential capacities for repair in the two cell lines in this study.

MATERIALS AND METHODS

Cell culture conditions

The HTori-3 cell line is a human thyroid epithelial cell line⁽⁸⁾. Cells were maintained in 75 cm² flasks (Nunc) in DMEM F12 medium supplemented with 7% fetal calf serum, penicillin (100 i.u. ml⁻¹), streptomycin (100 µg.ml⁻¹), and 2 mM L-glutamine. The HTori-3 cells were maintained at 37°C and 5% CO₂ in air.

The 340-RPE T53 cell line is a human retinal pigment epithelial cell line⁽⁹⁾. The cells were maintained in 75 cm² flasks in DMEM F 12 containing 2 mM L-glutamine, pyridoxine, and 15 mM HEPES buffer supplemented with 10% fetal calf serum, gentamycin 10 µg.ml⁻¹, hygromycin 10 µg.ml⁻¹ and sodium bicarbonate 0.348% (w/v). The cells were maintained at 37°C and in 10% CO₂ in air. Both cell lines were routinely trypsinised with 0.05% trypsin/EDTA for 5 min.

Irradiation conditions

Cells were irradiated with UVA or UVB at room temperature using two UV sources. UVA was delivered from a bank of four tubes which emit at a maximum at 365 nm (Philips, TLD 18 W) at a distance of 14 cm source to petri dish, with a dose rate of 20.5 J.m⁻².s⁻¹. UVB was delivered from a broad spectrum tube (UVP products, model UVLM-26) which emits UVA (46.7%), UVB (52.9%) and UVC (0.4%), at a dose rate of 7.09 J.m⁻².s⁻¹. Maximum emission is at 313 nm. The distance between the tube and the petri dish was 16.5 cm. Irradiation was measured by using a UVX digital radiometer (UVP products). The UV radiometer was calibrated against a double grating spectroradiometer (model DM150BC, Bentham Instruments Ltd). Before exposure, the cells were washed twice with PBS. For UVA irradiation, cells were always covered with 3 ml of PBS, whereas for UVB, cells were covered with a thin layer of PBS. During UVA irradiation cells were irradiated through the lid of the petri dish, whereas for UVB, the lid was removed.

Culture treatments

To measure the occurrence of DCD, cells in log phase were irradiated as above. After irradiation the cells were trypsinised and counted. Cells were assayed immediately for plating efficiency and in addition plated out in flasks to yield approximately 300 viable colony-forming

cells after the cloning efficiency and cytotoxicity of the particular dose was taken into consideration. The medium in these flasks was changed twice a week, and the plating efficiency of the cells from these flasks was measured once they had completed 10 and 20 population doublings (10 population doublings correspond to 10 days for the HTori-3 cell line and 20 days for 340-RPE T53 cell line). The number of population doublings was calculated by counting the total number of cells per flask.

For the DNA repair studies, ara-C was added immediately after UV-B irradiation at a final concentration of 150 µM for 6 h. The cells were in a confluent stage during the irradiation procedure because ara-C is cytotoxic to cells in DNA synthesis⁽⁶⁾.

Plating efficiency

Plating efficiency was measured by using a standard adherent colony assay. After irradiation, the cells were trypsinised, counted and plated out in order to give approximately 50–100 colonies per petri dish. After two (HTori-3) and three (340 RPE-T53) weeks respectively, the colonies produced were fixed with methanol and stained with Giemsa. Colonies containing more than 50 cells were counted.

RESULTS

As can be observed from Figure 1, there is a marked difference in radiosensitivity between the two cell lines, with HTori-3 being much more sensitive in contrast to the 340 RPE-T 53.

Ara-C, an inhibitor of DNA repair, reduced the cell

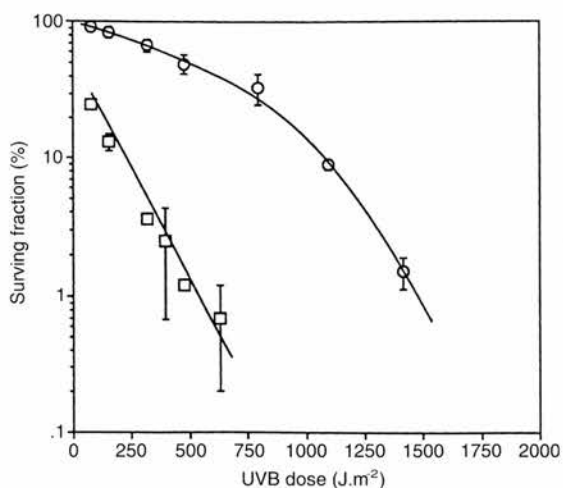


Figure 1. Survival curves for the 340-RPE T53 (O) and HTori-3 (□) cell line following exposure to UVB. Cells were in log phase during irradiation. The surviving fraction is plotted as a mean of three independent experiments \pm SE mean.

survival of both cell lines after UVB irradiation (Figure 2). It is important to note that when the cells are in log phase, both cell lines show increased radiosensitivity following UVB exposure, compared to cells irradiated when confluent (Figures 1 and 2).

A dose dependent reduction in plating efficiency was observed in both cell lines which is significant during the first passage (one passage corresponds to 10 population doublings) (Figure 3). HTori-3 cell line (Figure 3(a)) seems not to be able to recover completely after receiving large doses, even after 20 population doublings, in contrast with the 340 RPE-T53 cell line (Figure 3(b)). In addition, UVA irradiation was shown not to be able to induce the expression of this effect, over a range of doses giving similar initial surviving fractions (data not shown).

DISCUSSION

The results presented in this study, clearly show that irradiation of both cell lines with UVB was able to reduce the plating efficiency of the cells which survived after irradiation and have completed up to 20 population doublings. In addition, this reduction is dose dependent, suggesting that a specific cytotoxic threshold has to be reached in order for the phenomenon of DCD to be observed. UVA was not able to induce DCD in this study. The investigation of the role of DNA repair in DCD shows that at high doses of UVB the repair mechanism is compromised.

Our investigations show that only UVB was able to induce DCD in the two human epithelial cell lines. This

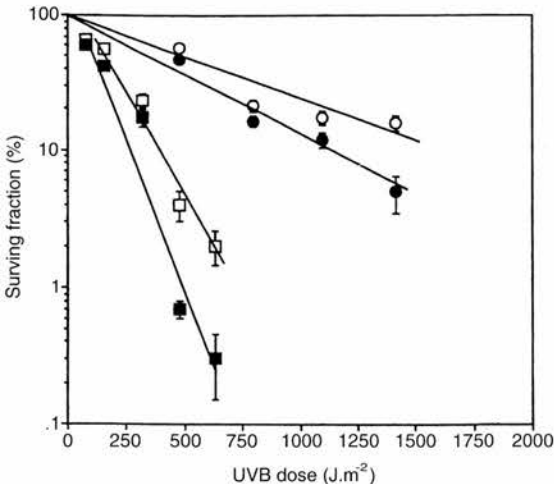


Figure 2. Survival curves obtained following UVB exposure and subsequent treatment with ara-C, for DNA repair studies. Cells were irradiated at confluence. The surviving fraction is plotted as a mean of three independent experiments ± SE mean. HTori-3: (□) no ara-C, (■) + ara-C. 340-RPE T53: (○) no ara-C, (●) + ara-C.

is in contrast with the work of Mothersill *et al*⁽¹⁰⁾ who showed that both UVA and UVB were able to induce DCD. However, it has been shown that UVB can produce a dose dependent DCD in two human epithelial cell lines which is similar to that of Mothersill *et al*⁽¹⁰⁾, who found a dose dependent DCD in the HACAT cell line by using a variety of chemical agents, and UVB and UVA. They suggested that a certain amount of damage can be tolerated by the cell population, either by repair or apoptosis, but once this threshold has been exceeded, a new mechanism begins to function within the cell population, preventing immediate death and promoting survival. Our results support this hypothesis but further investigation is required.

UVA was shown not to be able to induce DCD in this study, suggesting that the cells might have the ability to

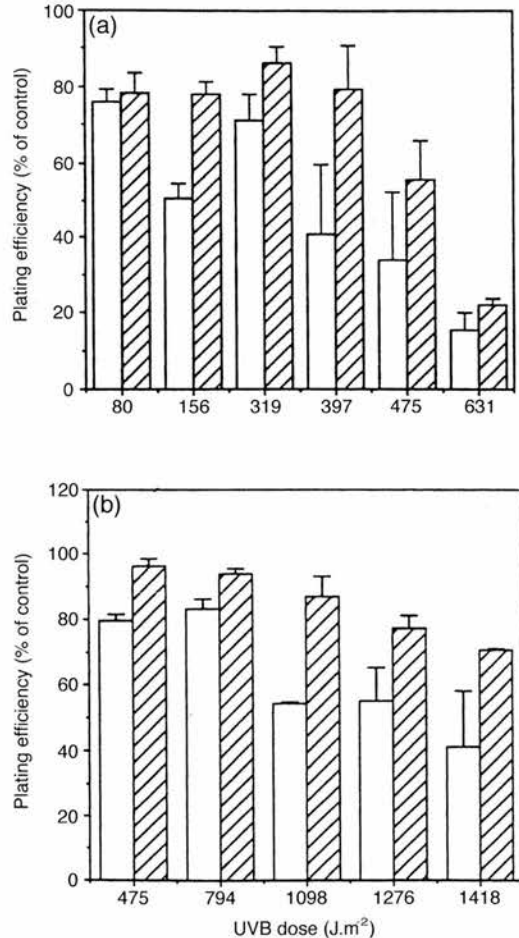


Figure 3. Plating efficiencies of the HTori-3 (a) and 340-RPE (b) cell lines following exposure to UVB, after the 1st and 2nd passage. Results are plotted as a mean of two independent experiments ± SE mean. White, 1st passage; shaded, 2nd passage.

repair oxidative damage more efficiently, in contrast to the bulky lesions caused by UVB (cyclobutane pyrimidine dimers, 6-4 photoproducts). One might speculate that the UVB photoproducts clearly must dominate over the generation of free radicals as far as the DNA damage and cell survival was concerned. On the other hand, the broad spectrum lamp which has been used emits about 52% of UVB and 48% of UVA supporting the mechanism suggested by Mothersill that DCD can be induced by a combination of strand breaks and oxidative stress, conditions which in theory can be met by using this specific UV lamp.

DNA repair studies failed to explain the marked difference in cell survival. Both cell lines exhibited the same response following ara-C treatment, indicating that differences in DNA repair capacity is not the causal factor of differences in radiosensitivity. Having observed a marked difference in the survival between the two cell lines following UVB exposure, it might be predicted that differences in radiosensitivity would be mirrored following ionising radiation. Surprisingly, the survival of both cell lines in response to ionising radiation are almost identical (data not shown). One possible explanation for this difference might involve the induction of p53 and subsequent differential expression of different downstream pathways by ionising compared to UV irradiation⁽¹¹⁾. The results presented here suggest the existence of a mechanism which coordinates the response of both cell lines towards UV but not ionising radiation. This mechanism probably involves p53 acti-

vation by UV, therefore further experiments need to be undertaken in order to elucidate any existing relationship between DNA repair and p53-induction by UV.

The concept of DCD introduces an interesting component of the current model of carcinogenesis. Mammalian cells which have the ability to survive after exposure to different kinds of mutagens, are not able to repair DNA damage completely. As a consequence, this inherited damage might cause long-term changes in the DNA sequence, increasing the spontaneous mutation frequency.

The change of a cell's phenotype from normal to malignant includes a whole series of individual steps which involves the serial accumulation of genetic changes over a considerable period of time. Therefore, DNA changes which take place over a period of a life time will decrease the stability of the genome and increase the probability of multiple mutations, resulting finally in tumour development.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Andrea Bodnar and the Geron Cooperation for kindly supplying the 340-RPE T53 cell line. They are also grateful to Dr H. Moseley and Dr N. Gibbs (The Photobiology Unit, Ninewells Hospital, University of Dundee) for performing the calibration of the UV meter, and estimating the spectrum of the UV sources. This work was supported by a scholarship from the University of St Andrews (G.B.).

REFERENCES

1. Little, J. B. *Radiation-induced Genomic Instability*. Int. J. Radiat. Biol. **74**(6), 663-671 (1998).
2. Stamato, T. D. and Perez, M. L. *EMS and UV-light-induced Colony Sectoring and Delayed Mutation in Chinese Hamster Cells*. Int. J. Radiat. Biol. **74**(6), 739-745 (1998).
3. Chang, W. P. and Little, J. B. *Evidence that DNA Double-strand Breaks Initiate the Phenotype of Delayed Reproductive Death in Chinese Hamster Ovary Cells*. Radiat. Res. **131**, 53-59 (1992).
4. Mothersill, C. and Seymour, C. *Lethal Mutations and Genomic Instability*. Int. J. Radiat. Biol. **71**(6), 751-758 (1997).
5. Alper, T., Mothersill, C. and Seymour, C. B. *Lethal Mutations Attributable to Misrepair of Q Lesions*. Int. J. Radiat. Biol. **54**(3), 525-530 (1988).
6. Iliakis, G. *Effects of ara-C on the Growth and Repair of Potentially Lethal Damage in Erlich Ascites Tumor Cells*. Radiat. Res. **83**, 537-552 (1980).
7. O'Reilly, J. P. and Mothersill, C. *Comparative Effects of UVA and UVB on Clonogenic Survival and Delayed Cell Death in Skin Cell Lines from Humans and Fish*. Int. J. Radiat. Biol. **72**(1), 111-119 (1997).
8. Lemoine, N. R., Mayall, E. S., Jones, T., Sheer, D., McDermid, S., Kendall-Taylor, P. and Wyndford-Thomas, D. *Characteristics of Human Thyroid Epithelial Cells, Immortalised in vitro by Simian Virus 40 DNA Transfection*. Br. J. Cancer **69**, 409-416 (1989).
9. Bodnar, A. G., Quелlette, M., Frokklis, M., Holt, S. E., Chiu, C. P., Morin, G. B., Harley, C. B., Shay, J. W., Lichtsteiner, S. and Wright, E. W. *Extension of Life-span by Introduction of Telomerase into Normal Human Cells*. Science **279**, 349-352 (1998).
10. Mothersill, C., Crean, M., Lyons, M., McSweeney, J., Mooney, R., O'Reilly, J. and Seymour, C. B. *Expression of Delayed Toxicity and Lethal Mutation in the Progeny of Human Cells Surviving Exposure to Radiation and other Environmental Agents*. Int. J. Radiat. Biol. **74**(6), 673-680 (1998).
11. Lu, Xin and Lane, D. P. *Differential Induction of Transcriptionally Active p53 Following UV or Ionising Radiation: Defects in Chromosome Instability Syndromes?* Cell **75**, 765-778 (1993).