The SbcCD Protein of *Escherichia coli*

Lucy A Kirkham

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

1999

Declaration

This thesis was composed by myself and the work described herein is my own, except where otherwise stated.

Abstract

The SbcCD complex mediates the inviability of long palindromes in *Escherichia coli*. SbcC is an SMC-like protein. SbcD belongs to the family of phosphodiesterases and is a single-strand endonuclease. SbcCD is a double-strand exonuclease and a hairpin nuclease. A previous model proposed that SbcCD hairpin nuclease regulates palindrome maintenance. Homologues of SbcCD exist in prokaryotes and eukaryotes. The eukaryotic homologue, Rad50/Mre11, is involved in DSB formation, DSB repair, V(D)J recombination and telomere maintenance.

In this work the hairpin nuclease activity of SbcCD has been characterised to help elucidate its cellular roles. The sites of hairpin cleavage have been mapped and the chemical nature of the cleaved termini determined. Hairpin cleavage occurs immediately 5' of the loop and is followed by stem degradation. Cleavage and degradation have been uncoupled. The cleavage products have 5' phosphate and 3' hydroxyl termini. The nuclease requirements for substrate structure were examined to test the model for palindrome maintenance. A dumbell substrate lacking termini was generated and found to be cleaved, supporting the model and suggesting that SbcCD is involved in DSB formation like its homologues. SbcCD has been shown not to possess a helicase activity on a short hairpin substrate. DNA binding has been demonstrated by gel filtration and studied by gel retardation. It is proposed that SbcCD may bind its substrate transiently.

Acknowledgements

Thanks go to my supervisor David Leach and to John Connelly who have inspired me with their superhuman optimism. I thank them for guidance, for sharing with me their knowledge, and for proof-reading. In addition, John supplied me with purified SbcCD protein, which saved me an indeterminate amount of time and for which I am very grateful. Thanks also to Ken Murray, my second supervisor, for valuable criticism and words of encouragement. I should like to thank David Pinder who made and purified mutant SbcCD proteins, and to all the other members of the Leach team, past and present, who have given me much help and support.

I thank Kevin Hiom and Lynn Powell for advice on studying DNA binding, Alex Alexandrov for advice on helicases, and Dave Dryden for advice on fluorimetry. Financial support for my project came from a BBSRC Postgraduate Studentship Award and a CASE Award from NBL Gene Sciences Ltd, and I should particularly like to thank Phil Eastlake for the latter.

I am grateful to my good friends, including Tom Pringle (*aka* Dr Bunhead) and members of the Glencorse Pipe Band, and to the Scottish countryside for much needed diversion, since too much of anything isn't good for you. And I thank my brother Peter for supplies of vitamin C and aiding me in time management during the final stages of this work.

Most of all I wish to thank my parents for their infinite support. More specifically, thanks to them for my early interest in science, and for my frequent posts back home, during which they cheerily bolstered my strength with substantial hunks of emotional and physical cake. I mention but a random couple of ways in which they have helped me.

Finally I acknowledge the Sales Manager for the Shoe Event Horizon, progeny of Douglas Adams, to whom I owe the precis of my abstract; Because.

Table of Contents

Abbreviations

%	percent
'	indicates polarity of DNA
Å	angstrom unit
A, T, G, C	adenine, thymine, guanine, cytosine
aa	amino acid
ADP	adenosine 5'-diphosphate
Amp	ampicillin antibiotic
AMPS	ammonium persulphate
ATP	adenosine 5'-triphosphate
ATPγS	adenosine 5'-O-(3-thiotriphosphate); non hydrolysable analogue of ATP
B. subtilis	*Bacillus subtilis*
BIME	bacterial interspersed mosaic element
bp	base pairs
BSA	bovine serum albumin
c	centi (10^{-2})
C-terminal	pertaining to the end of the peptide chain that carries the free alpha carboxyl group of the last amino acid
C. elegans	*Caenorhabditis elegans*
Ci	curie unit
Cm	chloramphenicol antibiotic
cpm	counts per minute
Δ	deletion
D, J, V, C	diversity, joining, variable, constant regions of immunoglobulin heavy and light chains
Da	daltons (1Da = 1 atomic mass unit)
dATP	deoxyadenosine 5'-triphosphate
DB	dumbell
dCTP	deoxycytosine 5'-triphosphate
ddAMP	dideoxyadenosine 5'-monophosphate
ddATP	dideoxyadenosine 5'-triphosphate
dGMP	deoxyguanosine 5'-triphosphate
DNA	deoxyribonucleic acid

DNA-PK	DNA protein kinase complex, including DNA, catalytic subunit, and Ku heterodimer
$DNA\text{-}PK_{CS}$	catalytic subunit of DNA-PK
DNase	deoxyribonuclease
DR	direct repeat DNA sequence
DSB	double-strand break
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	diaminoethanetetra-acetic acid disodium salt
ERIC	enterobacterial repetitive intergenic consensus
ES	embryonic stem
Exo	exonuclease
f	femto (10^{-15})
F-	recipient bacterial strain lacking F sex plasmid, capable of conjugational mating with Hfr
FPLC, HPLC	fast protein liquid chromatography, high pressure liquid chromatography
g	gram
g	force of gravity (9.8N)
G1	first gap phase of the cell cycle (between M and S)
G2	second gap phase of the cell cycle (between S and M)
GDP	guanosine 5'-diphosphate
GTP	guanosine 5'-triphosphate
Hfr	donor bacterial strain carrying F sex plasmid, capable of conjugational mating with F-
HO	endonuclease; initiates mating type switch in *S. cerevisiae*
HP	hairpin
HU	hydroxyurea
hyperrec	hyper-recombination
IR	inverted repeat DNA sequence
IR	ionizing radiation
IRIF	ionizing radiation induced foci
IRU	intergenic repeat unit
k	kilo (10^3)
kb	kilobases
Kd	dissociation constant
K_m	Michaelis constant

-l	(fractions of a) litre
L	litres
M	mega (10^6)
M	molar
-mer	indicates an oligonucleotide
m	metre
μ	micro (10^{-6})
m	milli (10^{-3})
m-AMSA	4'-(9-acridinylamino)methanesulphon-*m*-anisidide
MAT	mating-type locus
MC	mitomycin C antibiotic
MMS	methylmethanesulphonate
mol	moles
M_R	molecular weight
mRNA	messenger RNA
n	nano (10^{-9})
N-terminal	pertaining to the end of the peptide chain that carries the amino group of the first amino acid
NBS	Nijmegan Breakage Syndrome
nc, cc	nicked circle, closed circle
nt	nucleotides
NTP	nucleoside 5'-triphosphate
^oC	degrees celsius
ORF	open reading frame
ori	origin of replication
p	long arm of a eukaryotic chromosome
p	pico (10^{-12})
PAGE	polyacrylamide gel electrophoresis
pal	palindrome
PCR	polymerase chain reaction
pH	power of hydrogen (-log[H^+])
phage	bacteriophage
pI	isoelectric point
PU	palindromic unit
q	short arm of a eukaryotic chromosome
$^{R, S}$	resistant, sensitive
REP	repetitive extragenic palindrome

RF	ratio (distance migrated by a standard divided by distance migrated by DNA of interest) In this thesis, the substrate DNA was taken as the standard
RNA	ribonucleic acid
rpm	revolutions per minute
S	Svedberg unit
S- M-phase	DNA synthesis, mitosis phase of the cell cycle
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces cerevisiae*
SAR/MAR	scaffold/matrix associated region
Sbc	suppressor of *recBC*
SC	synaptonemal complex
SDM	site directed mutagenesis
SMC	structural maintenance of chromosomes
SMQ	sterile Milli-Q water
SSA	single-strand annealing
SSB	single-strand binding protein
TdT	terminal transferase
TLC	thin layer chromatography
T_m	melting temperature
Tn	transposon
TRED	trinucleotide repeat expansion disease
Tris	2-amino-2-(hydroxymethyl)-1, 3-propandiol
TEMED	N-N-N'-N'-tetramethyl-1, 2-diaminoethane
ts	temperature sensitive
U	enzyme unit (specifies enzyme activity in terms of the time taken for 1nmol substrate to be acted on by 1U enzyme at 37°C)
UV	ultra violet radiation
v/v	volume by volume
W	watt
w/v	weight by volume
X. laevis	*Xenopus laevis*

Standard abbreviations were used for elements of the periodic table and for amino acid residues.

CHAPTER 1

INTRODUCTION

Section 1

A General Overview of Genome Stability

Genetic Recombination

Genetic recombination provides genomic variability by generating new combinations of alleles, and sometimes new alleles. Variability is beneficial for adaptation to changes in the environment. However too much variability lowers genome stability. Therefore recombination occurs at a level which is an equilibrium between fitness and flexibility.

Four types of genetic recombination have been identified: Homologous, illegitimate, site-specific and transpositional (reviewed Leach 1996a). Homologous and illegitimate recombination are relevant to this thesis.

Homologous recombination functions in recombinational repair of double-strand lesions, programmed DNA rearrangements (*eg* antigenic variation of bacteria, mating-type switching in yeast), homologous chromosome pairing in eukaryotic meiosis, and priming of DNA synthesis in certain bacteriophage (*eg* T4, Section 3, λ rolling-circle). Illegitimate recombination occurs between sequences of little or no homology and therefore can occur anywhere in the genome. It consequently may be an important source of genomic instability.

Homologous Recombination in Prokaryotes

The RecBCD Pathway of Early Events

The RecBCD pathway occurs in wildtype *E. coli* (Wang and Smith 1985) and utilises the *recB*, *recC* and *recD* genes and the χ sequence. It is responsible for intermolecular recombination, *eg* between the chromosome and a linear DNA, *eg* plasmid oligomerisation (Summers 1996). It accounts for 99% of conjugational recombinants and for the post-replicational repair of radiation-induced DSBs (Horii and Clark 1973).

Homologous recombination is initiated by the potent RecBCD exonuclease V (reviewed Myers and Stahl 1994, Smith *et al* 1995). This is a member of the family of endo-exonucleases, which are involved in DNA repair and have either endo- or exonuclease activity depending on the availability of

subunits, cofactors and substrates (reviewed Fraser 1994). The RecD subunit confers exonuclease activity on the RecBC helicase. Exo V enters DNA via a blunt or nearly blunt double-strand end and degrades duplex DNA. Then it meets a χ site. χ is a hot spot for recombination. It is an aymmetric octanucleotide sequence, 5'- GCTGGTGG -3'. A properly oriented χ site temporarily inactivates RecD (confirmed Kuzminov *et al* 1994), which also requires RecA and SSB, and alters the affinity of the RecBC heterodimer for the RecD subunit. RecD dissociates to activate RecBC(D^-) as a recombinogenic helicase which proceeds to generate single-strand DNA.

Mutants in *recBC* have very low viability and are sensitive to a wide range of DNA damaging agents. They lack both RecBC(D^-) recombinase, required to repair the replication fork and resume replication, and RecBCD exonuclease V, required to degrade the displaced arm so that a new round of replication from *oriC* yields a viable (unbranched) chromosome. These double problems make them less viable than *recA* mutants, defective in a later stage of homologous recombination (below) (Ryder *et al* 1996).

χ acts in no other pathway of recombination than the RecBC pathway (Myers and Stahl 1994).

The RecF Pathway of Early Events

The RecF Variant Pathway

A variant of the RecF pathway occurs in wildtype *E. coli* and utilises the *recF*, *recO* and *recJ* genes and SSB (reviewed Summers 1996). It is responsible for intramolecular recombination, *eg* plasmid monomerisation. It accounts for the post-replicational repair of UV damage, *ie* the single-strand gaps which remain after removal of pyrimidine dimers (Horii and Clark 1973).

The RecF Pathway

The RecF pathway is activated in *sbcB* mutants and further activated by mutation of *sbcC/D* (Section 2). Indeed, *sbcC/D* mutants arise spontaneously in *sbcB* cells. When activated, the RecF pathway is responsible for inter- and intramolecular homologous recombination in *recBC* mutants (reviewed Summers 1996). Double mutation of *sbcBC* or *sbcBD* restores wildtype levels

of viability to *recBC* cells (Section 2). These genes are thus named "suppressor of *recBC*".

sbcB encodes exonuclease I (Horii and Clark 1973). Exo I is a processive 3' to 5' single-strand exonuclease (Lehman and Nussbaum 1964). Inactivation of Exo I stimulates the RecF pathway. It has been postulated that Exo I might degrade the putative recombination intermediate required for the RecF pathway (Horii and Clark 1973). SbcCD is also a 3' to 5' exonuclease (Connelly *et al* 1999, Section 3). Loss of SbcCD may therefore increase the effect of loss of Exo I.

The pathway utilises the *recF, recO, recR, recN, recJ, recQ* (reviewed Lloyd 1991, Summers 1996, Leach 1996a) and *rdgC* (Ryder *et al* 1996) genes. The *recR, recN,* and *recQ* genes are redundant in wildtype *E. coli.* Mutations in *recD, recJ* and *recN* are more deficient than the single mutants, suggesting that their functions overlap. *recF, recO* and *recR* are in the same epistasis group and therefore may work together. Several of these genes are regulated by LexA, which controls the cellular response to DNA damage (Lovett and Clark 1984).

The mechanism of RecF recombination is unknown. RecJ is a 5' to 3' single-strand exonuclease and RecQ is a DNA helicase. RecQ might act with RecJ as a recombinogenic helicase to generate duplexes with recombinogenic 3' single-strand tails at the free ends of a DSB (Ryder *et al* 1996). Such a putative recombination intermediate would be analogous to that generated by RecBC(D^-). RecN may protect these recombinogenic tails and RecF, RecO and RecR may facilitate strand pairing (Leach 1996a). RecF binds DNA and may facilitate RecA filament formation (Lloyd and Sharples 1992). RecF (and the β subunit of DNA polymerase III) are induced when cell metabolism slows down, suggesting that the cell induces proteins required to maintain viability in preparation for a prolongued period of starvation (Villarroya *et al* 1998).

The effect of *sbcC/D* on the RecF pathway was demonstrated (Lloyd 1991). The recombinant progeny of a cross between wildtype Hfr and wildtype F^- (indicative of the RecBCD pathway of recombination) were compared to those of a cross using *recBCsbcBC* F^- (indicative of the RecF pathway). In the

wildtype cross, the frequencies of multiple exchanges and mixed genotype colonies were lower, the frequency of inheritance of markers proximal to *sbcCD* was higher, and growth rate was higher. These differences could all be attributed to reduced efficiency of recombination upon transfer of *sbc*C$^+$ to the *recBCsbcBC* strain, reducing viability and resulting in a low recovery of *recBCsbcB* recombinants.

The pathway is also used to prime DNA replication by a recombination-dependent mechanism similar to that observed in bacteriophage T4 and in inducible stable DNA replication (iSDR) (Ryder *et al* 1996, Section 3).

The RecE Pathway of Early Events

The RecE pathway is induced in *sbcA* mutants (reviewed Summers 1996). When activated, the RecE pathway is responsible for inter- and intramolecular homologous recombination in *recBC* mutants. Hence *sbcA* is another "suppressor of *recBC*". The pathway requires the products of *recE*, *recF*, *recJ*, and *recT* (Gillen *et al* 1981).

sbcA mutation induces the RecE pathway by inducing *recE* (Gillen *et al* 1981). *recE* is located in the chromosomal segment identified as the Rac prophage, a cryptic lambdoid prophage. It encodes exonuclease VIII, an ATP independent double-strand 5′ to 3′ exonuclease. *recT* is an analogue of the λ *redβ* gene and catalyses the annealing of complementary single-strand DNA. Therefore Exo VIII might activate an analogous pathway to the λ *red* pathway (Lloyd and Sharples 1992). Alternatively it may act like RecBC(D$^-$).

RecE intermolecular recombination, as assayed by plasmid oligomerisation, is 20-fold more efficient than wildtype, *ie* more efficient than RecBC intermolecular recombination (reviewed Summers 1996).

Comparison of the RecF and RecE Pathways

There is a more stringent requirement for RecF in the RecF pathway than in the RecE pathway (Gillen *et al* 1981). Mutation of *recF* decreases recovery of conjugational recombinants 1000-fold by the RecF pathway, but only 10-fold by the RecE pathway. There is also a more stringent requirement for RecJ in the RecF pathway than in the RecE pathway (Lovett and Clark 1984).

It has been suggested that removal of long single-strand tails by single-strand nucleases such as SbcCD (later Sections), Exo I and RecJ provides blunt DSB entry sites for RecBCD (Thoms and Wackernagel 1998). However deletion of *sbcA* or *sbcCD* suppresses *recBC* (Section 2). The authors suggest that in this case the single-strand tails are stabilized and can participate in *recA* mediated homologous recombination at non χ sites.

Phage Pathways of Early Events

Certain bacteriophage also specify homologous recombination pathways, such as *red* and *gam* in phage λ. In addition, phage λ also possesses a gene which is involved in a RecF-like pathway in a *recBCsbcBC* host (Sawitzke and Stahl 1994).

Late Events

Late recombination events generate heteroduplex DNA from the recombinogenic single-strand DNA product of early recombination events. They require *recA, ruvA, ruvB, ruvC, recG* and SSB.

Single-strand DNA interacts with RecA (reviewed Rao and Radding 1995). RecA forms a right-handed helical nucleoprotein filament on single-strand DNA by head to tail polymerisation. The RecA-DNA filament promotes pairing with a second DNA molecule in a homology search. A DNA triplex or quadruplex is formed by non Watson-Crick base triplet bonds. This can be kilobases long and is coated by RecA. RecA promotes strand exchange by single-strand invasion of a homologous duplex and formation of a displacement (D)-loop intermediate to generate a 4-way crossover (Holliday) Junction.

The Holliday Junction undergoes branch migration which alters the extent of exchanged genetic material between the sister chromatids. RuvAB catalyses forward branch migration. The *ruvA* and *ruvB* genes are induced by the SOS system. The complex consists of a RuvA tetramer sandwiched between two ring-like RuvB hexamers which encircle the DNA (Yu *et al* 1997). RuvA confers Holliday Junction specific DNA binding and loads RuvB onto the Junction (Hiom and West 1995a, Hiom *et al* 1996). RuvB confers ATP dependent helicase activity which drives translocation (Mitchell and West 1994, Marrione and Cox 1995). The RuvB subunits have been suggested to function in pairs (Marrione and Cox 1996). Branch migration results from translocation as the DNA passes through the two RuvB rings (Yu *et al* 1997). Like RuvAB, RecG is a junction-specific helicase (Lloyd and Sharples 1993). It catalyses reverse branch migration, *ie* in the opposite direction to RuvAB (Whitby *et al* 1993). This may be particularly important in DNA damage repair (below) where transient crossover permits filling in of a single-strand gap without fixing of the crossover as a recombinant. RecG is the prototype of a new class of helicases (Whitby *et al* 1994).

The Holliday Junction is resolved by cleavage by RuvC endonuclease (Bennett and West 1995). Cleavage is ATP independent (Dunderdale *et al* 1994), Holliday Junction specific (Bennett *et al* 1993) and exhibits some sequence specificity (Bennett and West 1995). It occurs on the symmetrical, continuous, non crossover strands of the junction. RuvC dimer activity is dependent on RuvAB (Whitby *et al* 1996). It has been proposed that RuvA holds open the junction whilst RuvC scans for a recognition site. Then RuvC binding triggers RuvA dissociation.

RecA Independent Recombination

RecA is essential for the later stages of recombination (above), with one exception. Intramolecular recombination occuring by the RecE pathway, as assayed by plasmid monomerisation, is largely unaffected by mutations in *recA* or *recF* (reviewed Summers 1996). RecE intermolecular recombination is dependent on both gene products.

Homologous Recombination in Eukaryotes

Holliday Junction Pathways

Holliday Junctions can be intermediates in eukaryotic homologous recombination as well as in prokaryotes. This type of recombination may occur at a DSB site, as in DSB repair (below). Alternatively it may occur at a single-strand nick (Meselson and Radding 1975). In this case, gene conversion, crossover or non crossover products may be generated in various ratios. Rad51 is the eukaryotic homologue of RecA (Benson *et al* 1998).

The SSA Pathway

Another pathway of homologous recombination exists in eukaryotes, called single-strand annealing (SSA) (Lin *et al* 1984). Unlike the other mechanisms of recombination, SSA is inaccurate, *ie* genetic material is lost during the process. SSA may occur when two homologous sequences are repeated in direct orientation (DR) and flank a DSB site. Like the Holliday Junction pathways, it depends on 5' to 3' degradation of DSB ends. This degradation is bidirectional and exposes the complementary sequences, which then anneal. The non homologous 3' tails are then removed and the ends ligated. The complete product has a deletion of one of the repeats and of the intervening sequence.

Some, but not all, of the *RAD52* epistasis group proteins are required in SSA (Ivanov *et al* 1996): Rad52 may act in SSA, but is not essential. Deficiency is not as severe as in the Holliday Junction pathways, and the requirement is overcome if the region of homology is extended. The efficiencies of the Holliday Junction and SSA pathways are equally reduced in the absence of Rad50, Mre11 and Xrs2 by approximately 2- to 3-fold. It has been suggested that Rad50, Mre11 and Xrs2 are rate limiting for SSA. Rad51, Rad54, Rad55 and Rad57 do not function in SSA. It has been proposed that SSA competes with Holliday Junction pathways for DSB substrates and is favoured in the absence of Rad51, Rad54, Rad55 and Rad57.

The Rad52 Epistasis Group

The *RAD52* epistasis group was initially identified in *S. cerevisiae* (reviewed Tsubouchi and Ogawa 1998). It includes *MRE11*, aka *RAD58* and *Ngs1* (Chamankhah and Xiao 1998), *RDH54* (Trujillo *et al* 1998), *RAD50, RAD51, RAD52, RAD54, RAD55, RAD57* and *XRS2* (Milne *et al* 1996) and divides into two subgroups. The members of each subgroup function in the same pathway and may represent components of a multiprotein complex. *RAD51, RAD52, RAD54, RAD55* and *RAD57* are involved in homologous recombination but not in non homologous end-joining. However only *RAD52* is required for all types of homologous recombination (reviewed Haber 1998). Mutants in these genes may affect mating-type switching, meiotic recombination and mitotic recombination. These phenotypes are not alleviated by *spo13* mutation. *RAD50, MRE11* and *XRS2* are involved both in non homologous end-joining and homologous recombination. Null mutants are defective in meiotic recombination, proficient in mating-type switching and affect mitotic recombination (Section 4). They are alleviated by *spo13*. These three processes involve the repair of DSBs. The differing genetic requirements for DSB repair in mitosis, and DSB formation and processing in meiosis, indicate that these two processes are distinct.

Humans have highly conserved homologues of the yeast Rad52 epistasis group (Dolganov *et al* 1996). This suggests that the mechanisms of homologous recombination and non homologous end-joining may be conserved.

Yeast Meiotic Recombination

Yeast meiotic recombination occurs by homologous recombination, is initiated at DSBs (Section 4), and is an essential stage in meiosis (Padmore *et al* 1991). All *MRE* genes are essential for meiotic recombination and all *mre* mutants are suppressed by *spo13*. The yeast meiotic recombination genes have been classified into catagories (reviewed Ajimura *et al* 1992).

RAD50, MRE11 and *XRS2* are MMS sensitive, hyperrecombinational in mitosis and proficient in UV-induced mitotic recombination (Section 4, Ajimura *et al* 1992). Alterations in these genes generate "early block" mutants

which arrest at an early stage in meiotic recombination. *Sc*Rad50 and *Sc*Mre11 are implicated in DSB formation (Section 4). Other proteins which function in DSB formation are candidates for interaction with the yeast Rad50/Mre11 complex. These include Xrs2, Mer2, Mei4, Rec104, Spo11 and Mre3/Rec114 (Johzuka and Ogawa 1995).

RAD51, RAD52, RAD54 and *RAD57* are MMS sensitive and defective in UV-induced mitotic recombination (Ajimura *et al* 1992). These genes generate "intermediate block" mutants. *rad50S* mutants also generate an intermediate block in meiotic recombination.

Spo11 is implicated in meiotic DSB formation and may form a complex with Rad50, Mre11 and Xrs2 (Section 4). It is meiosis-specific and forms an intermediate block mutant in meiotic recombination. *spo11* mutants are defective in meiotic recombination (Keeney *et al* 1997). They exhibit severe nondisjunction of homologous chromosomes at meiosis, leading to aneuploid spores which are mostly inviable and which have altered morphology. A *sae2* mutant (also called *com1*) has a similar mutant phenotype to *rad50S* in meiosis, and is epistatic to *spo11*. It has been suggested to participate in the removal of Spo11 from DSBs (McKee and Kleckner 1997a).

Mutants in Mre4/Mek1 and Red1 confer neither early nor intermediate block phenotypes (Xu *et al* 1997). Both function in meiosis and mitosis. Mre4/Mek1 is homologous to serine/threonine protein kinases and mutants have almost wildtype synaptonemal complex (SC) structure. Red1 is a major structural component of chromosomes and mutants lack axial elements or SC. These two proteins may together determine a meiotic recombination checkpoint.

Illegitimate Recombination

Strand-Slippage

Strand-slippage is a hypothetical type of illegitmate recombination (Albertini 1982). It occurs at hairpin secondary structures. Hairpin stems are not easily penetrated by DNA polymerase. DNA polymerase is postulated to progress some distance into the stem before pausing. It then may slip from one short DR located just within the hairpin stem, the "donor," to another just past the hairpin structure, the "target". The main body of the hairpin sequence is deleted. Likewise, an insertion is generated if the hairpin forms in the daughter strand. Replication pausing also facilitates homologous recombination (Bierne and Michel 1994) (Section 3).

The model was tested in *E. coli* using an "asymmetric deletion insert," which consisted of a hairpin forming palindrome with asymmetric donor and target sequences inserted into a Cm^R cassette (Trinh and Sinden 1991). Reversion to Cm^R could occur by strand-slippage deletion of the palindrome between the DRs if the DRs were suitably aligned. Since the DRs were asymmetric, one being located just within and one just beyond the palindrome, they would be suitably aligned for strand-slippage deletion of the palindrome on one strand only. Inversion of the cassette would place the suitably aligned repeats on the other strand. Inversion of the cassette was observed to reduce reversion 20-fold when the DRs were suitably arranged in the leading strand. Inversion had no effect on a symmetric insert. This is consistent with strand-slippage occuring preferentially in the lagging strand. A later experiment revealed that reversion was reduced more than 25-fold compared to the symmetric insert when the hairpin was stabilized in the leading strand (Rosche *et al* 1995). The preference for strand-slippage in the lagging strand may result from the presence of single-stranded regions which can form secondary structures. Gaps do not occur during leading strand synthesis.

Consistent with the existence of strand-slippage in prokaryotes, the *TerB* hairpin at the terminus of *E. coli* is a deletion hot spot, as is a palindromic sequence (Bierne and Michel 1994). Excision of transposons Tn*10* (Ehrlich *et al* 1993), Tn*9*, and Tn*5* (Albertini *et al* 1982) has been proposed to occur by

strand-slippage since it occurs between short DRs. There is evidence for a different but related replication slippage event at a long palindrome in *E. coli* which occurs from the leading to the lagging strand (Pinder *et al* 1997).

Consistent with strand-slippage in eukaryotes, large deletions (greater than 7bp) of quasi-palindromes can occur between short DRs in *S.cerevisiae* (Tran *et al* 1995). In addition, a 122bp long palindrome was a hot spot for spontaneous deletion by illegitimate recombination between two 6bp flanking DRs in a human cell line (Kramer *et al* 1996).

DNA polymerase (pol3)-stimulated transposon excision in yeast is a putative strand-slippage event. It is reduced in *rad52* and *rad50* mutants (Gordenin *et al* 1992). *pol3* mutants exhibit a 100-fold increase in frequency for which *RAD50* and *RAD52* are required (Tran *et al* 1995). *rad52* mutants showed a 30-fold reduction in putative strand-slippage events, and *rad50, mre11* and *xrs2* mutants showed a 10-fold reduction (Tsukamoto *et al* 1996b). The products of these genes therefore appear to function in strand-slippage.

Non Homologous End-Joining

Non homologous end-joining is ligation of two non homologous double-strand DNA ends (reviewed Leach 1996a, Weaver 1995, Tsukamoto and Ikeda 1998). It involves some filling-in of single-strand DNA ends which occurs by nucleotide incorporation. Like SSA and DSB repair by homologous recombination, it initiates at a DSB. There are three mechanisms of non homologous end-joining in eukaryotes. Ends can be joined by pairing of short complementary sequences (microhomology). Similarly, untemplated addition of a few nucleotides primed at a 3' end of a blunt DSB on one side can enable pairing of microhomology. Alternatively, two protruding single-strand ends may be ligated and the gap filled in by nucleotide addition. The latter mechanism does not require microhomology. In prokaryotes end-joining requires micro-homology because single-strand DNA cannot be ligated in these cells. The mechanism is not fully characterised. The replication protein of phages M13 and f1 may participate in end-joining (Ehrlich *et al* 1993).

Evidence for Hairpin Intermediates

Short IRs of a few nucleotides may be generated at sites of non homologous end-joining. These P (palindromic) inserts have been proposed to be derived from a hairpin intermediate (Lewis 1994). The loop of the hairpin is nicked at a point a few nucleotides from the apex, and this distance determines the length of the P insert. Nucleotide insertion and ligation complete the process.

End-joining of a linear plasmid via a hairpin intermediate has been observed in murine fibroblastoid cells (Lewis 1994). The same mechanism was not detected in *E. coli*. P inserts are generated on transposition of certain plant transposons (reviewed Lewis 1994). A hairpin nicking activity has also been proposed to exist in murine germ-line cells (Akgun *et al* 1997). In these cells, a long palindrome undergoes one-sided deletion, unlike the two-sided deletion which results from strand-slippage in yeast and bacteria. Such products may arise from hairpin nicking followed by degradation of one arm.

The Role of DNA-PK

The DNA-PK protein kinase complex plays an essential role (Cary *et al* 1997). The DNA-PK holoenzyme is conserved between yeast and mammals, indicating that the pathway is conserved, and consists of three subunits; a catalytic subunit DNA-PK_{CS} and a Ku heterodimer (Cary *et al* 1997). These proteins are 460kDa, 70kDa and 86kDa respectively in mammals, and the larger Ku is 80kDa in yeast (Milne *et al* 1996). Ku, SIR2, SIR3, SIR4 and DNA ligase act epistatically in non homologous end-joining (Boulton and Jackson 1998).

Four families of mutations are associated with DNA-PK activity, *XRCC4*, *XRCC5*, *XRCC6* and *XRCC7* (reviewed Weaver 1995). Ku86/80 is encoded by the *XRCC5* group of genes, associated with *xrs*, *XRV15B*, *sxi-2* and *sxi-3* mutants in higher eukaryotes (Boubnov *et al* 1995) and the *HDF2* gene in yeast (Milne *et al* 1996, Tsukamoto *et al* 1996a). Ku70 is also encoded by the *XRCC5* group of genes (Boubnov *et al* 1995), *HDF1* in yeast, and its expression requires *sxi-1* in higher eukaryotes (Weaver 1995). The DNA-

PK_{CS} serine/threonine kinase is encoded by the *XRCC7* group of genes, associated with *scid* and *v-3* mutants in mouse (Milne *et al* 1996). Yeast does not have a DNA-PK_{CS} homologue, but the Tel1 and Mec1 proteins are related and may function with Ku (reviewed in Boulton and Jackson 1998). The product of *XRCC4* interacts with and increases the activity of DNA ligase IV (Weaver 1995).

DNA-PK is activated by DNA ends (Boubnov *et al* 1995). Avid, sequence independent binding is confered by the Ku heterodimer which recognises unprotected double-strand DNA ends, hairpin loops, single-strand nicks and gaps, single-strand to double-strand transitions, and distortions in the DNA helix (Cary *et al* 1997). This array of binding substrates suggests that Ku may act at an intermediate stage of non homologous end-joining (Boubnov *et al* 1995). Ku also has helicase/translocase activity (Tsukamoto *et al* 1996a). Ku protein clusters on DNA as a result of translocation (Cary *et al* 1997). End-joining in the absence of Ku is imprecise (Milne *et al* 1996).

A model for the role of DNA-PK in non homologous end-joining has been proposed (Cary *et al* 1997). Recombinogenic DNA DSB ends are each bound by a Ku heterodimer. Two heterodimers associate to form a tetramer, thereby bringing two ends together (end-tethering). In addition DNA-PK may facilitate the formation of a hairpin intermediate by Ku translocation which generates loops. The presence of other proteins on the DNA may block translocation. The catalytic subunit later phosphorylates Ku (DNA-PK autophosphorylation) which causes dissociation of DNA-PK. Other targets of the kinase include p53 and c-*myc* (Tsukamoto *et al* 1996a). Phosphorylation may signal DNA damage and regulate a cell cycle checkpoint. It may also activate an end-joining factor.

DNA-PK may also be involved in regulating transcription via a sequence specific DNA binding mode of Ku (Cary *et al* 1997).

V(D)J Recombination in Higher Eukaryotes

V(D)J recombination is responsible for immunoglobulin and T-cell receptor gene rearrangement which generates diversity in the immune response (reviewed Leach 1996a). The mechanism of V(D)J recombination differs from that of non homologous end-joining in that it occurs at a particular site. Otherwise these two mechanisms bear similarities. Hairpin nicking activity is central to the process. Some genetic information is lost during the process.

The genetic locus for the heavy (H) chain of the immunoglobulin is rearranged first. D_H (Diversity) is joined to J_H (Joining). Then V_H (Variable) is joined to D_HJ_H. V_H is attached to C_H (Constant) and therefore expression can occur at this stage. Expression stimulates rearrangement of the light (L) chain locus by the same mechanism. C_H is later rearranged by class switching, which is a different mechanism but also similar to non homologous end-joining.

The V(D)J coding regions are adjacent to signal regions. Recombination generates a coding joint product and a non coding signal joint. During the process, RAG1 and RAG2 cleave the signal sequence (McBlane *et al* 1995), and generate covalently closed hairpins from the coding ends (Hiom and Gellert 1997). Hairpin cleavage followed by nucleotide incorporation and ligation of the ends generates P inserts at the coding joint. This has been observed in murine lymphoid cells (Lewis 1994). It causes the coding joint to be imprecise. The hairpin cleaving activity is distinct from DNA-PK.

A Role for *sbcB* in Prokaryotic Illegitimate Recombination

Not all mutants in *sbcB* suppress both the recombination deficiency and the UV sensitivity of *recBC* mutants in the presence of *sbcC/D* (Allgood and Silhavy 1991). A family of non null mutants in *sbcB*, called *xonA*, suppress the UV sensitivity but not the recombination deficiency, and do so in the absence of *sbcC/D*. Both families of mutants are deficient in Exo I. *xonA* mutants were found to increase deletion formation in *recBC*$^+$ cells. Deletion formation is characteristic of illegitimate recombination. It was therefore suggested that Exo I may have an enzymatic role in illegitimate recombination.

Preferential Modes of Recombination in Eukaryotes

By far the major pathway of recombination in *S. cerevisiae* is homologous recombination, whereas the major pathway in mammals is non homologous end-joining (Roth and Wilson 1985, Petrini *et al* 1995, Dolganov *et al* 1996). Indeed Ku is not essential in yeast (Milne *et al* 1996). This may be because a small genome requires the fidelity of homologous recombination, whereas a large genome contains much repetitive DNA and a homology search risks ectopic recombination which may cause translocation (Milne *et al* 1996). These preferences account for the observation that in yeast, but not in mammals, G1 cells which lack sister chromatids are more sensitive to ionizing radiation than G2 cells. Illegitimate recombination in yeast is best studied under conditions where homologous recombination is non functional, *eg* in a *rad52* mutant.

DNA Repair

Some types of lesion in DNA can be chemically reversed. For the most part, lesions are repaired by specialised systems including recombinational repair, mismatch repair and several pathways of excision repair. Recombinational repair and mismatch repair are relevant here.

Recombinational Repair

Repair During Replication

Although repair of DSBs and single-strand gaps can occur by end-joining, homologous recombination is preferable since it retains the entire original sequence. The DSB repair pathway of mitotic homologous recombination (reviewed Shinohara and Ogawa 1995) can be used to repair DNA damage which might arise spontaneously or result from DNA damaging treatment, *eg* exposure to chemical mutagens or γ (ionizing IR), X or UV radiation. This can generate gene conversion or crossover products (Szostak *et al* 1983) or non crossover products (Thaler *et al* 1987).

An alternative pathway is Synthesis-Dependent Strand Annealing (reviewed Shinohara and Ogawa 1995). This occurs at a resected DSB. No Holliday Junction intermediate is formed and the donor duplex emerges intact after the process, having been used only as a template for DNA synthesis on the other duplex.

Post-Replicational Repair

Post-replicational repair occurs via a Holliday Junction intermediate, but involves only short tract DNA synthesis (reviewed Shinohara and Ogawa 1995). Replication skips a lesion in the template strand and this generates a single-strand gap opposite the lesion. After replication, the lesion is not repaired. But it is tolerated since the single-strand gap is repaired by branch migration and filling in. This can occur in the absence of excision repair.

DSBs

DSBs in Yeast

DSB repair mechanisms are essential because DSBs are potentially lethal. DSBs artificially introduced by HO endonuclease during meiosis stimulate meiotic recombination (Cao *et al* 1990).

DSB Hot Spots

Certain sites in the yeast genome are highly susceptible to DSB formation. They are hot spots for homologous recombination. Meiotic recombination hot spots were studied (Cao *et al* 1990). An artificial hot spot was generated by inserting a *LEU2* fragment beside the *HIS4* gene. Meiosis-specific recombination at the *LEU2-HIS4* hot spot is 6-fold higher than that at the unchanged *HIS4* locus, displays normal interference and kinetics and occurs at two sites, I and II. Diffuse DSBs were detected transiently at both sites. These had 3' single-strand overhanging ends of heterogeneous lengths up to 200bp. DSB formation at sites I and II is not sequence specific (Xu and Kleckner 1995).

Similar DSBs were observed at a natural hot spot, *ARG4* (Cao *et al* 1990). The frequency of DSB formation at a hot spot is determined by the accessibility of

the DNA to factors (Wu and Lichten 1995). This is affected by the nucleotide sequence at that site, transcriptional interference, an open chromatin configuration, and position effects. Position effects include locus dependence, by which a locus up to 17kb distant can affect DSB frequency at the hot spot, and local suppression, by which a strong hot spot can compete for limiting factors in a vicinity of up to 8.8kb.

MMS

Methyl methanesulphonate (MMS) is a genotoxic chemical which introduces a lethal 3-methyladenosine (3MeA) lesion (reviewed Xiao *et al* 1996). These lesions are repaired by an alkyl-specific excision repair pathway. A specific glycosylase removes 3MeA and the resulting apurinic gap is cleaved by a double-strand apyrimidinic/apurinic endonuclease (Kirtikar *et al* 1977). A DSB is generated which is repaired by recombinational DSB repair. Thus, like ionizing radiation, MMS generates DSBs (albeit indirectly) and has been used as a radiomimetric agent. DNA radiation repair mutants are sensitive to MMS.

Mismatch Repair

Replication Slippage Error

Occassionally DNA polymerase may misalign the newly synthesised strand and slip (Radman *et al* 1995). This replication slippage may generate small insertions or deletions, called frameshift mutations. Frameshifts are not recognised by the proofreading activity associated with the enzyme. Replication slippage may alternatively result in the incorporation of a non complementary base. In either case the unpaired bases loop out to create mismatches. Replication slippage occurs more frequently on microsatellite repeat arrays than on non repetitive DNA. It occurs more freqently on mono-, di- and trinucleotide repeat arrays than on tetranucleotide repeats. Strand-slippage is a type of replication slippage. In contrast to strand-slippage however, replication slippage is generally not stabilized by hairpin formation and generates loops of only one or a few nucleotides.

The Mismatch Repair System

Mismatches up to 3 nucleotides in length are recognised by the mismatch repair system (reviewed Modrich 1995). They provoke daughter strand incision at a nearby GATC sequence. The daughter strand is recognised by its lack of adenine methylation at a GATC site. This step requires the MutH, MutL and MutS proteins in *E. coli*. MutS ATPase binds the mismatch, MutL joins the complex which then activates MutH. MutH is a GATC-specific single-strand endonuclease. This step is not required in the presence of a single-strand break. Nicking is followed by excision. Single-strand exonucleolytic degradation occurs from the nick to a site approximately 100 nucleotides beyond the mismatch. This step requires 3' to 5' Exo I, or 5' to 3' Exo VIII or RecJ, exonucleases and DNA helicase II (aka MutU or UvrD). It also requires MutS and MutL. The gap is finally repaired by filling-in and ligation.

Functionally homologous mechanisms exist in yeast and humans. In these cells the repair tract can be long patch, as in *E. coli* (4kb), or short patch (10bp) (Nag and Kurst 1997). In *S. cerevisiae* there are six MutS homologues, MSH1-6. MSH4 is meiosis specific. A MLH1/PMS1 heterodimer functions as MutL. The short patch repair system involves some of the same proteins (Sugawara *et al* 1997).

Mismatch Repair Deficiency

Mismatch repair deficient human cells are hypermutable both spontaneously and in response to alkylating agents (Modrich 1995). This phenotype is called replication error (RER^+). It correlates with microsatellite repeat instability, since replication slippage occurs more frequently at microsatellites than non repeated DNA.

RER^+ can cause sporadic (non familial) cancers in a range of tissues (Radman *et al* 1995). RER^+ is also the cause of Hereditary non polyposis colon carcinoma (HNPCC) aka Lynch Syndrome. HNPCC is one of the most common human genetic diseases and occurs with a frequency of 1 in 200 individuals. Affected individuals often develop other epithelial cancers in addition to colon tumours. In this disease, di-, tri- and tetranucleotide repeat

arrays undergo small expansions (reviewed Lynch and Smyrk 1998). Where such expansions occur in certain genes, they effect tumourigenesis. Whether an individual develops colonic tumours alone (type I HNPCC), or extracolonic tumours as well as colonic tumours (type II), depends on the nature of the mutation.

DNA Conformation

Condensation

Chromatin

DNA condensation is the multi-tiered compaction of cellular DNA into a protein-associated network, chromatin. The mitotic chromosomes of eukaryotes are highly condensed into a mitotic chromosome scaffold (reviewed Koshland and Strunnikov 1996, Saitoh *et al* 1995, Hirano 1995). This allows chromosome separation at cytokinesis. Eukaryotic interphase chromosomes are less condensed and are generally much longer than the length of the cell. They associate with proteins in a nuclear matrix (reviewed Saitoh *et al* 1995). Prokaryotic DNA is similarly condensed into chromatin (reviewed Bianchi 1994).

The first tier of eukaryotic chromatin is the nucleosome "beads on a string" structure (reviewed Koshland and Strunnikov 1996, Saitoh *et al* 1995, Hirano 1995). The nucleosome core consists of an octamer of two each of histone proteins H2A, H2B, H3 and H4. Approximately 150bp DNA is wrapped solenoidally around this complex. Approximately 50bp linker DNA separates adjacent nucleosomes. Histone H1 assembles on linker DNA. H1 and H3 are hyperphosphorylated in mitosis, and the latter event is tightly coupled to condensation. Looping forms a higher tier of condensation. The mitotic chromosome scaffold organises chromosomes into loop domains of 50 to 100kb. The base of each DNA loop is a scaffold (or matrix) associated region (SAR or MAR). This is AT-rich and binds topoisomerase II and SMC2 (Section 2) which fasten the loops. Topoisomerase may act as a swivel which alters superhelical density, and catenates or decatenates chromatin fibres. It is unknown whether topoisomerase II and SMC2 are required in interphase

chromatin. Other protein components of chromatin include the HMG-box proteins. These high mobility group proteins bind bent DNA and a subset bind specific sequences of linear DNA and bend it.

Bacterial chromatin does not contain a regularly repeating unit (reviewed Bianchi 1994). It incorporates unequally spaced HU proteins, around which DNA is wrapped solenoidally. HUs bind linear DNA transiently and bend it, and bind bent DNA stably. They are functionally analogous to, and can be substituted by, HMG-box proteins HMG1 and HMG2. On account of their structural role in condensation, HMGs and HUs are implicated in many similar functions. These include cell division, DNA replication and transposition, and HU dimers protect DNA against cleavage by ionizing radiation (Boubrik and Rouviere-Yaniv 1995).

Supercoiling

Prokaryotic, and probably eukaryotic, genomes are under dynamic torsional stress (reviewed Lilley 1986). This results in two forms of supercoiling. Twisting is an increase in superhelical density by deformation about the long axis of the duplex. Writhing is an increase in superhelical density by deformation of the duplex widthways, *ie* the duplex axis crosses itself. These are interchangeable. Solenoidal wrapping of DNA by nucleosomes and HU complexes is a means of absorbing negative supercoils. Localised supercoils may be introduced by topoisomerase and helicase activities during transcription, replication, DNA looping by protein dimerisation and protein binding.

Topoisomerases introduce transient breaks into DNA to change the superhelicity (reviewed Bergerat *et al* 1997). Type I topoisomerases introduce single-strand breaks, wheras type II introduce double-strand breaks. The type II family is related to yeast Spo11 (Section 3). These are ATP dependent and relax both positive and negative supercoils. They function as heterotetramers of two subunits each of GyrA, GyrB , ParE and ParC. GyrA cleaves DNA and GyrB binds ATP.

Heterochromatin at Telomeres and in End-Joining

Telomeric DNA and centromeric DNA retain a highly condensed conformation throughout interphase, called heterochromatin. This results in repression of genes inserted proximal to the telomere, telomere position effect (TPE) (Boulton and Jackson 1998). The SIR proteins are required for this in yeast. SIR2, SIR3 and SIR4 interact with each other, and SIR3 and SIR4 interact with histones H3 and H4 (reviewed Tsukamoto and Ikeda 1998). SIR3 and SIR4 are required for the hypoacetylation of H4 which is necessary for the formation of heterochromatin. SIR2, SIR3 and SIR4 are also required for non homologous end-joining in yeast (above). Ku (and other non homolgous end-joining proteins, Section 4) is required for telomere length maintenance. It has been suggested that non homologous end-joining and telomere length maintenance both occur via a heterochromatin structure (Tsukamoto and Ikeda 1998). Ku is postulated to bind a DNA end and recruit SIR4 (which binds Ku70). SIR4 then recruits SIR2 and SIR3. The SIR proteins form a nucleofilament on DNA by interacting with the N-termini of H3 and H4. This results in DNA condensation. In the case of end-joining, the recombination proteins and ligase IV act on the heterochromatin DNA. The heterochromatin may be required to protect the DNA ends from degradation.

Secondary Structure

Conformational Anomalies

B-DNA is generally more thermodynamically stable than alternative conformations (reviewed van Holde and Zlatanova 1994, Leach 1999 in press). A-DNA is shorter and squatter and also right handed. However it is unphysiological and forms where there is little water of hydration. Certain other conformations become thermodynamically favourable under physiological conditions when the effective superhelical density is high: These conformations therefore are stabilized by unconstrained negative supercoiling.

Palindromes are sequences which are self-complementary in single-strand DNA and exhibit two-fold rotational symmetry in double-strand DNA.

They have often been likened to inverted repeats. Hairpins are fold-back structures which can potentially form from single-strand perfect palindromic sequences. They have a stem-loop structure. Cruciforms consist of two hairpins formed directly opposite one another from double-strand palindromic DNA.

Pseudo-hairpins are formed from imperfect palindromes (quasi-palindromes) with sufficient self-complementarity to fold-back. Their imperfect nature results in the formation of mismatches along their stems. They may include non palindromic spacer sequences at their centres. S-DNA is a cruciform-like structure in which each strand forms a pseudo-hairpin at a different position. The arms are therefore not directly opposite one another.

Z-DNA (zig-zag) consists of a very long and thin, left handed helix in which the two strands are wound in opposite directions. It potentially forms at alternating purine-pyrimidine dinucleotide repeats. H-DNA and *H-DNA are bi-molecular triplex forms. They are formed of a pseudo-hairpin juxtaposed alongside a single-strand. In H-DNA, the pseudo-hairpin is pyrimidine rich and the single-strand is purine rich, and *vice versa* in *H-DNA. Proton availability determines which is formed. Quadruplex DNA can be stabilized in various alignments *in vitro* either by G_4 or G_2C_2 quartets. It is very stable, inhibits DNA synthesis *in vitro* and it is unknown whether it exists *in vivo*. If it does exist, it may form at telomeres.

The Kinetics and Thermodynamics of Hairpin Formation and Palindrome Extrusion

Cruciform extrusion and hairpin formation differ in that cruciform extrusion additionally requires the melting of a region of duplex DNA. Other than this, the two processes bear similarities.

The hairpin arms of a cruciform extrude *in vitro* until the negative superhelical tension of the carrier DNA is reduced (Mizuuchi *et al* 1982). The rate of extrusion is dependent upon temperature, according to T_m, and increases with superhelical density of the carrier DNA (Mizuuchi *et al* 1982, Sinden *et al* 1983, Sinden and Pettijohn 1984, Warren and Green 1985). The

stability of the extruded form decreases with loop size (which may be determined by the size of a non-complementary spacer) and increases with stem length (Mizuuchi *et al* 1982, Xodo *et al* 1991).

Studies *in vivo* showed that stability is also increased by the presence of a pyrimidine 5' of the loop and complementary purine 3' of the loop, optimally GC (Davison and Leach 1994b), an AT rich loop and a GC rich stem (Davison and Leach 1994a). The reduction of cruciform extrusion by introduction of a small (8bp) central asymmetry indicated that extrusion occurs by a centre-dependent pathway (Chalker *et al* 1993). This supports the a model for extrusion in which a small bubble forms at the spacer or loop site, snaps into a protocruciform by base pairing and then develops into a cruciform by branch migration.

Extrusion is therefore governed by extrusion kinetics and thermodynamic stability of the hairpin or cruciform. Even where extrusion is thermodynamically feasible, it does not occur if there is a high kinetic barrier. Consequently short hairpins (*eg* a 33bp stem) rarely extrude and may do so only in the presence of a cofactor (Sinden *et al* 1983).

Detecting Hairpins and Cruciforms

The existence of hairpins at single-strand regions on a replication fork has been demonstrated *in vivo* by the methylation resistant state of GATC sites when inserted into the spacer region of a double-strand palindrome (Allers and Leach 1995). Double-strand GATC sequences are methylated by Dam methylase, but are not recognised by the enzyme when single-stranded as in a hairpin loop. Cruciforms have also been observed by atomic force microscopy (AFM) (Shlyakhtenko *et al* 1998). They appear highly mobile, adopting either extended perpendicular or compact parallel conformations.

Palindrome Instability and Inviability

Perfect palindromes up to approximately 40bp occur naturally in *E. coli* (reviewed Leach *et al* 1987). Quasi-palindromes longer than 40bp and palindromes separated by long spacers also occur naturally. However long palindromes, greater than approximately 100bp or 150bp total length, cannot

be propagated in *E. coli* (Lilley 1981, Warren and Green 1985). They undergo either of two fates; instability or inviability.

Instability is the partial internal deletion of a palindrome during replication. This process is likely to occur by strand-slippage. The product may be sufficiently short to be stable.

Inviability is the loss of the carrier replicon from the population. It occurs when the presence of the palindrome prevents replication of the carrier replicon. Inviability is determined by cell genotype (Section 2).

Kinetic and thermodynamic parameters which reduce the probability that a palindrome will extrude also reduce instability and inviability (Warren and Green 1985). However instability has been detected for palindromes as short as 22bp (Das Gupta *et al* 1987). Instability is also a function of position because it depends on the precise effects of local sequence. Both processes have been detected across the phyla, although there appear to be different levels of palindrome tolerance in different organisms and at different locations (Akgun *et al* 1997, reviewed Leach 1996b). Repetitive DNA and IRs are much more common in the genomes of higher eukaryotes than in prokaryotes.

Palindrome Cleavage Activities

Holliday Junctions are similar in structure to cruciforms. They can be cleaved by RuvC and Rus in *E. coli* (above), CCE1 of *S. cerevisiae* (Kleff *et al* 1992), X-solvase of mouse B cells (Solaro *et al* 1995), endonuclease VII of phage T4, endonuclease I of phage T7 and others. Each of the latter two introduces two single-strand nicks, nick and counter-nick, in a non concerted fashion (Picksley *et al* 1990, Mueller *et al* 1990, Pottmeyer and Kemper 1992, Solaro *et al* 1993). T7 Endo I cleaves three way (Y) junctions as well as four way (X) junctions (deMassy *et al* 1987). Junction resolving enzymes are proposed primarily to recognise secondary structure (Bhattacharyya *et al* 1991).

Hairpins can be opened by the activity of non specific single-strand endonucleases at the loop (Kabotyanski *et al* 1995). A better candidate would

be an enzyme with a higher degree of specificity than this, because it would allow for a degree of regulation in the process. However such an activity need not be hairpin specific *per se*. A candidate for a hairpin opening activity is presented in this thesis.

Genetic Disease

Cancer

p53 in Malignant Transformation

Malignant transformation of eukaryotic cells can result from gain-of-function mutation of an oncogene or loss-of-function mutation of a tumour suppressor. p53 is a tumour suppressor gene product (Kim *et al* 1996). It is implicated in the maintenance of genome stability. It preferentally binds double-strand DNA with insertion or deletion mismatches. It has a direct role, and may have an indirect role, in nucleotide excision repair. Li-Fraumeni Syndrome results from mutation of p53 and is characterised by gene amplification.

p53 accelerates cell death in response to ionizing radiation (reviewed Leach 1996a): p53 mutants are more resistant to ionizing radiation than wildtype. This allows damaged chromosomes to enter S phase, which leads to chromosome breakage (breakage-fusion-bridge cycles) and gene amplification. Amplification can cause further mutagenesis, including mutagenesis of genes which affect tumour progression.

Checkpoint Control

Cell cycle progression of a normally dividing eukaryotic cell is regulated by cell cycle checkpoints (reviewed Leach 1996a). The G1/S checkpoint prevents DNA replication before DNA damage repair is complete. The G2/M checkpoint prevents cell division before DNA damage repair.

Loss of checkpoint control may be associated with DNA repair and immune deficiencies. Activation of p53 results in defective G1/S checkpoint control

(above). Abnormal V(D)J recombination can activate oncogenes for lymphoid cancers, including B- and T-cell lymphomas and leukaemias (Cox and Sinclair 1997).

Genetic Predisposition to Cancer in Humans

Genetic predisposition to cancer results either from alleles which contribute directly to malignant transformation which are unmasked by a second mutation in postnatal life, or from alleles which increase the rate of somatic mutation (Cox and Sinclair 1997). Alleles in the latter class include defects in checkpoint controls and are usually recessive.

Nijmegan breakage syndrome (NBS) confers sensitivity to ionizing radiation, predisposition to early-onset cancers, immune system defects which lead to respiratory infections, and developmental abnormalities including stunted growth and microcephaly (small head) (reviewed Featherstone and Jackson 1998). The gene responsible for this rare autosomal recessive disease, *NBS1*, maps to 8q21.3. The DNA damage dependent cell cycle checkpoint is defective in NBS patients; p53 is not induced at G1/S and DNA synthesis is not suppressed in response to ionizing radiation.

Ataxia telangiectasia (AT) has identical cellular characteristics as NBS, however its clinical features differ (reviewed Featherstone and Jackson 1998). It features progressive hypersensitivity to ionizing radiation, chromosomal instability (especially of chromosomes 7 and 14 where immunoglobulin and T cell genes are located), and loss of the G1/S cell cycle checkpoint. It results in lymphomas. The allele responsible, *ATM*, maps to 11q23 and encodes a relative of the protein kinase involved in V(D)J recombination. This protein signals the presence of DNA damage to p53.

In HNPCC, the somatic mutation rate is increased by mismatch repair deficiency, as described above.

Bloom Syndrome is a type of dwarfism inherited as an autosomal recessive in which patients have decreased immunity, increased sun sensitivity and develop early-onset cancers of several types (reviewed Cox and Sinclair 1997). It is characterised at the cellular level by elevated sister chromatid

recombination, chromosome fragility and hypermutation. The gene responsible, *BLM*, maps to 15q26.1 and encodes a RecQ homologue which interacts with DNA ligase I (reviewed Petrini *et al* 1995).

TREDs

Characteristics of TREDs

A growing number of hereditary neurological and neuromuscular diseases are known to arise as a result of expansion of trinucleotide repeat arrays (reviewed Leach 1996a, Cox and Sinclair 1997, Leach 1998). These diseases are called trinucleotide repeat expansion diseases, TREDs (Mitas *et al* 1995). Expansion generally occurs in trinucleotides containing G and C bases (below). It occurs explosively, from a few repeats to a very large number rapidly. Expansion can be polar, localised to one end of the affected gene, *eg* in Fragile X syndrome (Pearson *et al* 1997). It can occur in specific tissues or at specific times in development. Transmission is affected by parental origin, in some cases by imprinting. TREDs are associated with unusual inheritance patterns. These include incomplete penetrance, where not all the carriers of a dominant mutant gene exhibit the mutant phenotype, and anticipation, where each new generation of carrier exhibits increased severity or earlier onset of the disease.

Of all TREDs, the best studied is Huntington's Disease (HD) (reviewed Leach 1996a, Cox and Sinclair 1997, Leach 1998). This is an autosomal dominant progressive degenerative disorder of the basal ganglia and cerebral cortex which causes dementia, choreic movement, and emotional and cognitive impairment, including hallucinations and paranoia, and death within 20 years of onset. It has a frequency of 1 in 2500 adults. The age of onset is variable, usually approximately 40 years, sometimes as early as 20 years. Juvenile onset is associated with larger repeat arrays and is paternally determined. The 5'- CAG -3' trinucleotide repeat of the translated region of the *IT15 (HD)* gene normally occurs as 6 to 34 copies. This is amplified in affected individuals to 35 to 121 copies.

Other TREDs include Fragile X (aka Martin-Bell Syndrome, FRAX-A), Friedreich's ataxia (FA), Spinobulbar muscular atrophy (SBMA), Myotonic

dystrophy (DM), Spinocerebellar ataxia type 1 (SCA1), Mild mental retardation (FRAXE) and Dentatorubro-pallidoluysia atrophy (DRPLA) (Cox and Sinclair 1997).

Pseudo-hairpin Formation by Trinucleotide Repeats

All but one of the known TREDs result from expansion of $d(CXG)_n$ repeats. These are either $d(CAG)_n.d(CTG)_n$ or $d(CGG)_n.d(CCG)_n$ repeats. These repeats can form pseudo-hairpins (Mitas *et al* 1995, Marquis Gacy *et al* 1995, Darlow and Leach 1995). They can also form S-DNA or tetraplexes. These structures are stabilized by G≡C pairing. The exception is FA which results from expansion of a $d(GAA)_n.d(TTC)_n$ repeat. This repeat can form H-DNA (reviewed Leach 1999) and the $d(GAA)_n$ strand has been shown to form a pseudo-hairpin *in vitro* at low temperature (5°C) (Suen et al 1999). The $d(TTC)_n$ strand does not form a pseudo-hairpin. Repeats are written in the frame in which they fold. Both $d(CTG)_n$ and $d(CCG)_n$ pseudo-hairpins fold with an even number of nucleotides in the loop (Darlow and Leach 1995).

Longer arrays have the same T_m as shorter ones, indicating that they form several short pseudo-hairpins rather than few long ones (Petruska *et al* 1996). $d(GAC)_n.d(GTC)_n$ repeats do not form pseudo-hairpins and are not observed in expanded arrays (Darlow and Leach 1995).

Dynamic Mutation

The amplification of trinucleotide repeats is called dynamic mutation. In humans it occurs with a bimodal pattern, either by small incremental expansion or by large discrete saltatory expansion (Sarkar *et al* 1998). Incremental expansion predominates when the tract is of a length associated with premutation, *ie* the lower disease lengths. Saltatory expansion predominates when the tract is of a length associated with anticipation, *ie* very long disease lengths. Deletion normally predominates in *E. coli* but the bimodal pattern can be reproduced by maintaining the culture at a low temperature (Sarkar *et al* 1998).

Expansion and deletion is hypothesised to occur by strand-slippage across the base of pseudo-hairpins formed by $d(CAG)_n.d(CTG)_n$, $d(CGG)_n.d(CCG)_n$

and $d(GAA)_n.d(TTC)_n$ trinucleotide repeats and certain tetranucleotide repeats (reviewed Leach 1999). According to the results of Trinh and Sinden (above), strand-slippage would be expected to occur preferentially when the pseudo-hairpin (pyrimidine rich strand) is on the lagging strand. Incremental expansion occurs below a threshold tract length. Large discrete saltatory expansion occurs at tract lengths equal or above the threshold. The threshold is equal to the length of an Okasaki fragment (Sarkar *et al* 1998). It has therefore been suggested that the bimodal expansion pattern results from secondary structure formation in the lagging strand Okasaki fragment. Incremental expansion results when small pseudo-hairpins form in the lagging strand Okasaki fragment by strand-slippage. Saltatory expansion results when large loops containing pseudo-hairpins are formed. Large slippage loops form at longer tract lengths because the chances are lower that the end of the Okasaki fragment is anchored by adjacent non repetitive sequence.

This mechanism would account for some of the unusual characteristics of TREDs, including the explosive nature of expansion, incomplete penetrance and anticipation. Pseudo-hairpin extrusion might be affected by expression of the gene, thus explaining the observed polarity of expansion. That expansion may occur in specific tissues or times in development may similarly be related to gene expression.

Most trinucleotide repeat expansion in a mismatch deficient cell occurs as a result of small insertions and mismatches from replication slippage (Schmidt *et al*). Therefore the mismatch repair system is extremely important and prevents most of the potential expansion. It is unsurprising that HNPCC is often associated with multiple cancers.

Section 2

The Identification and Genetic Analysis of SbcCD

Seminal Experiments

The Identification of SbcCD

sbcBC is Required for RecF Suppression of *recBC*

recBC mutants of *E. coli* are deficient in homologous recombination, which is reduced to approximately 1% of the wildtype level, and have low viability (Section 1). They can be rescued by activation of either the RecE or the RecF pathway of alternative homologous recombination. These pathways restore wildtype phenotype. The RecE pathway is activated by *sbcA* and the RecF pathway was thought to be activated by *sbcB* mutation, each in addition to *recBC*.

The genetic requirements for RecF pathway activation were reexamined (Lloyd and Buckman 1985). *recBC* recombination deficiency is assayed as reduced recombinant formation in conjugation and transduction, lethal sectoring (visible as cell death at the edges of a colony), and increased sensitivities to UV and ionizing radiations, and to the chemical mutagen mitomycin C (MC^S). Transmission of $sbcB^+$ from a wildtype Hfr donor to a *recBCsbcB* recipient would be expected to restore the MC^S phenotype. However when this experiment was performed, MC^S progeny were obtained before transmission of *sbcB*. The early locus reponsible was mapped by a series of interrupted matings and P1 transductions to the region between *proC* and *phoR*, and was named *sbcC*. (However the location of *sbcC* was initially confused with that of *rdgC*, Section 3). It was verified that the effect was due to loss of *sbcC* function by Tn*10* insertion into *phoR*. Increased deletion of the region was obtained concomitant with restoration of MC^R phenotype. The locus was similarly verified by restriction mapping of Tn*1000* mutants.

recBCsbcBC mutants were completely recombination proficient. *recBCsbcB* was partially recombination deficient, exhibiting some lethal sectoring, MC^S, lower (30%) viability than *recBCsbcBC*, and abnormally long filament formation, and being outgrown by *recBCsbcBC*. However *recBCsbcB* is UV^R and therefore significantly more proficient than *recBC*.

recBC was not suppressed by *sbcC* alone; *sbcB* is the critical factor for suppression. *recBCsbcB* spontaneously reverted to *recBCsbcBC* at a high frequency, presumably because selection is strong under these conditions. It is likely therefore that *sbcC* mutants arise from *sbcB* mutants in a *recBC* cell.

sbcC is the Primary Determinant of Palindrome Inviability

As detailed in Section 1, perfect palindromes longer than approximately 30bp are subject to instability in *E. coli* and those longer than approximately 150bp are inviable. To elucidate the mechanism of palindrome inviability, a very long perfect palindrome of 3.2kb was constructed of inessential DNA and inserted into a λ phage vector (Leach and Stahl 1983). Palindrome viability of infected strains was scored as efficiency of plating. The efficiency of plating of the palindrome-containing phage on *recBCsbcB* approached that of the wildtype control phage. Plating efficiency was similarly high on *recBCsbcBrecF*, indicating that the activation of the RecF pathway was not responsible for the loss of palindrome inviability. Plating on *recBCsbcBrecF* has since been repeated and the same viability observed on *recBCsbcB* strains mutant in other genes of the RecF pathway (Leach *et al* 1987). Approximately 98% of the first generation progeny of plating on *recBCsbcB* retained the parental inviability on wildtype, indicating that there had been no extensive palindrome rearrangement or deletion in the majority of phage. However a low proportion of the first generation and a much higher proportion (16% red^+, 57% red^-) of the second generation progeny grew on wildtype. This was attributed to a progressive loss of palindromic sequence. Palindrome instability was confirmed by restriction analysis of the progeny and it was found that all progeny had undergone deletions.

In the light of the observation that *recBCsbcB* commonly mutates to *recBCsbcBC*, the requirement for *recBCsbcB* for loss of palindrome inviability was reexamined. A λ *redgam* phage carrying a 571bp nearly perfect palindrome was plated on test strains (Chalker *et al* 1988). A high efficiency of plating (corresponding to loss of palindrome inviability) was obtained on *sbcC* hosts, irrespective of their genetic backgrounds. This indicates that SbcC alone is responsible for palindrome inviability. In addition, reversion of the progeny to viability on wildtype (corresponding to instability) was

very low on *sbcC*. A perfect palindrome of 2x 1.1kb has been propagated in *sbcC*, although it is susceptible to instability (Malagon and Aguilera 1998).

sbcD Maps to the Same Region as *sbcC*

The *proC* to *phoR* interval to which *"sbcC"* had been mapped consisted of a 4.3kb *Bam*HI to *Eco*RI region contained in plasmids (Lloyd and Buckman 1985). This could not be detailed by assaying loss of suppression of *recBCsbcBC*, since this phenotype is very adverse. Instead loss of palindrome inviability was used (Naom *et al* 1989). A λ phage containing a 571bp palindrome was used to infect *sbcC* single mutant cells carrying plasmids containing restriction fragments of the *Bam*HI to *Eco*RI region. Restoration of palindrome inviability was scored to map the *sbcC* locus within the region. A range of Tn*1000* insertion mutants in the *sbcC* region were derived from plasmids containing the *Bam*HI to *Eco*RI region and identified by their palindrome insensitive phenotype. They were analysed by restriction to define the locus more closely. Finally, deletion derivatives were used to map the locus to an interval extending from a 0.1kb *Alu*I to *Hae*III region to a 0.6kb *Xho*I to *Pst*I region.

This interval was expressed in maxicells (which have highly irradiated chromosomes and in which replication and gene expression are limited to the unirradiated plasmids present) (Naom *et al* 1989). It was found to encode a weakly expressed polypeptide of approximately 120kDa and a polypeptide of approximately 45kDa. Nucleotide sequencing revealed that the region is translated in a counter-clockwise direction and contains two large ORFs. The larger was predicted to give a polypeptide of 118.730kDa, assumed to be SbcC. The smaller was upstream of *sbcC* and predicted to give a polypeptide of 44.717kDa, named *orf-45*. The two ORFs appeared to be co-expressed as an operon from a single mRNA.

The palindrome inviability assay was used to investigate the function of *orf-45* (Gibson *et al* 1992). A kanamycin resistance (*kan*) cassette was used to disrupt chromosomal *orf-45*. Like *sbcC* mutants, the *orf-45::kan* insertion mutants caused loss of inviability of the 571bp palindrome contained in λ. To determine whether the loss of palindrome inviability was due to loss of Orf-45 function, or to a reduction in SbcC function by a polar effect of the

insertion on downstream *sbcC* expression, these genes were expressed from plasmids in chromosomal mutants. As expected, expression of *sbcC*$^+$*orf-45*$^+$ *in trans* caused palindrome inviability in both chromosomal *sbcC* and chromosomal *orf-45::kan*. Expression of *sbcC*$^+$ *in trans* caused palindrome inviability in a chromosomal *sbcC* mutant, as expected, but not in the chromosomal *orf-45::kan* disruption strain. This indicates that loss of function of Orf-45 is at least partially responsible for the loss of palindrome inviability in an *orf-45* mutant. (Actually, expression of *orf-45*$^+$ *in trans* caused palindrome inviability in neither the chromosomal *sbcC* mutant nor the *orf-45::kan* disruption strain, indicating that the *orf-45::kan* disruption has a polar effect on *sbcC* expression. The authors suggest that it may reduce *sbcC* expression below a limit required for palindrome inviability. However a point mutation in *orf-45* did not affect *sbcC* expression).

The similarity of effect of SbcC and Orf-45 on palindrome inviability inspired a test of the effect of Orf-45 on recombination (Gibson *et al* 1992). Fast growing excision revertants were sought among *recBCsbcB* cells carrying a Tn*10* insertion in the *sbcCorf-45* region. By transforming with complementing plasmids and infecting with the palindrome-containing λ phage, three classes of revertant were found; *sbcC*, *orf-45*, and *sbcCorf-45*.

Mutants in *orf-45* appeared identical in effect to *sbcC* mutants. The *orf-45* gene was consequently renamed *sbcD*.

SbcCD Mediates Palindrome Inviability During DNA Replication

Long Palindromes Block Replication

When a λ lysogen is superinfected, the incoming phage DNA circularises and becomes supercoiled but is not active. This was used to test for palindrome inviability in an inactive carrier replicon (Leach and Lindsey 1986). Control λ phage or λ *redgam* phage containing a 530bp palindrome were used to infect non lysogens under conditions preventing rolling-circle replication, packaging or burst. Recovery of the different phages was scored by restriction and agarose gel analysis. As expected, palindrome recovery from the *rec*$^+$*sbc*$^+$ strain was poor, but that from *recBCsbcB* (presumed *sbcC* or

sbcD) was efficient. When the same phage were used to infect λ lysogens, recovery of palindrome-containing phage was as high as wildtype in both strains. This indicates that the palindrome causes loss of its carrier replicon under conditions where the DNA is active. The authors proposed that the aspect of DNA activity responsible was likely to be replication. The efficient recovery of palindrome-containing phage from the *rec*$^+$*sbc*$^+$ lysogen also indicates that palindrome inviability is not mediated by supercoiling. Recovery of palindrome-containing phage from *rec*$^+$*sbc*$^+$non lysogens after 60 minutes growth was poorer than after 10 minutes growth, indicating that inviability is mediated at a time well into the lytic cycle.

The restriction and agarose gel assay was repeated using *rec*$^+$*sbc*$^+$ and *sbcC* non lysogens and a λ phage containing a 571bp palindrome. Palindrome-containing phage were again recovered efficiently from both strains, confirming that *sbcC* alone is required for loss of inviability (Chalker *et al* 1988). The palindrome-containing phage plated as efficiently on *recD* as *sbcC*. However plating on *recAD* was as low as on *rec*$^+$*sbc*$^+$. The *recD* host has elevated recombination; the lack of functional RecD reduces translocation of RecBC and increases the formation of recombinogenic ends. The observed efficient plating on *recD* might be explained by a block to replication in the presence of a palindrome since this would cause the strain to be more dependent on homologous recombination for phage growth.

That palindrome inviability is mediated during replication was confirmed by a density labelling assay (Shurvinton *et al* 1987). Heavy isotopes 13-C and 15-N were incorporated into λ phage containing a 8.4kb or a 530bp palindrome. These were used to infect *rec*$^+$*sbc*$^+$ strains under conditions of replication block (*dnaBts* mutant at 42°C) or free replication. Recovered phage were separated by density gradient centrifugation and scored. After a single round of replication, recovery of phage containing the 8.4kb palindrome was reduced 250-fold and that of phage containing the 530bp palindrome was reduced 20-fold, as compared to wildtype phage. Under conditions of replication block, recovery of palindrome-containing phage was as efficient as wildtype.

Long Palindrome Replication is Slow

The block to long palindrome replication could be explained by the Holliday Junction cleavage model of inviability if a palindrome were cleaved under conditions of replication. Alternatively the block might be a result of slow replication of a palindrome, which would support the suggestion that inviability is caused by interference with replication by the palindrome. To determine which model was correct, the restriction and agarose gel assay was repeated using fully Dam-methylated λ phage in a *dam* host (Lindsey and Leach 1989). Phage containing a 571bp palindrome and wildtype phage were constructed. The fully methylated status of this initial DNA was verified by *Dpn*I and *Mbo*I restriction. *Dpn*I cleaves fully methylated DNA only, and *Mbo*I cleaves fully unmethylated DNA only. A mixture of these phage was used to infect *rec*$^+$*sbc*$^+$*dam* hosts lysogenic or non lysogenic for λ. Growth was permitted for 15 minutes only to prevent late gene expression, packaging and burst. The extracted DNA was restricted with *Eco*RI to identify wildtype and palindrome-containing phage, and with *Dpn*I and *Mbo*I to remove all but once replicated DNA. There was no loss of palindrome-containing phage from the lysogenic host. The total recovered palindrome DNA from the non lysogen was less than the total recovered wildtype DNA, as scored in the fraction not restricted with *Dpn*I and *Mbo*I, demonstrating the previously observed block to replication. The recovery of once replicated palindrome DNA was equal to that of wildtype phage DNA, indicating that there had been no destruction of the initial palindrome DNA. This disproves the Holliday Junction cleavage hypothesis. Recovery of once replicated DNA from several cycles of replication would be unaffected by the rate of replication. Therefore this suggests that palindrome inviability might result from slow palindrome replication, and need not involve palindrome cleavage. However the two hypotheses are not necessarily mutually exclusive, especially if replication helps in the extrusion of palindromes.

SbcCD May Be Involved in RecBC Inhibition of Rolling-Circle Replication

The Mode of Plasmid Replication Can Be Altered in *recBC* Cells

Plamids normally replicate by the *ori* dependent (theta θ) mode (reviewed Summers 1996). The products of replication of ColE1-type (high copy number) plasmids in *E. coli* of various genotypes were studied by agarose gel electrophoresis (Cohen and Clark 1986). ColE1 plasmid DNA recovered from *rec*$^+$*sbc*$^+$ cells was either supercoiled closed covalent circle or relaxed open circle. In contrast, most of that recovered from *recBCsbcB* (presumed *sbcC* or *sbcD*) cells was of high molecular weight. High molecular weight ColE1 plasmid DNA recovered from *recBC* was reduced to 15%, and none was recovered from *sbcB* (presumed *sbcC* or *sbcD*). Therefore *sbcB* (presumed *sbcC* or *sbcD*) enhanced the effect obtained in *recBC*. The effect of *recBC* was reproducible using *rec*$^+$ in the presence of the λ RecBC repressor, Gam (expressed from *gamts* under temperature control). The same effect was observed using several different ColE1 derivatives, and a λ derivative which also belongs to the family of high copy number plasmids. The high molecular weight DNA could be degraded by RecBC double-strand exonuclease but not by S1 single-strand endonuclease, indicating that it consisted of linear double-strand multimers. This was confirmed by electron microscopy. Similarly examined were a low copy number plasmid (Silberstein and Cohen unpublished), a range of medium copy number plasmids and a minichromosome (consisting of a plasmid containing the major part of the *oriC* sequence). These were all found to undergo linear multimer synthesis in a *recBCsbcB* (presumed *sbcC* or *sbcD*) background (Silberstein and Cohen 1987).

The kinetics of linear multimer synthesis were studied by extracting DNA from exponentially growing *gamtssbcB* (presumed *sbcC* or *sbcD*) cells at designated intervals after derepression of *gam*. Linear multimers appeared after 15 minutes of growth at 37°C, formed the majority of plasmid DNA after 45 minutes, accumulated linearly with respect to chromosomal DNA after 1 hour and exceeded chromosomal DNAs by 25-fold after 3 hours. Circular monomer DNA did not increase during this time. Unlike theta

plasmid replication therefore, linear multimer synthesis was not subject to plasmid copy number control mechanisms which usually arise from plasmid functions (Cohen and Clark 1986). The minichromosome and the medium copy number plasmids continued in linear multimer synthesis in a *dnaA* replication initiation mutant, albeit at a reduced rate, but this was prevented in *dnaB* replication progression mutant (Silberstein and Cohen 1987).

All these observations were consistent with the model proposed (Cohen and Clark 1986) that the linear multimer synthesis corresponded to plasmid rolling-circle replication. It was proposed that theta replicating forms can sometimes sustain single-strand nicks. In the presence of RecBC, these would be degraded. However in the absence of RecBC, they would initiate sigma replication. It was proposed that any theta form, including plasmids and the chromosome, could potentially undergo a replication transition under suitable conditions.

The Role of SbcCD in Replication Mode Transition

Linear multimer synthesis was RecA dependent, reduced 35-fold in *recABCsbcB* (presumed *sbcC* or *sbcD*), but unaffected in *lexArecBCsbcB* (presumed *sbcC* or *sbcD*), suggesting that homologous recombination is required but not the SOS response. In apparent contradiction however, linear multimer synthesis was unaffected by the presence of *recF* or *recJ* mutations (Cohen and Clark 1986). The requirement for homologous recombination was clarified when it was found that the rate of linear multimer synthesis was reduced in these mutants (Silberstein and Cohen 1987). *sbcA* was found to substitute for *sbcB* (presumed *sbcC* or *sbcD*), but synthesis in *recBCsbcA* was RecA independent (Cohen and Clark 1986). Therefore although RecBC must be inactivated, homologous recombination is required, and the probable function of the *sbcB* (presumed *sbcC* or *sbcD*) and *sbcA* mutations is to activate alternative pathways of homologous recombination. Consistent with this, plasmid monomerisation by intramolecular RecE recombination is RecA-independent, however RecF recombination is RecA-dependent (Section 1).

Sequence Analysis of SbcC and SbcD

Gene Organisation

Expression of *sbcC* and *sbcD*

The DNA sequence of *sbcDsbcC* was determined (Naom *et al* 1989). A putative minus 10 promoter sequence was identified with a 6/6 match to consensus. However only a 2/6 match to consensus was identified at minus 35. It had been postulated that a 15 to 20 base match to a repeated 10 base degenerate palindrome sequence was sufficient for promoter activity (O'Neill 1989). Sequences flanking the minus 10 site were observed to match this requirement, and it was therefore predicted that the promoter was active. This was confirmed by a nuclease protection hybridisation assay which mapped the site of initiation of transcription to the position predicted (Gibson *et al* 1992). An AGG sequence 8 bases upstream of the translational start codon was suggested to constitute the ribosome binding site (Naom *et al* 1989).

sbcD is upstream of *sbcC*. The last translated *sbcD* codon overlaps the start codon of *sbcC* by one base. There are no independent promoter sequences nor independent ribosome binding sites for *sbcC*. It is suggested that *sbcDsbcC* are transcribed as a single operon from the same promoter, and that the expression of *sbcC* is dependent on ribosome translocation from *sbcD*. The protein purified from maxicell extracts was found to contain much more SbcD than SbcC, which is consistent with inefficient ribosome translocation from *sbcD* to *sbcC*. Also consistent with this, a lack of bias against rare codons was noted in both *sbcC* and *sbcD* (Naom *et al* 1989).

Location of *sbcDC*

phoB is located upstream of *sbcDC*. Downstream of *sbcDC* there are no sequences of dyad symmetry, characteristic of transcriptional terminators. However the sequence of the 3' end of *sbcC* has been identified as the beginning of an arabinose inducible promoter. Transcription of the arabinose inducible operon starts approximately 30 bases downstream of the *sbcC* stop codon (Naom *et al* 1989).

SbcC and SbcD Form a Complex

The identical nature of the mutant phenotypes of *sbcC* and *sbcD* led to the suggestion that the two gene products might be components of the same enzyme (Gibson *et al* 1992). The coexpression of the two genes as an operon (Naom *et al* 1989) is consistent with this.

The closest homologues of SbcC and SbcD form complexes together (T4 *gp46/47* Gram and Ruger 1985, T5 *gp D13/D12* Blinov *et al* 1989, Rad50Mre11 Johzuka and Ogawa 1995, below). SbcC is related to the SMC family of proteins, and these commonly form multisubunit complexes with non-SMC and/or other SMC proteins (*eg* Jessberger *et al* 1998). During purification the two peptides eluted from a gel filtration column in the same fraction, indicating that SbcCD is a complex (Connelly *et al* 1997, Section 3).

Conserved Sequences and Motifs

Debunking the RecBC SbcCD Divergence Myth

Initial computer alignments indicated that SbcC shares localised weak homology with *E. coli* RecB, RecN and UvrD, bacteriophage T4 *gp46* and bacteriophage T5 *gpD13*, and that SbcD shares homology with *E. coli* RecC, phage T4 *gp47* and phage T5 *gpD12* (Naom *et al* 1989). The authors suggested that RecBC and SbcCD may have evolved by divergence from common ancestors. The poor expression of both complexes, the common interaction with λ Gam protein (Section 3) and the position of the two operons on the genome (on opposite sides, consistent with the *E. coli* genome duplication hypothesis) seemed to support the suggestion. However further analysis demonstrated that homology between RecBC and SbcCD existed only over short regions, that these regions were not colinear and that overall homology was low; *ie* sequence similarity in certain regions was high, but identity was low (Leach *et al* 1992). This indicated that divergence was unlikely.

Closest Homologues of SbcC and SbcD

Regional homology between T4 *gp46/gp47* (sequenced by Gram and Ruger 1985) and T5 *gpD13/gpD12* (sequenced by Kaliman *et al* 1988) had already been demonstrated. Homology between T4 *gp46* and T5 *gpD13* was stronger than that between T4 *gp47* and T5 *gpD12* (Blinov *et al* 1989). SbcC was found also to exhibit stronger homology with *gp46* and *gpD13* than SbcD with *gp47* and *gpD12* (Leach *et al* 1992). T4 *gp47gp46* are two of eleven genes in the early gene region and, unlike *sbcDC*, appear to possess independent ribosome binding sites (Gram and Ruger 1985). The proteins have predicted molecular weights of 39.1 and 63.5kDa respectively. T5 *gpD12D13* have independent promoters (Kaliman *et al* 1988).

A further computer search found that the Mre11 protein of *S. cerevisiae* and Rad32, the equivalent protein in *S. pombe*, also shared homology with the SbcD subgroup of proteins, and this was localised mostly in the N-terminal domain (Sharples and Leach 1995). Mre11 and Rad32 exhibited strong sequence conservation in the C-terminal region. The Rad32 protein is predicted to be 73.5kDa in size and to have a pI of pH5.52 (Tavassoli *et al* 1995). The two-hybrid system was used to identify an interaction between *S. cerevisiae* Mre11 and Rad50 (Johzuka and Ogawa 1995). As a result the Rad50 sequence was aligned and found to share localised homology with SbcC and its relatives (Sharples and Leach 1995).

Numerous protein sequences from phyla as diverse as archaea, bacteria and mammals show similar colinear, localised homology to SbcC and SbcD (Swiss Prot). Many of these have not yet been cloned or analysed in detail. Close homologues in prokaryotes are generally named SbcC and SbcD. *B. subtilis* ORF3 was found to display significant localised homology to SbcD (Sharples and Lloyd 1993). *ORF3* is much shorter than *sbcD* but insertion of a single nucleotide extends the homology. A *sbcC* homologue exists downstream of *ORF3*, called *yisA* (Medina *et al* 1997). The two ORFs possess independent Shine-Dalgarno sequences and are predicted to encode proteins of 245 and 170 amino acids respectively. A SbcC homologue exists in *Clostridium perfringens* (Medina *et al* 1997). Close homologues in archaea have also been called SbcC and SbcD, *eg* SbcD in *Synechocystis* (Gary Sharples

unpublished). In higher eukaryotes they are called Rad50/Mre11, *eg* mouse and human (Section 4).

A Note on Homologues

The "close homologues" of SbcC and SbcD share only limited homology. This is restricted to short colinear regions which specify recognised sequence motifs. It is not completely accurate to call them homologues. However they are commonly refered to as such.

Motifs in SbcC and SbcD

In describing amino acid consensus sequences, the standard single letter abbreviations for amino acids have been used, with + to represent hydrophobic amino acids, x to represent any amino acid, and small letters to represent less well conserved amino acids.

The consensus sequence,

<hydrophobic stretch>(G/A)xxgxGK(T/S)

is present in the N-terminus of SbcC (Naom *et al* 1989, Leach *et al* 1992). This creates a fold which is predicted to be preceeded by a β-sheet and followed by an α-helix (Naom *et al* 1989). This secondary structural motif is definitive of a Walker NTP- (purine-) binding motif "A" and the G-rich part binds the γ-P of NTP (Walker *et al* 1982). Present in the C-terminus of SbcC is the weakly conserved consensus

++++D

in a D-rich region, definitive of a Walker B motif (Leach *et al* 1992, Walker *et al* 1982). The D of this consensus is critical and conservative substitutions abolish enzyme activity (Saitoh *et al* 1995). The intervening sequence consists of two long stretches of heptad repeats separated by a short spacer, and includes a high concentration of glutamine residues (Leach *et al* 1992). These two stretches are predicted to form α-helices (Sharples and Leach 1995) characteristic of a coiled-coil motif (below). The spacer contains the consensus

CPxC

which is half the zinc finger motif

$CxxC\text{-}x_n\text{-}CxxC$

(Leach *et al* 1992). These consensuses and motifs are common to the closest homologues of SbcC (Gram and Ruger 1985, Kaliman *et al* 1988).

The amino acid sequence of SbcD matches the split consensus for phosphoesterases and also has the basic C-terminal region common to phosphoesterases (Sharples and Leach 1995). The closest homologues of SbcD also possess these characteristics, although the yeast homologues have a much longer basic region and a longer linker between phosphoesterase motifs II and III (Sharples and Leach 1995). The additional region of the C-terminus shared by the eukaryotic homologues is charged and acidic (Section 4).

Phylogenetic Analysis of SbcC and SbcD

SMC Proteins and SMC-like Proteins

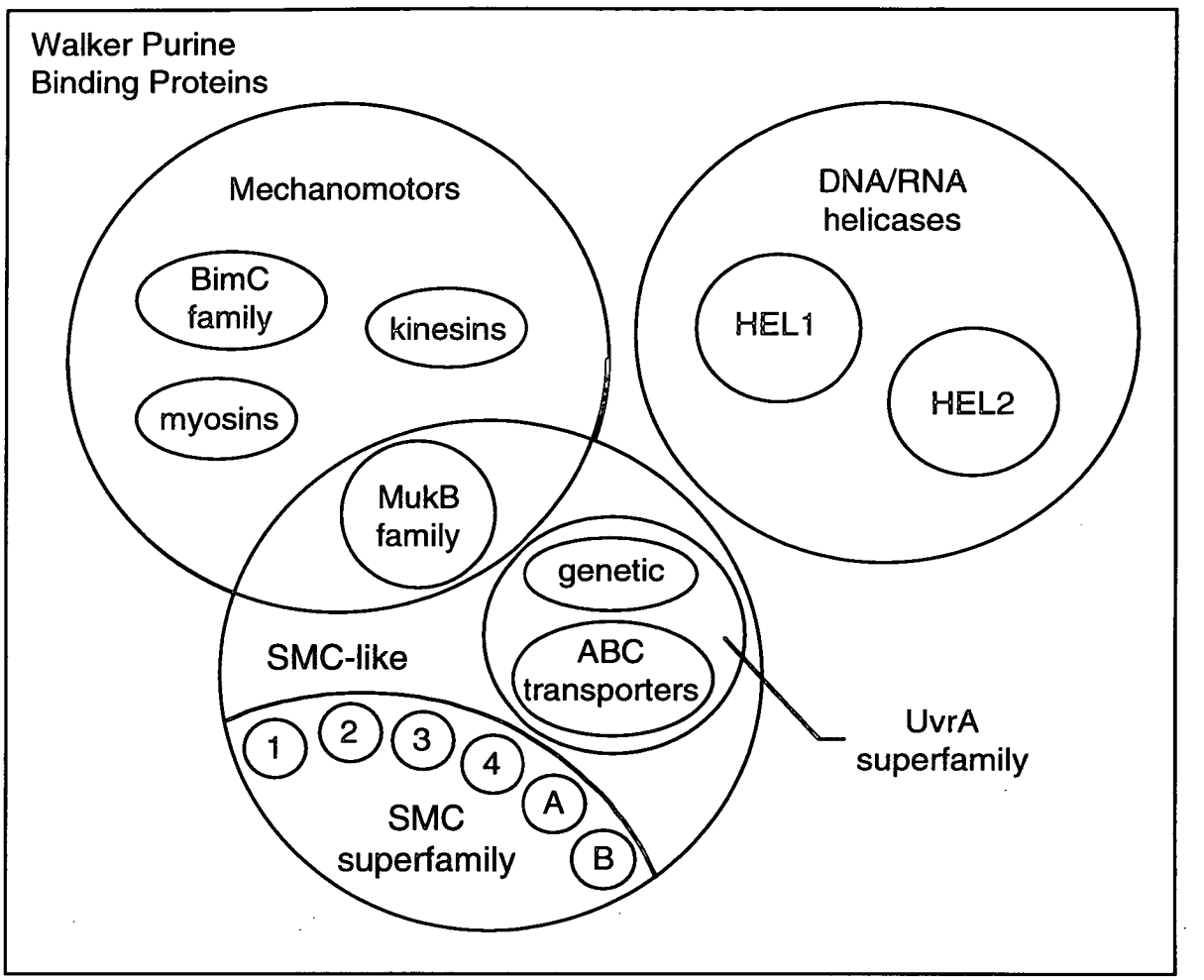
Superfamilies of ATP-binding Proteins

Walker NTP Binding Proteins

The vast range of enzymatic activities in a cell is mostly carried out by purine-binding proteins which couple ATP or GTP hydrolysis to a biochemical activity. Of the vast range of purine-binding proteins, the vast family of Walker purine-binding proteins is a subset. Within the Walker purine-binding proteins, three superfamilies are relevant to the study of SbcCD; the UvrA superfamily, the mechanochemical protein superfamily, and the SMC superfamily. These overlap as indicated in figure 1.2.1 (based on information from Gorbalenya and Koonin 1990, Leach *et al* 1992, Koshland and Strunnikov 1996, Melby *et al* 1998). Gp46/47 and gpD13/D12 have been placed in the family of ABC membrane transporters. They have xSxG instead of LSGG at the DA box (below) but are known to act at the membrane, and are otherwise homologous to the group.

SbcC and its closest homologues are members of the UvrA superfamily (Leach *et al* 1992) and the outlying group of SMC-like proteins (Melby *et al* 1998). The NTP-A and NTP-B domains of Walker proteins must juxtapose

Figure 1.2.1



SMC1: *Sc*SMC1, bovine SMC1

SMC2: *Sc*SMC2, *Ce*MIX-1, chick ScII, *Xl*XCAP-E, *Sp*cut14

SMC3: *Sc*SMC3, *Sp*SMC3, bovine SMC3

SMC4: *Sc*SMC4, *Ce*DPY-27, *Xl*XCAP-C, *Sp*Cut3

SMCA: *Synechocystis* SMC

SMCB: *B. subtilis* SMC

SMC-like (outliers): *E. coli* SbcC, *Sc*Rad50, mouse Rad50, human Rad50,
E. coli MukB, *Sc*RHC18, *Sp*Rad18

Melby *et al* 1998, Gorbalenya and Koonin 1990, Strunnikov *et al* 1993, Hirano and Mitchison 1994, Strunnikov *et al* 1995, Jessberger *et al* 1996, Chuang *et al* 1996, Sutani and Yanagida 1997

for GTP or ATP hydrolysis and are commonly located near one another in the primary peptide sequence (Walker *et al* 1982). However in SMCs, mechanochemicals and some UvrA proteins the NTP-A and NTP-B motifs are located at opposite termini, and the intervening region is coiled-coil. Monomers of these proteins have a characteristic Head-Rod-Tail structure, corresponding to the NTP-A: coiled-coil: NTP-B organisation. The "Tail" regions vary between the superfamilies (Peterson 1994, Niki *et al* 1992, Melby 1998).

Coiled-Coil Proteins

A coiled-coil motif consists of a strongly amphipathic heptad repeat pattern of hydrophobic and hydrophilic residues (Lupas *et al* 1991). The sequence, called "a" to "g," forms an α helix which may associate with two or three others in parallel to form a wound filament structure. The coiled-coil packs at an angle of approximately 20° in a structurally complementary manner known as "knobs into holes". The residues at positions a and d are hydrophobic (Zhou *et al* 1992). They fall on the same face of the helix, generating a hydrophobic interface between the helices and thereby stabilizing the structure. Residue a contributes even more to stability than residue d. Computer simulation has confirmed the importance of hydrophobicity at these positions and shown that leucine is a good coiled-coil former at position d (Zhang and Hermans 1993).

Coiled-coil proteins have diverse functions as muscle regulators, DNA binding (including leucine zipper) proteins, growth regulators, transactivating proteins, and cytoskeletal proteins (Zhou *et al* 1992). An algorithm has been used to identify regions likely to form coiled-coils (Lupas *et al* 1991). A sequence is scored, and its score used to calculate the probability of coiled-coil formation. This algorithm has been used successfully to delineate the hinge region in myosin. It has been used to calculate that the region of SbcC between the termini has a coiled-coil forming probability of 1.0 (Connelly unpublished).

The UvrA Superfamily

The UvrA protein of *E. coli* functions in a complex with UvrB and UvrC as a helicase and single-strand nuclease in UV damage nucleotide excision repair of DNA (Claasen and Grossman 1991). It consists of two modules of similar organisation. Members of the UvrA superfamily share homology with a single module of UvrA and possess four conserved motifs (Gorbalenya and Koonin 1990). Motifs I and III are extended versions of the Walker NTP-A and NTP-B binding motifs, motif II is located between motifs I and III and motif IV is located on the C-terminal side of motif III. The consensus sequences are;

motif I	+x+++GxxgxGKst++x+
motif II	lsgGxxqxr+x+
motif III	++++Depxxxld
motif IV	+++xh

The motifs are organised;

I (35aa-1200aa) II (~10aa) III (~20aa) IV

In addition, some members also possess zinc finger motifs; UvrA possesses two complete zinc fingers, one in each of its two modules, and T4 *gp46*, T5 *gpD13* and SbcC each possess half zinc fingers. Disruption, by substituting cysteine into the N-terminal zinc finger motif of UvrA, had only a minor effect on activity, suggesting that the motif is not required for DNA binding (Gorbalenya and Koonin 1990). The authors speculated that a half motif from each of the two subunits of a dimer might together form an inter-subunit lock by Zn^{2+} co-ordination. Supporting this, the central bulk of the protein is involved in dimerisation (Claasen and Grossman 1991).

The UvrA superfamily includes two families of ATPases (Gorbalenya and Koonin 1990): The ABC transporters (*eg* CFTR, FtsE, GlnQ) are located in the plasma membrane, and T4 *gp46* and T5 *gpD13* are homologous to these. A family of proteins with genetic roles in DNA repair and recombination (*eg* UvrA, MutS, HexA, RecF, RecN) includes Rad50 and SbcC (Sections 3 and 4). The distance between motifs I and II is variable in both families, but tends to be shorter in the ABC transporters.

The DNA and RNA Helicase Superfamilies

Two vast families of putative RNA and DNA helicases, HEL1 and HEL2, were previously described (Gorbalenya and Koonin 1989). These have a further extended NTP-B motif, the DEAD box. HEL1 includes RecA, RecB, RecD and UvrD, and HEL2 includes RecQ and UvrB. They are related to the UvrA superfamily, but are not included in it.

The Mechanochemical Protein Superfamily

Mechanochemical proteins are molecular motors. They are force generating enzymes fuelled by purine NTP hydrolysis which generally translocate along cytoskeletal filaments and slide them relative to one another. The sliding is probably engineered by a "walking" or "rachet" mechanism. The motor domain is also involved in NTP hydrolysis. Myosin is a small protein which crosslinks actin filaments and slides them relative to one another during muscle contraction and cytokinesis (Huxley 1998). It has motor domains at both ends, a Motor-Rod-Motor structure. Kinesin is a flexible protein which crosslinks organelles and vesicles to microtubules to move them during cytokinesis, and possesses a motor domain at one end, Motor-Rod-Tail. Dynamin is another microtubule associated mechanochemical protein. The BimC family is required for spindle function at cell division, and has been suggested to crosslink microtubules to push them apart (Kashina *et al* 1996). Members form bipolar homotetramers of kinesin-related subunits which are assembled as two parallel homodimers arranged in antiparallel. The MukB family members have non essential roles in chromosome segregation (Niki *et al* 1992). MukB is a large Motor-Rod-Tail protein with a globular hinge in the rod, and two zinc finger-like sequences in the tail. Zinc is required for NTP binding. Unlike the other mechanochemical proteins described here, it has DNA binding activity contained in the C-terminus.

The SMC Superfamily

As indicated above, SbcC and its closest homologues fit into both the UvrA superfamily and the outlying group of SMC-like proteins. RecN is also a member of both the UvrA superfamily and the outlying group of SMC-like proteins. However it is more closely related to the SMCs than is SbcC or its

closest homologues (Sharples and Leach 1995). (RecN is involved in recombinational repair of DSBs). *E. coli* MukB is a member of both the mechanomotor proteins and the outlying group of SMC-like proteins (Melby *et al* 1998).

The SMC proteins are interesting because they are essential, ubiquitous through the phyla, and participate in a huge range of DNA reactions. During the past five years, they have been studied with fervour and patterns are beginning to emerge regarding the characteristics and cellular roles of the subfamilies. It is therefore pertinent to discuss the current understanding of this superfamily.

Identification of the SMC Superfamily

Two mutants in *S. cerevisiae* were obtained which increased minichromosome nondisjunction more than 10-fold and were hence named *smc* stability of minichromosomes (Larionov *et al* 1985). *smc1* was cloned and characterised (Strunnikov *et al* 1993). Smc1p was found to be essential for mitotic cell division, and for centromere based, short artificial chromosome and natural chromosome segregation. It was shown to act during mitosis only. The ORF specified a protein of 141kDa, but immunoreactive polypeptides were 165kDa and 178kDa in size, and assumed to be post-translationally modified. *smc2* displayed the transient morphology characteristic of defective chromosome condensation (Strunnikov *et al* 1995). The cloned product, of predicted size 134kDa, was demonstrated to be required for both the establishment and maintenance of the condensed state, unlike topoisomerase II which functions in establishment only. Therefore although the protein shared colinear homology with Smc1p, and a rapidly growing number of other proteins (in a range of organisms) which also functioned in chromosome maintenance, its role was distinct from that of Smc1p. The Smc proteins were renamed SMC, structural maintenance of chromosomes, and the family was named after them.

Chromosome condensation is subject to cell cycle regulation (Section 1). Studies on chick cells determined that the mitotic scaffold consisted of two major components, ScI and ScII. The most abundant protein, ScI, was identified as topoisomerase II (Earnshaw *et al* 1985). ScII was identified as a

homologue of *S. cerevisiae* SMC1 (Saitoh *et al* 1994). Genetic studies indicated that topoisomerase interacts with SMC1 in the mitotic scaffold (Peterson 1994). There is also evidence to suggest that SMC proteins may have a role in maintaining the nuclear matrix (Hirano 1995a).

Domain Structure of SMC Proteins

SMC proteins are large (115 to 160kDa) and have three defining structural features: They have an extended, N-terminal Walker NTP-A binding motif called a P-loop (Melby *et al* 1998);

FKS (~16aa) GxNGSGKSN (~102aa) QG

They have an extended C-terminal Walker NTP-B binding motif called a DA box (Strunnikov *et al* 1993);

LSGG (~17aa) PxP++++DE+DAALD

And the intervening sequence is coiled-coil. The P-loop and the DA box are essential for function, and mutations in the coiled-coil cause conditional lethality (Koshland and Strunnikov 1996). Of the three defining features, the DA box is the most conserved in SMC proteins (Saitoh *et al* 1995).

The DA box: ATP Hydrolysis or DNA Binding?

It is unclear as to whether the DA box functions in ATP hydrolysis or as a DNA binding motif. It was originally predicted to form a helix-loop-helix structure, which would implicate this region in DNA binding activity (Strunnikov *et al* 1993). However it could alternatively form a β strand (Saitoh *et al* 1995). If the DA box were to act as a Walker NTP-B motif it would interact with the NTP-A motif to form the ATP hydrolysis pocket, and for this it would be required to form a β strand (Walker *et al* 1982). Consistent with a role in ATP hydrolysis, mutations in the DA box abolish ATP hydrolysis (Saitoh *et al* 1995). The authors propose that D co-ordinates Mg^{2+}, which is itself co-ordinated by the β- and γ-P of NTP (as defined by Walker *et al* 1982). This proposed function of the DA box in ATP hydrolysis implies that the termini of SMC proteins interact and therefore must be aligned. However, consistent with a role in DNA binding, it has been shown that the C-termini of SMC1 and SMC2 each bind double-strand DNA, and that the N-termini, but not the C-termini, bind a photoaffinity ATP analogue (Akhmedov *et al* 1998). Perhaps it is significant that the authors were unable

to demonstrate ATP hydrolysis by the N-termini. They suggest that the N-terminus may be responsible for ATP binding, but the C-terminus might be required for hydrolysis. Neither experiment is totally unambiguous, *eg* mutations may affect dimerisation rather than DA box function alone, and the termini may have differing activities independently and in the intact molecule.

The extent of homology to the DA box of SbcC (Naom *et al* 1989), T4 *gp46* (Gram and Ruger 1985) and T5 *gpD13* (Kaliman *et al* 1988) (aligned Leach *et al* 1992) is as follows, where *n* represents a non-conservative substitution, and *c* represents a conservative substitution;

SbcC LSGG ... *nxnn+n*+DE+*ncn*LD
gp46 *n*Sn*G* ... *nxnn*+++DE+*n*D*cnn*D
gpD13 *n*Sn*G* ... *nxn*++*n*+DE+*nnnn*D

The P-loop

The P-loop was so named because it is thought to form a phosphate binding loop (Hirano *et al* 1995b) during ATP hydrolysis. Outliers to the SMC proteins have mostly conservative substitutions in the P-loop. The extent of homology to the P-loop of SbcC (Naom *et al* 1989), T4 *gp46* (Gram and Ruger 1985) and T5 *gpD13* (Kaliman *et al* 1988) is as follows, where brackets indicate that the sequences have not been aligned specifically with regard to this part of the consensus (Leach *et al* 1992);

SbcC *(ccn)...* Gx*n*G*n*GK*cn* ... (QG)
gp46 *(ccn)..* GxNG*n*GKS*n* ... *(Qn)*
gpD13 *(ccn)* ... GxNG*n*GKS*n* ... *(nn)*

The Coiled-Coil

The two α helical, heptad repeat sequences characteristic of a coiled-coil have been studied in SMC proteins (Melby *et al* 1998). They are long, approximately 300 amino acids, and separated by a spacer of several hundred amino acids. The spacer includes several isolated, highly conserved amino acids. RHC18, of *S. cerevisiae*, and Rad18, the equivalent in *S. pombe* are members of the SMC outliers and show sequence homology to the SMC spacer.

Physical Structure of SMC Proteins

SMC2 can form homomultimers and heterodimers with SMC1, and this has been suggested to occur via interaction of the coiled-coil domains (Strunnikov *et al* 1995). Indeed eukaryotic SMCs generally function as heterodimers of SMC monomers (Hirano *et al* 1995b, Jessberger *et al* 1998). In contrast, prokaryotic SMCs are predicted to act as homodimers since prokaryotic genomes each appear to encode only one *bona fide* SMC protein (Melby *et al* 1998). SMCs act as high molecular weight complexes which include non-SMC accessory proteins (Koshland and Strunnikov 1996, Jessberger *et al* 1998, Melby *et al* 1998).

The *E. coli* MukB protein, although classified as an outlier to the SMC superfamily, has a similar domain structure to the SMC proteins and has mutant phenotypes similar to those of the *B. subtilis* SMC protein. It forms homodimers independently of accessory proteins. It was therefore chosen as a model with which to study the physical structure of SMC and SMC-like proteins (Melby *et al* 1998). Electron microscopy of the homodimer revealed two globular domains separated by a rod. The rod was identified as the coiled-coil. By examining different molecules, the products of deletion mutagenesis and a fusion protein, evidence was obtained for the following model of MukB structure: The larger of the two globular domains was suggested to consist of the terminal regions, either closely juxtaposed or connected. The smaller globular domain was suggested to consist of the coiled-coil spacer. The spacer was demonstrated to be flexible, allowing scissor-like movement of the coiled-coil arms from 0° to at least 180°, and was therefore referred to as a hinge. Where both termini were intact they could be brought into alignment at 180° by bending at the hinge. The distance between the three globular domains was consistently equal, and the extra bulk of the fusion protein was visible at both ends of the homodimer. This indicated clearly that the homodimer is of antiparallel alignment.

SMC dimers therefore provide an exception to the rule that long coiled-coils associate in parallel alignment (Melby *et al*1998). However this structure is consistent with the prediction outlined above that the DA box interacts with the P-loop to form an ATP binding pocket (Saitoh *et al* 1995). Heterodimeric

SMCs are predicted to pair four different sequences, but this is feasible for those examined (Melby *et al* 1998).

A solution of SbcCD was studied by electron microscopy (Connelly *et al* 1998). However it is unknown whether sufficient Mn^{2+} remained in the preparation for the complex to remain intact throughout the procedure, so the protein was referred to as SbcC(D). Two globular termini were observed separated by a gently curved, long rod. No structures were observed in which the termini were juxtaposed. SbcC lacks a large spacer (Naom *et al* 1989). It is likely that the two globular regions correspond to the termini, and that SbcC lacks a flexible hinge (Melby *et al* 1998). The structure observed is that expected for an antiparallel homodimer of SbcC which does not bend at a hinge.

SMC Families

Several alignments of SMC proteins have been performed, and the authors of the reports have named the families differently (Koshland and Strunnikov 1996, Jessberger *et al* 1998, Melby *et al* 1998). In addition, SMCs group differently according to whether the alignment is based on the P-loop or the DA box (Melby *et al* 1998). The nomenclature used here is that of Melby *et al* 1998. There are four families of eukaryotic SMC proteins, named after their counterparts in *S. cerevisiae*, SMC1, 2, 3 and 4. There is one family of archaeal SMC proteins, called SMCA (for archaeal), and one family of prokaryotic SMC proteins, SMCB (for bacterial). In addition there is a family of outliers, or divergent SMCs, which have lesser homology. Outliers exist in a range of organisms. These are here called SMC-like proteins. Examples of members of the families are shown in figure 1.2.1 (Strunnikov *et al* 1993, Hirano and Mitchison 1994, Strunnikov *et al* 1995, Jessberger *et al* 1996, Chuang *et al* 1996, Sutani and Yanagida 1997).

There is one known exception to this phylogeny; *Synechocystis* is a cyanobacterium but its SMC protein falls into the SMCA family. Of all the genomes completely sequenced, only three do not include a protein with sufficient homology to enter this phylogeny; *Methanococcus thermoautotrophican* has a distantly related protein, *Borelia burgdorferi* has a

short SMC-like protein, and *Helicobacter pylori* lacks any SMC or SMC-like protein.

Functions of SMC Families

SMC proteins are involved in a wide range of cellular activities. In eukaryotes, these fall into four main categories; dosage compensation, mitotic chromosome condensation, sister chromatid cohesion, and DNA recombination and repair (Jessberger *et al* 1998).

The diverse roles of SMC proteins was suggested to be regulated by the formation of different complexes with distinct specificities (Strunnikov *et al* 1995). Evidence has since been obtained that an SMC protein can form different heterodimers with different activities (*eg* MIX-1 in Chuang *et al* 1996). Two principle types of heterodimers have been identified, SMC1/3 and SMC2/4 (Jessberger *et al* 1998).

Dosage Compensation

C. elegans DPY-27 (SMC4) is responsible for dosage compensation between XX hermaphrodites and XO males (Chuang *et al* 1994). Transcription of one X chromosome is reduced 2- or 3-fold in hermaphrodites by the action of a complex including DPY-21, 26, 27, 28 and 30. It was later found that the complex also includes a second SMC protein, MIX-1 (SMC2) (Chuang *et al* 1996). The authors proposed that this was achieved by repression of X-linked transcription by X chromosome condensation. If this is so, the process is obviously mechanistically related to mitotic chromosome condendation.

Mitotic Chromosome Condensation

Mitotic chromosome condensation is generally achieved by a multisubunit complex called condensin which includes SMC2/4 homologues. As indicated above, chick ScII (SMC2) is the second most major constituent of the mitotic chromosome scaffold. The XCAP-E (SMC2) and XCAP-C (SMC4) proteins of X. *laevis* 13S condensin were found to be essential for chromosome condensation (Hirano and Mitchison 1994). This multisubunit complex also includes non SMC proteins XCAP-D2, XCAP-G and barren, which are phophorylated in a cell cycle dependent way. It interacts with

topoisomerase II. *S. pombe cut3* and *cut14* mutants are defective in cell division, exhibiting a "cut" phenotype in which uncondensed chromosomes are severed by the invaginating cell membrane at mitosis (Saka *et al* 1994). Cut14 (SMC2) was purified as a complex with Cut3 (SMC4) (Sutani and Yanagida 1997). The Cut14/3 complex was found to possess an ATP independent DNA annealing activity which was approximately 70% more efficient than *E. coli* RecA. This activity was proposed to contribute to DNA supercoiling and to be regulated by topoisomerase II. Mitotic chromosome condensation in *C. elegans* occurs by a multisubunit complex including MIX-1 (Chuang *et al* 1996). Since this SMC2 protein is also involved in dosage compensation, this examplifies that the role of an SMC protein can be regulated by heterodimerisation. It is predicted that MIX-1 heterodimerises with an SMC4 protein in mitotic chromosome condensation, but this protein has not yet been identified. *dpy-26* and *dpy-28* mutants exhibit a cut phenotype, suggesting that they may encode subunits of the *C. elegans* condensin (Chuang *et al* 1994).

Sister Chromatid Cohesion

Mitotic sister chromatids must be held together during the time between DNA replication and the metaphase to anaphase transition, which is the correct time for separation. Cohesins are required for this. In *S. cerevisiae* SMC1 and SMC3 form a cohesin complex with SSC1, a non SMC protein (Michaelis *et al* 1997). This binds sister chromatids once synthesised (mid S phase) until anaphase (M phase). Localisation of the complex is dependent on SMC1, suggesting that complex assembly is ordered and occurs at sites determined by SMC1/3. Cohesion is lost when SSC1 dissociates.

DNA Recombination and Repair

Bovine RC-1 is a complex involved in homologous recombination and includes bovine SMC1, SMC3, DNA polymerase ε and DNA ligase (Jessberger *et al* 1996). D-loop formation is indicative of an early stage in recombination, and both *S. cerevisiae* SMC1/3 and bovine RC-1 promote this reaction *in vitro* if the target DNA is homologous and supercoiled (Jessberger *et al* 1998). *S. cerevisiae* SMC1/3 has a DNA annealing activity, which could

be implicated in DNA recombination or repair, as well as sister chromatin cohesion or chromosome condensation (Jessberger *et al* 1998).

In addition, a number of the SMC outliers have roles in recombination and repair (Jessberger *et al* 1998): *S. cerevisiae* RHC18 and *S. pombe* Rad18 are involved in an unusual post-replicative repair pathway (Lehmann *et al* 1995) which involves the respective homologues of *E. coli* RecA. Rad51 is a member of the Rad52 epistasis group. A human homologue of RHC18 has been identified, but its activity has not yet been analysed. SbcC and Rad50 are also implicated in recombination (Section 3).

Rad21, the *S. pombe* homologue of SSC1 is important for DSB repair (Birkenbihl and Subramani 1992). RC-1 appears to be involved in both mitotic sister chromatid cohesion and DNA recombination (above). Therefore these two processes may be mechanistically related.

Other SMC Heterodimers

The ability of *S. cerevisiae* SMC1 and SMC2 to heterodimerise (Strunnikov *et al* 1995) suggests that other SMC heterodimers may exist and have cellular roles (Melby *et al* 1998). Supporting this, the C-termini each of SMC1 and SMC2, and the Cut3/14 (SMC2/4) complex independently promote DNA annealing (Jessberger *et al* 1998). Annealing is promoted 10-fold more efficiently by the C-termini of SMC1 and SMC2 than the N-termini, indicating that the C-terminus is the primary mediator of the reaction, and 2-fold more efficiently than by *E. coli* RecA protein (Akhmedov *et al* 1998).

Binding of SMC Proteins to Secondary Structures in DNA

To investigate the specific roles of SMC proteins in their functional complexes, the biochemical activity of X. *laevis* 13S condensin (SMC2/4) was studied (Kimura and Hirano 1997). Bacterial topoisomerase I and eukaryotic topoisomerase II were used to score for positive or negative supercoils introduced into a substrate DNA. Topoisomerase I removes negative supercoils only, and topoisomerase II removes both negative and positive supercoils. Relaxed circle DNA was incubated with ATP, condensin and topoisomerase *in vitro* and the products were analysed by 2d electrophoresis.

Condensin was found to be a double-strand DNA stimulated ATPase which introduces positive supercoils. The activity was shown to be stoichiometric, not catalytic, increasing with the availability of condensin. The activity is stimulated by long (600bp) duplex DNA more than shorter (100bp) duplex. Moreover the activity is stimulated by cruciform DNA in comparison to the double-strand substrate.

The N- and C-terminal domains of SMC1 and SMC2 were purified individually and their DNA binding preferences analysed by gel retardation (Akhmedov *et al* 1998). Both C-termini bind double-strand DNA with a 100-fold preference over single-strand, and a cruciform structure with a 20-fold preference over linear DNA. SARs (Section 1) are AT-rich and have a high potential to form secondary structures. Binding to AT-rich sequences was therefore also analysed (Akhemedov *et al* 1998). Binding to the poly(dA-dT)poly(dT-dA) heteropolymer is 200-fold higher than that to less flexible hetero- and homopolymers. Binding to centromere sequences is also strong. These results are consistent with preferential binding to DNA secondary structures.

Models for SMC Activity

DNA Motor Model

The similarity of SMC proteins to mechanomotor proteins inspired an early model of DNA supercoiling by SMC proteins in which SMCs were novel motor proteins (Peterson 1994). They were proposed to facilitate chromatin loop formation by successive cycles of binding and translocation in a mechanism similar to the racheting of myosin along muscle fibres. This could involve either reeling in or twisting of the chromatin fibre, either of which might require ATP hydrolysis (Hirano 1995a).

Chromatin Clamp Model

A model was suggested in which certain sites on the DNA of eukaryotes were clamped by topoisomerase II, itself bound by SMC proteins (Gasser 1995). In this way, SMC proteins would form cross-bridges between tethered sites, and the chromosome would be reeled in. However the existence of SMCs in non-eukaryotes suggested direct interaction with DNA, rather than

with eukaryotic scaffold proteins (Hirano 1995). The evidence for DNA binding by SMC proteins is consistent with this (Kimura and Hirano 1997, Akhmedov *et al* 1998). A similar structure has been suggested to form at telomeric heterochromatin and in non homologous end-joining (Sections 1 and 4).

A less specific self assembly model was proposed in which multimerised SMCs formed filaments to determine chromosome structure (Hirano 1995a). This was predicted to require ATP hydrolysis.

Wrapping Model

It was suggested that DNA wraps around the SMC condensin complex to introduce positive supercoils and that binding of the complex to the DNA constrains these supercoils (Kimura and Hirano 1997). The negative supercoils introduced concomitantly would then be removed by topoisomerase, which would not be able to access the constrained supercoils. Deproteinisation would reveal the positive supercoils. This is a novel mechanism of supercoiling unlike the mechanisms of either DNA helix tracking proteins or RecA-like helicases, both of which form unconstrained supercoils. The observed preference for secondary structure was suggested to stimulate interaction with long DNA, which is more likely to contain distortions than short DNA. It was suggested that condensin binding might initiate at the high affinity sites of DNA distortion. A possible reason for this function of secondary structure binding was suggested (Akhmedov *et al* 1998): Perhaps a major point of activity is at the completion of DNA replication, when two replication bubbles collide. The entangled sister chromatids may carry distortions. It is proposed that resolution of the sister chromatids by topoisomerase is concomitant with the initiation of SMC binding and the beginning of condensation.

An alterantive function of secondary structure binding might be in recombination (Akhmedov *et al* 1998). In support of this, SMC1/3 was shown to catalyse the formation of joint molecule recombination intermediates. And stem-loops sometimes coincide with recombination sites *in vivo* (Section 3).

As noted above, SMC proteins have been shown to bind flexible AT-rich DNA including centromeric DNA (Akhmedov *et al* 1998). Other sites of DNA secondary structure at which SMC proteins have been suggested to act include sites bound by proteins which induce bends, *eg* HMG proteins, and the base of the nucleosome (Akhmedov *et al* 1998).

Phosphoesterase Proteins

The Phosphoesterase Signature Motif

Three regions of amino acid sequence identity common to all known eukaryotic and bacteriophage protein phosphatases were identified, and predicted to form loop structures (Zhuo *et al* 1994). Part of this consensus was found to be conserved in a range of other phosphoesterases, including phosphodiesterases (Koonin 1994). The split consensus was then found to be weakly conserved throughout diverse phosphoesterases from eubacteria to humans (Zhuo *et al* 1994). A fourth motif was identified (Sharples and Leach 1995), and a fifth (Tsubouchi and Ogawa 1998), as follows;

motif I	lrilxxxDxH
motif II	GDxxD
motif III	GNHD/E
motif IV	IvlxH
motif V	++xGHxH

I (10-50aa) II (16-56aa) III (36-94aa) IV (21-100aa) V

In addition phophoesterases have a basic C-terminal region, the sequence of which is not conserved (Sharples and Leach 1995).

The signature suggests that some aspects of the catalytic mechanism and structure are common to all phosphoesterases, however certain residues are more highly conserved in certain families of phosphoesterases (Zhuo *et al* 1994). By site directed mutagenesis it was found that all the conserved residues tested were required for catalysis, that certain residues were important for metal binding, and others for substrate phosphate binding (Zhuo *et al* 1994).

Families of Phosphoesterases

Phosphoesterases, also called phosphatases, form a superfamily of members with diverse functions. Families include serine/threonine protein phosphatases, tyrosine protein phosphatases, exonucleases (5' phosphodiesterases), endonucleases (5'-3' phosphdiesterases), 2'-3' cyclic phosphodiesterases, diadenosine tetraphosphatases, sphingomyelinases and an RNA lariat (2'-5') phosphodiesterase (Zhuo *et al* 1994, Sharples and Leach 1995). Within the serine/threonine protein phosphatase family are four subfamilies; 1, 2A and 2B, which share a high level of sequence identity, and 2C (Walton and Dixon 1993). Some phosphoesterases form heterodimers with targeting subunits which localise the protein or regulate its acitivity by inhibition (Hubbard and Cohen 1993).

Structural and biochemical analyses of one family may be informative of other families. *S. pombe* Rad32, *S. cerevisiae* Mre11, *E. coli* SbcC, T4 gp47 and T5 gpD12 are members of the exonuclease family of phosphoesterases (Sharples and Leach 1995).

Metal Ion Binding by Phosphoesterases

With the exception of the tyrosine protein phosphatases (Goldberg *et al* 1995), most phosphoesterases require divalent transition metals or Mn^{2+} for activity (Zhuo *et al* 1994). The phosphoesterase of λ (1, 2A, 2B) requires Mn^{2+} or Ni^{2+} for activity, and Zn^{2+}, Cu^{2+} and Hg^{2+} are inhibitory (Zhuo *et al* 1993). It is the smallest member of the serine/threonine family 1, 2A, 2B, lacking most of the C-terminal domain (Cohen and Cohen 1989). It probably binds only one metal ion (Zhuo *et al* 1994), unlike other members which bind two divalent cations (Jenny *et al* 1995, Goldberg *et al* 1995). Mammalian phosphatase 1 is tentatively predicted to co-ordinate Mn^{2+} (Goldberg *et al* 1995), and human calcineurin A (2B) Fe^{2+} and Zn^{2+} (Jenny *et al* 1995). SbcCD has an absolute requirement for Mn^{2+} *in vitro*, with 10% activity in Cu^{2+} (Section 3).

The two catalytic metal ions are co-ordinated close to one another at the catalytic centre of the mammalian phosphatase 1 enzyme (Goldberg *et al* 1995, and below). They are required for substrate interaction: With certain

residues in the active site, they create a positive electrostatic potential which stabilizes the negatively charged phosphate oxygens of the substrate and makes them more susceptible to nucleophilic attack (Beese and Steitz 1991, Goldberg *et al* 1995). It has been predicted that the exact nature of the metal ions therefore might not be significant (Beese and Steitz 1991). The metals have also been suggested to effect a conformational change to a catalytically active form (Zhuo *et al* 1994).

Phosphoesterase Structure

When mammalian phosphatase 1 was used as a model, it was found to have two domains (Goldberg *et al* 1995): The N-terminal domain includes a β–α–β–α–β tertiary structure motif which forms a shallow pocket constituting the centre of the catalytic site. The metal ions are co-ordinated at the C-end of the β-strands in this motif, at the bottom of the pocket. This structure is formed by the conserved residues of the phosphoesterase motif, which was consequently identified as a metallophosphoesterase motif. The C-terminal domain forms an irregular structure which is juxtaposed against the N-terminal domain and which binds substrates and inhibitors. Each metal ion is co-ordinated by five amino acids of the metallophosphoesterase motif, and the two sets of five are organised in different spatial arrays. The active site pocket is positioned at the junction of a Y-shaped surface groove. One arm of the groove is hydrophobic. Another is acidic and is contained in the C-terminal domain. The third, the C-terminal groove, runs from the pocket to the C-terminus. Inhibitors lie along the grooves and contact the catalytic centre at the pocket.

Section 3

The Functions of SbcCD

The Activities of SbcCD and its Prokaryotic Homologues

Prokaryotic Homologues

Bacteriophages

T4

T4 DNA metabolism genes are expressed from a polycistronic cluster (Hsu and Karam 1990). Genes *46* and *47* are expressed with similar polarity. Unterminated messages have been detected from an early promoter upstream of the cluster, a middle mode (MotA dependent) promoter upstream of *47* and an early promoter within the sequence of *47*. In addition there is a terminated message which derives from a middle mode promoter between *47* and *46* (Hsu and Karam 1990). Gp46 and gp47 are essential (Gram and Ruger 1985). They are thought to form a complex which is located in the bacterial membrane and has exonuclease activity. Neither polypeptide is hydrophobic, and it has been suggested that another of the early gene products may confer hydrophobicity as an auxillary factor (Gram and Ruger 1985).

On infection, T4 bacteriophage degrades host chromosomal DNA, presumably to provide nucleotides for phage DNA synthesis. This occurs in two steps (Kutter and Wiberg 1968). T4 endonuclease II makes single-strand nicks in native cytosine-containing host DNA. T4 gp46/47 then degrades these products. This is the major role of gp46/47. Gp46/47 also has exonucleolytic activity on T4 DNA (Prashad and Hosoda 1972). DSBs and single-strand nicks are converted to gaps and then degraded in the absence of T4 DNA polymerase. Active polymerase competes with gp46/47 to prevent degradation. This self cleavage function may explain the requirement of gp46/47 for T4 recombination (Bernstein 1968) and T4 replication (Hosoda *et al* 1971). It has been suggested that the conversion of nicks to gaps generates single-strand regions which are substrates for joint molecule formation and recombination, and that recombinational repair of

T4 DNA in late infection may be responsible for preventing DNA synthesis arrest (Prashad and Hosoda 1972).

Mosig suggested a model of recombination-dependent DNA replication for phage T4 (Luder and Mosig 1982, reviewed in Kreuzer *et al* 1995). Due to the requirement for a primer, replication of the T4 DNA results in duplex with 3' overhangs. These single-strand ends were proposed to invade a bubble which forms transiently within a duplex. A homologous recombination intermediate is generated which primes DNA synthesis on both strands. Consistent with the model, plasmid replication can be initiated by T4 3' overhangs if the plasmid bears homology to these ends, or DSBs in the T4 genome if the plasmid bears homology to sequence flanking the DSBs (Kreuzer *et al* 1995). This mechanism stabilizes DSBs, whether in plasmid or T4 genomic DNA. T4 gp46/47 are required for maximum replication induction and are predicted to have roles in degradation and recombination.

T5

The high identity between T4 gp46/47 and T5 gpD13/D12, and the ability of one to complement the other during superinfection, indicates that D13/D12 may be a similar nuclease (Blinov *et al* 1989).

Bacteria

The activities of bacterial homologues have not been studied. Initial sequencing of the *B. subtilis sbcD* gene identified a single base deletion which was suggested to cause premature termination of translation (Sharples and Lloyd 1993). However complete homologues of both *sbcC* and *sbcD* genes have since been identified in *B. subtilis* (Medina *et al* 1997).

SbcCD is a Nuclease

Circumstantial Evidence That SbcCD is a Nuclease

The closest homologues of SbcCD are nucleases; T4 gp46/47 and T5 gp D13/D12 (Leach *et al* 1992). SbcC and SbcD also share regional sequence homology with RecB and RecC respectively, which together form a nuclease (Naom *et al*). SbcD belongs to the phosphoesterase family (Sharples and Leach 1995).

In addition there is evidence that SbcCD is functionally homologous to RecBC (despite lacking sequence homology, Section 2): λ phage replicate by the sigma mode and use the phage encoded Red protein in phage-specific recombination (Kulkarni and Stahl 1989). Sigma replication is facilitated in a *recBC* host. The product of the λ *gam* gene, γ "Gam" protein, specifically inhibits RecBC (Karu *et al* 1975). Gam is a homodimer and inhibits both the nuclease and ATPase activities of RecBC in an ATP independent manner by reversible binding. It was found that λ phage containing a long palindrome were viable on a wildtype host if they also contained *gam*$^+$ (Kulkarni and Stahl 1989). This suggested that Gam protein inactivated SbcC or SbcD and that these proteins bear some structural similarity to RecBC. Structural similarity is indicative of functional similarity between SbcCD and RecBC nuclease.

Purification of SbcCD

The method of overexpression and purification of SbcCD (Connelly *et al* 1997) is outlined here because it was used to generate the wildtype and mutant protein required for the work described in this thesis. The *sbcCD* operon was inserted into a plasmid, pTrc99A, under the control of the isopropyl-1-thio-β-D-galactopyranoside (IPTG) inducible P_{trc} promoter. The constructed plasmid was introduced into a chromosomal *sbcCD* deletion mutant *E. coli* host and used to overexpress SbcCD. Cells were sonicated in sucrose cell buffer and the lysate retrieved by centrifugation. Nucleic acid was removed by polyethyleneimine to generate Fraction I. The peak of SbcCD eluted in a NaCl gradient from a DEAE sepharose gel filtration column in buffer A (200mM-225mM) was collected as Fraction II.

Ammonium sulphate was added and the resulting pellet resuspended in buffer A/5mM Mn^{2+} to generate Fraction III. The protein eluted from a sephacryl 500 gel filtration column in buffer A/5mM Mn^{2+} and dialysed against buffer A to remove exogenous Mn^{2+} was Fraction IV. This was 98% homogeneous. It was dialysed against buffer P, passed through a hydroxyapatite gel filtration column in buffer P, eluted in a phosphate buffer gradient (130-160mM), dialysed against buffer A and concentrated to generate Fraction V. This was stored at -70°C.

SbcCD is a Double-Strand Exonuclease and a Single-Strand Endonuclease

SbcCD was tested for nuclease activity with a variety of candidate substrates (Connelly and Leach 1996). Circular double-strand pUC19 DNA was not cleaved whether supercoiled or relaxed. The linearised double-strand pUC19 DNA was cleaved in the presence of 1mM ATP and 5mM Mn^{2+} to generate an oligomeric limit product. SbcCD therefore has a double-strand ATP dependent exonuclease activity. Covalent closed circle single-strand M13 DNA was cleaved in the absence of ATP and the presence of 5mM Mn^{2+}. Therefore SbcCD has a single-strand ATP independent endonuclease activity. The SbcD peptide, as distinct from the SbcCD complex, was required for this activity. This suggests that SbcD contains the nucleolytic centre of the enzyme, as predicted from sequence analysis, and SbcC modulates this activity. The requirement for Mn^{2+} ions is consistent with SbcCD belonging to the phosphoesterase family of proteins which requires divalent metal ions for activity. SbcD contains the structural and sequence motifs common to the phosphoesterase family (Connelly and Leach 1996).

Parameters for Exonuclease Activity

The optimum conditions for the double-strand exonuclease activity have been investigated (Connelly *et al* 1997).

Cofactors

Optimal ATP concentration was dependent on the concentration of Mn^{2+}. It was suggested that ATP chelates free Mn^{2+} ions. 30% activity was obtained in ATPγS and 23% in GTP, but none of the other cofactors tested (dATP, ADP, GDP) supported activity. This suggests that SbcC co-ordinates purine NTPs, as expected for an SMC protein, and that NTP binding, but not hydrolysis, is required for cleavage.

The nuclease has an absolute requirement for Mn^{2+}. 10% activity was obtained in 5mM Cu^{2+}, and activity was not stimulated in Co^{2+}, Fe^{2+}, Zn^{2+}, Mg^{2+}, Ni^{2+} nor Ca^{2+}. Combinations of divalent metal ions did not stimulate the activity (John Connelly and Erica de Leau unpublished).

Buffer Conditions

Optimal cleavage was obtained in 50mM NaCl. The activity above 100mM NaCl was less than that at 0mM. The exonuclease was functional between pH7 and pH9.5, and pH7.5 was chosen for assays. Acid was extremely inhibitory and alkali slightly inhibitory. SbcCD aggregates at pH5 (John Connelly personal communication).

Stoichiometry of the SbcCD Complex

Similar estimates for SbcCD molecular weight were obtained using FPLC gel filtration (sensitive to protein morphology) and sedimentation equilibrium analysis (insensitive to morphology) (Connelly *et al* 1997). However electron microscopy pictures suggested that the protein (or a part of it) is not simply globular, but of the Head-Rod-Tail structure characteristic of SMC-like proteins. This is consistent with the Head-Rod-Tail prediction for SMC-like proteins. Sedimentation equilibrium analysis found the SbcCD complex to be 720kDa in the absence of Mn^{2+} and 1210kDa in the presence of Mn^{2+}. SbcC and SbcD co-eluted in Mn^{2+} on analytical HPLC gel filtration but

separated in the absence of Mn^{2+}, indicating that Mn^{2+} is required for complex structure (as well as catalysis). The existence of the 720kDa species indicates that some association takes place in the absence of Mn^{2+}. If this consists of a hexamer of SbcC, the active complex, of M_R 1210kDa, may have a stoichiometry of $SbcC_6SbcD_{12}$. SbcC and SbcD were present in equimolar amounts on HPLC analysis in Mn^{2+}. This has led to the tentative alternative suggestion that the M_R of 1210kDa may arise from a $SbcC_8SbcD_8$ complex (John Connelly unpublished).

The Hairpin Nuclease of SbcCD

The hairpin nuclease activity of SbcCD was identified and studied as described in Chapter 3. It was found not to require DNA ends. Subsequent to this work, the hairpin nuclease has been further investigated (Connelly *et al* 1999). In addition to cleavage 5' of the loop (Chapter 3), a blunt hairpin substrate is cleaved at the 3' end. The activity has 3' to 5' polarity, shown by time course digestion of 5' and 3' end-labelled substrates using denaturing PAGE and TLC. The minimum length requirement of the hairpin nuclease for the hairpin stem is between 8bp and 16bp. The activity appears to be non sequence specific since a range of stem-loops were cleaved.

The best substrate for SbcCD nuclease was a long double-strand DNA. A short blunt-ended duplex (oligonucleotide) competed equally with a short loop-ended duplex (hairpin oligonucleotide). A short single-strand linear DNA was a very poor substrate unless attached to a duplex (to form a single-strand overhang) where it was efficiently cleaved. No evidence was observed for cleavage at this double-strand to single-strand boundary. It was suggested that SbcCD recognises and cleaves the double-strand region adjacent to a single-strand region, but does not require this single-strand to double-strand boundary to be active. A range of blunt and overhang duplexes and a branched substrate were digested to a limit product. Where short substrates were used, the limit product was equal to half the length of the double-strand stem, irrespective of the ends. It was suggested that SbcCD measures the half-length of its substrate (below).

SbcCD Forms DSBs at Hairpins in Replication Forks

DSBs at Replication Forks

DSB Formation

DSBs are Induced at Arrested Replication Forks

DSBs in foreign DNA can result from cleavage of both strands of the duplex by restriction endonucleases (Kuzminov *et al* 1994). However most DSB formation is proposed to result from the susceptibility of arrested replication forks to breakage (Kuzminov 1995b). Breakage results in replication fork disintegration in two ways. "Replication fork collapse" occurs as a result of fork arrest at pre-existing single-strand interruptions in the template. Single-strand interruptions may arise from the replisome skipping over unrepaired lesions in the previous round of replication. "Replication fork breakage" results from the rupture of a fork arrested at intact DNA. The replisome may arrest at intact DNA where there is a protein mediated block to progression (Kuzminov 1995b), under conditions of thymine starvation (Kuzminov 1995b), in strains mutant for replicative helicases (Rep and DnaB, in Michel *et al* 1997), at termination sites (Kuzminov 1995b) or secondary structures (Leach 1994), at sites of template DNA damage *eg* by UV irradiation (Kuzminov 1995b), or in the presence of gyrase inhibitors (Kuzminov 1995b). A protein-mediated block to replisome progression may be the reason why transcription opposing the direction of replication slows down a replication fork and most genes are aligned with replication (Kuzminov 1995b).

Experimental evidence for the involvement of replication arrest in DSB formation has accumulated (Kuzminov 1995b). When replication was inhibited by replication fork arrest (using the various methods listed above), half the DNA at the point of replication was degraded. This was interpreted to mean that a DSB had been introduced at the arrested fork and one strand of the replication fork had consequently been degraded. In *lig* and *polA* mutants, which have a high frequency of single-strand DNA interruptions, the chromosome was observed to fragment (Kuzminov 1995a).

RuvABC Generates DSBs at Arrested Replication Forks

The replication fork progresses slowly in mutants lacking a replicative helicase and there is frequent replication pausing (Michel *et al* 1997). Replicative helicases are proposed to remove DNA-bound proteins which would otherwise cause the fork to pause and DSBs to form. In support of this, an essential homologue of Rep exists in *Staphylococcus aureus* and a functional homologue exists in T4 (Dda helicase) (Michel *et al* 1997). Mutants in Rep are lethal and mutants in DnaB accumulate linear DNA in combination with mutations in *recBC* (Seigneur *et al* 1998). In this instance, DSBs which form at the replication fork cannot be repaired by recombination. Comutation of *ruvABC, ruvA* or *ruvB* rescues these strains, but comutation of *ruvC* alone is lethal (Seigneur *et al* 1998). Linear DNA does not accumulate in mutants lacking a replicative helicase and either RuvABC or RuvC. This indicated that DSB formation occured in the presence of, and was dependent on, the RuvABC proteins.

The authors proposed that RuvABC-dependent DSB formation is part of a mechanism for RuvABC/RecBCD-mediated rescue of broken replication forks: A Holliday Junction is formed at an arrested replication fork by annealing of the two newly synthesised strands. This junction may be stabilized by RuvAB binding (or form spontaneously). The replication fork may be reconstituted either by RuvC resolution and RecABCD-mediated homologous recombination, RecABCD-mediated recombination or RecBCD-mediated degradation.

DSB Processing and Repair

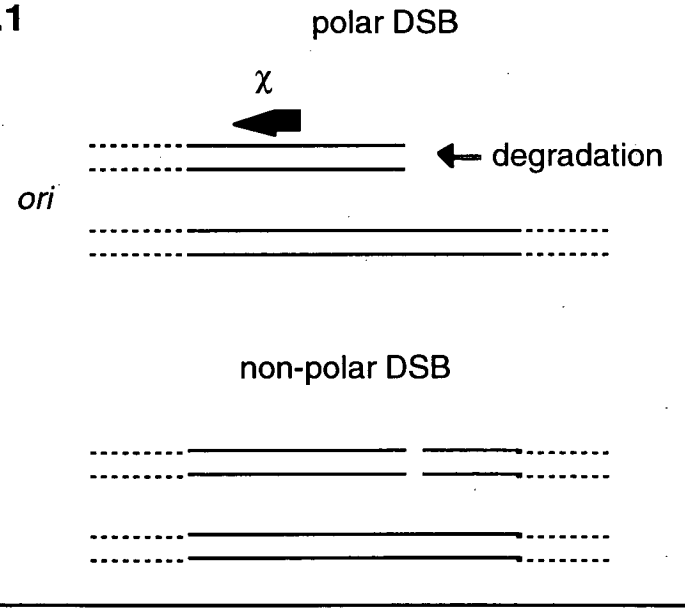
DSBs and the Orientation of Chi Sequences

Homologous recombination is initiated by RecBCD exonuclease V which enters DNA via a blunt or nearly blunt double-strand end and degrades duplex DNA (Section 1). A DSB can be viewed as two blunt double-strand ends. DSB repair in prokaryotes is believed to occur by homologous recombination (Section 1).

χ sequences in *E. coli* are approximately 7-fold more frequent than random, presumably because a large number are required to prevent genome degradation by RecBCD (Kuzminov *et al* 1994). It was observed that 90% χ sequences in *E. coli* are oriented towards *oriC, ie* the opposite orientation to replication, such that they would protect against degradation of DNA in the terminus to origin direction. This suggests that most DNA damage repaired by RecBCD recombination takes the form of a "polar" DSB, rather than the "non-polar" DSB which arises from ionizing radiation damage, figure 1.3.1. Such a structure could arise from the arrest of a replication fork, which implies that the primary role of RecBCD-χ recombination is the repair of broken replication forks. The extremely low viability of *recBC* mutant would therefore result from defective repair of DSBs. Homologous recombination was postulated to reconstitute the replication fork at a stage prior to the interruption.

Consistent with the evolution of the RecBCD system primarily for repair of broken replication forks, it was observed that the repair of broken forks required RecA and RecBCD (Kuzminov 1995b, Michel *et al* 1997). *rec*⁻ strains exhibiting a high frequency of replication fork arrest were inviable (Kuzminov 1995a). Hyper-recombination was detected in the region of a replication block, and amplification and deletion frequencies at inverted repeats in the vicinity of a replication block were increased, as expected if the block were repaired by recombination (Kuzminov 1995b). Replication fork

Figure 1.3.1



arrest increased the rate of localised homologous recombination, illegitimate recombination and mutagenesis (Bierne and Michel 1994). In addition, DSBs in a haploid cell could only be repaired in regions of the chromosome which had been replicated, and fast replicating strains (on concentrated nutrient media) were less sensitive to γ-irradiation than slower replicating strains (on less concentrated media) (Kuzminov 1995a).

iSDR

DSBs can also be repaired by iSDR. This is a pathway of recombination dependent DNA replication similar to that originally proposed for T4 (above). Initiation of DNA replication by the iSDR system occurs in *E. coli* independently of concomitant protein synthesis and is induced by thymine starvation or treatment with DNA synthesis inhibitors (Kogoma and Lark 1975). Upon induction of the SOS response, a DSB is introduced near *oriM* (Asai and Kogoma 1994). 3′ ended single-strand DNA is generated from the DSB by the activity of either RecBC(D) helicase (RecBCD pathway) or Exo VIII (RecE pathway) (Asai and Kogoma 1994), or presumably RecJ/RecQ (RecF pathway). The 3′ single-strand tail becomes coated with RecA and enters a homologous duplex, displacing one strand to form a D-loop. The D-loop then acts as a site of duplex opening for replication initiation.

DSBs at Palindromes

In summary, DSBs mostly arise at replication forks as a result of fork arrest, and are repaired by various mechanisms of recombination. Most notable of these is RecBCD homologous recombination which is likely to have evolved primarily for the repair of DSBs at replication forks.

Palindromic sequences are one cause of replication fork arrest (Weaver and DePamphlis 1984). It is likely that the hairpin structure adopted by these sequences blocks replication. Palindromes are sites of DSB formation and hotspots for recombination (Warren and Green 1985). DSBs might be induced at these sites by the activity of RuvABC. However cleavage of the secondary structures formed at these sites would increase DSB formation and recombination. All of the closest homologues of SbcCD are involved in the formation and/or processing of DSBs (above, Section 4).

A Model for the Regulation of Palindrome Maintenance by SbcCD Hairpin Nuclease

Palindrome Maintenance is Regulated

Long palindromes would consistently fold back into hairpins were they not to confer inviability (Allers and Leach 1995). They would threaten genome stability by promoting mutagenic illegitimate recombination by replication slippage (Section 1) as demonstrated in *sbcCD* mutant strains (Section 2). As the primary determinant of palindrome inviability, SbcCD may be involved in preventing this mutagenesis (below). However this is an insubstantiated hypothesis.

Other palindromes up to 40bp in length are maintained in *E. coli* (Section 1) and have important functions resulting from their hairpin forming ability. Many transcripts include short hairpin sequences (reviewed Higgins 1986). In the 3' untranslated region these act as barriers to 3' to 5' exonuclease digestion of the mRNA, *eg* by RNase III, which stabilize the message and increase expression of upstream genes. Rho independent transcriptional terminators end with a GC-rich stem-loop followed by a U-rich sequence which causes pausing and dissociation of RNA polymerase. Some promoters include short hairpins which function as protein binding sites or as regulatory elements which respond to metabolic conditions, *eg* tryptophan attenuation. In addition, the terminus region of the *E. coli* chromosome contains hairpin sequences which are postulated to cause DNA polymerase to arrest when it encounters them from one side (Kuzminov 1995b). Several bacteriophage replication origins form hairpin structures for activity, *eg* phage M13, G4, ϕ29.

Prokaryotic genomes include two types of repetitive DNA both of which has a distinct imperfect palindromic sequence consensus. The PU (palindromic unit, Gilson *et al* 1984) or REP (repetitive extragenic palindromic, Stern *et al* 1984) sequence is 35 nucleotides long and exists in *E. coli* and *S. typhimurium* only. The IRU (intergenic repeat unit, Sharples and Lloyd 1990) or ERIC (Enterobacterial Repetitive Intergenic Consensus, Hulton *et al* 1990) sequence is 126 nucleotides long, but shows better conservation in the central 40 nucleotides, and exists in a range of prokaryotic genomes. REPs and IRUs

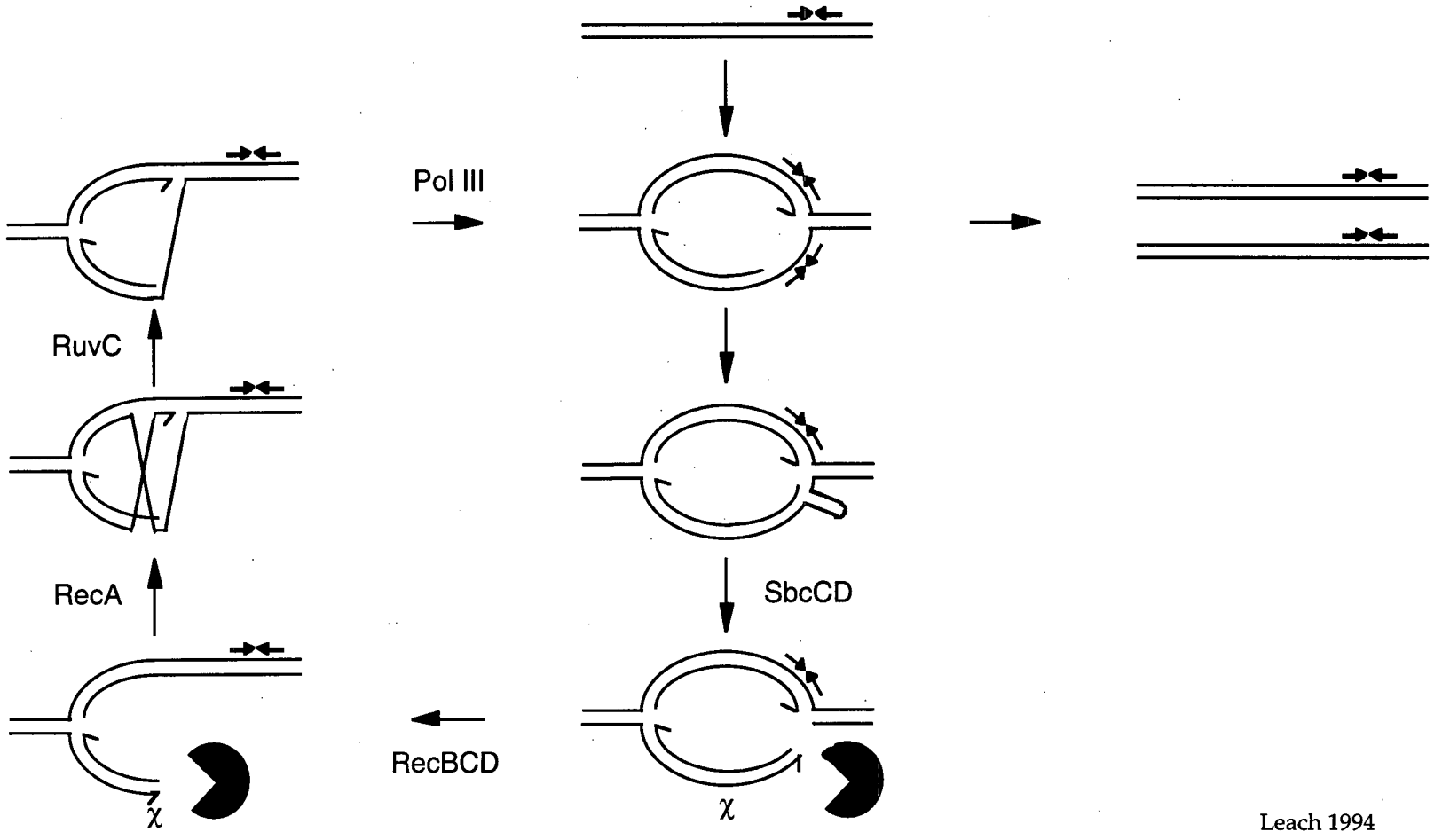
are present in the intergenic regions of operons in either orientation with respect to transcription, and in different locations in different species. Their secondary structure is likely to be important, and mutations in one arm of an IRU give rise to spontaneous compensatory changes in the other arm (Hulton *et al* 1991). Some function in regulating gene expression, but these roles do not explain the sequence conservation and are therefore likely to be secondary. They may constitute a form of "selfish" DNA which has been recruited for certain functions in certain locations. REPs bind DNA gyrase and DNA polymerase I and have been suggested to constitute the apeces of the bacterial chromosome (nucleoid) scaffold. Multiple REPs often exist in tandem arrays, BIMEs (bacterial interspersed mosaic elements, reviewed Espeli and Boccard 1997). BIME-1 arrays include converging Y and Z1 REPs, and BIME-2 arrays consist of alternating Y and Z2 REPs. DNA gyrase binds and cleaves BIMEs-2 at every second Y REP if it diverges from its Z2 partner (Espeli and Boccard 1997). The authors proposed that these are sites at which DNA gyrase removes positive supercoils introduced by DNA or RNA polymerase. They are often at the 3' ends of genes and might prevent the transference of supercoils into adjacent genes by acting as swivels. This is consistent with a role in organising chromosomal domains in the scaffold, analogous to SARs in eukaryotes.

It is therefore important that palindrome maintenance is regulated. Beneficial palindromes must be maintained and detrimental palindromes must be removed.

The Model for Regulation of Palindrome Maintenance

A model for the regulation of palindrome maintenance was proposed (Leach 1994, Connelly *et al* 1998), figure 1.3.2. The model postulates that a palindromic sequence could form a hairpin structure at the replication fork. This stalls the replication fork. SbcCD recognises and cleaves the hairpin, removing the loop and forming a DSB. The DSB is repaired by homologous recombination which collapses the replication fork to provide another opportunity for unhindered replication of the region. This is consistent with the model for DSB repair at replication forks by homologous recombination (above).

Figure 1.3.2



Leach 1994

It is the frequency with which a palindrome forms a hairpin secondary structure that determines whether it will be maintained or confer inviability on the cell. If the palindrome is long (longer than 150bp) it will always form a hairpin in the replication fork, the replication fork will always collapse and the strain will be inviable. If the palindrome is short (less than 40bp) it will not form a hairpin and will be replicated unhindered. Strains carrying medium sized palindromes may or may not survive. If a hairpin forms during replication, the fork will collapse and be reconstituted at an earlier stage prior to replication of the palindrome, thereby providing another opportunity for palindrome replication. This is consistent with the observation that SbcCD cleaves relatively short hairpins *in vitro* (above) which suggests that the viability conferred by palindromes shorter than approximately 150bp is due to secondary structure stability rather than nuclease preference.

This is a model of arm directed repair, distinct from mechanisms of strand directed repair, *eg* mismatch repair. Strand directed repair may operate to remove regions of sequence amplification up to 3 or 4 nucleotides in length (Section 1). These mechanisms are insufficient for the repair of palindromes since palindromes are longer (and cause DNA polymerase to stall).

Requirements of the Model

The model depends on the other arm of the replication fork remaining free of secondary structure to act as a repository of correct genetic information. This is likely becuase the template strand of the lagging arm forms secondary structures more frequently than the other three strands (Trinh and Sinden 1991).

It also depends on the SbcCD hairpin nuclease not requiring DNA ends to recognise and cleave its hairpin substrate. This has been demonstrated in Chapter 3.

Evidence for the Model

Two predictions can be made from the model: RecBC is required for the repair of DSBs formed at a palindrome, and the palindrome template sequence is preserved in a cell, although it may be selectively lost from a population depending on its stability as a hairpin. These were tested using a 240bp palindrome which is sufficiently short to allow wildtype *E. coli* to be viable (Leach *et al* 1997). This palindrome is expected to form a hairpin at replication commonly, but not always. A λ phage containing the palindrome lysogenised *recA*, *recB* and *recC* hosts at a 10^4- to 10^5-fold lower frequency than a wildtype host, indicating that the palindrome exhibited some inviability in the *rec*$^-$ host. The inviability resulted from a requirement for Rec products for repair after lysogeny of the *rec*$^-$ host. When the palindrome containing phage were induced, palindrome containing progeny were recovered with 40-fold lower frequency from *sbcC/D*$^+$ than from *sbcC/D*$^-$ but the survivors retained the palindrome. When the prophage was amplified by PCR, a very low level of palindrome deletion was observed, indicating that the palindrome sequence is retained as predicted by the model. If homologous recombination occurs at the palindome it would be expected that a dilysogen (consisting of two tandemly repeated prophages) would frequently undergo recombination mediated deletion to a monolysogen. Consistent with this, all dilysogens were converted to monolysogens in *rec*$^+$*sbcC/D*$^+$ and none in *rec*$^+$*sbcC/D*$^-$.

The Significance of the Model for Regulation of Palindrome Maintenance

The Antimutator Effect of SbcCD

It can be hypothesised that the pathway has two antimutator functions. It may remove mutagenic secondary structures, and it may prevent genetic instability at sequences which only occassionally form secondary structures. However this has not been substantiated experimentally.

The "Anti-Recombination" Effect of SbcCD

The model suggests that *rec*⁻ cells have low viability partly because the DSBs formed by SbcCD at a replication fork are not repaired. SbcCD was initially believed to have an "anti-recombination" effect since *sbcB(C/D)* mutants rescued *recBC* (RecF pathway Sections 1 and 2). It may have such an effect in the absence of RecBC if it degrades recombinogenic ends. However the model suggests that SbcCD facilitates recombination in the presence of RecBC by allowing the entry of recombination enzymes at the DSB it forms (Connelly and Leach 1996, Section 1).

The Anti-Rolling Circle Replication Effect of SbcCD

The pathway functions to restore DNA replication when stalled at a secondary structure. However SbcCD has an inhibitory effect on rolling circle replication; RecBC acts alone or in conjunction with SbcCD on certain plasmids to maintain the theta mode of plasmid replication (Section 2). A mechanism for this effect has been suggested (Cohen and Clark 1986). This suggests that default replication occurs by the sigma mode, and theta replication only results when the rolling circle tail is degraded. A rolling circle structure bears similarities with a replication fork. On the basis of the model for regulation of palindrome maintenance, it can be envisaged that SbcCD may act on a rolling circle similarly to a replication fork. SbcCD would act at secondary structures to allow reconstitution of the intact plasmid by homologous recombination.

Deprotection of DSBs by SbcCD and its Homologues

The loop of a hairpin provides a block to exonuclease degradation of the stem. The proposed action of SbcCD at the hairpin can be viewed as end deprotection (Connelly *et al* 1998). Deprotection exposes the stem to recombination enzymes and recombination ensues. This may be a feature common to the functions of the homologues of SbcCD: *m*-AMSA is an inhibitor of topoisomerase because it induces the formation of covalent topoisomerase-DNA cleavage complexes. It can be used in humans for chemotherapy as an anti-tumour agent, or in bacteria as an antibiotic. T4 gp46/47 provides resistance to *m*-AMSA, and has been suggested to do so by

recombinational repair of the covalent complexes (Woodworth and Kreuzer 1996, Neece *et al* 1996). This would imply that gp46/47 acts to deprotect the bound ends to expose them for recombination (Connelly *et al* 1998). There are additional similar mechanisms implied in eukaryotes (Section 4).

SbcC as an SMC-Like Protein

SMC proteins preferentially bind secondary structures and a subset has roles in genetic recombination (Section 2). The proposed binding of SbcCD to hairpin structures at replication forks to facilitate homologous recombination is consistent with the classification of SbcC as a SMC-like protein. It may provide a rationale for the function of certain other SMCs in more complex organisms.

Effect of SbcCD on the RecF Pathway

The gene originally isolated which appeared to restore *sbcC*$^+$ to *recBCsbcBC* mutants, as scored by recombination and growth characteristics (Lloyd and Buckman 1985), was subsequently found to map approximately 3kb downstream of *sbcCD* (Naom *et al* 1989, Section 2). This gene, named *rdgC* for recombination dependent growth, expresses a protein of 33kDa with a coiled-coil domain, and its function has been studied (Ryder *et al* 1996). As initially observed, multiple copies of RdgC (thought at the time to be SbcC) are detrimental to the growth of *recBCsbcBC*. *ΔrdgC* alone has a wildtype phenotype. The combination of *recBCsbcBC* with *ΔrdgC* has approximately wildtype viability (growth rate and conjugational recombination) although is slightly MC^S. However the combination of *recBCsbcB* with *ΔrdgC* has lower viability than *recBCsbcB*, indicating that *ΔrdgC* makes these cells exceptionally vulnerable to SbcCD protein. The combination of *recBCsbcBCΔrdgC* with *recA*, *recF* or *ruvC* reduced viability (growth rate, conjugational recombination and DNA repair). This indicated that RdgC may function in the RecF pathway.

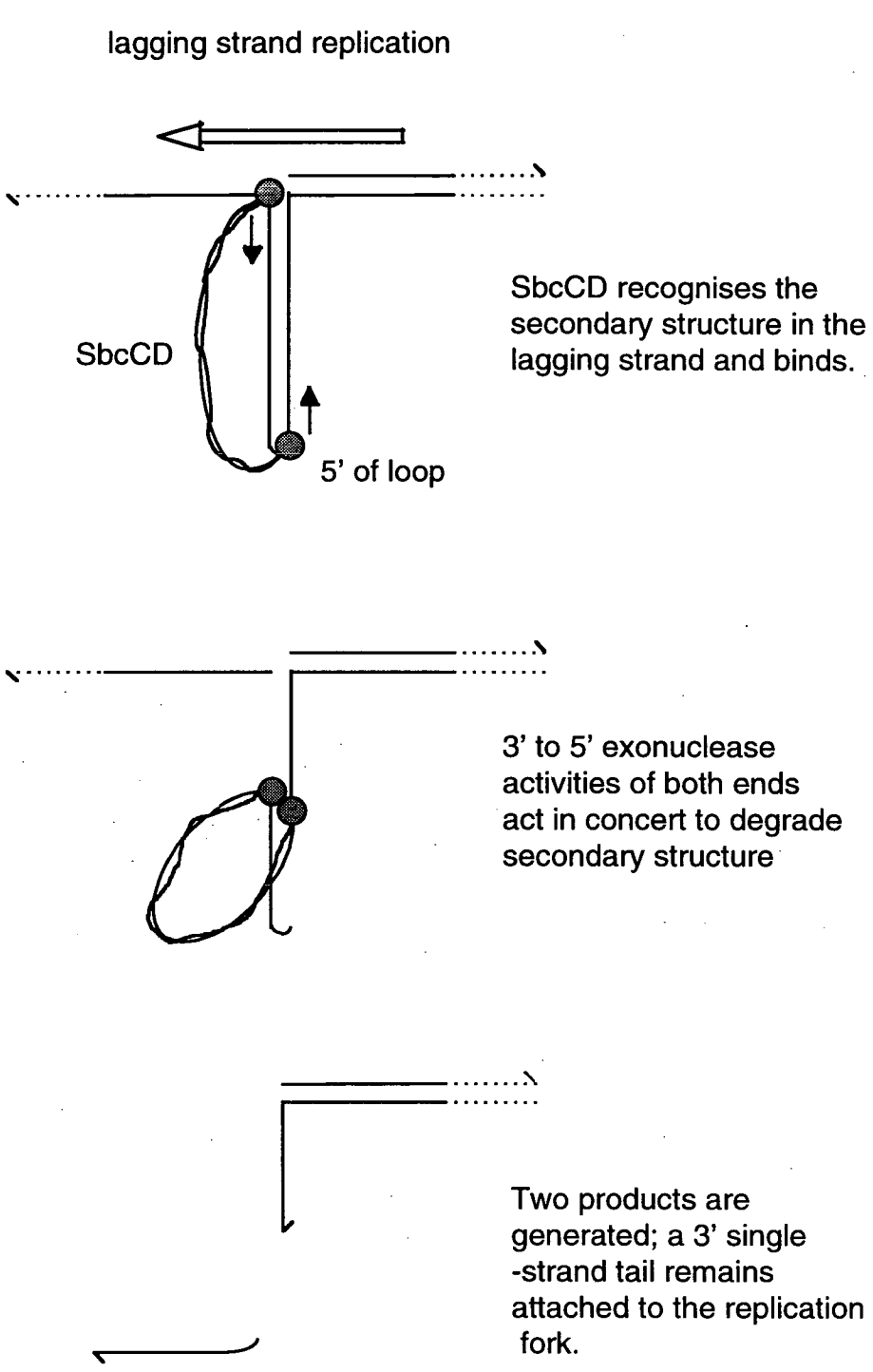
SbcCD increases the frequency of replication fork collapse by forming DSBs at replication forks, and may degrade the displaced arm in the absence of RecBC (above). In a RecF activated strain, this degradation would reduce the chances of fork reconstitution by recombination, and the broken arm of the

fork would remain. Initiation would then result in the synthesis of inviable, branched chromosomes. The authors (Ryder *et al* 1996) proposed that RdgC functions to remove the displaced arm to allow successful replication. They proposed that RdgC is (or activates) an exonuclease. The exonuclease would presumably compete with SbcCD for the broken arm. This may explain why *recBCsbcBC* strains are so much more viable than *recBCsbcB*.

A Model for Substrate Half-Length Measurement by SbcCD

The "Head" and "Tail" regions of SbcCD are approximately 3.4nm in diameter which could cover 10 nucleotides of DNA (Connelly *et al* 1999). These resemble the terminal regions of SMC proteins, shown to act in DNA binding (Akhmedov *et al* 1998). Substrates containing a duplex region are digested to a limit product equal in size to half the duplex length, indicating that SbcCD measures substrate half-length (above). This is proposed to occur at a replication fork as the result of the concerted action of the two DNA-binding termini of the complex (Connelly *et al* 1999), figure 1.3.3. The Head and Tail regions initially bind the 3' ends of the duplex and then processively degrade, in the 3' to 5' direction, towards the centre of the substrate. When the Head and Tail meet, the protein dissociates from the substrate. This leaves two 3' overhangs; one is a disconnected fragment, and the other is attached to the replication fork. The latter could provide a substrate for homologous recombination, as implicated in the models for DSB repair at replication forks and for the regulation of palindrome maintenance. This is a possible solution to the exonuclease polarity conundrum of the eukaryotic homologues (Section 4). Alternatively it could provide a substrate for non homologous end-joining. The authors suggest that the Head of SbcCD may occasionally bind a different substrate to the Tail. This is also predicted to result in the generation of recombinogenic ends.

Figure 1.3.3



Connelly *et al* 1999

Other Possible Functions of SbcCD

SbcC and SbcD are Expressed at Low Level

SbcC and SbcD are both expressed at low level (Section 2). T4 gp46 and gp47 are also expressed at low level from middle mode and early mode promoters in tandem, which are thought to mediate transcriptional regulation (Gram and Ruger 1985). The significance of this is unknown. It may be that a low level of SbcCD expression is necessary for the regulation of palindrome maintenance. Perhaps the proportion of secondary structures which escapes cleavage at a replication fork is critical, or expression is cell-cycle regulated.

Plasmid Maintenance

pBR322 was lost rapidly in a *recBCsbcB(C/D)* strain but slower in the single mutant *sbcC/D* (Chalker *et al* 1988). This effect may be related to replication mode control. Consistent with this, the instability of multicopy plasmids in *recBCsbcBC* strains has been attributed to rolling circle replication from collapsed replication forks which generates multimers which, in turn, interfere with plasmid segregation at cell division (Ryder *et al* 1996).

Certain phage use a secondary structure to prime replication (above). Activity of SbcCD at these sites might be expected to provide a level of immunity to these phage.

Trinucleotide Repeats

A CTG trinucleotide repeat oligonucleotide with a GC clamp and no spacer region, which forms a pseudo-hairpin with G≡C and T=T base pairs, was designed (Connelly *et al* 1999). It was digested identically to the other hairpin substrates tested, indicating that SbcCD potentially may cleave repeat expansions. Consistent with this, CTG amplification in the bimodal *E. coli* system (Section 1) requires loss of SbcC (Sarkar *et al* 1998). However mutation of *sbcC* does not affect trinucleotide repeat expansion in mismatch proficient cells (Schmidt *et al* 1999, Chapter 6).

Hairpins as Intermediates in DNA Hydrolysis

DSBs commonly form by DNA hydrolysis. It has been proposed that, in reactions where hydrolysis is followed by transesterification in obligatory succession, a hairpin may be the common intermediate (Kennedy *et al* 1998). The cited example of this is in Tn*10*, which transposes by a conservative (non-replicative) mechanism proposed to act in four steps. The 3' ends of the transposon are nicked to leave 3' hydroxyl termini in the transferred strands (hydrolysis). The 3' hydroxyl termini attack the 5' ends in the non-transferred strands, cleave them forming a DSB, and join onto the scissile phosphate to generate a hairpin structure and release the adjacent donor duplex DNA (transesterification). The hairpin is then attacked at the linkage, *ie* the loop, to regenerate the 3' hydroxyl terminus (hydrolysis). And the transposon joins onto the target duplex by strand transfer (transesterification). The model is supported by evidence for a transient hairpin intermediate. It was proposed to help explain how a single transposase monomer (carrying a single active site) catalyses all the chemical events which occur at one end of the transposon.

Similar mechanisms are suggested to occur in certain other prokaryotic transposons. These hairpins are resolved by the transposase monomer. However the authors suggest that the mechanism may be conserved throughout a wide range of hydrolysis-transesterification reactions. If this is so, other enzymes will be involved in other reactions, perhaps including SbcCD. Speculative evidence of this is offered in Section 4.

Section 4

The Functions of Rad50/Mre11

Identification and Analysis of Rad50/Mre11

Biochemical Characterisation of Rad50 and Mre11

Biochemical Nature

Eukaryotic Rad50 shares homology with SbcC, and Mre11 with SbcD (Section 2). Rad50 is a member of the outlying group of SMC proteins, and Mre11 a member of the phosphodiesterase proteins. Homologues in *S. cerevisiae*, mouse and human are here referred to as *Sc*, m, and h respectively.

Homologues in Yeast

*Sc*Rad50 is encoded by an open reading frame of 1312 amino acids and the product is 153kDa in size (Alani *et al* 1989). The protein is predominantly hydrophilic, the longest hydrophobic region being a 35 amino acid stretch near the C-terminus. The N-terminus contains a short glycine-rich sequence. This has the potential to form a flexible loop which has been suggested to allow protein conformational change. This loop is flanked by a hydrophobic β-strand and an α-helix, and together these three motifs are characteristic of a purine-binding site, particularly ATP-binding. The region also contains candidate sites for the co-ordination of divalent metal ions. The central part of the protein contains two heptad repeat stretches, I and II, each approximately 250 amino acids long, separated by a spacer of 322 amino acids. These are predicted to form coiled-coils. Heptad II bears significant homology to the S-2 coiled-coil domain of rabbit myosin at a level above that of other coiled-coils. Heptad I has been predicted to form a rod of 364Å and heptad II a rod of 385Å if fully extended.

ScMRE11 (meiotic recombination) was identified in a screen for mutants defective in meiotic recombination (Ajimura *et al* 1992). The open reading frame encodes a protein of 643 amino acids with a predicted molecular weight of 72.4kDa (Johzuka and Ogawa 1995). The protein has two putative 5' promoters and two putative 3' polyadenylation signals. It has a highly acidic C-terminal region which contains a heptad repeat of aspartate. This

has been suggested to be important for protein-protein interactions (below), similarly to the acidic regions of *S. cerevisiae* transcriptional activator proteins.

The Mre11 homologue in *S. pombe* is *RAD32* (Tavassoli *et al* 1995).

Homologues in Higher Eukaryotes

mRad50 was initially identified as a phosphoprotein which cross-reacts with monoclonal antibodies raised against the p53 tumour suppressor (Kim *et al* 1996). Its epitopic homology to p53 may reflect functional homology if it describes a functional domain. The reactive monoclonal antibodies recognise the C-terminus of p53. This region of p53 is required for oligomerisation, binding to single-strand DNA and RNA, and double-strand DNA containing insertion and deletion mismatches.

The mRad50 ORF is of the same length as that of yeast *Sc*Rad50 and contains a 3' untranslated region and a poly(A) addition site (Kim *et al* 1996). It encodes a protein of predicted size 153kDa. The observed size is approximately 180kDa, the difference probably being due to addition of carbohydrate moieties at some of the seven potential glycosylation sites. mRad50 bears homology to *Sc*Rad50 in three colinear regions (amino acids 1 to 81, 129 to 192 and 1117 to 1271). These specify an N-terminal Walker A NTP-binding motif, a C-terminal Walker B NTP-binding motif, and an intervening region of heptad repeats predicted to form a coiled-coil. mRad50 also contains leucine zipper motifs, which may facilitate protein-protein interactions. mRad50 expressed in yeast partially rescues a *Sc*Rad50 mutant, indicating that these are functional homologues. The incompleteness of the rescue could be due to low level expression or to imperfect interaction with other proteins in the multisubunit complex.

hRad50 was isolated on the basis of its genetic homology to *Sc*Rad50 (Dolganov *et al* 1996). It possesses three potential nuclear localisation signals, and its identity as a nuclear protein was confirmed by immunofluorescence analysis. Like SbcC and *Sc*Rad50, hRad50 has three conserved motifs. A homologue in *C. elegans* has also been identified.

hMre11 was isolated using the two-hybrid system to screen for proteins which interact with DNA ligase I (Petrini *et al* 1995). The ORF encodes a polypeptide predicted to be 708 amino acids long and 81kDa in size. Another group found hMre11 to be slightly shorter than this, 680 amino acids long, with a sequence more closely matching that of mMre11 (Paull and Gellert 1998). hMre11 exhibits extensive homology to *Sc*Mre11, especially in the N-terminal half (Petrini *et al* 1995). mMre11 was isolated on the basis of its homology to hMre11 (Xiao and Weaver 1997). It is approximately 80kDa in size. Like *Sc*Mre11 it possesses an acidic C-terminal tail.

Biochemical Activities of the Proteins

Multimerisation

In high (200mM) NaCl, *Sc*Rad50 homogeneously formed an asymmetric homodimer with a molecular weight of 282kDa, as predicted from sedimentation equilibrium analysis. In low (20mM) salt, it formed a heterogeneous mixture of species, suggested to be less asymmetric dimers and higher order multimers. Phosphorylated smooth muscle myosin also has two conformations under different NaCl concentrations; the transition occurs as a result of bending in the S-2 coiled-coil domain and is biologically significant. It was suggested that the conformational change in *Sc*Rad50 is likewise the result of bending in a coiled-coil region.

Two mutants were generated in *ScMRE11* (Furuse *et al* 1998). *mre11D16A*, contains Ala in place of the N-terminal Asp 16. This residue is important in other phosphoesterases; for metal ion binding in calcineurin A and for K_{cat} in λ phosphoprotein phosphatase. *mre11ΔC49* lacks the 49 C-terminal residues which represent the cluster of charged amino acids absent in SbcD but present in eukaryotic homologues. Both purified *Sc*Mre11 mutant proteins were recovered with molecular weights slightly greater than twice those expected for monomers. This indicates that the domain required for multimerisation is not present in the termini.

DNA Binding Activity

*Sc*Rad50 was purified and its biochemical activities studied (Raymond and Kleckner 1993a). Double-strand DNA binding was demonstrated in the presence of ATP, and also in dATP, ddATP and ATPγS, but not in the absence of a nucleotide cofactor nor in the presence of any other nucleotide cofactor. *Sc*Rad50 bound double-strand linear and circle DNA in ATP. It also bound single-strand circle DNA, but the nucleotide cofactor requirements for this were not examined. The tested substrates had different sequences, indicating that DNA sequence is not important for substrate recognition. ATP was bound with relatively low affinity, and ATPase activity upon DNA binding was negligable.

*Sc*Mre11 has DNA binding activity (Furuse *et al* 1998). It binds linear pUC118 and supercoiled double-strand DNA (and therefore doesn't require ends). It binds double-strand oligonucleotide DNA better than single-strand, a substrate with a 3' overhang better than one with a 5' overhang, and a substrate with cohesive ends better than one with non cohesive ends. The latter suggests DNA annealing and concatemer formation may facilitate binding. Binding is weak in the absence of metal, and occurs in the presence of divalent transition metals or Mg^{2+}. Wildtype binding activity is obtained with *mre11D16A*, but only very weak binding is obtained with *mre11ΔC49*. This indicates that DNA binding activity is confered by the C-terminus, including the acidic C-terminal tail, which is lacking in SbcD.

*Sc*Mre11 has two DNA binding sites, called A and B (Usui *et al* 1998). Both are in the C-terminal half of the protein, and B is at the C-terminus. A is required for meiotic DSB processing and B is required for DSB formation (below).

DNA substrate binding by hMre11 was studied by competition assays (Paull and Gellert 1998). The 3' overhang and hairpin substrates were bound more stably than the preferred substrates of the double-strand exonuclease (below).

Nuclease Activities

The nuclease activity of *Sc*Mre11 was also studied (Furuse *et al* 1998). The nuclease is active on linear double-strand plasmid and single-strand circle M13 DNA. Supercoiled and relaxed closed circular double-strand DNA are not digested indicating that digestion, unlike binding, requires ends. Substrates with blunt ends and 5' overhangs are cleaved at non discrete sites, but those with 3' overhangs are cleaved only very inefficiently. This is suggestive of a 3' to 5' exonuclease activity *in vitro*. The nuclease is active in Mn^{2+} only and is ATP independent. Wildtype nuclease activity is obtained with *mre11ΔC49* but none with *mre11D16A*. This indicates that the nuclease activity is confered by the N-terminus.

Four different nuclease activities of *Sc*Mre11 in Mn^{2+} ions have been reported (Usui *et al* 1998). It has a single-strand endonuclease (Usui *et al*1998, Trujillo *et al* 1998, Moreau *et al* 1999), a slow double-strand endonuclease activity (Trujillo *et al* 1998), a 3' to 5' single-strand exonuclease (Usui *et al*1998), and a 3' to 5' double-strand exonuclease (Usui *et al*1998, Trujillo *et al* 1998). The latter acts on a blunt or 3' recessed end (Usui *et al*1998).

hMre11 has exonuclease activity in the presence of Mn^{2+}, but not Mg^{2+} or Ca^{2+} (Paull and Gellert 1998). This exonuclease degrades single-strand, double-strand, branched and hairpin DNA: It releases single nucleotides rather than oligonucleotide products. It degrades branched DNA from the tip rather than at the fork (like Rad1/Rad10). Single-strand DNA may be cleaved via an intramolecular base paired secondary structure. A dumbell substrate is cleaved at sites within the loop and on the 3' side of the loop, but less efficiently than a linear substrate.

*Sc*Mre11 therefore has two functional domains which can act independently. In order for the nuclease domain to function independently of the DNA binding domain, it must contain its own separate binding determinants, although these may permit only weak binding.

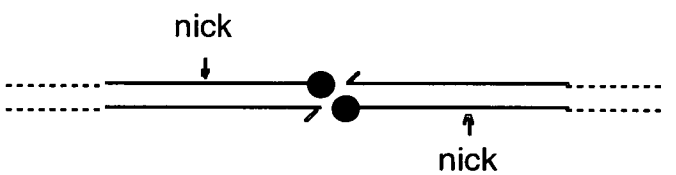
The Exonuclease Conundrum

The Mre11 exonuclease activity is paradoxical in three ways (reviewed Haber 1998): It requires no high energy cofactor or ATP, unlike the *Sc*Rad50/*Sc*Mre11 complex. It is stimulated only 3- or 4-fold by hRad50, which is not sufficient to account for the null phenotype of *rad50* (below). And it is specific to 3' ends, binds 3' ends stably and acts with 3' to 5' polarity. Yeast DSB resection occurs in a 5' to 3' direction (below), and since *Sc*Rad50/Mre11 is a nuclease involved in this step, it might have been expected to have 5' to 3' polarity.

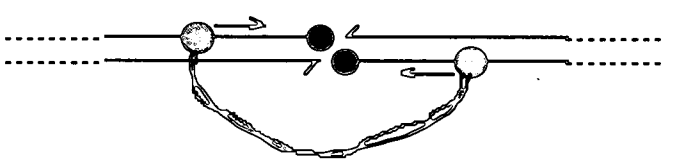
One possible solution to the exonuclease polarity conundrum cites a hairpin as an intermediate in DSB resection. Two 3' to 5' exonuclease domains acting in concert, as suggested for SbcCD (Section 3), could result in 5' to 3' resection, figure 1.4.1.

Figure 1.4.1

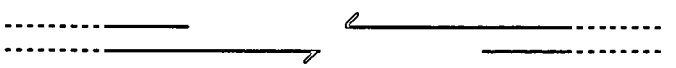
Spo11 binds the 5' end of the DSB and a nick is introduced (independently) on the same strand



SbcCD degrades the nicked strand in a 3' to 5' direction



A 5' to 3' resected DSB is generated



All three paradoxes can be reconciled if the activity of the exonuclease is significantly modified *in vivo*. This may occur as a result of binding of the other components of the active complex, in the context of chromatin, or by the presence of a cofactor which is missing *in vitro*. A cofactor requirement would be consistent with the requirement for Mn^{2+} *in vitro*, characteristic of a nuclease which requires a different divalent cation and an additional cofactor *in vitro*. A model for the activity of hMre11 in chromatin was proposed in which hMre11 binds a region of chromatin which will become a DSB. It may recognise or induce secondary structure in this region of chromatin. After DSB formation, hMre11 may bind the 3' end. Then it might recruit additional factors necessary for 5' to 3' resection. Perhaps it is relevant that the 3' end of a DSB can be extended prior to homologous recombination (Sugawara and Haber 1992).

The Acidic C-terminal Tail of Mre11

The acidic tail which is present at the extreme C-terminus of eukaryotic Mre11 but absent in SbcD is of special interest in this thesis. It is required for DNA binding (above). It is not essential for Mre11 function in yeast and can be compensated for by overexpression of Mre11 (Chamankhah and Xiao 1998). It has been proposed as a protein-protein interaction site (above) and binds three proteins identical in size to meiosis-specific proteins Rec102, Rec104 and Rec114 (Usui *et al* 1998). This may suggest that these proteins aid DNA binding, as will be discussed in Chapter 5.

The Rad50/Mre11 Complex

Rad50/Mre11 Interaction

The two-hybrid system was used to investigate the protein-protein interactions in which *Sc*Mre11 participates (Johzuka and Ogawa 1995). Using the full length *Sc*Rad50 or *Sc*Mre11 protein, the LexA DNA-binding domain and the Gal4 activation domain, LexA-Rad50 and Gal4-Mre11 fusions were constructed and tested for the ability to induce *lacZ* expression *in vivo*. Expression was induced, indicating that *Sc*Rad50 and *Sc*Mre11 interact physically *in vivo*. In addition, LexA-Mre11 induced expression with Gal4-Mre11, consistent with an ability of *Sc*Mre11 to form homodimers or higher order multimers *in vivo*.

Similar to the SbcCD complex, the *Sc*Rad50/*Sc*Mre11 complex is therefore likely to consist of multiple subunits. SbcD is the nuclease subunit of SbcCD and the SbcC subunit is required with SbcD for ATP dependent exonuclease activity; likewise *Sc*Rad50 binds DNA in ATP and *Sc*Mre11 has nuclease and DNA binding activities.

Other Proteins in the Complex

Mutations in *ScRAD50*, *ScMRE11* and *ScXRS2* have pleiotropic and indistinguishable phenotypes in mitotic (vegetative) growth and in meiosis and are epistatic to one another (Johzuka and Ogawa 1995, Raymond and Kleckner 1993a, Ivanov *et al* 1992). This is consistent with the activity of the three proteins together in a complex. *Sc*Xrs2 is 96.3kDa in size and has a very hydrophilic C-terminus (Ivanov *et al* 1994). It has been suggested that meiosis-specific proteins are recruited to form a multiprotein complex active in meiosis (Johzuka and Ogawa 1995). Consistent with this, the *Sc*Rad50/*Sc*Mre11/*Sc*Xrs2 complex also includes two meiosis-specific proteins in meiosis (Usui *et al* 1998). No other proteins are present in the *Sc*Rad50/*Sc*Mre11/*Sc*Xrs2 complex during mitosis.

*Sc*Mre11 has two binding sites for *Sc*Rad50 (Usui *et al*1998). These are in the first and second thirds of the protein. *Sc*Mre11 is the binding core of the yeast Rad50/Mre11/Xrs2 complex.

Co-immunoprecipitation of hRad50 and hMre11 under increasing salt concentrations show that they are in highly stable association (Dolganov *et al* 1996). This 1600kDa human complex includes at least one other protein. This protein, 95kDa in size and called p95, NBS1 or nibrin, is identical to that absent in NBS patients (Trujillo *et al* 1998, Carney *et al* 1998). It bears limited homology to Xrs2 but is assumed to perform a similar function on account of its identical mutant phenotypes and molecular weight (Carney *et al* 1998, Featherstone and Jackson 1998). It is substoichiometric to hRad50 and hMre11 (Trujillo *et al* 1998). Two other large proteins may be associated with this complex (Dolganov *et al* 1996, Carney *et al* 1998), or not (Trujillo *et al* 1998). One of these two is probably artefactual (Carney *et al* 1998).

NBS1

The N-terminus of NBS1 contains a forkhead-associated FHA domain and a breast cancer BRCT domain (Featherstone and Jackson 1998). The FHA domain is characteristic of nuclear protein kinases involved in DNA replication and repair. The BRCT domain can facilitate protein-protein interactions and is characteristic of proteins involved in DNA repair and DNA damage checkpoints. NBS1 is not required for hRad50 interaction with hMre11. It is expressed from an abundant, ubiquitous 4.4kb transcript and a 2.6kb testis specific transcript in a similar pattern to hMre11 (Carney *et al* 1998).

Genetic Analysis of Rad50 and Mre11

Mutant Phenotypes in Yeast Mitotic Growth

Null Mutants

A *rad50::ura3* disruption allele, which does not express *rad50*, conferred sensitivity to MMS, indicative of defective mitotic DNA repair (Malone *et al* 1990). It is defective in the repair of spontaneous DNA damage and ionizing radiation damage. In addition growth is slow, and mitotic recombination is increased 10-fold above wildtype levels (hyperrecombination). A null point mutation and a deletion in *ScRAD50* both conferred the same phenotypes. As expected for hyperrecombination, they are capable of homologous linear plasmid integration (Alani *et al* 1990). A *rad50* allele was obtained with site directed mutations in Heptad II of the ATP binding consensus, *ie* the region which shares homology to the myosin S-2 domain. This displayed the same phenotype as the null *rad50* mutant. These mutants are capable of mating-type switching, but do not perform this function at full wildtype levels (Raymond and Kleckner 1993a). The null mutant cells have increased size, abnormal shape and abnormal nuclear segregation (Kironmai and Munyappa 1997). The abnormal nuclear segregation results from inappropriate entry into M-phase and causes low viability. In HU (hydroxyurea, DNA damaging agent) which reversibly arrests the cell cycle at the S-phase checkpoint control, 80% of cells display defective cytokinesis, indicating that there is no morphologically defined S-phase checkpoint

control. The S-phase checkpoint is therefore defective. The slow growth phenotype results from a shorter cell cycle combined with a high level of cell death.

Partial Function Mutants

Nine *rad50S* (separation of function) non-null mutants were obtained, all of which map to the N-terminal ATP binding domain on either side of the 22 amino acid ATP-binding consensus sequence (Alani *et al* 1990). These plated with wildtype efficiency at weak and intermediate levels of MMS, but were sensitive to high concentrations of MMS. Mitotic growth was otherwise normal, exhibiting wildtype levels of spontaneous mitotic recombination. *mre11S* was isolated with the same phenotypes as *rad50S* (Nairz and Klein 1997). This has mutations Pro (84) to Ser, and Thr (188) to Ile, and the mutant protein is slightly longer than the wildtype on account of a frameshift. The N-terminal region (amino acids 63 to 117) contains several dispersed amino acids conserved throughout eukaryotes of which Pro 84 is one. However these mutations are not in the conserved phosphoesterase motifs, suggesting that *mre11S* is defective in its interaction with other proteins.

A previously identified mutant was renamed *mre11-58* when found to be allelic to *mre11* (Tsubouchi and Ogawa 1998). This includes changes at two sites; His (213) to Tyr within the conserved phosphoesterase motif IV (of five), and a homologous amino acid change. Only the His (213) to Tyr mutation affects function, and this change alone was named *mre11-58S*. *mre11-58S* differs from *rad50S* in its mitotic phenotype, exhibiting sensitivity to MMS and repairing DSBs relatively slowly. In addition, processing of DSBs formed by HO endonuclease was supressed, although mitotic repair remained proficient. This suggests that DSB processing by Mre11 depends on the phosphoesterase domain.

Mutant Phenotypes in Yeast Meiosis

Homologous Recombination

The *rad50::ura3* disruption (null) mutant shows reduced spore formation and the spores produced are inviable. Both phenotypes result from abolished meiotic recombination which causes aberrant chromosome segregation at the first meiotic division (Malone *et al* 1990). The null point mutants and the mutant in the ATP binding consensus have the same phenotype. Further mutation of *spo13*, which bypasses the first meiotic division, relieves this meiotic inviability (Cao *et al* 1990).

The *rad50S* mutants display partial progression through meiosis resulting in no meiotic recombination. Spore formation is reduced and the spores produced are inviable (Cao *et al* 1990). The meiotic lethality is not alleviated by *spo13* alone (Alani *et al* 1990). Therefore defects are not restricted to segregation. Lethality is alleviated by *spo11Δspo13* (Cao *et al* 1990). The authors suggest that *spo11Δ* prevents recombination by acting at a very early stage prior to the *rad50S* block. DSBs are formed but not processed in *rad50S* mutants. The *mre11S* mutant displayed similar phenotypes to *rad50S* (Nairz and Klein 1997). *mre11-58S* is like *rad50S* with respect to its meiotic phenotype (Tsubouchi and Ogawa 1998). The unprocessed DSBs were observed to be resistant to λ exonuclease, as expected if they are covalently bound.

Chromosome Synapsis

A SC is a tripartite proteinaceous structure (Alani *et al* 1990). It consists of two lateral elements and a central element. Lateral elements are dense and run parallel to and along the length of a bivalent. The central element consists of intervening material which connects the lateral elements. A lateral element alone, before it becomes associated with another element, is called an axial element.

Neither the *rad50Δ* nor the *rad50* ATP-binding null mutants form SCs (Alani *et al* 1990). Only very early axial elements can be observed. In the *rad50S* mutant DNA replication occurs but progression through meiosis is partial

and asynaptic (Cao *et al* 1990). Axial development progresses beyond that in the null mutants or the ATP-binding mutant (Alani *et al* 1990). Spindle pole body and spindle morphogenesis is abnormal. Epistatic analysis using *spo11Δ* indicated that the *rad50S* block occurs during meiotic prophase. The *mre11S* mutant displays a similar phenotype (Nairz and Klein 1997). It was also observed that homologous pairing is reduced in this mutant, and that some non-homologous SCs are formed.

Mutant Phenotypes in Higher Eukaryotes

Mre11 is Essential for ES Cell Proliferation

A conditional knock-out system was used to study the requirement for mMre11 in ES cells (Xiao and Weaver 1997). *mmre11* was integrated into the murine ES cell line between LoxP direct repeats. A Cre recombinase dependent site-specific recombination event between the repeats generated deletions in *mmre11*. No ES cell lines lacking any of the exons of *mmre11* survived beyond day 10, whereas those retaining intact *mmre11* underwent normal mitoses. mMre11 is therefore essential for normal murine ES cell proliferation.

Genetic Disease in Humans

The yeast mutants of Rad50, Mre11 and Xrs2 have phenotypes of chromosomal instability (above). Similar such phenotypes might be expected to cause heritable chromosomal instability syndromes in humans. Chromosome instability would potentiate tumourigenesis by increasing the rate of somatic mutation, *eg* in Bloom Syndrome and AT (Section 1). It has been suggested that Bloom Syndrome might feature replication fork arrest followed by DSB formation, as in prokaryotes (Michel *et al* 1997).

However no known genetic diseases arise from *mre11* or *rad50* mutation. This is likely due to the essential requirement for these genes in ES cell development (above), which presumably prevents null mutants from arising even as embryonic lethals. Mutation of the p95 component of the human Rad50/Mre11 complex, thought to substitute for *Sc*Xrs2, is associated with NBS, and is called NBS1 (Carney *et al* 1998, above, Section 1).

Regulation of Gene Expression

In Yeast

ScRAD50 is 3.9kb long (Raymond and Kleckner 1993b). It has two major transcripts, 4.2kb and 4.6kb, with different 5' ends. Both are generated in mitosis and meiosis and polyadenylated. However the 4.2kb transcript is presumed to be the major transcript since it is more abundant and includes the entire gene. The 4.2kb transcript has several alternative start sites, including three major ones. The steady state levels of both transcripts are co-ordinately transient through meiosis. They increase immediately prior to sporulation (concomitant with transfer to sporulation medium), reach a maximum after four hours and decrease thereafter. The peak occurs prior to meiosis I, at approximately the time of SC formation. The transcripts of other essential meiotic genes rise and fall with the same temporal pattern, including *SPO13, SPO11, MER1, RED1, HOP1* and *MEI4*. The meiotic peaks of the 4.2kb and 4.6kb transcripts are 10-fold and 3-fold greater than the basal vegetative level respectively. Both peaks are low compared with those of the other meiotic transcripts. These fluctuations are observed only in sporulating diploids (**a**/α mating type). The exact timing and magnitude of fluctuation is dependent on the growth medium prior to sporulation, due to differences in meiotic synchrony. By comparing with *SPO13* and *URA3*, the mid-meiotic transcript level of *ScRAD50* was estimated as 1 transcript per 10 to 40 cells. Despite the fluctuation in transcription, the steady state level of *Sc*Rad50 protein remains constant during meiosis, at or below the mitotic level. The reason for this is unknown. The frequency of meiotic recombination is 4-fold higher than that of mitotic recombination (McKee and Kleckner 1997a).

The mitotic and meiotic transcripts of *ScMRE11* are equal in size (Johzuka and Ogawa 1995). Meiotic *ScMRE11* mRNA steady state levels are transient. They rise and fall with the same kinetics as *ScRAD50* transcripts. S1 nuclease mapping of *ScMRE11* mRNA showed that two transcripts are generated. These share the same start site, have different termination sites and differ in length by only 0.1kb. *RAD32* is expressed at moderate level (Tavassoli *et al* 1995). Like *ScMRE11*, two transcripts are generated. This is characteristic of genes involved in meiosis.

The similar transience of expresion of *ScRAD50* and *ScMRE11* suggests that the two genes may be regulated by a common mechanism. The sequence context of *ScXRS2* is also suggestive of low expression (Ivanov *et al* 1994).

In Higher Eukaryotes

The steady state levels of mRad50 are influenced by confluence of cell culture, as expected for a tumour suppressor (Kim *et al* 1996). Three transcripts are generated from *mRAD50*. The full length transcript is 5.3kb. In the embryo, it is expressed in most tissues, and at especially high levels in heart, liver and thymus. In neonatal mice, it is expressed at much reduced levels in brain, gut, liver and skin, is induced in spleen and remains high in heart and lung. In the adult it is generally expressed only at low level, slightly more in heart, lung and aorta, and is expressed at very high level in testis. A 3.1kb transcript is specific to heart and lung and is not expressed before birth. Its product, predicted to be 50kDa, is colinear wth the C-terminus of the full protein. A 1.6kb transcript is expressed in the adult testis only. Its product, predicted 58kDa, is colinear with the N-terminus and includes the NTP-binding domain. If mRad50 is required for DNA damage repair, it is not surprising that the full length transcript is generally expressed at lower level in adult cells than cells of high mitotic index. Expression in adult heart and lung is unexpected, since adult cardiac myocytes are terminally differentiated and lung cells undergo little division. mRad50 may have a DNA protective role in these cells. The high level expression in testis is discussed below.

hRAD50 generates two transcripts by alternative 3' end processing (Dolganov *et al* 1996). The major transcript, 5.5kb, is expressed maximally in adult testis, and at low level in activated adult T and B lymphoid cells and embryonic liver. It is also expressed abundantly in fibroblastoid (Maser *et al* 1997) and lymphoid cell lines, but at a level lower than in testis (Dolganov *et al* 1996). The transcript is stable. A 7kb transcript is expressed at low level. hRad50 expression is dependent on hMre11 expression (Paull and Gellert 1998).

hMRE11 generates two transcripts (Petrini *et al* 1995). A 2.5kb transcript is expressed in all tissues examined at a level proportional to the proliferative

capacity of the cell type; expression is highest in testis and spleen. A 6kb transcript is expressed with the same pattern but at a low level in brain. hMre11 is expressed independently of hRad50 (Paull and Gellert 1998). It is expressed abundantly in fibroblastoid cell lines (Maser *et al* 1997).

Mre11 and Rad50 have complex patterns of gene expression. In summary, expression of both increases with cell mitotic index, and expression is high in adult mammalian testis. The pattern is shared by human and mouse Rad51 (Petrini *et al* 1995, Dolganov *et al* 1996). *Sc*Rad51 is involved in meiotic recombination and expression in testis may reflect a similar conserved function in mammals.

The steady state level of hRad50 and hMre11 transcripts does not change on exposure to ionizing radiation, despite the probable function of these proteins in repair of DNA damage (Dolganov *et al* 1996, Maser *et al* 1997). In contrast, mRad50 transcript steady state levels increase on exposure to MMS (although cell viability is reduced 25%) and UV radiation (in MOPS buffer) suggesting that transcription increases in response to DNA damage (Kim *et al* 1996).

Chromosomal Location of Human Homologues

mRAD50 is located in the A5 to B1 region of mouse chromosome 11 (Kim *et al* 1996). *hRAD50* is located at 5q31, where q is the short arm and p is the long arm of chromosome 5 (Dolganov *et al* 1996). It spans 100 to 130kb. The 5q23 to 5q31 region is significant because here map the chromosomal breakpoints of translocations which are implicated in several human genetic diseases (Hill *et al* 1996, Fairman *et al* 1996). These include limb girdle type 1A muscular dystrophy LGMD1A, acute nonlymphocytic leukaemia ANLL (aka acute myeloid leukaemia AML) and myelodysplastic syndrome MDS (myeloma is cancer of plasma cells). A 1000bp region at 5q31 is the minimal "critical region" common to deletions causing all these disorders (Le Beau *et al* 1986). It was therefore proposed that a tumour suppressor gene maps to this site. *hRAD50* is one of several candidate genes, including those for IL9, testican, α-e-catenin, mortalin, Cdc25 and selenoprotein P (Hill *et al* 1996). More detailed mapping will be required to determine which of these is responsible.

hMRE11 maps to approximately 11q21 (Petrini *et al* 1995). It lies between 11q14 and 11q22. This region includes genes which, when mutated, are associated with cancer-related chromosomal abnormalities, including sporadic breast cancer, cervical cancer, lymphoid neoplasms, squamous cell carcinomas and two AT complementation groups. If mutant hMre11 also potentiates tumourigenesis, this would be consistent with a role of hMre11 in maintaining genomic stability. A region homologous to *hMRE11* was also detected between 7q11.22 and 7q11.23. This may represent a pseudogene or a closely related functional gene.

Cellular Roles of the Rad50/Mre11 Complex

Roles in Homologous Recombination

Yeast recombination is principally homologous (Section 1), and yeast homologous recombination occurs principally in meiosis. However homologous recombination also occurs in mitotic growth during mating-type switching and as a mechanism of DNA repair. Most recombination in higher eukaryotes occurs by non homologous end-joining (Section 1).

Homologous Recombination in Yeast Meiosis

The N-terminal, eukaryotic-conserved amino acids of *Sc*Mre11 have been suggested to serve meiosis-specific functions (Nairz and Klein 1997).

DSB Formation in Meiotic Recombination

The initial step in meiotic recombination is the formation of discrete DSBs (Cao *et al* 1990). Most meiotic recombination in *S. cerevisiae* occurs at hot spots (McKee and Kleckner 1997b) (Section 1). *rad50* and *spo11* null mutants abolished DSB formation at the *LEU2HIS4* hot spot, indicating that *Sc*Rad50 and Spo11 are essential for this (Cao *et al* 1990). The fact that meiotic inviability could be overcome by *spo13* supports an early function of *Sc*Rad50 (and Spo11) in meiotic recombination in DSB formation.

DSB Processing in Meiotic Recombination

DSBs occur transiently in wildtype meiotic recombination and are resected in the 5' to 3' direction (Alani *et al* 1990). The non-null mutant *rad50S* generated DSBs at the same time and at the same sites as the wildtype. However these were stable, long and discrete and accumulated in meiosis (Alani *et al* 1990). This suggests that, in addition to its role in initiating meiotic recombination by forming DSBs, *Sc*Rad50 acts at an intermediate step in meiotic recombination. When HO endonuclease was expressed under the control of a meiosis-specific promoter in *rad50Δ* mutants, some DSBs that it generated persisted until the next mitotic cycle (Malkova *et al* 1996). This supports the suggestion that *Sc*Rad50 acts after forming DSBs in meiosis.

*Sc*Mre11 was demonstrated to resect the 5' ends of meiotic DSBs, leaving long single-strand overhanging 3' tails (Nairz and Klein 1997). This suggests that *Sc*Mre11 acts as a 5' to 3' exonuclease or single-strand endonuclease *in vivo*. This is consistent with its homology to SbcD, but inconsistent with the 3' to 5' exonuclease activity identified *in vitro* (the exonuclease polarity conundrum, above). It is not yet understood how the 3' to 5' Rad50/Mre11 exonuclease is responsible for 5' to 3' digestion.

The tails then participate in a homology search, strand-invasion and are further processed during recombination to generate mature recombinants.

Yeast Rad50/Mre11 Acts with Spo11

The molecular mechanism of meiotic DSB formation and processing by the yeast Rad50/Mre11 complex were studied in more detail (Keeney and Kleckner 1995). Although DNA obtained from *rad50S* mutants was expected to contain unresected DSBs, it was found to be largely intact unless treated with proteinase K. This suggested that DSB sites are bound covalently by a protein, and that DSBs are introduced when this protein disattaches. Protein binding was confirmed by CsCl density gradient centrifugation. Covalent binding was confirmed using a range of treatments which disrupt non-covalent interactions. 70% to 90% of DSBs were protein associated and a proportion of these bound protein on both sides of the break. By denaturing the protein associated DSBs and hybridising to strand-specific probes, it was

deduced that the protein binds the 5' end of the DSB. This is the strand which is resected in wildtype cells.

The unresected, protein associated DSB DNA copurified with 45kDa Spo11 (Keeney *et al* 1997). Spo11 was found to bind the DSB termini specifically, by using immunoreactive protein. The Spo11-DNA complex was stable only in genetic backgrounds in which DSBs remain unresected, indicating that it exists transiently where DSBs are transient. Since Spo11 is required for cleavage, it was suggested that Spo11 has at least two subunits, each of which effects the cleavage of a single DNA strand. Spo11 was proposed to form the catalytic subunit of the meiotic cleavage complex.

Spo11 is a member of a phylogenetically widespread subfamily of type II topoisomerase proteins (Bergerat *et al* 1997). The prototype of this family is topoisomerase VI of *Sulfolobus shibatae*, a hyperthermophilic archaeon. This is a heterotetramer of two each of subunits A, 45kDa, and B, 60kDa. The A subunit cleaves DNA and is homologous to *S. cerevisiae* Spo11, *S. pombe* Rec12, a putative protein in *C. elegans*, the A subunit of a subclass member in *Methanococcus jannaschii* and the A subunit of topoisomerase IV in *Pyrococcus furiosus*. These share a conserved 200 amino acid domain with five motifs. The B subunit binds ATP and its homologues include the Hsp90 family of heat shock proteins and the MutL family of mismatch repair proteins.

The homology of Spo11 to topoisomerases suggests that it cleaves DNA by a transesterification mechanism (Keeney *et al* 1997). Proteins which cleave DNA via a protein-DNA intermediate attack using the hydroxyl group of serine, threonine or tyrosine as a nucleophile. A tyrosine residue (Tyr 135 in Spo11) is conserved in the topoisomerase VI family (Bergerat *et al* 1997). This has been shown by SDM to be essential for spore formation and viability in Spo11. In addition, the phosphodiester link has been shown to involve a tyrosine residue by its alkali resistance (Keeney *et al* 1997). It is unknown whether Spo11 interacts with a B subunit for this function, but Hsp82 (a Hsp90 member) is induced in meiosis.

It therefore appears that Tyr 135 of Spo11 attacks the 5' end of a DSB site. It is significant that, unlike hydrolysis, phosphodiester cleavage by transesterification is reversible until the topoisomerase has been removed

from the DSB (Keeney *et al* 1997). The yeast Rad50/Mre11/Xrs2 complex is a candidate for this removal since it is a nuclease and its second point of action in the meiotic recombination pathway is in DSB resection. It is unknown whether Spo11 removal and DSB resection occur in the same process. It is also unknown how Rad50/Mre11 is required by Spo11 for DSB formation.

Meiotic recombination does not occur in *Drosophila melanogaster* and this may be because the Spo11 homologue acts as a topoisomerase rather than as a nuclease only (Keeney *et al* 1997).

The Sae2 (Com1) protein is likely also required for Spo11 excision in meiotic DSB processing (McKee and Kleckner 1997a).

Two Possible Modes of Yeast Rad50-Mre11 Interaction

On the basis of the domain structure of Mre11, it has been proposed that meiotic DSB formation utilises a loose Rad50/Mre11 complex, perhaps not even containing Rad50, and meiotic DSB processing utilises a tight Rad50/Mre11 complex (Usui *et al*1998, reviewed Haber 1998). *Sc*Mre11 is suggested to act as the binding core of the DSB forming complex, recruit meiosis-specific factors via its C-terminal domain, and alter chromatin conformation in a "pre DSB complex". After DSB formation, the complex reorganises; *Sc*Mre11 is no longer the binding core, one or more of its nuclease activities becomes active in resection, and *Sc*Rad50 and Xrs2 are present in the "post DSB complex". Binding of Spo11 may create a single-strand region adjacent to it which can be cleaved by *Sc*Mre11. If the complex has helicase activity, it may translocate and extend the single-strand region during resection.

The Mechanistic Connection between Meiotic Recombination and Chromosome Synapsis

DSBs form after DNA replication and before the appearance of mature recombinants (Cao *et al* 1990). This led the authors to speculate that DSB formation might be concurrent with chromosome pairing and synapsis, and that the two processes might share common early events. Null mutants in *rad50* block both chromosome synapsis and meiotic recombination at a very early stage, and *rad50S* mutants block both processes at intermediate stages,

suggesting that *Sc*Rad50 is involved in events common to early and intermediate stages of both processes (Alani *et al* 1990, Padmore *et al* 1991).

DNA metabolism in meiosis was described visually (Padmore *et al* 1991). Five stages were described. At the level of cytology, stage I is characterised by DNA replication. SC development occurs in three steps; stage II is characterised by SC precursors (A) which are short axial core segments, stage III involves the elongation of the tripartite SC (B), and stage IV is characterised by full length SCs at pachytene. At the end of pachytene, at stage V, spindle formation is contemporaneous with SC disappearance, suggesting that SC dissociation is very rapid. The two processes do not overlap. Chromatin becomes increasingly condensed through prophase. At the level of recombination, stages I and II involve the formation of non discrete DSBs. DSBs start to disappear during stage III, and are gone by stage IV. The disappearance of DSBs is closely followed by the formation of mature recombinants in stage V. These five stages strongly support an intimate functional connection between SC development and recombination.

DSBs *per se* are not required for SC formation (Weiner and Kleckner 1994). A model was proposed whereby homologue interaction precedes DSB formation which is required for SC formation and recombination. The homology search results in pre-meiotic pairing and early meiotic pairing. This occurs by close paranemic DNA-DNA interactions between intact dupleces. (A paranemic spiral consists of two parallel threads coiled in opposite directions and can be easily separated). These are then converted to recombination intermediates via DSBs. The DSBs convert the paranemic interactions to plectonemic interactions, thereby stabilizing them. (A plectonemic spiral consists of threads coiled in the same direction and cannot be separated unless uncoiled). The SC then forms, representing stable joint molecules. These are then converted to mature recombinants.

As further evidence for this model, homologous contacts were observed during logarithmic growth, and in *S. pombe* , in which there are no SCs (Loidl *et al* 1994). In addition, hotspots for meiotic recombination interact before DSB formation (Xu and Kleckner 1995).

It was also observed that pairing occurs throughout a genome simultaneously in one cell, but at different times in different cells. This all-or-nothing effect may result from co-operativity or a threshold concentration of a required factor. The SC might regulate recombination by facilitating crossovers at the correct time (Padmore *et al* 1991).

Yeast Mitotic Homologous Recombination

Yeast mitotic homologous recombination consists of homologous recombinational DNA repair and mating-type switching. Homologous recombinational DNA repair involves the stimulation of homologous recombination at DSBs formed spontaneously or by DNA damaging agents (Section 1). The mating-type switch is a specialised mitotic gene conversion event initiated by a site-specific DSB in the *MAT* locus which is introduced by the activity of HO endonuclease (Ivanov *et al* 1994). Thus both mechanisms involve the processing of DSBs introduced by other means.

Multiple Effects on Mitotic Homologous Recombination

When HO endonuclease was used in a model system to introduce DSBs into mitotic *rad50Δ* cells the range of recombinants obtained differed from that in wildtype cells (Malkova *et al* 1996). This suggests that *Sc*Rad50 has a role in the processing of mitotic DSBs. The authors suggested that *Sc*Rad50 might be involved in connecting sister chromatids to regulate mitotic recombination.

*Sc*Xrs2 (and the mitotic *Sc*Rad50/*Sc*Mre11/*Sc*Xrs2 complex) is not required for the mating-type switch (Ivanov *et al* 1994). However the null mutant does delay switching. Despite the epistatic relationship between Rad50 and Xrs2, a *rad50Δxrs2Δ* double mutant displays a greater delay in mating-type switching than either single mutant. In contrast, Rad51 and Rad52 are both essential for the mating-type switch (Tsubouchi and Ogawa 1998). The delay may be partly due to a decrease in 5' to 3' resection of DSBs which would delay strand invasion (Ivanov *et al* 1994).

In some cases *ScRAD50, ScMRE11* and *ScXRS2* are required for homologous recombinational repair. The *rad50* null mutant is as sensitive to HU as *rad51*

and *rad52* (Allen *et al* 1994). Under some circumstances null mutants delay the completion of homologous recombinational repair but the gene products are not essential for the process (Lee *et al* 1998). In still other circumstances the null mutants confer a hyperrecombination phenotype during mitosis. Spontaneous interchromosomal recombination within a nutritional allele increases 10-fold in *rad50, mre11* or *xrs2* mutants, whereas it decreases 5- to 10-fold in *rad51* and 1000-fold in *rad52* (Ivanov *et al*1994).

The Mitotic Homologous Recombination Conundrum

It is not understood how *Sc*Rad50, *Sc*Mre11 and *Sc*Xrs2 have multiple effects on mitotic homologous recombination (reviewed Haber 1998).

The mitotic hyper-rec phenotype of *rad50* and *mre11* mutants may result from blockage to non homologous recombination (below). The reduction in DSB religation might stimulate recombination. However mutation of Ku does not confer a hyperrecombination phenotype. Alternatively, loss of 5' to 3' DSB resection might extend the region of heteroduplex DNA formed during branch migration and increase gene conversion. Thirdly, this phenotype might result from accumulation of recombinogenic lesions in G2 in the absence of *Sc*Rad50, *Sc*Mre11 or *Sc*Xrs2 (Malone *et al* 1990, Ivanov *et al* 1992). Homologous recombinational repair of DSB normally occurs in G2 in the presence of sister chromatids. Accumulation of lesions might stimulate interhomologue, interallelic recombination.

Comparing Meiotic and Mitotic Homologous Recombination Pathways

Yeast mutants *mre11D56N* and *mre11H125N*, which have substitutions in two phosphodiesterase motifs, are partially sensitive to ionizing radiation but otherwise proficient in mitotic homologous and non homologous recombination (Moreau *et al* 1999). However these mutants accumulate unresected DSBs and fail to sporulate in meiosis. It was proposed that the nuclease activity of *Sc*Mre11 is redundant in mitotic recombination, but is essential for processing of Spo11-bound meiotic DSBs.

The homology search mechanism has been suggested to be conserved between meiotic and mitotic homologous recombination (Alani *et al* 1990).

The interesting suggestion has also been made that the meiotic homology search may have evolved from the mitotic DSB repair process (Padmore *et al* 1991). In support of this, Spo11, Hop1, Mre3/Rec114, Mre4/Mek1, Mei4, Red1, Sae2(Com1) and the members of the *ScRAD52* epistasis group are involved both in mitotic and meiotic homologous recombination in yeast (McKee and Kleckner 1997b, Nairz and Klein 1997). The mutant phenotype of *sae2* in mitosis is much like wildtype (McKee and Kleckner 1997a). Perhaps Sae2 removes protein bound to a small fraction of mitotic DSBs.

Alternatively, the roles of *Sc*Rad50/*Sc*Mre11 may differ in mitotic recombination repair and meiotic recombination, and these pathways may be mechanistically different despite both using enzymes of the *RAD52* epistasis group (Malone *et al* 1990).

Roles in Non Homologous End-Joining

Non homologous end-joining predominates as the favoured mechanism of recombination in higher eukaryotes and occurs only at low level in yeast (Section 1). It occurs during normal mitotic growth. In yeast *Sc*Rad50 is required for erroneous DSB repair by the non homologous mechanism (Malkova *et al* 1996). Null mutants in *ScRAD50, ScMRE11* and *ScXRS2* cause a 100-fold reduction in non homologous DNA repair (Lee *et al* 1998).

Non Homologous End-Joining

DNA damage can generate different types of DNA ends, rather than simple blunt DSBs. It is conceivable that the different nuclease activities of Mre11 allow efficient repair of a range of substrates, and prevent the need for assembly of different repair complexes (Usui *et al*1998). This applies to homologous recombinational repair as well as non homologous end-joining. In yeast, Rad50, Mre11 and Xrs2 are epistatic to Ku, SIR2, SIR3, SIR4 and DNA ligase with respect to non homologous end-joining (Boulton and Jackson 1998).

DNA Repair by Non Homologous End-Joining In Yeast

A system for the study of non homologous end-joining in yeast involves the detection of deletions formed on a dicentric plasmid (Tsukamoto *et al* 1997). DSBs, commonly formed on a dicentric plasmid by breakage-fusion bridge cycles, may be repaired by non homologous end-joining to generate deletions. Deletion formation was reduced in *rad50, mre11, xrs2* and *hdf1* mutants by 50, 80, 30 and 50-fold respectively. *Sc*Rad50, *Sc*Mre11 and *Sc*Xrs2 were independently found to be required for non homologous end-joining (Moore and Haber 1996).

The two mitotic phenotypes of *mre11-58S* (above) initially appear contradictory. However the two phenotypes may simply indicate that DSB repair normally occurs both by homologous recombination (gene conversion and SSA) and by illegitimate recombination (non homologous end-joining), and all three processes require *Sc*Mre11 (Tsubouchi and Ogawa 1998). The retardation of DSB repair may result from a lack of another required factor, suggesting that *Sc*Mre11 normally interacts with another factor during non homologous end-joining.

When the *mre11D16A* mutation was inserted into the genome, it confered some deficiencies in mitotic DSB repair and some deficiencies in meiotic recombination (Furuse *et al* 1998). The mutant was deficient in integration into a homologous target sequence, but proficient in end-joining repair of linear DNA. It was suggested that *Sc*Mre11 acts at the gateway step of the non homologous end-joining and homologous recombination (gene conversion) pathways. The N-terminal nuclease domain may be required for some steps of gene conversion, but appears not to be required for non homologous end-joining.

Non Homologous End-Joining In Higher Eukaryotes

The biochemical and genetic analyses described above provide convincing evidence that a major function of the Rad50/Mre11 complex in higher eukaryotes is in recombination. The major pathway of recombination in these cells is non homologous end-joining. This functions in generating

diversity, in DNA repair and, in the case of V(D)J recombination, in programmed DNA rearrangement (Section 1).

hMre11 promotes joining of different types of non homologous ends (Paull and Gellert 1998). Joining of two blunt ends, mismatched 5' overhangs, and one blunt and one 5' overhang all generated deletions of short sequences from the tip of one recombining end. These are the expected products of non homologous end-joining.

hMre11 was essential for ligation (Paull and Gellert 1998). Ligation occurs after opening of the postulated hairpin intermediates formed of the recombining ends by RAG1 and RAG2 (Section 1). The hairpin cleavage activity of hMre11 is similar in many ways to SbcCD hairpin nuclease and suggests hMre11 as a candidate for hairpin opening in V(D)J recombination. Hairpin opening by hMre11 is most efficient on substrates containing a loop of mismatched nucleotides. The hairpin intermediates in V(D)J recombination do not contain mismatches and are not opened by hMre11 alone or by hMre11 with RAG proteins. Activity may depend on recruitment of other factors after hMre11 binding. It also may be restricted by proteins which protect the DNA ends. In support of this proposed role of hRad50/hMre11 in ligation, the phenotypes and expression patterns of *cdc9* (DNA ligase I) and *hrad50* mutants are similar (Petrini *et al* 1995).

De-Protection of Ends

Ku protects DNA ends by binding them avidly (Section 1). Like yeast Rad50/Mre11 in Spo11 removal, mammalian Rad50/Mre11 may be implicated in de-protecting DNA ends if it has a role in removing Ku during non homologous end-joining. This model is consistent with the possibility that Ku may act during formation of the hairpin intermediate (Section 1), whereas Rad50/Mre11 may act during resolution of the hairpin intermediate (above). However Ku is not known to interact with Rad50, Mre11 or Xrs2 (Maser *et al* 1997).

A Role in Telomere Maintenance

Telomeres of Yeast

Telomeres of *S. cerevisiae* consist of a 300bp heterogeneous array of the following sequence, called "TG_{1-3}" (reviewed in Kironmai and Munyappa 1997);

$$C_{2-3}A(CA)_{1-6}/(TG)_{1-6}TG_{2-3}$$

The array includes multiple copies of the Rap1 binding site, and telomere-associated middle repetitive elements Y' and X in the distal tract. Length maintenance is dynamic. It occurs by two mechanisms (reviewed in Nugent *et al* 1998): End protection requires Cdc13. Telomere replication requires telomerase. Telomerase contains Est1, Est3, the Est4/Cdc13 dimer, the catalytic subunit Est2, and the telomere RNA Tlc1. Mutations in Ku worsen the phenotypes of mutants in either group, indicating that it belongs to a third epistasis group for telomere maintenance (Nugent *et al* 1998).

Cdc17 and Pif1 lengthen, and Tel1, Tel2, Est1, Kem1, Tlc1 and Hdf1 shorten telomeres. Mutants in Tel1 and Tel2 adopt a new stable telomere length, whereas those in Est1 and Tlc1 undergo a progressive change in length. Senescence occurs when telomere length shortens progressively.

Telomere Homeostasis in Yeast

Null mutants in *rad50* display striking chromosomal rearrangements and senescence in yeast; the distal tract of the telomere is progressively shortened by approximately 150bp over 30 generations (Kironmai and Munyappa 1997, Boulton and Jackson 1998). The *mre11D16A* mutation confers deficiencies in telomere maintenance and growth (Furuse *et al* 1998). The *rad50S* mutant undergoes distal tract elongation by 300bp (Kironmai and Munyappa 1997). Rap1 is one of the very few other proteins that effects both telomere lengthening and shortening. The *rad50S* allele is dominant over *ScRAD50* wildtype and *rad50Δ* for growth and telomere length. Telomere associated regions show a high level of recombination. *Sc*Rad50 may control telomere length by recombination. Rad52 dependent recombination has a role in telomere maintenance, however not all members of the *RAD52* epistasis group are involved.

Double mutants in Ku and Rad50, Mre11 or Xrs2 have been found to confer a phenotype no more severe than the single mutants (Boulton and Jackson 1998). This indicates that Ku and the Rad50/Mre11 complex are epistatic with respect to telomere maintenance. Independently however Rad50, Mre11 and Xrs2 were found to be epistatic to Est1 and Est2 for telomere length maintenance, and not to Ku nor Cdc13 (Nugent *et al* 1998). This suggested that the complex functions in the telomerase-mediated pathway of telomere maintenance, perhaps by generating single-strand ends for telomerase activity. These contradicting results might be reconciled if Ku has two different activities in telomere maintenance and functions in both pathways.

Ku and the Rad50/Mre11 complex are epistatic with respect to non homologous end-joining (above). However unlike Ku and the SIR proteins, mutations in *rad50, mre11* and *xrs2* do not affect TPE. It has been suggested that Ku and the Rad50/Mre11 complex might bind telomeres directly and recruit other proteins to determine telomere length homeostasis and TPE (Boulton and Jackson 1998). Other than Ku, proteins in the telomere end binding and the telomerase pathway do not affect TPE (Nugent *et al* 1998).

Roles in Checkpoint Regulation

The G1/S Checkpoint: DNA Damage Response

DNA damage sensed in G1 prevents progression of the cell cycle into S phase.

Sub-Nuclear Relocalisation

Both hRad50 and hMre11 are uniformly distributed in the nucleus, except in the nucleolus, of unirradiated fibroblastoid cells (Maser *et al* 1997). In response to ionizing radiation, both proteins relocalise in the nucleus and form discrete nuclear foci, IRIFs, which are visible using immunofluoresence labelling. The IRIFs formed include hRad50, hMre11 and NBS1 (Maser *et al* 1997, Carney *et al* 1998, Usui *et al*1998). The number of DSBs induced is proportional to the radiation dose, the number of foci formed per nucleus and the number of nuclei containing foci (Maser *et al* 1997). Other DNA

lesions induced by other DNA damaging agents do not induce foci. A cell line defective in repair of DSBs displays greater foci induction. This all suggests that the foci are located at DSBs, as expected if the complex is involved in recombinational DSB repair. After DSB repair, the IRIFs disappear and hRad50 and hMre11 redistribute uniformly (Carney *et al* 1998).

Ionizing radiation may alternatively induce hRad51 foci (Maser *et al* 1997). hRad51 foci presumably represent the homologous recombinational DNA repair mechanism, and hRad50/hMre11 foci the non homologous end-joining mechanims of repair. hRad51 and hRad50/hMre11/NBS1 foci do not occur in the same nucleus, indicating that these pathways of recombinational repair of DSBs are mutually exclusive. The functional difference is supported further by the observation that hRad51 foci additionally form in undamaged cells at S phase.

The DNA damage response is therefore regulated. The choice as to which pathway is followed may depend on cell cycle status (Maser *et al* 1997): A cell in S/G2 may adopt homologous recombination, but a cell in G1 must adopt non homologous end-joining since it lacks sister chromatids. It may also depend on dose.

The Rad50/Mre11/NBS1 DNA Damage Sensor

ATM is involved in the signal transduction pathway which regulates the G1/S DNA damage response by sensing and signalling the presence of DSBs. An AT cell line, mutant in *ATM* (Section 1), displays reduced hRad50/hMre11 foci induction and elevated hRad51 foci induction. This indicates that DSBs are repaired by homologous recombination instead of non homologous end-joining in these cells, and that the functional AT signal transduction pathway triggers non homologous end-joining. IRIF in normal cells is independent of p53 and DNA-PK, indicating that signalling occurs upstream of these factors.

In addition to ATM, NBS1 is required for the hRad50/hMre11 IRIF DNA damage response (Carney *et al* 1998). This supports speculation that NBS1 and ATM act in the same signal transduction pathway. It has been

suggested that NBS1 may be required to assemble the hRad50/hMre11 complex at foci (Trujillo *et al* 1998). The role of NBS1 as a DNA damage sensor would account for the deficient G1/S checkpoint in NBS patients. The carcinogenesis observed in NBS patients would be due to the accumulation of unrepaired chromosomal damage in the absence of functional NBS1. The immune deficiencies might indicate that NBS1 has a role in immunoglobulin rearrangement by V(D)J recombination. The developmental abnormalities might result from slow cell growth or apoptosis.

The limited nature of Xrs2 and NBS1 sequence conservation across the phyla may indicate differences in the DNA damage response signal transduction pathway between different organisms (Carney *et al* 1998).

The G2/M Checkpoint

DNA damage sensed in G2 prevents cell division, M. The cell cycle pauses (arrest) until the damage has been repaired (adaptation) when it resumes (reviewed in Lee *el al* 1998). Arrest depends on checkpoint control genes including *RAD9, RAD17, CDC5* and *CKII.* Adaptation depends on checkpoint control and DSB repair genes. Arrest is permanent and adaptation is prevented in *hdf1* mutants (Lee *et al* 1998). In these cells, the normal 5' to 3' degradation of DSB ends occurs more rapidly than in wildtype. The effects of *hdf1* on the G2/M checkpoint are suppressed by *mre11.* 5' to 3' degradation is not increased in *hdf1 mre11* cells. Therefore adaptation depends on the sensing of single-strand DNA, and the DNA-PK and hRad50/hMre11 complexes have antagonistic effects on the generation of this single-strand DNA. SIR2, SIR3 and SIR4 do not affect the G2/M checkpoint.

A Role in the Regulation of Palindrome Maintenance

It has been suggested that systems for the regulation of palindrome maintenance may exist across the phyla (Connelly *et al* 1999). If a system analogous to palindrome maintenance in *E. coli* exists in eukaryotes, the

Rad50/Mre11 complex might be expected to perform the requisite cleavage of hairpins.

Regulation of Palindrome Maintenance in Yeast

Stimulation and Inhibition of Recombination

rad50/mre11 mutants are recombination deficient in meiosis and hyperrec in mitosis (above), indicating that the wildtype product normally stimulates recombination in meiosis and inhibits in mitosis. This is analogous to SbcCD, which can have either stimulatory or inhibitory effects on recombination (Section 3).

Hairpins Form Meiotic Recombination Hot Spots

Hairpin structures are sites of DSBs and homologous recombination in *E. coli* (Section 3). The effect of hairpin structures on homologous recombination in *S. cerevisiae* was investigated by tetrad analysis (Nag and Kurst 1997). Hairpin structures are acted on inefficiently by the mismatch repair system in yeast. A long (140bp) hairpin-forming palindrome inserted into yeast was not recognised by the mismatch repair system. Unlike other inserted mutations, no 5:3 (post-meiotic segregation) ascospores were obtained after meiotic recombination, but a high level of 6:2 (gene conversion) ascospores were obtained. Gene conversion in these spores was independent of the mismatch repair system and mostly occured in the palindrome to wildtype direction. This indicated that the hairpin-stimulated gene conversion did not occur via a heteroduplex intermediate. Restriction mapping demonstrated that the hairpin was a hot spot for DSB formation, independent of the gene into which it was inserted. It was proposed that the 6:2 products arose from double-strand cleavage of the hairpin, degradation of the palindrome sequence to generate a resected DSB, and homologous recombinational repair using the intact homologue.

Secondary Structures in Mitotic Recombination

A site for replication fork arrest in *S. cerevisiae* coincides with a mitotic recombination hot spot (Kuzminov 1995b). This supports the idea that replication pausing has a role in DSB formation in eukaryotes as well as in

prokaryotes. Unlike in prokaryotes, where replication forks meet at the terminus, the high multiplicity of eukaryotic replication forks means that forks can meet anywhere on the genome and there is consequently no bias in the distribution of DSBs generated in this manner (Kuzminov 1995a).

Palindrome Metabolism in Higher Eukaryotes

A long inverted repeat DNA (approximately 6kb) was inserted into mice as part of a multicopy transgene (Collick *et al* 1996). The inverted repeat showed instability both in mitosis and in the germline. A variety of insertion and deletion events occured, including small rearrangements, partial deletions and complete deletions of the transgene. The frequency of the events varied between male and female germline cells. Another long palindrome (15.3kb), introduced into mice in a transgene, was also rearranged in the germline to generate a range of products (Akgun *et al* 1997). Deletions occured both by illegitimate and homologous recombination. In both experiments, the long palindromes were viable but underwent rearrangement to more stable forms. Tolerance to palindromes varies across the phyla (Section 1). It therefore remains to be determined whether a mechanism of regulation of palindrome maintenance operates in higher eukaryotes.

Models for the Mechanism of Cellular Function

De-Protection of Ends

As discussed above, the *Sc*Rad50/*Sc*Mre11/*Sc*Xrs2 complex is a likely candidate for Spo11 removal from meiotic DSBs at or prior to resection. The studies on *mre11S* supported this, indicating that *Sc*Mre11 nuclease is directly responsible for 5' resection of DSBs (Nairz and Klein 1997). However there was residual resection in the absence of *Sc*Mre11. This was taken to suggest that Spo11 can occassionally be liberated by a different nuclease. This function is likely to require other proteins too, since other genes are needed for DSB processing in meiosis, *eg SAE2(COM1).*

This process can be viewed as end de-protection (Connelly *et al* 1998). End protection could regulate meiotic recombination, preventing untimely DSB

resection (Keeney and Kleckner 1995). As mentioned above, the transesterification nuclease reaction of Spo11 is reversible until Spo11 is removed. If the two DNA-bound Spo11 subunits are regulated, religation could be mediated by Spo11 protein-protein interaction.

The role of SbcCD on cleavage of hairpins can similarly be viewed as a type of end de-protection (Section 3).

Chromatin Conformation

The results of the DNA binding studies above are consistent with *Sc*Rad50 contacting DNA directly in its function (Raymond and Kleckner 1993a). It might load co-operatively onto DNA. It might recognise chromatin sites that will become DSBs, bind these sites and then recruit the other proteins involved in DSB formation and processing. Binding could remove inhibitory proteins or change DNA conformation. The *mre11ΔC49*mutant is approximately wildtype for mitosis but displays deficiencies in meiotic recombination and meiotic chromatin configuration (Furuse *et al* 1998). This suggests that the eukaryotic specific C-terminal domain of *Sc*Mre11 is required for meiosis. It was proposed that the positively charged amino acids in the C-terminus interact with partially exposed DSB sites in DNA.

Chromatin conformation has been shown to undergo a phase transition in meiosis which is dependent on *Sc*Mre2 and *Sc*Mre11. Pre-DSB sites in meiosis (Ohta *et al* 1994) and DSB sites in both meiosis and mitosis (Johzuka and Ogawa 1995) are DNase I and micrococcal nuclease hypersensitive, suggesting that the chromatin is exposed. However not all micrococcal nuclease hypersensitive sites are DSB hot spots, and the level of DSBs at a site depends on competition from other sites and the extent of homology between paired homologues at that DSB site (Ohta *et al* 1998).

It was proposed that *Sc*Rad50/*Sc*Mre11/*Sc*Xrs2 binds DNA at exposed sites to further open the chromatin structure (Johzuka and Ogawa 1995). These exposed sites may be nucleosome-free sites. In mitosis, this complex represses spontaneous DSB formation and reduces recombination. Hence *rad50* and *mre11* null mutants have a mitotic hyperrec phenotype. In meiosis, it recruits additional meiosis-specific factors to form a "pre-DSB complex".

The pre-DSB complex mediates interhomologue interactions and interhomologue recognition (Ohta *et al* 1998) and activates the (Spo11) nuclease which is responsible for DSB formation (Johzuka and Ogawa 1995). Hence meiotic recombination is abolished in the null mutants.

This model was tested (Ohta *et al* 1998). Null mutations in *ScRAD50, ScMRE11, ScXRS2* and *ScMRE2* all slightly increase premeiotic micrococcal nuclease sensitivity at DSB hot spots, but do not affect sensitivity at a non hot spot. The sensitivity at DSB hot spots of *mre11* and *mre2* mutants increases through meiosis similarly to wildtype, except the peak sensitivity is less than wildtype. This suggests that these "type 1" proteins may participate in the chromatin transition which occurs between premeiosis and meiosis, and therefore supports the model. The sensitivity of *rad50* and *xrs2* mutants increases through meiosis to a level above wildtype. Mutation of these "type 2" proteins therefore does not affect the chromatin transition. The *mre11* mutant is epistatic to the *rad50* mutant with respect to meiotic chromatin conformation at DSB hot spots. This implies that type 2 functions (*Sc*Rad50, *Sc*Xrs2) are dependent on type 1 functions (*Sc*Mre11, *Sc*Mre2), which probably act at an earlier step.

CHAPTER 2

MATERIALS AND METHODS

Section 1

Materials

General Solutions

General Technique

Solutions to be used for DNA manipulations were generally sterilized in an autoclave at 15lb/in for 20 minutes, or made up from autoclaved stock solutions. Such solutions were autoclaved immediately they had been made up, to prevent changes in concentration or pH. The pH of a solution was adjusted before autoclaving. Such solutions that could not be autoclaved were filter-sterilized; the solution was passed from a sterile syringe through a sterile 0.2µm Acrodisc (Gelman Sciences) which had been prewetted with sterile Milli-Q water (SMQ).

Solutions for use in contact with protein were not necessarily sterile, but were made up with Milli-Q purified (MQ) water (Millipore Milli-Q 50), and the glass- and plasticware used in contact with them were rinsed thoroughly with MQ water.

Certain solutions were made up in dH_2O and were not sterilized, *eg* gel fix solution, stocks of running buffers.

Solutions were made up in beakers, rather than conical flasks, to prevent inaccuracy of measurement due to loss of materials on the sides of the container. Where solutions were stirred, a magnetic hot plate stirrer (Corning PC-351) was used at room temperature, unless otherwise stated. Adjustments in pH were made using a pH meter (Mettler Toledo 320) and a magnetic stirrer at room temperature.

Solutions for General Use

Water

Solutions were generally made up in MQ water. This was found to be purer than standard dH_2O. Where such solutions were not to be autoclaved, SMQ was used. SMQ was aliquoted and stored for up to a week at 4°C.

10mM (1x) TrisCl pH8.0, pH7.5

A 1M stock was made by dissolving 121.1g Tris base (Boehringer Mannheim) in 800ml MQ with 40ml concentrated HCl. This was aided by covering with saranwrap and stirring-heating gently. The solution was allowed to cool to room temperature and the pH was adjusted. The volume was made up to 1L, aliquoted and autoclaved. This stock was then diluted 100-fold in SMQ, aliquoted and stored for up to a week at 4°C.

10x Tris EDTA pH7.5

This stock contains 100mM TrisCl pH7.5 and 10mM EDTA pH7.5. It was made in MQ and autoclaved. Once opened, it was stored at 4°C for up to several weeks. 1xTE was made by diluting 10-fold in SMQ and stored at 4°C for up to a week.

0.5M EDTA pH8.0, pH7.5

93.05g EDTA. $2H_2O$ was stirred vigorously and heated gently until it was partly dissolved in 400ml dH_2O. It did not dissolve completely until the pH had been adjusted with 10M NaOH, the volume had been made up to 500ml, and it had been heated for a further 30 minutes. The solution was aliquoted and autoclaved, aliquoted again and stored for up to several weeks at 4°C.

10mM ATP pH7.0

A 100mM stock was made by dissolving 60mg ATP (Sigma) in 0.8ml SMQ in an eppandorf tube. The pH was adjusted to 7.0 with 0.1M NaOH. The volume was adjusted to 1ml with SMQ, and the solution aliquoted and stored at -70°C. 10mM ATP was made by 10-fold dilution of once-thawed 100mM stock in SMQ and stored for up to several months at -20°C.

0.5M DTT

0.77g DTT (Sigma) was dissolved in 10ml MQ water and autoclaved. This was aliquoted and stored at -20°C for up to several months.

Materials for the Preparation of DNA Substrates

Materials for Making Substrates

Oligonucleotides

Oligonucleotides were synthsised by Oswel Oligonucleotide Service. 78mer, 65mer, 81mer, 56mer, 98mer, 37mer, 41mer, 26Tmer and 26Bmer were synthesised on a 200nmole scale. Thio78mer was synthesised on a 40nmole scale. The sequences of these oligonucleotides are shown, figure 2.1.1. 77Pmer was also used. This has the same sequence as 78mer, but lacks the sugar and base (C) of the 3' terminal nucleotide. The oligonucleotides were not HPLC purified. These stocks were stored at $^{-}20^{o}C$.

Ligation of Oligonucleotides

10x Ligase Buffer

10x ligase buffer (New England Biolabs) was stored at $^{-}20^{o}C$ for up to several months. For reference, it contains 500mM TrisCl pH7.8, 100mM $MgCl_2$, 100mM DTT, 10mM ATP and 0.5mg/ml BSA.

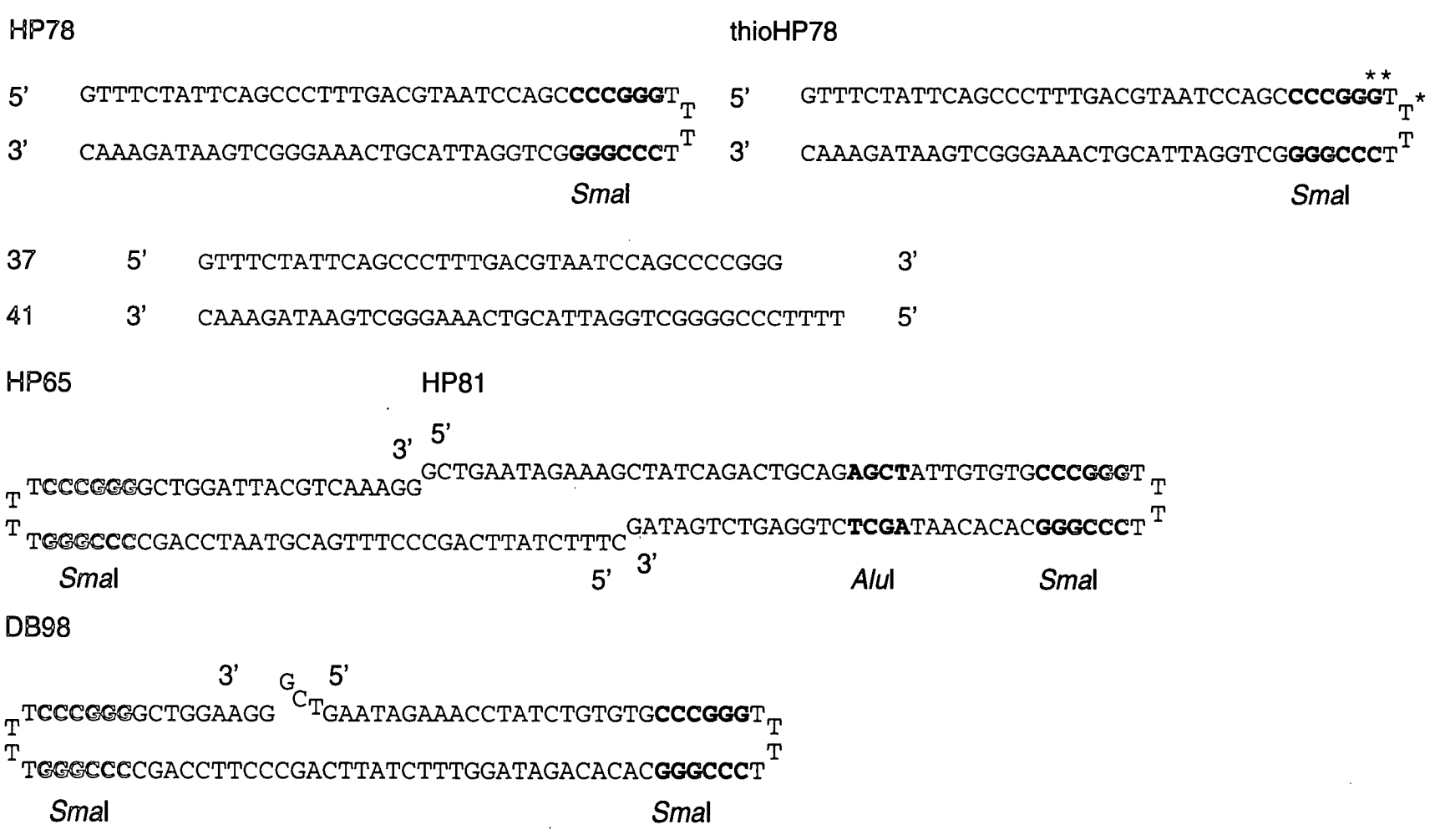
1x Ligase Buffer

4µl 10x ligase buffer were mixed with 4µl 0.5mg/ml BSA and made up to 40µl with SMQ. This was not stored, but used to dilute T4 ligase 30-fold in a precooled eppandorf immediately before use.

T4 DNA Ligase Enzyme

T4 DNA ligase (New England Biolabs) was stored at $^{-}20^{o}C$ for up to several months.

Figure 2.1.1


HP78
5' GTTTCTATTCAGCCCTTTGACGTAATCCAGCCCCGGGT T
3' CAAAGATAAGTCGGGAAACTGCATTAGGTCGGGGCCCT T
SmaI
thioHP78
5' GTTTCTATTCAGCCCTTTGACGTAATCCAGCCCCGGGT T *
3' CAAAGATAAGTCGGGAAACTGCATTAGGTCGGGGCCCT T
SmaI
37 5' GTTTCTATTCAGCCCTTTGACGTAATCCAGCCCCGGG 3'
41 3' CAAAGATAAGTCGGGAAACTGCATTAGGTCGGGGCCCTTTT 5'
HP65
HP81
3' 5'
T TCCCGGGGCTGGATTACGTCAAAGG GCTGAATAGAAAGCTATCAGACTGCAGAGCTATTGTGTGCCCGGGT T
T TGGGCCCCGACCTAATGCAGTTTCCCGACTTATCTTTC GATAGTCTGAGGTCTCGATAACACACGGGCCCT T
SmaI
5' 3'
AluI
SmaI
DB98
3' G 5'
T TCCCGGGGCTGGAAGG CTGAATAGAAACCTATCTGTGTGCCCGGGT T
T TGGGCCCCGACCTTCCCGACTTATCTTTGGATAGACACACGGGCCCT T
SmaI
SmaI

HP56

5' GCCCTTTGACGTAATCCAGCCCCGGGT$_T$ 27T 5' GTTCCAAGCTGGATTACGTCAAAGGGC 3'

3' CGGGAAACTGCATTAGGTCGGGGCCCT^T 27B 3' CAAGGTTCGACCTAATGCAGTTTCCCG 5'

Pal569

5'- **GAATTC**ATTTCAGCATTTATTGGTTGTATGAGAGTAGATAGAAAAAGACAACTCTGGCTTGAAGCTATCAAAAAACTAAGTAGTGATGAA
3'- **CTTAAG**TAAAGTCGTAAATAACCAACATACTCTCATCTATCTTTTTCTGTTGAGACCGAACTTCGATAGTTTTTTGATTCATCACTACTT
*Eco*RI

AACTTTTCAAATATGGAACTCATCAGCCTCATTTCTAAATATGAAGAGTTAAGACGTAATGAACCACAGATTCAAGTGGACGATGATAAA
TTGAAAAGTTTATACCTTGAGTAGTCGGAGTAAAGATTTATACTTCTCAATTCTGCATTACTTGGTGTCTAAGTTCACCTGCTACTATTT

TTCACTAAATTGTTTTATGACAATATCCAGAAATATCTGCTTCGAAT**GAGCTC**GGGACATGCAATTGTTTTATTTACTATCACAAGATTA
AAGTGATTTAACAAAATACTGTTATAGGTCTTTATAGACGAAGCTTA**CTCGAG**ACCTGTACGTTAACAAAATAAATGATAGTGTTCTAAT
*Sac*I

GTAGATG**ACTTTTCGCCAACGA**CATCTACTAATCTTGTGATAGTAAATAAAACAATTGCATGTCCA**GAGCTC**ATTCGAAGCAGATATTTC
CATCTAC**TGAAAAGCGGTTGCT**GTAGATGATTAGAACACTATCATTTATTTTGTTAACGTACAGGT**CTCGAG**TAAGCTTCGTCTATAAAG
15bp loop *Sac*I

TGGATATTGTCATAAAACAATTTAGTGAATTTATCATCGTCCACTTGAATCTGTGGTTCATTACGTCTTAACTCTTCATATTTAGAAATG
ACCTATAACAGTATTTTGTTAAATCACTTAAATAGTAGCAGGTGAACTTAGACACCAAGTAATGCAGAATTGAGAAGTATAAATCTTTAC

AGGCTGATGAGTTCCATATTTGAAAAGTTTTCATCACTACTTAGTTTTTTGATAGCTTCAAGCCAGAGTTGTCTTTTTCTATCTACTCTC
TCCGACTACTCAAGGTATAAACTTTTCAAAAGTAGTGATGAATCAAAAAACTATCGAAGTTCGGTCTCAACAGAAAAAGATAGATGAGAG

ATACAACCAATAAATGCTGAAAT**GAATTC** -3'
TATGTTGGTTATTTACGACTTTA**CTTAAG** -5'
*Eco*RI

Hendrix *et al* 1983

Bacteriological Materials

Strain DL952

DL952 (*ara Δ(lac pro) rpsL thi* Φ80 *lacZ* ΔM15 *ΔsbcCD* pMS7) is a *rec*$^+$ *sbcCD*$^-$ strain containing the plasmid pMS7. pMS7 is a derivative of the pUC18 high copy number vector containing a 571bp palindrome (David Pinder, PhD thesis).

Glycerol Storage Medium

Strain DL952 was obtained from a -70°C frozen stock consisting of 1ml liquid culture to which 5 drops of sterile glycerol had been added.

LAmp Agar

10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl and 15g Bacto-agar (Difco) were dissolved in 800ml dH_2O, and the pH was adjusted to 7.2 with 10M NaOH. The volume was made up to 1L and autoclaved. After cooling to approximately 50°C, ampicillin (Beecham Pharmaceuticals) was added at 100μg/ml.

LAmp Broth

This was made up identically to LAmp agar, except Bacto-agar was not added, and ampicillin was added at a temperature between 50°C and room temperature.

Restriction Digestion

Restriction endonucleases and their incubation buffers were supplied by Boehringer Mannheim and are detailed here for reference. Buffers were stored for up to several months and enzymes were stored for up to several years at -20°C.

10x *Eco*RI Buffer (H)

1mg/ml BSA, 500mM TrisCl pH7.5, 100mM $MgCl_2$, 1M NaCl, 10mM DTT

10x *Alu*I Buffer (B)

60mM TrisCl pH7.5, 60mM $MgCl_2$, 10mM DTT

Restriction Enzymes

Where / indicates the position of cleavage, the 5' to 3' target sequence of *Eco*RI is

G/AATTC

and the 5' to 3' target sequence of *Alu*I is

AG/CT

The incubation temperature of both enzymes is 37°C.

Annealing of Oligonucleotides

100mM TrisCl pH8.0

100mM TrisCl pH8.0 was made as a 10-fold dilution in SMQ from 1M stock (General Solutions). It was stored at 4°C for up to several weeks.

100mM, 1M NaCl

1M NaCl was made by dissolving 29.22g NaCl in 500ml MQ, stirring-heating gently. The solution was autoclaved. It could be stored at room temperature for up to several months. 100mM NaCl was made as a 10-fold dilution in SMQ from this.

Mineral Oil

Mineral oil (Sigma) was stored at room temperature for up to several months.

Materials for Labelling DNA Substrates

10x PNK Forward Buffer, 1x PNK Forward Buffer

10x PNK buffer was obtained from Boehringer Mannheim and contains, for reference, 500mM TrisCl pH8.2, 100mM $MgCl_2$, 1mM EDTA, 50mM DTT and 1mM spermidine. It was stored at -20°C for up to several months. 1x buffer was made from this as a 10-fold dilution in SMQ and not stored.

T4 PNK Enzyme

T4 PNK enzyme (Boehringer Mannheim) was stored for up to several months at -20°C.

^{32}P-γATP 4000Ci/mmol (Easytide), ^{32}P-αdCTP 3000Ci/mmol

Radiolabels were purchased from ICN and generally used within their first half-life (approximately 14 days). Twice the amount indicated was used if the radioisotope had reached its second half life. Radiolabels were stored at -20°C.

Klenow

Klenow (Boehringer Mannheim) was stored at -20°C for up to several months.

dNTPs

These were made up in SMQ. Those at concentrations greater than 50mM were treated as stocks and stored at -20°C for up to several months.

Materials for DNA Purification

Phenol and Chloroform Extractions

1xTE/0.2% β-MSH

4µl β-MSH (0.2% v/v) was added to 2ml 1xTE buffer. This was used immediately.

Phenol

Phenol (88%) was purchased in distilled, liquefied form from Rathburn Chemicals. It was stored as 50ml aliquots, and protected from light, in aluminium foil wrapped polypropylene tubes at -20°C. 50mg 8-hydroxyquinoline, 0.1% w/v (Sigma), was dissolved into the phenol. The phenol was then equilibrated with TE buffer: It was split into two tubes of 25ml, and to each aliquot an equal volume 10x TE pH7.5 was added. The mixture was shaken to emulsify on a rotating wheel (Voss model 4400) for 10 minutes at room temperature, centrifuged at 4200rpm for 5 minutes (MSE Centaur-2 bench centrifuge) at room temperature, and the aqueous (top) layer removed. More TE was added and the equilibration repeated twice more. 2ml 1x TE/0.2% β-MSH was carefully layered onto the phenol. The prepared phenol was protected from light and stored at 4°C for up to several weeks.

25:24:1 Phenol:Chloroform: Isoamylalcohol

Phenol was prepared as above, except it was not layered with 1x TE/0.2% β-MSH. Instead an equal volume of 24:1 chloroform:isoamylalcohol was added to the phenol, the mixture was shaken and centrifuged at 4200rpm (MSE Centaur-2 bench centrifuge) for 5 minutes at room temperature, and the aqueous layer was removed. Then the solution was layered with 1x TE/0.2% β-MSH, and protected and stored as for phenol.

24:1 Chloroform:Isoamylalcohol

24 volumes chloroform were mixed with 1 volume isoamylalcohol. The solution was protected from light in a brown glass bottle with a ground glass stopper, and stored at room temperature for up to several months.

Gel Filtration

Sephadex® G-25 Slurry

Sephadex® G-25 (medium) powder was supplied by Pharmacia. 10g G-25 powder were dissolved in 15ml SMQ. The solution was allowed to equilibrate for 1 to 24 hours at 4°C before being used to make a gel filtration column. The solution was stored at 4°C for up to a week. Columns were not stored.

QIAGEN® Tip100, Tip20 Kits

Large scale plasmid DNA preparation was carried out using a QIAGEN® Plasmid Midi Kit with a QIAGEN® Tip100. Purification of DNA using a QIAGEN® Tip 20 utilised the same solutions. The solutions were provided by the manufacturer and are detailed here for reference. They were stored for up to several months.

Buffer P1

100µg/ml RNase A, 50mM TrisCl pH8.0, 10mM EDTA, stored at 4°C

Buffer P2

200mM NaOH, 1% SDS, stored at room temperature

Buffer P3

3M KAc pH5.5, stored at 4°C

Buffer QBT

750mM NaCl, 50mM MOPS pH7.0, 15% ethanol, 0.15% Triton X-100, stored at room temperature

Buffer QC

1M NaCl, 50mM MOPS pH7.0, 15% ethanol, stored at room temperature

Buffer QF

1.25M NaCl, 50mM TrisCl pH8.5, 15% ethanol, stored at room temperature

QIAquick™ Nucleotide Removal Kit

Buffer PE

The buffer PE concentrate was supplied by QIAquick™. Before use, 24ml ethanol was added to 6ml. The prepared buffer was stored at room temperature for up to several months.

Buffer PB

Buffer PB was supplied by QIAquick™.

Solutions for Concentrating DNA

70% Cold Ethanol in SMQ

7 volumes ethanol were mixed with 3 volumes SMQ and stored at -20°C for up to several months.

3M NaAc pH5.3

40.8g NaAc was dissolved in 80ml MQ water and the pH adjusted to 5.3 using glacial acetic acid. The solution was aliquoted and autoclaved, and stored at 4°C for up to several months after opening.

Solutions for Measuring DNA Concentration

Measurement of the Concentration of Labelled DNA

0.5M Na_2HPO_4 pH7.0

142g Na_2HPO_4 (BDH) were dissolved in 2L MQ water and the pH adjusted to 7.0 with concentrated HCl. 150ml aliquots were autoclaved. Opened stocks were not stored.

EcoScint

EcoScint (National Diagnostics) was protected from light with aluminium foil and stored at room temperature for up to several months.

Materials for the Preparation of DNA Size Markers

End-labelling of Markers

Marker V

Marker V (Boehringer Mannheim) was stored for up to several months at -20°C.

Marker 8-32

Marker 8-32 (Gibco BRL) is a mixture of oligonucleotides. It was stored for up to several months at -20°C.

5x Exchange Buffer

A 10ml stock of 250mM imidazole pH6.3, 60mM $MgCl_2$ and 75mM β-MSH was made up in SMQ. Aliquots were stored at -20°C for up to several months. 1.5mM ADP and 2.5mM ATP were added to 0.5ml aliquots, and stored at -20°C for up to several months.

Maxam Gilbert Cleavage Ladders

1mg/ml Calf Thymus Carrier DNA

1mg/ml calf thymus DNA was stored at -20°C for up to several months.

1M Piperidine

10M piperidine (Sigma) was stored at 4°C for up to several months. 1M piperidine was made freshly by dilution of this stock 10-fold in SMQ at room temperature in a fume hood. This was carried out promptly since 10M piperidine dissolves the plastic eppandorf.

1% SDS Solution

20% SDS (w/v) (Amresco) was aliquoted and stored at room temperature for up to several months. SDS crystallises at approximately 15°C. If this occured, the aliquot was heated to 25°C in a water bath (Grant Instruments) until the solution became homogeneous again. 1% SDS was made by 20-fold dilution, filter-sterilized, and stored at room temperature for up to a week.

Solutions for Gel Electrophoresis

General Solutions for Gel Electrophoresis

Gel Running Buffers

50x TAE

242g Tris base was dissolved in 57.1ml glacial acetic acid, 100.0ml 0.5M EDTA pH 8.0 and dH_2O to 1L, by gently heating on a hot plate stirrer. The stock was protected from light by storing in a brown glass bottle, and stored at room temperature for up to a month. The 1x solution was approximately pH8.0 and was not stored.

50x Low Ionic Strength TAE pH8.0

335ml 1M TrisCl pH8.0 was mixed with 55ml 3M NaAc and 200ml 0.5M EDTA pH8.0 and made up to 1L with dH_2O. This gives a 50x stock concentration of 335mM TrisCl pH8.0, 165mM NaAc, 100mM EDTA pH8.0. This was stored at room temperature and protected from light in a brown glass bottle for up to a month. The 1x solution was approximately pH5.5 and was not stored. This buffer has been used to study RuvA-RuvC-Holliday Junction binding (Whitby *et al* 1996).

10x TBE

108g Tris base and 55g boric acid (Fisher) was dissolved in 20ml 0.5M EDTA pH8.0 and dH_2O to 1L by heating gently on a stirrer-heater. This 10x stock was stored at room temperature and protected from light in a brown glass bottle for up to a month. The 1x buffer was approximately pH7.0 and was not stored.

1x Alkaline Electrophoresis Buffer

5ml 10M NaOH was mixed with 2ml 0.5M EDTA pH8.0 and dH_2O water to a total volume of 1L. This 1x solution, 50mM NaOH, 1mM EDTA pH8.0, was used fresh.

10x Tris-glycine pH8.3

30.2g Tris base was mixed with 188g glycine (Sigma) and 50ml 20% (w/v) SDS solution and made up to 1L with dH_2O water. This stock was pH8.3 and contained 250mM Tris, 2.5M glycine, 1% SDS. It was stored for up to several months protected from light in a brown glass bottle at room temperature.

10x Tris-glycine pH9.4

60.5g Tris base was mixed with 7.5g glycine and 3.72g EDTA and made up to 1L with dH_2O water. This 10x stock contained 500mM Tris, 100mM glycine and 1mM EDTA It was pH9.8 and stored at room temperature protected

from light in a brown glass bottle for up to several months. The 10-fold diluted solution was pH9.4.

50x NaH_2PO_4/KOAc pH6.2, pH6.75

7.5g NaH_2PO_4 and 9.8g KOAc were dissolved in dH_2O water to 100ml to make a 0.5M NaH_2PO_4, 1M KOAc solution. The pH was adjusted with 10M NaOH. The stock was autoclaved, the bottle wrapped in tin foil to protect it from light, and stored for up to several weeks at 4°C.

50x Tris/EDTA pH7.0, pH7.5

12.1g Tris base and 3.7g EDTA were dissolved in dH_2O water to 100ml to make a 1M Tris, 100mM EDTA solution. The pH was adjusted with concentrated HCl. The stock was autoclaved and stored as for 50x NaH_2PO_4/KOAc.

50x TBE pH8.3

53.8g Tris base, 27.6g boric acid and 3.7g EDTA were dissolved in dH_2O water to 100ml. The solution was 4.45M Tris, 4.45M borate and 100mM EDTA, and its pH was 8.3. The stock was autoclaved and stored as for 50x NaH_2PO_4/KOAc pH6.2.

Gel Loading Buffers

4x Denaturing Gel Loading Buffer

10ml formamide (Fluka) was mixed with 200µl 0.5M EDTA pH8.0 and 10mg each of dyes bromophenol blue (Sigma) and xylene cyanol (Sigma). The solution, containing 0.1% (w/v) dyes and 10mM EDTA, was filter-sterilized and stored at 4°C for up to several months. A less prominent dye front could be obtained by using a 10-fold dilution of this loading buffer in formamide, which was stored similarly. The loading buffer was diluted up to 4-fold in the sample to be loaded.

6x Native Gel Loading Buffer with Dye

15mg each of dyes bromophenol blue and xylene cyanol were added to 10ml 30% glycerol in SMQ, to make a solution, containing 0.15% (w/v) dyes, which was diluted up to 6-fold in the samples to be loaded. This was sterilised and stored as for denaturing gel loading buffer.

10% Glycerol Solution

Glycerol (BDH) was diluted 10-fold in SMQ to make 10% glycerol stock, filter-sterilized and stored at 4°C for up to several weeks. This solution was included in binding reactions at a final concentration of 5% for use both in stabilizing a DNA-bound complex and as a gel loading buffer.

6x Alkaline Gel Loading Buffer

To 9.58ml 18% Ficoll (Sigma) in SMQ, 300µl 10M NaOH, 120µl 0.5M EDTA pH8.0 and 15mg each of dyes bromocresol green (Sigma) and xylene cyanol were added. The resulting solution contained 300mM NaOH, 6mM EDTA, 15% (w/v) each of the dyes and approximately 18% Ficoll. This was filter-sterilized and stored for up to several months at 4°C.

1x SDS Gel Loading Buffer pH6.8

1M TrisCl pH6.8 was made in the same way as 1M TrisCl pH8.0, except the pH was adjusted with a different volume of concentrated HCl. To 7.5ml 10% glycerol in SMQ were added 500µl 1M TrisCl pH6.8, 1ml 20% SDS and 10mg bromophenol blue dye. This solution was stored for up to several weeks at room temperature. Immediately before use, 1ml 1M DTT was added, so the final concentration of the solution was 50mM TrisCl pH6.8, 100mM DTT, 2% SDS, 0.1% (w/v) dye and approximately 10% glycerol.

Developing of Autoradiographs

Concentrates for X ray developer solutions (Photosol), XF2A, XF2B, XD2A, XD2B and XD2C, were made up according to the manufacturer's instructions.

Software

ImageQuant™

Phosphorimages of radioactive gels were analysed using ImageQuant™ v3.3 software (Molecular Dynamics).

GrabIt™

Non radioactive agarose gels containing DNA intercalated with ethidium bromide were viewed using a UV transilluminator and Grab-IT™ v2.0 software.

Solutions for Polyacrylamide Gel Electrophoresis

General Solutions

Acrylamide stocks

Acrylamide stocks were supplied by Scotlab. A 40% (w/v) stock of 19:1 acrylamide: N,N'-methylenebisacrylamide was purchased for DNA electrophoresis gel mixtures, and a 30% (w/v) stock of 37.5:1 acrylamide: N,N'-methylenebisacrylamide was used for protein electrophoresis gel mixtures. These were stored at 4°C for up to several months.

5% KOH/MeOH

5g KOH pellets (BDH) were dissolved in 100ml methanol in a fume hood to make a 5% (w/v) solution. This was stored for up to several months at room temperature.

10% AMPS

0.5g AMPS were dissolved in 5ml SMQ to make a 10% (w/v) solution. This was stored for up to a week at 4°C.

Polyacrylamide Gel Fix

250ml glacial acetic acid and 250ml methanol were mixed and made up to 2.5L with dH_2O in a fume hood, to make a 10%/10% (v/v) solution. The solution was stored for up to several months, or until it had been thrice used, in a solvents cabinet at room temperature. 20% polyacrylamide gels were fixed with a similar solution which also contained 5% glycerol to aid drying, *ie* 125ml glycerol were included.

Denaturing Acrylamide Gel Mixtures

Standard Denaturing Acrylamide Gel Mixtures

The volume of 40% 19:1 acrylamide stock, v, required to make an acrylamide mixture for DNA electrophoresis of concentration n% in a mixture volume, V, was calculated as follows:

$$v = (n/40) \times V$$

Denaturing acrylamide gel mixtures were made up in a 75ml final volume by dissolving 31.5g urea in 1x TBE containing the required concentration, n%, of acrylamide. Thus 31.5g urea were added to a volume of 19:1 40% acrylamide, v, 7.5ml 10x TBE and dH_2O to 75ml. The mixture was heated and stirred gently until the urea had dissolved and allowed to cool before being used to pour a gel. The gel mixture was stored for up to a week at 4°C. Mixtures with acrylamide concentrations of 10%, 12%, 15% and 20% were made.

Formamide Denaturing Acrylamide Gel Mixtures

Formamide denaturing acrylamide gel mixtures were made and stored identically to standard denaturing acrylamide gel mixtures, except 10% formamide was added, *ie* 7.5ml formamide (Fluka) was included in the 75ml total volume.

12% SDS-Acrylamide Resolving Gel Mix pH8.8

1.5M TrisCl pH8.8 was made in the same way as 1M TrisCl pH8.0, except 181.7g Tris base was added to 1L, and the pH was adjusted with a different volume of concentrated HCl. To 25ml 1.5M TrisCl pH8.8 was added 0.5ml

20% SDS and 34ml 30% (w/v) 37.5:1 acrylamide, and the volume was made up to 100ml with dH_2O. The mixture was stored for up to a week at 4°C.

5% SDS-Acrylamide Stacking Gel Mix pH6.8

0.5M TrisCl pH6.8 was made as a 2-fold dilution from 1M TrisCl pH6.8 ("1x SDS gel loading buffer"). To 25ml 0.5M TrisCl pH6.8 was added 0.5ml 20% SDS and 13ml 30% (w/v) 37.5:1 acrylamide and the volume made up to 100ml with dH_2O. The mixture was stored for up to a week at 4°C.

Native Acrylamide Gel Mixtures

Native acrylamide mixtures were made up identically to standard denaturing acrylamide gel mixtures, except no urea was added and the mixtures were generally made in 50ml volumes. Thus a volume of 40% (w/v) 19:1 acrylamide *v*/2 was used with an appropriate dilution of running buffer in dH_2O. Mixtures of acrylamide concentrations of 3.5%, 5%, 6.8%, 10% and 12% were used.

Solutions for Agarose Gel Electrophoresis

Native Agarose Gel Mixtures

The concentration of agarose gels was calculated as a (w/v) percentage. Agarose gel mixtures were made by melting an appropriate mass of agarose in an appropriate volume of 1x running buffer in a microwave oven at high power for 2 minutes and swirling regularly. The mixture was allowed to cool to about 50°C before being poured into the perspex mould to prevent cracking of the perspex. Standard agarose (Flowgen) gel mixtures were generally made at a concentration of 0.8%. Standard NuSieve® GTG® agarose (Flowgen) low*Tm* gels were made at a concentration of 3%.

Other Solutions for Agarose Gel Electrophoresis

Denaturing NuSieve® GTG® Agarose Mixture

A 3.5% NuSieve® GTG® alkaline agarose gel mixture was made by dissolving 3.5g NuSieve® GTG® agarose in 100ml dH_2O in a microwave oven at high power for 2 minutes and swirling regularly. Once the solution had cooled to 60°C, 0.5ml 10M NaOH (to 50mM) and 0.2ml EDTA pH8.0 (to 1mM) were added. This mixture was poured as for native agarose gels.

10mg/ml EtBr Stock

Ethidium bromide (Sigma) was made up to a 10mg/ml solution. This was freshly diluted 20-fold for use.

Neutralising Fix

7% (w/v) TCA was made up to 2.5L in dH_2O.

Materials for Protein Analysis and Detection

Solutions for SbcCD

General Solutions

Wildtype SbcCD Protein

SbcCD protein was provided by John Connelly (Chapter 1 Section 3), and stored for up to several years in 1x Buffer A/Mn/BSA at -70°C. It was thawed once to aliquot, and no more than twice after that. Protein to be refrozen was snap-frozen in liquid nitrogen before being placed in -70°C storage.

Mutant SbcCD Protein

Preliminary preparations of mutant SbcCD proteins were prepared by David Pinder (Chapter 5 Section 3).

Buffer B/Mn/BSA

121.1g Tris base, 7.44g EDTA and 58.44g NaCl were dissolved in a total volume of 2L MQ water. The pH of the solution was adjusted to 7.5 using concentrated HCl. This 10x stock contained 500mM Tris, 1mM EDTA and 500mM NaCl and was stored at 4°C for up to several months. The solution was diluted 10-fold to 1x concentration for use: 10ml 10x buffer B was mixed with 10ml glycerol, 65µl β-MSH and SMQ to 100ml. This solution was filter-sterilized and stored at -20°C for up to several months. Immediately before use, 1µl 50mM fresh Mn^{2+} and 1.25µl 20mg/ml BSA stock were added to 50µl the glycerol/β-MSH stored solution.

50mM Mn^{2+}

A 50mM solution of $MnAc.4H_2O$ (Sigma) was made by dissolving 0.122g in 10ml SMQ. The solution was maintained at 4°C and used fresh.

10mM ATPγS

10mM ATPγS (Sigma) was made up as for ATP, except 54.7mg were initially dissolved in SMQ, rather than 60mg.

SbcCD Assays

10x Tris Nuclease Buffer

5ml 1M TrisCl pH7.5 and 4ml 100% glycerol (BDH) were mixed and filter-sterilized. The solution was aliquoted and stored at -20°C for up to several months. To an aliquot of 450µl was added 50µl 20mg/ml BSA (Boehringer Mannheim), 25µl 0.5M DTT and 475µl SMQ. The final solution contained 250mM Tris buffer pH7.5, 20% glycerol, 1mg/ml BSA and 12.5mM DTT. It was stored for up to several months at -20°C.

10x NaAc Binding Buffer pH5.5

A 1M solution of NaAc was made by dissolving 13.6g NaAc (Fisher) in 100ml MQ, adjusting the pH with glacial acetic acid and autoclaving. 10x

NaAc binding buffer was made and stored identically to 10x Tris nuclease buffer, except 5ml 1M NaAc pH5.5 was used in place of 1M TrisCl pH7.5.

10x Tris-Maleate Binding Buffer pH5.5, pH 6.0, pH6.5, pH7.0, pH7.5, pH8.0

A 1M solution of Tris-maleate was made by dissolving 12.12g Tris base and 11.6g maleate (Hopkin and Williams) in 100ml MQ water, adjusting the pH with 1M NaOH and autoclaving. 10x Tris-maleate binding buffer was made and stored identically to 10x Tris nuclease buffer, except 5ml 1M Tris-maleate was used in place of 1M TrisCl pH7.5.

10x MOPS Binding Buffer pH7.0, pH7.5

A 1M solution of MOPS was made by dissolving 20.92g MOPS (Sigma) in 100ml MQ water, adjusting the pH with 1M KOH and autoclaving. MOPS binding buffer was made and stored identically to 10x Tris nuclease buffer, except 5ml 1M MOPS was used in place of 1M TrisCl pH7.5.

10x HEPES Binding Buffer pH7.5, pH8.0

A 1M solution of HEPES was made by dissolving 23.8g HEPES (Sigma) in 100ml MQ water, adjusting the pH with 1M NaOH and autoclaving. HEPES binding buffer was made and stored identically to 10x Tris nuclease buffer, except 5ml 1M HEPES was used in place of 1M TrisCl pH7.5.

10x Tris Binding Buffer pH8.0, pH8.5

A 1M solution of TrisCl was made by dissolving 12.12g Tris base in 100ml MQ water, adjusting the pH with concentrated HCl and autoclaving. TrisCl binding buffer was made and stored identically to 10x Tris nuclease buffer, except 5ml 1M TrisCl pH8.0 or pH8.5 was used in place of 1M TrisCl pH7.5.

10x Glycine Binding Buffer pH8.5, 9.0, 9.5, 10.0

A 1M solution of glycine (Sigma) was made by dissolving 7.52g in 100ml MQ water, adjusting the pH with 1M NaOH and autoclaving. Glycine binding

buffer was made and stored identically to 10x Tris nuclease buffer, except 5ml 1M glycine was used in place of 1M TrisCl pH7.5.

10x NaH_2PO_4/KOAc Binding Buffer pH6.2, pH6.75

This was made and stored identically to 10x Tris nuclease buffer, except the initial mixture was made with 4ml each 50x NaH_2PO_4/KOH buffer and 20% glycerol, a 400µl aliquot of this was taken, and 525µl SMQ was added with the DTT and BSA.

10x Tris/EDTA Binding Buffer pH7.0, pH7.5

This was made and stored identically to 10x NaH_2PO_4/KOH binding buffer, except 50x Tris/EDTA buffer was used.

10x TBE Binding Buffer pH8.3

This was made and stored identically to 10x NaH_2PO_4/KOH binding buffer, except 50x TBE buffer was used.

50mM *p*-APB

12mg *p*-APB (Fluka) were dissolved in 1ml methanol on ice. The eppandorf was covered in aluminium foil to protect the solution from light, and the solution was used fresh.

100mM Na Bicarbonate

4.2g Na_2HCO_3 (BRL) was dissolved in 500ml MQ water. The pH was adjusted to 9.0 with 80ml 100mM Na_2CO_3. The solution was autoclaved. Once opened, the solution was stored at 4°C for up to a week.

100mM Na_2CO_3

5.3g Na_2CO_3 were dissolved in 500ml MQ water. The solution was pH12. The solution was autoclaved. Once opened, the solution was stored at 4°C for up to a week.

50mM Mg^{2+}

0.123g $MgSO_4.7H_2O$ (Fluka) was dissolved in 10ml SMQ and used fresh at 4°C.

50mM Zn^{2+}

0.068g $ZnCl_2$ (Sigma) was dissolved in 10ml SMQ and used fresh at 4°C.

50mM Ni^{2+}

0.131g $NiSO_4$ (Sigma) was dissolved in 10ml SMQ and used fresh at 4°C.

50mM Ca^{2+}

0.074g $CaCl_2$ (Sigma) was dissolved in 10ml SMQ and used fresh at 4°C.

50mMFe^{2+}

0.099g $FeCl_2.4H_2O$ (Sigma) was dissolved in 10ml SMQ and used fresh at 4°C.

50mM Co^{2+}

0.119g $CoCl_2.6H_2O$ (Sigma) was dissolved in 10ml SMQ and used fresh at 4°C.

50mM Cu^{2+}

0.125g $CuSO_4.5H_20$ (Sigma) was dissolved in 10ml SMQ and used fresh at 4°C.

10mM GTP

10mM GTP (Sigma) was made up as for ATP, except 58.9mg were initially dissolved in SMQ, rather than 60mg.

10mM ADP

10mM ADP (Sigma) was made up as for ATP, except 48.7mg were initially dissolved in SMQ, rather than 60mg.

10mM GDP

10mM GDP (ICN) was made up as for ATP, except 50.7mg were initially dissolved in SMQ, rather than 60mg.

10mM dATP

10mM dATP (ICN) was made up as for ATP, except 89.6mg were initially dissolved in SMQ, rather than 60mg.

Single Nucleotide Addition using TdT

5x TdT solution 1

5x TdT reaction buffer pH6.6 was supplied by Boehringer Mannheim. For reference, it contains 1M potassium cacodylate, 125mM TrisCl, 1.25mg/ml BSA. It was stored at -20°C for up to several months.

10x TdT solution 2

TdT solution 2 was supplied by Boehringer Mannheim. For reference it contains 25mM $CoCl_2$. It was stored at -20°C for up to several months.

TdT

TdT was supplied by Boehringer Mannheim at a concentration of $25x10^3$ U/ml.

10μM ddATP

10mM ddATP (ICN) was made up as for 10mM ATP, except 55mg were initially dissolved in SMQ, rather than 60mg. Then the solution was diluted 1000-fold in SMQ. This was stored at -20°C for up to several months.

Solutions for the Detection of SbcCD

Silver Staining

A BioRad Silver Stain Plus Kit was used.

Stain Solution

Silver Complex solutionReduction Moderator solution 35ml MQ water in a 2L beaker was stirred continuously with a magnetic stirrer, whilst 5.0ml Silver Complex solution (solution I, BioRad) were added. Then 5.0ml Reduction Moderator solution (solution II, BioRad) were added. And then 5.0ml Image Development Reagent (solution III, BioRad) were added. 50.0ml Development Accelerator solution (BioRad) was then mixed into the staining solution which was immediately added to the gel.

Silver Stain Stop Solution

5% (v/v) acetic acid was made up in MQ water and used fresh.

Western Blotting

1x Transfer Buffer

Transfer buffer was made without methanol. 11.64g Tris base and 5.86g glycine were dissolved in 3.75ml 20% SDS and MQ water to 2L, to make a solution containing 48mM Tris, 39mM glycine and 0.037% (v/v) SDS, at pH9.2. The solution was stored for up to a week at 4°C.

Ponceau S Stain

Ponceau S was supplied by Sigma. It was stored at room temperature for up to several months.

Coomassie Blue Stain

Coomassie Blue Protein Assay Reagent (Pierce) was stored at room temperature for up to several months.

Coomassie Blue Destain

450ml methanol and 100ml acetic acid were mixed and made up to a volume of 1L in dH_2O in a fume hood. This was stored for up to several months at room temperature in a solvents cabinet. Initial gel rinses were performed in destain which had be used up to 3 times. The overnight destain was performed in unused destain.

10x PBS

A 10x stock of PBS was made with 80g NaCl, 2g KCl, 14.4g Na_2HPO_4, 2.4g KH_2PO_4 and 800ml MQ water. These were dissolved by gently stirring-heating, and the pH adjusted to 7.4 with concentrated HCl. The volume was made up to 1L with MQ water and the solution was aliquoted and autoclaved. Once opened, this was stored for up to a week at 4°C. PBS was used fresh at a 1x concentration, 10-fold diluted from the stock in MQ water.

Blotto/Tween Buffer

10g dried milk powder (Marvel) were dissolved in 200ml 1x PBS with 400µl Tween 20 (polyoxyethylene sorbitan monolaurate, Sigma) by stirring. This solution contains 5% Marvel and 0.2% (v/v) Tween 20. It was used fresh. Antifoam A was not used; it can be added to reduce bubble formation. Sodium azide was not used; it can be added to prevent fungal growth, but is poisonous.

Tween Buffer

40µl Tween 20 was mixed with 20ml 1x PBS and used fresh.

NaCl/TrisCl pH7.5

150ml 1M NaCl and 100ml 0.5M TrisCl pH7.5 were mixed with MQ water to 1L to make a solution of 150mM NaCl and 50mM TrisCl pH7.5. This was used fresh.

Chromogenic Reagents

NBT and BCIP were supplied by Promega and stored at -20°C for up to several months.

Alkaline Phosphatase Stop Solution

200µl 2mM EDTA pH8.0 was made as a 250-fold dilution from 0.5M EDTA pH8.0 in MQ water (0.8µl in 192µl) and used fresh.

Polyclonal Antibodies

Anti-SbcC (664) and anti-SbcD (663) were supplied by John Connelly and stored at 4°C for up to several months. Rabbit anti-immunoglobin G (Promega) was stored at 4°C for up to several months.

Section 2

Methods

Experimental Technique

General Technique

All DNA manipulations were performed on ice and in non-siliconised eppandorf tubes unless otherwise stated. Other manipulations were performed at room temperature unless otherwise stated. In all reactions, reaction volume was made up with SMQ water; water was added first, buffer second and enzyme last; the reaction was mixed by pipetting up and down. All dilutions of enzyme stocks were carried out by adding solution to protein, except when initiating reactions when protein was added to the reaction volume.

Safety Measures

Ionizing Radiation

All radioactive manipulations were carried out with care using suitable screening and containment devices. Where possible, manipulations were conducted at a distance. A lead apron was worn for the excision of radioactive material from polyacrylamide gels. Exposure to ionizing radiation was monitored using a chest badge and finger monitors. Radioactive waste was screened and contained during storage, which was for a minimum duration, before disposal in a manner approved by the university. All aspects of handling of radioactive material were approved by the Scottish Office HM Industrial Pollution Inspectorate.

UV Radiation

Protective clothing and headgear was worn during all manipulations involving exposure to UV light.

Hazardous Chemicals

Treatment of hazardous chemicals was in accord with the manufacturers' recommendations. Suitable treatment of particularly hazardous chemicals (*eg* piperidine, acrylamide) has been noted in the method descriptions.

Preparation of DNA Substrates

Making Substrates

Synthetic Oligonucleotide Substrates

Oligonucleotide were synthesised by Oswel, figure 2.1.1, purified by UV shadowing, end-labelled, and purified again before use as DNA substrates for SbcCD. Those of these oligonucleotides that are palindromic have been called HP or DB (depending on the secondary structures formed).

Annealing of Oligonucleotides

Double-strand DNA was formed from a labelled oligonucleotide and its unlabelled complement. A concentration of unlabelled oligonucleotide equal to twice that of the labelled oligonucleotide was used to draw the reaction equilibrium towards complete annealing of the labelled strand.

15pmol labelled, purified oligonucleotide was mixed with 30pmol unlabelled, UV shadowed complementary oligonucleotide, 4µl Tris-Cl 100mM pH8.0 and 4µl NaCl 100mM in a total volume of 40µl. Alternatively, if labelled oligonucleotide was less concentrated, 1.5pmol labelled, purified oligonucleotide was mixed with 3.0pmol unlabelled, UV shadowed oligonucleotide and 3µl NaCl 1M in a total volume of 30µl. The annealing reaction was set up in a 0.5µl eppandorf tube and sealed with a drop of fresh mineral oil. It was subjected to the following thermocycles in a PCR machine: 3 minutes at 98°C, 10 minutes at 65°C, 10 minutes at 37°C and 10 minutes at 28°C. It was then allowed to cool to room temperature. 10µl native gel loading buffer was added and the reaction was native gel purified to remove the remaining single-strand DNA. The final preparation of annealed DNA was not heated above 4°C. Its concentration was measured.

Making Dumbells

Significant time was devoted to inventing and optimising a method for making dumbell DNA substrates. These methods have been detailed in Chapter 3.

Making Pal569

Pouring Plates

Approximately 40ml L Amp Agar was poured into each petri dish, and allowed to set at room temperature.

Growth of Bacteria

A permenant stock of *E. coli* DL952 was stored in glycerol at -70°C. To generate temporary stocks, this was streaked to single colonies on L Amp plates and incubated overnight at 37°C. The plates were stored at 4°C for up to 2 weeks. Starter cultures were grown by inoculating a single colony into 5ml L Amp broth in 1/2oz Bijou bottles and shaking overnight at 37°C. 12 overnight cultures were each set up with 50µl starter culture in 25ml LAmp broth (a 500-fold dilution) and shaken in a 100ml conical flask with a loosely fitting metal cap overnight at 37°C.

Large Scale Plasmid DNA Purification

Large scale purification of plasmid pMS7 was performed using a QIAGEN® plasmid midi kit. Fresh 25ml overnight cultures of DL952 were poured into 8 sterile polypropylene centrifuge tubes, pairs were balanced by adding more culture, and the tubes were centrifuged at 5000rpm (Sorvall centrifuge, SS-34 rotor) for 15 minutes at 4°C. The supernatants were decanted and the pellets each resuspended in 4ml buffer P1. Once the solution was homogeneous, 4ml buffer P2 was added and the mixture was inverted 6 times at room temperature, but left for no more than 5 minutes. Then 4ml buffer P3 was added, the mixture inverted 6 times and left on ice for 15 minutes. Tubes were balanced with SMQ, inverted again and centrifuged at 12000rpm (Sorvall, SS-34) for 30 minutes at 4°C. The supernatants were transferred to 8 fresh polypropylene tubes, balanced with SMQ and respun

at 12000rpm (Sorvall, SS-34) for 15 minutes at 4°C to remove suspended material. During this final spin, 8 QIAGEN® tips 100, supported vertically over a collecting tray, were equilibrated in buffer QBT by applying 4ml buffer to each and allowing them to run through under gravity. This took 5 to 10 minutes. Immediately the spin was complete, the supernatants were applied to the columns and allowed to flow through. The tips were each washed twice with 10ml buffer QC, and then the bound material was eluted with 5ml buffer QF. The elutant was collected in sterile corex tubes and isopropanol precipitated. It was helpful to mark the outside of the tubes before spinning to identify the pellets. Air drying was limited to 10 minutes to increase recovery since, although resuspension was aided by shaking the tubes, pipeting was avoided to prevent shearing of the plasmid DNA. The aliquots were each resuspended in 200µl Tris-Cl 10mM pH8.0, and then combined.

*Eco*RI Digestion of pMS7

20µg pMS7 DNA was mixed in a 200µl volume with 20µl 10x*Eco*RI buffer (Boehringer Mannheim buffer H) and 10µl (100U) *Eco*RI enzyme (Boehringer Mannheim). The reaction was incubated for 2 hours at 37°C, and then stopped with 8µl EDTA 500mM pH8.0 (20mM).

Labelling and Purification of Pal 569

*Eco*RI digested pMS7 was purified from digestion reactants using a QIAGEN® Tip 20, and resuspended in 100µl Tris-Cl 10mM pH8.0. The concentration of DNA in this preparation was calculated from its absorbance. The digested DNA was then 3' end-labelled using Klenow. The DNA was purified from unincorporated radionucleotides with a QIAGEN® Tip 20. This reduced handling of highly radioactive material prior to separation of pal569 from the remainder of pMS7 by gel purification. Separation of pal569 from the remainder of pMS7 before labelling was found to generate pal569 of lower purity, which was end-labelled with lower efficiency.

Snap-cooling of Substrate DNA

Prior to its use in a ligation reaction or in binding or nuclease assays, the native conformation of palindromic substrate DNA was assured by boiling and snap-cooling. The substrate DNA, generally in TrisCl 10mM pH8.0, was heated above a beaker of boiling water for 4 minutes, left at room temperature for 2 minutes to avoid formation of networks, and then immediately plunged into ice, where it was left for at least 10 minutes. The DNA was then not reheated above 4°C.

Labelling DNA Substrates

5' End-Labelling of Oligonucleotides

The 5' ends of synthetic oligonucleotides lack a 5' phosphate group and were labelled using the forward reaction of bacteriophage T4 polynucleotide kinase. PNK catalyses the transfer of the gamma phosphate of a nucleotide 5' triphosphate to the 5' hydroxyl terminus of the polynucleotide. If the gamma phosphate of the dNTP is radiolabelled, the DNA becomes labelled. Phosphylation of protruding 5' ends is more efficient than that of blunt ends. Phosphorylation of recessed 5' ends is inefficient.

Hairpin and Single-Stranded Oligonucleotides

20pmol UV shadowed oligonucleotide were mixed in a 40µl reaction volume with 4µl 10x PNK forward buffer (Boehringer Mannheim, supplied with the PNK enzyme), 4µl (approximately 40µCi) ^{32}P-γATP 4000Ci/mmol within its first half-life (easytide, ICN), and 3.5µl (35U) PNK (Boehringer Mannheim). This mixture was incubated for 30 minutes at 37°C. Then 4µl cold ATP 10mM was added, and the mixture incubated for a further 30 minutes at 37°C. This ensured that labelled and unlabelled DNAs were of similar molecular weight. Samples were taken for DNA concentration measurements. The reaction was then stopped by adding EDTA 500mM pH8.0 to a concentration of 10mM (5µl), or denaturing gel loading buffer if the sample was to be loaded directly onto a denaturing gel. Single-strand oligonucleotides were labelled more efficiently than blunt hairpins.

Dumbell Oligonucleotide

To compensate for the inefficient labelling of recessed ends, the labelling concentration of DB98 was twice that of hairpin or single-strand oligonucleotides: 40pmol UV shadowed DB98 oligonucleotide were mixed in a 40µl reaction volume. In all other respects the reaction was identical to single strand or hairpin 5' end labelling.

3' End-Labelling of DNA

Klenow is the C-terminal, larger domain of *E. coli* DNA polymerase I. It has the 5' to 3' DNA polymerase and the 3' to 5' exonuclease activities of DNA polymerase I, but lacks the 5' to 3' exonuclease which is contained in the N-terminal domain. The 3' end of a DNA substrate was filled in with radionucleotides by Klenow, using the 5' overhanging complementary sequence as a template.

Hairpin Oligonucleotide

The 77Pmer oligonucleotide was synthesised with a 5' phosphate. It has the same sequence as 5' end-labelled HP78 except it lacks the 3' terminal cytidine monophosphate, so it has a 5' overhanging end. Klenow generated a 3' end-labelled HP78 from HP77P.

1x PNK forward buffer was made by diluting 10x buffer in SMQ. 1µl Klenow enzyme (Boehringer Mannheim) was diluted in 9µl 1x PNK buffer. 5pmol UV shadowed HP77P oligonucleotide were mixed in a 100µl reaction volume with 10µl 10x PNK forward buffer, 5µl dCTP 60µM, 8µl (80µCi) ^{32}P-αdCTP 3000Ci/mmol within its first half life (Amersham), and 4µl (1U) 10-fold diluted Klenow. The mixture was incubated 6 minutes at 37°C. Samples were then taken for DNA concentration measurements. The reaction was stopped by adding EDTA 500mM pH8.0 to a concentration of 20mM (4µl), or denaturing gel loading buffer if the sample was to be loaded directly onto a denaturing gel.

Pal 569

Pal 569 was generated by *Eco*RI cleavage of pMS7. It was 3' end-labelled by filling in of the 3' recessed ends at the cleaved *Eco*RI sites with Klenow. The reaction was carried out on a mixture of pal 569 DNA and the remainder of the pMS7. The quantity of Klenow used was calculated on the basis that there were four 3' termini to be labelled for every pal 569 molecule, and 1U Klenow was required to end-label 4.2µg of the singly cleaved plasmid (two 3' termini).

Klenow enzyme was diluted 10-fold. 4.2µg DNA mixture (pal569 and pMS7 backbone) were mixed in a 200µl reaction volume with 20µl 10x PNK buffer, 10µl dATP 60mM, 10µl each dCTP, dGTP and dTTP 600mM, 10µl (100µCi) ^{32}P-αdATP 3000Ci/mM within its first half life (Amersham), and 8µl (2U) 10-fold diluted Klenow. In all other respects the reaction was identical to that for 3' labelling of the hairpin oligonucleotide.

Purification of Labelled Oligonucleotides

The labelling mixtures were purified to remove unincorporated radionucleotides, enzyme, ATP or dCTP, and inappropriate salts. One of several different methods was chosen, depending on the circumstances: Gel purification, which additionally removes any remaining termination products, followed by ethanol precipitation, or phenol, phenol/chloroform, chloroform and ethanol precipitation if gel contaminants were suspected; sephadex® G-25 spin column purification; nucleotide removal spin column purification.

DNA purification

Gel purification

The advantage of gel purification over other methods of DNA purification is that it allows a DNA species to be separated from another.

Standard Gel Purification

Samples to be gel purified were subjected to denaturing polyacrylamide gel electrophoresis. The unincorporated nucleotides separated from the major oligonucleotide species within 30 minutes if the gel was run at approximately 50W. Residual termination products took up to 60 minutes to separate completely. On dismantling the gel apparatus, the gel was allowed to peel away from the back (reservoir) plate and to stick to the front (plain) plate. Care was taken to prevent wrinkles from forming under the gel and the gel from stretching. The gel and its supporting glass plate were sealed with saranwrap. Small pieces of tracker tape were numbered with a ball-point pen and stuck to the sides of the gel (*eg* 2 down the left and 3 down the right) to serve as orientation markers. A 30 second autoradiograph was obtained from which the position of the desired labelled oligonucleotide could be determined. The desired band was excised from the gel using a sterile scalpel and mashed up in an eppandorf with a sterile needle. Assuming approximately 2 lanes per eppandorf, 500µl Tris-Cl 10mM pH8.0 were added per eppandorf. The tube was sealed with parafilm and a lid-lock, vortexed, and incubated overnight at 56°C. The polyacrylamide was pelleted by spinning at 12000*g* (Sorvall Microspin 24 centrifuge) for 15 minutes, and the supernatant carefully removed. DNA was recovered by ethanol precipitation.

Native Gel Purification

Native gel purification was used where it was important to retain the secondary structure of the DNA being gel purified. It was undertaken in a similar way to standard gel purification, with the following exceptions: DNA was mixed with native gel loading buffer on ice. It was loaded directly onto a 15% sequencing-sized native polyacrylamide gel which had been

prerun for only 20 minutes at 15W and which was at a temperature of less than 40°C. After separation of the different DNAs and excision of the required DNA, DNA elution from the mashed polyacrylamide was carried out on a turning wheel overnight at 37°C. This was possible because the melting temperatures, T_m, of the oligonucleotides annealed during this work were above 37°C.

UV Shadowing of Oligonucleotides

This method was used to separate the desired oligonucleotides from other products generated during the process of chemical synthesis. These other products include termination products which result from failed extensions, and products carrying protecting groups which result from incomplete deprotection.

100µl Oswel synthetic oligonucleotide and 50µl denaturing gel loading buffer were mixed on ice. This was equal to between approximately 1µg and 30µg DNA (37mer and DB98 respectively), depending on the particular oligonucleotide; 5µg for the hairpin oligonucleotide HP78. The mixture was strongly heated by suspending the eppandorf containing it in tin foil above a beaker of boiling water for 4 minutes. The DNA became fully denatured during this time. The hot mixture was then subjected to denaturing PAGE using a 10% sequencing-sized gel. 5 lanes of 30µl were used, the maximum capacity of the wells. The gel was run until the termination products had separated from the desired oligonucleotide, approximately 45 minutes for HP78. It was useful to note the orientation of the gel at this stage if more than one oligonucleotide had been loaded onto the gel.

The gel was then removed from the gel assembly and placed between two layers of Saranwrap. A TLC plate (Sigma) containing a polyethyleneimine cellulose 254nm fluorescent indicator was positioned under the wrapped gel, fluorescent side upwards, and both were supported on a glass gel plate. The gel was viewed briefly (5 seconds maximum) with a shortwave UV lamp. Care was taken to avoid damaging the DNA with any longer an exposure, and to protect the lamp by first removing drops of buffer with a tissue. The DNA generally appeared as a dark smear on a green background; the desired oligonucleotide formed the major band (excluding material running

at the dye front), termination products formed a light smudge running ahead, and oligonucleotides still carrying protecting groups ran behind. The major band was excised with a sterile scalpel and mashed up with a sterile needle. Allowing appoximately 2 to 2 1/2 lanes per eppandorf, 500µl Tris-Cl 10mM pH8.0 was added, the mixture was vortexed, the cap was sealed with parafilm, and the DNA was incubated overnight at 56°C to elute. The tube was then spun 15 minutes at 12000*g* to pellet the polyacrylamide. The supernatant was carefully removed. The concentration of DNA recovered was estimated by measurement of optical density. The DNA was not ethanol precipitated unless lower in concentration than approximately 18ng/µl.

Phenol and Chloroform Extractions

Phenol denatures proteins and chloroform removes phenol from the nucleic acid solution. In most cases it was sufficient to extract with phenol:chloroform and chloroform, but this was preceded by phenol extraction if contaminants remained. Phenol burns can be treated with PEG300 solution before medical attention is sought.

Phenol Extraction

The DNA to be purified was made up to a volume of 500µl with Tris-Cl 10mM pH8.0, and 500µl phenol was added to it. The mixture was vortexed 5 seconds so that an emulsion formed and spun at 12000*g*, room temperature for 15 seconds. The DNA was contained in the clear, aqueous phase. This layered above the yellow, organic phase because the salt concentration in the DNA was not excessive (less than 0.5M). The upper, DNA phase was removed, care being taken not to remove any interface material, phenol:chloroform extracted and ethanol precipitated.

Phenol:Chloroform Extraction

Phenol:chloroform extraction was carried out identically to phenol extraction except that a 25:24:1 solution of phenol: chloroform: isoamyl alcohol was used in place of phenol. It was not necessary to back-extract, *ie* Tris-Cl was not added to increase DNA yield from the phenolic phase. This was followed by chloroform extraction and ethanol precipitation.

Chloroform Extraction

Chloroform extraction was carried out identically to phenol extraction except that chloroform was used in place of phenol. DNA was concentrated by ethanol precipitation.

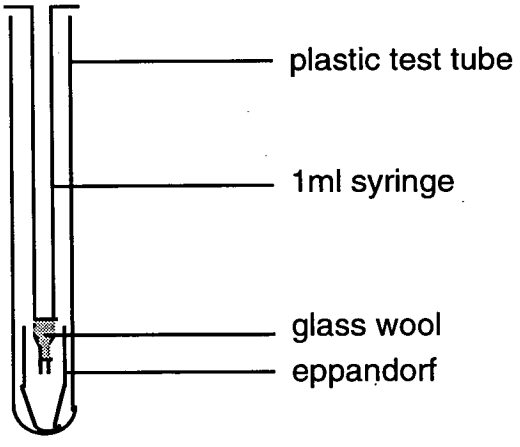
Gel Filtration

Sephadex® G-25 Spin Column

This method of gel filtration was used to remove unincorporated nucleotides from a DNA labelling reaction, but does not remove protein. Recovery of HP78 was found to be approximately 95%.

A thick sephadex® G-25 slurry was made up from G-25 powder in SMQ water and left at 4°C for 1 to 24 hours. G-25 beads in a gel filtration column absorb DNAs shorter than 10 nucleotides. The bottom of a 1ml syringe barrel (Becton Dickinson) was plugged with sterile glass wool, figure 2.2.1. The G-25 slurry was layered into the syringe using a Pasteur pipet to fill it, care being taken to avoid bubbles. The syringe was placed in a lidless eppandorf in a 15ml disposable plastic centrifuge tube (Sterilin). A sephadex column was prepared in this way for each DNA sample, plus one as a balance if required. The columns were spun in a swinging-bucket rotor in a bench centrifuge for 6 minutes at 1600*g*, room temperature, to compact the

Figure 2.2.1



resin. If the volume of the compacted resin did not reach the 0.9ml mark on the syringe, more slurry was added, the column was spun again, 100µl Tris-Cl 10mM pH8.0 was added, and the column was spun a third time.

The eppandorfs were replaced with new ones. The DNA solutions were loaded into the tops of the columns. The columns were sealed with parafilm and spun for 3 minutes at 1600*g*, room temperature. The fraction collected in the eppandorfs contained DNAs longer than 10 nucleotides, and the columns retained unincorporated nucleotides.

Nucleotide Removal Spin Column

QIAquick™ nucleotide removal spin columns (QIAGEN®) were used to purify DNA from unincorporated radionucleotides, contaminant protein, and salts. The columns have a capacity of 10µg.

The volume of DNA solution to be purified was measured. It was mixed in a plastic disposable 20ml tube (Sterilin) with a volume of PN buffer (QIAGEN®) equal to ten times its own volume. It was spun through a column, or more than one if the total amount of DNA exceeded 10µg, in aliquots of 700µl, each spun for 1 minute at 12000*g*, room temperature. The DNA bound the column. The column was then washed with 500µl buffer PE (QIAGEN®), and spun 1 minute at 12000*g*, room temperature, and this wash step repeated once. The residual wash buffer was removed with an additional 1 minute spin at 12000*g*. The DNA was eluted at room temperature by adding a suitable volume of Tris-Cl 10mM pH8.0 to the column (taking care to avoid the sill), allowing this to stand for 1 minute, and spinning 1 minute at 12000*g*. The eluted fraction was reapplied to the column and the elution repeated twice more to maximise recovery. In the case of purification of a labelling reaction, DNA was eluted in a volume of 20µl to 40µl.

Tip 20 Column

The binding capacity of the QIAGEN® Tip 20 is 20µg DNA. DNA recovery varied between 25% and 75%. This method was used to purify the products of *Eco*RI digestion of pMS7 from restriction reagents before radiolabelling.

The Tip 20 was supported in an upright position above a tray at room temperature. It was equilibrated with 1ml buffer QBT (QIAGEN®) for 3 minutes. 200µl DNA sample was mixed with 800µl QBT and loaded to the column. The column was washed 4 times with 1ml buffer QC (QIAGEN®), each time allowing the buffer to flow through the column under the force gravity for 3 minutes. DNA was eluted with 800µl buffer QF (QIAGEN®) into 2 eppandorfs. The recovered DNA was concentrated by ethanol precipitation and resuspended in 100µl Tris-Cl 10mM pH8.0.

Concentrating DNA

Ethanol Precipitation

The volume of the solution to be ethanol precipitated was measured. It was made up with Tris-Cl 10mM pH8.0 or divided between eppandorfs to provide a starting volume of 400µl to 500µl. To 1 volume of the DNA solution were added 2 volumes of cold (-20°C) 100% ethanol followed by 0.1 volume of sodium acetate 3M pH5.3. The solutions were mixed by inversion. If the DNA to be precipitated was shorter than 50 nucleotides, the ethanolic mixture was left for at least 1 hour at -20°C. The mixture was then centrifuged at 12000*g*, 4°C for 45 minutes. Care was taken on removal of the tube from the centrifuge to avoid dislodging the DNA pellet. The ethanolic supernatant was decanted and replaced with 1ml cold (-20°C) 70% ethanol to wash the pellet. This was spun for 15 minutes at 12000*g*, 4°C. The tube was carefully removed from the centrifuge, the supernatant carefully removed from the tube, and the DNA pellet allowed to air dry at room temperature for 15 minutes. In the absence of salt the DNA was translucent and colourless. The pellet was resuspended in a suitable volume of Tris-Cl 10mM pH8.0 for 20 to 40 minutes at room temperature, with regular mixing. For a freshly labelled, gel purified oligonucleotide, a 30fmol/µl solution could be made by resuspending in approximately 40µl. Longer drying reduced the effectiveness of resuspension.

Ethanol precipitation using NH_4Ac, instead of NaAc, can be used as a method of purification since NH_4 ions reduce coprecipitation of dNTPs. However this method was not used because NH_4 ions also inhibit T4 PNK.

Isopropanol Precipitation

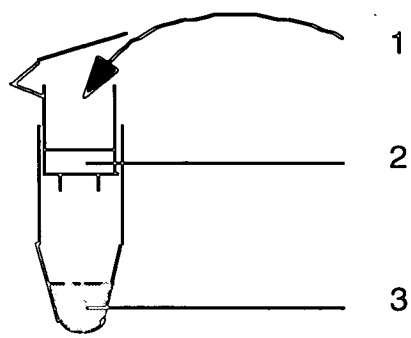
Where a large volume was impractical, DNA was isopropanol precipitated instead of ethanol precipitating. This was carried out identically to ethanol precipitation, except 0.7 volume isopropanol at room temperature was added in place of 2 volumes cold ethanol and a bench top centrifuge (Centra-3, IEC) or Sorvall centrifuge (SS-34 rotor) was used.

Microconcentration

Amicon "Microcon" microconcentrator spin columns were used as a quick method of concentrating ligated dumbell DNA before gel purification. A green cap column (20bp/30nucleotide cut off size) was found to concentrate DB98 DNA approximately 7-fold at 4°C, with a recovery of approximately 60%. A 25µl volume was recovered from a 300µl starting volume.

200µl Tris-Cl 10mM pH8.0 were passed through the column initially to wet it, spinning for 10 minutes at 12000*g*. The collected liquid was removed. A 1µl sample of pre-column DNA was taken, and the rest (at least 200µl) placed into the column, figure 2.2.2 (1). The tube was spun at 12000*g* for 15 minutes, and the collected material removed, figure 2.2.2 (3). The microcon column was inverted over a fresh tube, and the concentrated DNA recovered by pulse spinning, figure 2.2.2 (2). A 1µl sample of recovered DNA was taken and its concentration calculated by comparing its radioactivity (cpm) with that of the pre-column material, the concentration of which was already known.

Figure 2.2.2



Measuring DNA concentration

Estimation of the Concentration of Unlabelled DNA

The absorbance A (or optical density) of a solution is defined as

$$A = \log_{10} (I_0/I)$$

in which I_0 is the incident light intensity and I is the transmitted light intensity (Sambrook *et al* 1989). The absorbance is related to the molar absorption coefficient (extinction coefficient) ε (cm^{-1} M^{-1}), concentration *c* (M), and path length *l* (cm) by

$$A = \varepsilon lc$$

Since DNA absorbs light of wavelength 260nm, the concentration of a solution of unlabelled DNA was estimated by measuring its absorbance of UV light of wavelength 260nm. For double-stranded DNA, A_{260} = 1 represents a concentration of 50μg/ml, and for single stranded DNA, A_{260} = 1 represents a concentration of 20μg/ml (Sambrook *et al* 1989).

Unlabelled DNA suspended in Tris-Cl 10mM pH8.0 was diluted 10-fold in Tris-Cl. 1ml Tris-Cl 10mM pH8.0 was added to each of two quartz cuvettes, and used to calibrate the baseline on a spectrophotometer (Perkin-Elmer Lambda 15). Quartz cuvettes were used because these do not absorb UV light. The contents of one cuvette was replaced with the 10-fold dilution of sample DNA, and the absorbance measured against the blank Tris-Cl control. The approximate concentration of DNA was calculated from the A_{260}, the double- or single-strand nature of the DNA, and the dilution factor.

The purity of the sample could be roughly judged by plotting absorbance over a range of wavelengths from 200nm to 300nm. Protein absorbs at 280nm, and salt contamination absorbs at wavelengths below approximately 250nm.

Measurement of the Concentration of Labelled DNA

Two 1μl samples of labelled DNA were taken immediately after stopping the labelling reaction with EDTA, A and B. A third 1μl sample was taken after purification and concentration of the labelled DNA, C. These samples were

each dropped onto a small piece of filter paper (Fords Gold Medal Demy), approximately 1cm^2. The filters were allowed to air dry for 5 minutes at room temperature whilst pinned to a polystyrene block covered in aluminium foil. Filter B was then washed at room temperature to remove unincorporated radionucleotides: It was washed thrice for 2 minutes in 50ml stirred disodium hydrogenphosphate 500mM pH7.0, and once for 1 minute in 50ml unstirred 70% ethanol, and allowed to air dry. The filters were each placed in a disposable scintillation vial in 2ml Ecoscint, and their radioactivity measured in cpm using a liquid scintillation counter (Packard Tri-Carb 2100TR) set to count ^{32}P decay on a 1 minute programme. A control vial containing 2ml Ecoscint alone was also counted to measure background.

Where b is background radioactivity (cpm), A, B and C refer to the radioactivity of the filters (cpm), U is the radioactivity of the total labelled DNA plus unincorporated nucleotide in the labelling reaction (cpm), W is the radioactivity of total labelled DNA in the labelling reaction (cpm), R is the radioactivity of total recovered DNA (cpm), *mass* is the mass of DNA in the labelling reaction, n is the number of nucleotides in the labelled DNA molecule, *mol* is the number of moles DNA in the labelling reaction, and v is the volume of the labelling reaction (µl), the following calculations were performed:

$U = A - b$

$W = B - b$

$R = C - b$

mass for a single-stranded DNA = $n \times mol \times 330$

mass for a double-stranded DNA = $n \times mol \times 660$

$$\text{incorporation (\%)} = W/U \times 100$$

$$\text{specific activity of recovered DNA (cpm/g)} = W/mass$$

$$\text{concentration of recovered DNA (mol/µl)} = (R \times mol) \;/\; (W \times v)$$

Preparation of DNA Size Markers

End-Labelling of Markers

The dye front interfered with calibration in the region corresponding to DNAs shorter than 10bp on most gels.

Marker V

Marker V was used to calibrate native gels between 8bp and 587bp double-stranded DNA and denaturing gels between 8 and 587 nucleotides. Marker V DNA has 5' phosphate groups and was 5' end-labelled by the exchange reaction of PNK. 2µl DNA was mixed in a 10µl reaction volume with 2µl 5x exchange buffer, 2µl (20µCi) ^{32}P-γATP 4000Ci/mmol within its first half life (easytide, ICN) and 0.5µl (5U) PNK enzyme (Boehringer Mannheim). The reaction was incubated 30 minutes at 37°C and stopped with 1µl EDTA 500mM pH8.0. It was diluted 100-fold before use.

Marker 8-32

Marker 8-32 was used to calibrate native and denaturing gels between 8 and 32 nucleotides single-strand DNA. It has 5' hydroxyl groups and was labelled by the forward reaction of PNK. 1µl DNA was mixed in a total volume of 10µl with 1µl 10x PNK buffer, 1µl (10µCi) ^{32}P-γATP 4000Ci/mmol within its first half life and 1µl (10U) PNK enzyme. The mixture was incubated 30 minutes at 37°C and stopped with 1µl EDTA 500mM pH8.0. 8-32 was diluted 10-fold before use.

Maxam Gilbert Cleavage Ladders

This method is based on that of Maxam and Gilbert (Maxam and Gilbert 1980) and was used to generate a ladder of oligonucleotides terminating in A or G, figure 3.4. Treatment with formic acid depurinates the DNA to be sequenced; hydrazine can be used in an analogous method to remove pyrimidines. Depurinated DNA is then treated with piperidine which cleaves the sugar-phosphate bond either side of a gap. 5' end-labelled product DNAs thus carry 5' and 3' phosphate groups, and 3' end-labelled

products carry 5' phophate and 3' hydroxyl groups. The quantity of DNA used is critical; the ratio of DNA to formic acid determines the range of product sizes. Target DNA must be of high specific activity (containing at least $2x10^4$ cpm) to be visualised overnight by autoradiography.

Piperidine Cleavage

Depending on the ladder required, the oligonucleotide was 5' or 3' end-labelled. After purification, the DNA was resuspended in SMQ, not Tris-Cl 10mM pH8.0 because Tris suppresses the depurination of A and G. 100fmol (4000cpm) DNA was mixed in a 10µl reaction volume with 3µl calf thymus carrier DNA 1mg/ml and placed on ice for 5 minutes. 1.35µl formic acid 100% was mixed in, and the reaction incubated at 37°C for 14 minutes and then placed on ice for 10 minutes. 150µl 1M fresh piperidine was mixed in, the tube closed, and sealed with parafilm, a lidlock, and a lead block weight. 1M piperidine was made freshly by dilution from a 10M stock at room temperature in a fume hood. This was carried out promptly since 10M piperidine dissolves the plastic eppandorf. It was essential to seal the reaction since many of the reagents are highly toxic and volatile. However the reaction was carried out in the supervised radioactive area, not the fume hood. The piperidine reaction was incubated at 90°C for 30 minutes, and then quenched on ice for 10 minutes. It was pulse centrifuged (Stratagene picofuge) and transferred to a fresh eppandorf.

For good resolution of a Maxam Gilbert sequencing ladder, the piperidine must be thoroughly removed by washing. The butanol method was quicker than the water method, but resolution was generally poorer. After the wash step had been completed, a walk in the fresh Pentland air was taken.

Butanol Wash

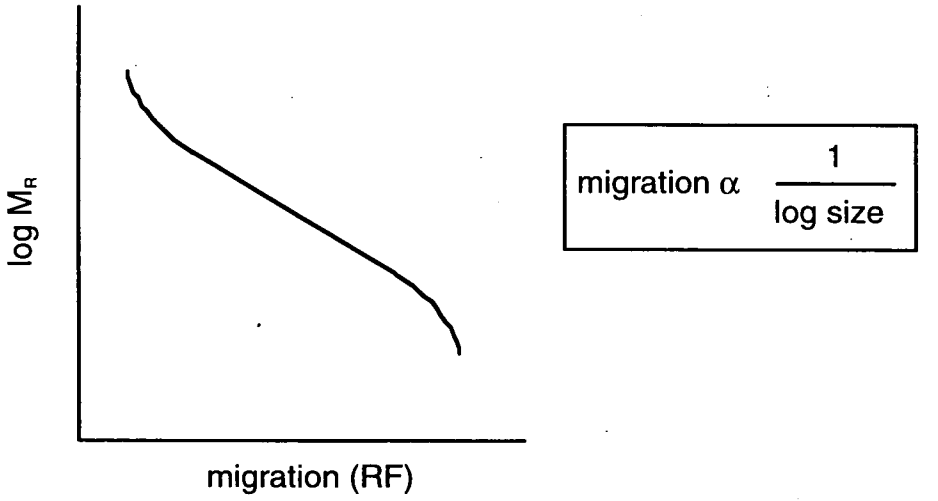
1ml butanol was added. The mixture was vortexed 10 seconds to single phase and centrifuged for 25 minutes at room temperature to pellet the DNA. The tube was carefully removed from the microfuge, and the supernatant carefully and thoroughly removed from the DNA "pellet". (The DNA is not easily visible, partly because it does not form a distinct pellet, but rather a smear). The pellet was dried 3 minutes in a speedvac (GeneVac

SF50). 150µl SDS 1% was added and vortexed 10 seconds to mix. 1ml butanol was added and the tube inverted gently 10 times to mix. And the mixture was centrifuged 10 minutes. The supernatant was removed, or if there were 2 phases, the top phase was removed. The tube was pulse spun again, and the last traces of liquid or top phase were removed. The tube was dried in the speedvac, 5 minutes for a pellet or 15 minutes for a lower phase. All the butanol must be removed because this carries piperidine which prevents good resolution. DNA was resuspended in 10µl denaturing gel loading buffer. For use, 2µl was diluted in 8µl SMQ and loaded onto a denaturing gel.

Water Wash

The reaction was dried in a speedvac for 2 hours. The "pellet" was resuspended in 100µl SMQ, and the solution was transferred to a fresh tube and dried in a speedvac for 30 minutes. The pellet was resuspended in 100µl SMQ and dried again for 30 minutes. The pellet was resuspended a third time in 100µl SMQ and dried again for 30 minutes. The pellet was resuspended in 10µl denaturing gel loading buffer. 2µl was diluted in 8µl SMQ for use, and loaded onto a denaturing gel.

Figure 2.2.3



Resolution of DNA by Gel Electrophoresis

Unpolymerised acrylamide was handled with care. Gels were poured over a sheet of blotting paper to contain the acrylamide. The electrophoretic mobility of a nucleic acid in a matrix (RF) is inversely proportional to the logarithm of its size (kb), figure 2.2.3.

Denaturing PAGE

DNA samples for denaturing PAGE were mixed with denaturing gel loading buffer and heated for 4 minutes by suspension above a beaker of boiling water.

Standard Method for Denaturing PAGE

A sequencing-sized (40cm x 21cm x 0.4mm) denaturing polyacrylamide gel was poured and run in a Sequi-Gen™ Nucleic Acid Sequencing Cell (BioRad). The glass plates were thoroughly cleaned with distilled water and ethanol to remove hydrophobic regions. The back plate (to which the reservoir was attached) was siliconised with a thin layer of Repelcote (BDH) to aid removal of the gel after runnning and of hydrophobic regions. The silicon layer was applied in the fume hood and allowed to dry completely before the plates and 0.4mm spacers were clamped together. This prevented the transfer of silicon onto the front (plain) plate, which would have caused gel handling problems. Any silicon on the front plate could be removed by treatment with KOH/MeOH solution followed by thorough cleaning with distilled water and ethanol.

A 10%, 12%, 15% or 20% urea-acrylamide gel mixture was made up in 1x TBE buffer. For very denaturing gels, 10% formamide (Fluka) was included in ther mixture. 150µl fresh 10% AMPS and 50µl TEMED (Sigma) were added to 10ml urea-acrylamide gel mixture, mixed, and promptly poured over a strip of blotting paper in a casting tray. The bottom of the glass plate assembly was immediately pressed into the casting tray mixture and held there for up to a minute to allow the mixture to move up between the plates by capillary action and seal them as it polymerised. The assembly was laid at an angle of approximately 20° to the horizontal, top edge raised, and on

blotting paper to contain any small splashes of acrylamide solution. 250µl fresh 10% AMPS and 25µl TEMED were added to 40ml urea-acrylamide gel mixture, mixed, and used to fill the cavity between the glass plates; the mixture was poured slowly down the side, using a 25ml pipet and taking care to avoid bubble formation. A square-tooth comb, with 16 teeth of 9mm width and 0.4mm thickness, was inserted into the top of the gel, parallel to the top of the plates and the top 1mm of the teeth remaining outside the gel cavity. The top of the gel was wrapped with saranwrap containing a piece of blotting paper soaked in 1x TBE. The gel was left to polymerise overnight.

The wrap was removed, the polymerised gel overflow cut off from the top of the gel, and the assembly placed in the gel rig according to the manufacturer's instructions. The reservoir and top tank (cathode) were filled with 500ml 1x TBE buffer, and another 500ml buffer were poured into the bottom tank (anode). The assembly was connected to a power supply (positive terminal to the bottom of the gel) and prerun for approximately 1 hour until it had reached a temperature of 60°C. Samples were heated as described above. Meanwhile the gel was temporarily disconnected, the comb removed (slowly to prevent deformation of the wells), and the wells thoroughly rinsed with a jet of 1x TBE buffer from a 50ml syringe fitted with a narrow gauge needle to remove urea and bubbles. The hot samples were immediately loaded into the wells using sequencing gel loading tips (Sigma), care being taken to distribute the material evenly across the width of the well, and to layer it from the bottom of the well. Precise loading technique was essential since uneven loading is a major cause of poor legibility of banding on gel electrophoresis. More distinct banding was obtained with lower loading volumes *eg* 5µl. However the gel was loaded as quickly as possible to avoid cooling of the gel below 60°C. The gel was reconnected, rapidly reheated to 60°C if necessary, and then run at 60°C at a steady power (approximately 50W, depending on the power supply and the weather) until the DNAs had been effectively separated. This was estimated from the running position of the dye front (Sambrook *et al* 1989).

After electrophoresis, the assembly was dismantled, and the gel allowed to peel off of the back plate onto the front plate. Stretching of the gel was avoided since this distorts the banding pattern. The gel was fixed and dried, and imaged by autoradiography or phosphorimaging.

Denaturing PAGE for the Separation of Ligated Dumbells

Denaturing PAGE was carried out identically to standard denaturing PAGE, except formamide was added to the gel mixture to increase denaturation.

Mini Protean Denaturing PAGE

Small (8cm x 10cm x 1.5mm) denaturing gels were set up using a Mini Protean II (BioRad) gel kit. The glass plates were cleaned with distilled water and then ethanol, and clamped together with the spacers in the casting assembly. 100µl fresh 10% AMPS and 10µl TEMED were mixed with 10ml 15% urea-acrylamide gel mixture made in 1x TBE buffer, and the mixture used to pour the gel with a 10ml pipet. A square-tooth comb was inserted into the top of the gel. The gel was left, uncovered, for 1 hour to polymerise. The gel was set up in the electrophoresis assembly according to the manufacturer's instructions, and the anodic and cathodic chambers both filled with 1x TBE buffer. The tank takes two gels; if only one gel was required, the second postion was filled with 3 glass plates clamped together. The comb was removed and the gel prerun at 50mA on a BioRad 300 power pack for 30 minutes. Samples were heated. Meanwhile the wells of the prerun gel were rinsed thoroughly with a jet of 1x TBE buffer from a 50ml syringe attached to a narrow gauge needle. The heated samples were immediately loaded onto the gel using sequencing gel loading tips. The gel was run at 50mA on a BioRad 200/2.0 power pack until the DNAs had been suitably separated, the time for this being estimated from the running position of the dye front. The electrophoresis assembly was dismantled and the gel dried and imaged. The fixation step was found to be unnecessary for these small gels.

Discontinuous SDS PAGE

An ATTO gel assembly was clamped together as for a native polyacrylamide gel, except the thick (1.5mm) spacer was used. 40ml 12% SDS-acrylamide resolving gel mixture pH8.8 was mixed with 0.5ml fresh 10% AMPS and 40µl TEMED, and poured into the assembly to approximately 1cm below the position of the bottom of the wells. This was allowed to polymerise at room temperature for 1 1/2 hours. 25ml 5% SDS acrylamide stacking gel mixture

pH6.8 was mixed with 0.5ml fresh 10% AMPS and 50µl TEMED, and poured into the assembly on top of the polymerised resolving gel. A square tooth comb with thick teeth was inserted into the top of the gel, which was allowed to polymerise for 1 1/2 hours at room temperature. The gel was set up as for a native ATTO gel, except 1x Tris-glycine pH8.3 running buffer was used and the gel was not prerun since this would disturb the discontinuity of the gel system. The samples were contained in SDS gel loading buffer pH6.8 and heated above boiling water for 4 minutes before being loaded into the gel with sequencing pipet tips. The gel was run at 200V until suitable separation had been attained. It was generally used for Western Blotting.

Native PAGE

Samples for native PAGE were either in a 5% glycerol solution, or were mixed with native gel loading buffer. They were not heated before loading. If dye was absent from the samples, at least one track containing dye was included.

Sequencing-Sized Native Gels

A sequencing-sized native polyacrylamide gel was poured and run in a Sequi-Gen™ Nucleic Acid Sequencing Cell (BioRad). The glass plates were treated and assembled as described for denaturing PAGE. The gel was poured identically to a denaturing gel, except native acrylamide gel mixture (*eg* 12%) in 1x TAE buffer was used instead of urea-acrylamide gel mixture in 1x TBE. The electrophoresis assembly was set up in the same way as for a denaturing gel, except 1x TAE was used in the buffer chambers instead of 1x TBE, and the gel, rig and buffer were pre-chilled and set up at 4°C. The gel was prerun 10 minutes at 15W. The comb was removed and the wells rinsed in 1x TAE using a 50ml syringe and fine gauge needle, although the rinsing step was not as critical as for urea-polyacrylamide gels. The samples were loaded using sequencing tips and the same precautions for precise loading were taken as for denaturing PAGE. The gel was run at a temperature below 40°C at approximately 15W for as long as it took for DNAs to be separated, *eg* 4 hours. The gel was dried and imaged.

Native Hoefer Gels

A Hoefer SE 600 Vertical Slab Unit was used to run native gels (18cm x 16cm 1.5mm). The glass plates were cleaned with distilled water and ethanol, and clamped together with 1.5mm spacers into the casting assembly. For a single gel, 50ml native acrylamide gel mixture in 1x TAE (*eg* 7%) was mixed with 250µl fresh 10% AMPS and 25µl TEMED and poured into the casting unit using a 25ml pipet. Or a similar volume of native acrylamide gel mixture in an alternative buffer was used. The gel was allowed to polymerise and mature uncovered at 4°C overnight. The running buffer was prechilled at 4°C. The comb was removed and the electrophoresis assembly was set up according to the manufacturer's instructions, either at 4°C or at room temperature. If only a single gel was required, a glass plate was clamped in the second postion of the electrophoresis assembly. The anodic and cathodic chambers were filled with running buffer to their full capacities of 1.2L and 0.8L respectively. The buffer was recirculated at the rate of 230ml/hour by a peristaltic pump set to 90% full speed (Pharmacia). The gel was prerun at 180V for 5 minutes on a 200/2.0 power pack (BioRad). The wells were rinsed with buffer and samples were loaded in a similar way as for sequencing sized native gels. The gel was run at 180V, dried and imaged.

Native ATTO Gels

An ATTO gel kit was used to run native gels (16cm x 16cm x 1.5mm). The glass plates were cleaned with distilled water and ethanol. The rubber spacer was positioned on the ridged back plate as it lay horizontal on the bench, the front plate was positioned over this, and the assembly was clamped together using bulldog clips, 2 each for the bottom and each side. Fold-back bulldog clips were used along the bottom so that the assembly could be balanced upright on them. For each gel, 25ml native acrylamide mixture in 1x TAE was mixed with 250µl fresh 10% AMPS and 25µl TEMED, and poured into the mould using a 25ml pipet. Alternatively, a similar volume of native acrylamide gel mixture in a different buffer was used. A square toothed comb was inserted into the top of the gel. The gel was allowed to polymerise uncovered for 1 to 24 hours. The gel was kept at 4°C during overnight maturation. The gel and buffer were prechilled at 4°C. The electrophoresis assembly was set up according to the manufacturer's

instructions, and if only a single gel was to be run, 2 glass plates were clamped in the second position. The electrophoretic chambers were filled with 1L running buffer, and the gel was prerun at 4°C, 170V for 10 minutes (BioRad pac 300). The wells were rinsed with buffer and samples were loaded in a similar way as for sequencing-sized native gels. The gel was run at 170V at 4°C, dried and imaged.

Native Mini Protean II Gels

The Mini Protean II (BioRad) gel kit was poured and set up identically to its use in small denaturing gels, except native acrylamide gel mixture in 1x TAE was used, and the electrophoretic chambers were filled with 1x TAE running buffer. The gel was prerun at 4°C, 5mA (approximately 45V) for 10 minutes on a BioRad 200/2.0 power pack. The wells were rinsed with buffer and samples were loaded in a similar way as for sequencing-sized native gels. The gel was run at 5mA at 4°C, dried and imaged.

Agarose Gel Electrophoresis

Native Midi Agarose Gels

Native midi agarose gels were run for the purpose of visualising candidate SbcCD-DNA complexes.

A concentration of 0.8% agarose in 1x TAE was chosen for optimal resolution of a candidate complex; it resolves DNA approximately 2500bp long (molecular mass 1.6 x 10^6), and the estimated molecular mass of an intact SbcCD protein complex is 1.2MDa. 0.8g standard grade agarose (Flowgen) was added to 100ml 1x TAE and dissolved by heating in a microwave oven for 2 minutes on a high setting, regularly interrupting to mix by swirling. A perspex gel mould (14cm x 11cm x 10mm) was set up with a square tooth comb with 14 thin teeth. The molten agarose was cooled to about 50°C before being poured into the mould to prevent its cracking. The gel was allowed to set at room temperature for 30 minutes. It was then placed in a horizontal gel electrophoresis tank (BRL model H5) and immersed in 1L 1x TAE buffer. The comb was removed, samples were loaded into the wells using normal pipet tips, and the gel was run at 80mA, 170V on a BioRad 200/2.0 power pack until good separation had been attained. For a DNA-

binding test using HP78 substrate, this would take up to 3 hours. The gel was dried and imaged.

A 3% NuSieve® midi agarose gel was poured and run identically to the 0.8% agarose gels, but 3g NuSieve® GTG® agarose was dissolved in 100ml 1x TAE.

Mini Agarose Gels

Diagnostic mini agarose gels were run as stepwise checks in certain DNA manipulations, *eg* after restriction digestion, or during a large scale DNA preparation. They were made at a concentration of 1% agarose in 1x TAE. 0.25g agarose was added to 25ml 1x TAE. The mixture was dissolved and poured as for midi gels, except heating was for 1 minute only, the mould (6cm x 7cm x 7mm) had a comb with thin teeth, and setting took 15 minutes. The gel was immersed in 500ml 1x TAE buffer in a horizontal gel electrophoresis tank (BRL model H6). Samples were loaded as for midi gels, and the gel was run at 20V to 120V (maximum) for as long as required, *eg* 30 minutes. The gel was transferred to a tray containing 100ml 1x TAE and 200µg ethidium bromide (20µl 10mg/ml), covered, and agitated gently for 10 minutes to stain the DNA. The gel was destained by washing 4 times with tap water, and imaged.

Alkaline Agarose Gels

A 3.5% alkaline agarose gel was made for denaturing gel electrophorsis. 3.5g NuSieve® GTG® (low *Tm*) agarose was dissolved in 100ml distilled water by heating in a microwave on high power for 2 minutes and swirling regularly. Once the solution had cooled to 60°C, 0.5ml NaOH 10M (to 50mM) and 0.2ml EDTA 0.5M pH8.0 (to 1mM) were added. The gel was poured and allowed to set as for native midi agarose gels. It was set up and run as for native midi agarose gels, except prechilled 1x alkaline TBE electrophoresis running buffer was used, the samples were loaded in alkaline loading buffer, and the gel was run at a power of 1V/cm (25V) at 4°C. A glass plate was laid on the gel after the samples had moved out of the wells to prevent the gel from detaching from its baseplate, which otherwise may occur as a result of the alkalinity of the buffer. The low temperature prevented the gel from

melting (alkaline gels draw more current than neutral gels). The running buffer was replenished several times during the run. This gel provided good separation of single-stranded DNAs between 70 and 300 nucleotides in length. The gel was fixed in neutralising fix, dried and imaged.

Fixing Gels

Denaturing gels only were fixed. Fixation of urea-polyacrylamide gels removes urea, which otherwise prevents complete drying, to increase resistance to cracking at the low temperatures of autoradiography. Fixation of alkaline agarose gels neutralises them. Fixation of gels also increases resolution. Supported on a gel plate, the gel was lowered into and immersed in a tray containing 2.5L fixing solution. It was soaked for 35 minutes.

Drying Gels

All gels run for the purpose of visualisation were dried, except diagnostic mini agarose gels. Drying increases resolution by diminishing the thickness of the gel and by preventing diffusion which can occur in a wet gel over time.

Polyacrylamide gels were dried using a gel drier: A sheet of blotting paper was wetted with distilled water and placed onto the gel, which was supported on a glass plate. A second piece of blotting paper, this one dry, was placed over the first and pressed firmly onto the gel to blot out some of its water. The sandwich was inverted, so that the plate was on top, and the plate was carefully removed from the gel, which was allowed to peel away and land on the blotting paper. Distortion by stretching or wrinkling of the gel was avoided. The gel was covered in a layer of saranwrap, and dried on a gel drier (BioRad model 583) connected to a pump (BioRad HydroTech™ Vacuum Pump) pulling a vacuum of 48.8 Torr (28" Hg) at a steady temperature of 80°C. A sequencing sized gel took 40 to 60 minutes to dry (10% to 20% respectively), a Hoeffer gel 25 minutes, an ATTO gel 20 minutes, and a Mini Protean II gel 10 minutes.

Agarose gels were dried by placing on a sheet of DE81 membrane (Whatman diethylaminoethyl cellulose ion exchange chromatography paper), itself on a

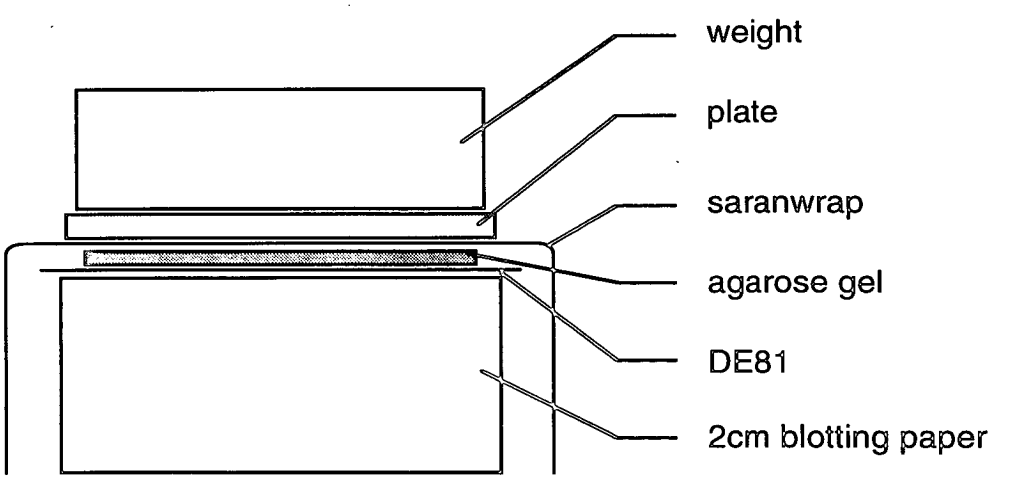
2cm stack of blotting paper sheets, covered with saranwrap and weighted down by a two lead blocks positioned on a plate to distribute the weight evenly, figure 2.2.4. The stack was left at room temperature for 1 1/2 to 2 hours. The dried gel on its DE81 support with two sheets of blotting paper was then removed and sealed in saranwrap.

Imaging DNA on Gels

Autoradiography

Radioactive polyacrylamide and midi agarose gels were visualised by autoradiography. The dried gel was taped into a Cronex autoradiograph cassette (DuPont) fitted with an Xtra-life intensifying screen (DuPont). A sheet of Cronex 4 X-ray film (DuPont), 30cm x 40cm and not preflashed, was fitted over the gel. The exposure was normally overnight, but varied from 10 seconds for certain labelling reactions, to several weeks. Exposures longer than 1 hour were carried out at -70°C to improve resolution. Films were developed, fixed, washed and dried using an X-OGRAPH Compact X2 automatic film processor.

Figure 2.2.4



Phosphorimaging

Radioactive polyacrylamide and midi agarose gels were visualised by phosphorimaging. The dried gel, covered in saranwrap to protect the screen, was taped onto a blanked phosphor storage screen (Molecular Dynamics). The screen was blanked by exposure to an intense homogenous light source (Molecular Dynamics) for 6 minutes. Exposure of the screen to the gel varied from 30 minutes to several days depending on the strength of the signal, but was usually overnight. The image was scanned using a Series 425 PhosphorImager™ (Molecular Dynamics) and ImageQuant™ software. Band intensities were measured and compared using the ImageQuant™ software, by volume integration of rectangular sections of the image. The default settings were generally used, except the intensity range and curve were altered as necessary to visualise weak bands.

Photography

A diagnostic mini agarose gel was examined under 300nm UV light on a transilluminator (UVP Incorporated model TM-20). DNA intercalated by ethidium bromide fluoresced at this wavelength. The image was visualised using GrabIt™ software. Photographs were taken on Polaroid 667 film using a Wratten 25 red filter, an exposure of 1 second and an aperture of f5.6.

Methods for Protein Analysis and Detection

Analysis of SbcCD Activity

SbcCD Protein Preparation

John Connelly provided SbcCD protein which had been purified in an identical manner to Fraction V (Connelly *et al* 1997, Chapter 1 Section 3). This protein stock was stored at -70°C in 1x buffer A with 1mM Mn^{2+} ($MnAc.4H_2O$). Its concentration was 1.8mg/ml and, assuming a protein stoichiometry of $SbcC_8SbcD_8$, its molarity was 1.38μM. Protein was generally not refrozen after being once thawed. Where it was refrozen, this

was done by plunging into liquid nitrogen. Mutant protein was provided by David Pinder (Chapter 5 Section 3).

Nuclease Assay

Unless otherwise stated, nuclease assays were set up as follows: 1µl (69nM) SbcCD protein stock, which had been thawed a maximum of twice, was diluted 10-fold in Buffer B/Mn/BSA. 1µl (6.9nM) was added to a 20µl reaction containing approximately 1.5nM substrate DNA. The reaction also contained 2µl 10x Tris nuclease buffer pH7.5 (1.25mM DTT, 2% glycerol, 100µg/ml BSA, 25mM Tris pH7.5), 5mM Mn^{2+} and 1mM ATP. The reaction was incubated at 37°C for 30 minutes before being stopped by the addition of denaturing gel loading buffer. The reactions were examined by denaturing PAGE using a 10% sequencing sized gel containing formamide. The gel was dried and imaged.

DNA Binding Assays

Binding Reactions

Unless otherwise indicated, protein was preincubated in 1mM Mn^{2+} on ice for 10 minutes prior to its addition to substrate DNA in order to increase stability of the active protein complex. Where protein was diluted, this was done in Buffer B/Mn/BSA. Several methods were used in developing a DNA-binding assay for SbcCD, and a large amount of time was devoted to devising and optimising them. Therefore these methods have been described in Chapter 5.

The Microspin Gel Filtration Assay

Sealed Pharmacia S-400 columns were stored at 4°C in 1x TE buffer (10mM Tris, 1mM EDTA pH 7.6). Before use they were vortexed for 5 seconds at room temperature, to homogenize the liquid bead suspension, their caps were loosened by a quarter of a turn, their plugs were removed (but kept), and the columns were placed in lidless eppandorfs. It was important to ensure that the caps of all columns were equally loosened. The columns were spun at 750 x *g* for 1 minute at 4°C (using an Eppandorf centrifuge model 5145C at 3000rpm). The columns were plugged and 350µl

equilibration buffer was added to each one. Except where stated, equilibration buffer contained 25mM Tris pH7.5, 1.25mM DTT, 5% glycerol, 100μg/ml BSA, 5mM Mn^{2+} and 1mM ATPγS. The caps were tightened, the columns were again vortexed for 5 seconds, and they were sealed (Parafilm) into lidless eppandorfs and rotated on an eppandorf wheel (Voss model 4400) for 2 to 5 hours at 4°C to equilibrate. The plugs were then removed, the caps loosened again and the columns were centrifuged again at 750 x *g* for 1 minute at 4°C. Binding reactions were immediately applied to the top centre of the columns, care being taken not to disturb the bed, and the columns were spun in fresh eppandorfs at 750 x *g* for 2 minutes at 4°C. The columns and the collected fractions were placed in dry scintillation vials and counted in a scintillation counter to measure the approximate radioactivity of unbound and protein bound DNA respectively. A 1 minute program of ^{32}P counting was used. The residual material in the binding tube was also counted to check this was similar between reactions. Experiments were performed at least in duplicate and the mean counts analysed.

Single Nucleotide Addition using TdT

A single nucleotide was added onto the 3' hydroxyl terminus of a 5' end-labelled, snap-cooled hairpin substrate (HP78) using calf intestinal terminal transferase: An 80μl nuclease reaction was set up using the same concentrations of reagents as in the nuclease assay, except 1mM ATPγS was substituted for ATP. The reaction was incubated at 16°C for 30 minutes and then stopped with 20mM (0.8μl) EDTA 0.5M pH 8.0. The reaction was purified by passing through a QIAquick™ Nucleotide Removal Column and eluting in 12μl TrisCl 10mM pH7.5. The sample was heated above a beaker of boiling water for 4 minutes and snap-cooled. 3μl of this DNA was included in two 20μl reaction volumes each containing 4μl (1x) 5x terminal transferase solution 1 (Boehringer Mannheim), 2μl (2.5mM) terminal transferase solution 2 ($CoCl_2$ 25mM, Boehringer Mannheim) and 1μl (0.5μM) ddATP 10μM. 2μl (50U) terminal transferase enzyme (Boehringer Mannheim) was added to one of the reactions. The volumes were mixed, and incubated at 37°C for 15 minutes. Products were analysed by denaturing PAGE, gel drying and imaging.

DNA-Protein Crosslinking using *p*-APB

p-APB Coupling to a Thiophosphate Substrate

A 100mM sodium bicarbonate stock buffer was made up using 4.2g sodium bicarbonate in 500ml SMQ. The pH was adjusted to 9.0 with 100mM sodium carbonate, itself made up using 5.3g sodium carbonate in 500ml SMQ. Both solutions were autoclaved before use. *p*-APB was prepared by dissolving 12mg in 1ml methanol to generate a 50mM solution. This was performed in a aluminium foil-covered eppandorf, to protect it from light. The coupling reactions were set up in aluminium foil-covered eppandorfs using 2.5µM unlabelled oligonucleotide, 0.5nM 5' end-labelled oligonucleotide, 20mM sodium bicarbonate pH9.0 buffer, 45% (v/v) DMSO and 5mM *p*-APB in a 1ml volume made up with SMQ. Duplicate reactions were performed with thioHP78 and HP78 substrate DNA. The coupling reactions were incubated at room temperature for one hour. They were each split into two aliquots of 500ml and ethanol precipitated, still protected from light by aluminium foil. These were resuspended in 25µl Tris-Cl pH8.0. In order to minimise the time that the labile compound was left, the concentrations of recovered DNA were not measured. 50% recovery was assumed.

UV Crosslinking of *p*-APB-Thiophosphate to Protein

The crosslinking reactions were set up in SMQ in 30µl volumes with 200nM DNA, 1x nuclease buffer pH7.5, a total of 5% glycerol, 1mM ATPγS, and 5mM Mn^{2+}. Duplicate reactions were set up using untreated HP78, *p*-APB coupled HP78, untreated thioHP78 and *p*-APB coupled thioHP78 in the presence or absence of 69nM preincubated protein. The reactions were incubated, as for binding reactions, at 16°C for 20 minutes and were protected from light. Crosslinking was performed whilst the samples were supported in open eppandorf tubes at 0°C. A transilluminator (UVP Inc model TM-20) was inverted 4cm above the tops of the eppandorfs and the samples exposed to 300nm wavelength UV light for 7 minutes. The original method advises a 2 minute exposure to 302nm light (Yang and Nash 1994). Long wavelength light reduces non-specific crosslinking and DNA damage

by shorter wavelengths. The eppandorf lids were closed after crosslinking, and the samples no longer protected from light.

Detection of SbcCD

Protein contained in reactions which had been run on native gels was detected by silver staining or by immunodetection. All glassware, plasticware and gloves which were to contact the gel were first rinsed with MQ water. Handling of the gel was minimised. It was useful to run gels in duplicate so that one could be used for protein detection and the other for DNA detection. It was also useful to carry out the reactions in the absence of BSA since this interfered with protein detected but did not adversely affect the activity of the protein. SbcC and SbcD polypeptides were used as protein size standards. Stop solution was prepared before the development step, for optimal development.

Silver Staining

Protein in a native 6.8% polyacrylamide Mini Protean II gel was detected using a BioRad Silver Stain Plus kit. 200ml Fixative Enhancer solution (BioRad) was poured into a 2L glass beaker. The gel was placed into this, and gently agitated for 20 minutes at room temperature on a rotatest shaker (Luckham model R100). The solution was decanted, and the gel was rinsed thrice with 200ml fresh MQ water gently agitated for 10 minutes. During the final 5 minutes of the final rinse, the staining solution was prepared: 35ml MQ water in another 2L beaker was stirred continuously with a magnetic stirrer, whilst 5.0ml Silver Complex solution (solution I, BioRad), then 5.0ml Reduction Moderator solution (solution II, BioRad), and then 5.0ml Image Development Reagent (solution III, BioRad) were added. The solution continued to be stirred as the final rinse was decanted from the gel. 50.0ml Development Accelerator solution (BioRad) was then mixed into the staining solution which was immediately added to the gel. The gel was swirled gently manually until the desired level of staining had been achieved without excessive background. This took 13 minutes. The reaction was stopped by decanting the staining solution and replacing it with 200ml acetic acid 5%. The gel was left in the stop solution for 20 minutes, before agitation in MQ water for 5 minutes. The gel was dried between 2 sheets of gel dryer

cellophane (Cellophane 50, BioRad) which had been wetted with warm tap water.

Western Blotting

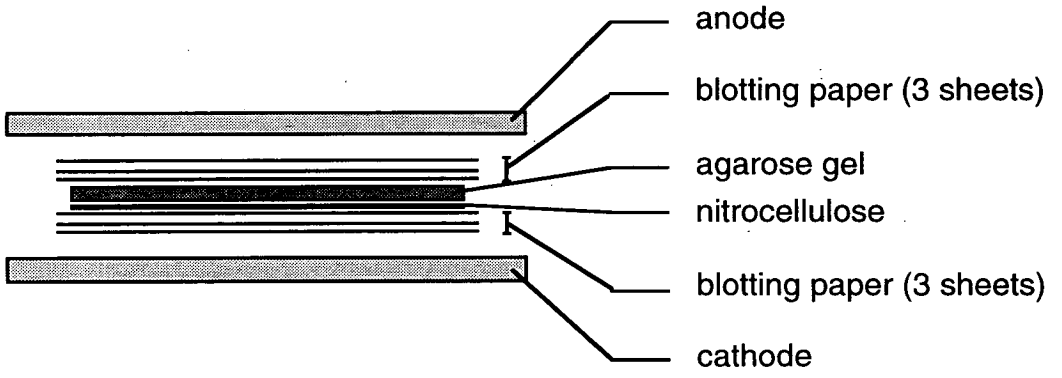
The method is based on that described (Harlow and Lane 1988). A discontinuous SDS gel in an ATTO gel assembly was used, with a 12% SDS-polyacrylamide resolving gel, and a 5% SDS-polyacrylamide stacking gel above it. A protein concentration of up to 4.5µg per well was loaded in SDS gel loading buffer. SbcCD was loaded alongside as a size marker. Where the protein had been purified from a native polyacrylamide gel, it was necessary to take the elutant from excised matter from several tracks. Two gels, or two sets of tracks were loaded if Western Blots for both SbcC and SbcD were desired. The samples were electrophoresed in 1x Tris-glycine buffer pH8.3. On dismantling the gel, the orientation was marked by removing the bottom left corner. The stacking gel was discarded. The gel was soaked in 1x transfer buffer for 30 minutes at room temperature, agitating gently on a Belly Dancer, in a plastic container reserved especially for Western Blots. Meanwhile a Semi Dry Transfer Cell (BioRad) was set up, figure 2.2.5: With minimal handling, a nitrocellulose filter (Gelman Sciences, BioTrace NT) was cut to the size of the gel and the corner which was to align with the bottom left of the gel was removed. The gel was soaked for 5 minutes in distilled water in a second container reserved especially for Western Blots. Where the filter had to be handled, forceps were used. The water was decanted and the filter was soaked in 1x transfer buffer for 10 minutes. The 6 sheets of blotting paper, also cut to size, were soaked in 1x transfer buffer.

Bubbles under the nitrocellulose filter were avoided, but any that formed under blotting paper could be smoothed out. The power applied to the Semi Dry Cell was calculated according to the gel area: A constant current of 3mA was applied per $1cm^2$ to a maximum of 25V. The ATTO gel measured 9cm x 15cm. The current was applied for 30 minutes using a BioRad 200/2.0 power pack. An initial check of protein transfer was obtained by Ponceau S (Sigma) staining the filter for 10 minutes (and destaining with distilled water for 15 minutes) and Coomassie Blue staining the gel for 10 minutes (and destaining with MeOH/acetic acid solution overnight). The position of marker proteins was recorded in pencil. The nitrocellulose filter was rinsed

twice for 10 minutes in 200ml 1x PBS, and then overnight in 200ml Blotto/Tween covered with aluminium foil, agitating gently at room temperature.

The Blotto/Tween was decanted and replaced with a 40ml Blotto/Tween containing 80µl of either anti-SbcC immunoglobin (664) or anti-SbcD immunoglobin (663), both obtained from John Connelly. The solution was agitated gently for 1 1/2 hours at room temperature, whilst the antibodies bound SbcC or SbcD, and the milk protein in the Blotto bound the filter non-specifically to reduce background. The solution was decanted and the filter was washed 4 times, each with 200ml 1x PBS for 5 minutes. This was decanted and replaced with a 20ml solution of 0.2% Tween in 1x PBS containing 2µl anti-immunoglobin and agitated for 1 hour at room temperature. This binds the antibodies and bears alkaline phosphatase *in situ*. The filter was washed thrice with 250ml NaCl/TrisCl solution pH7.5 for 10 minutes each. This solution removes background phosphate before addition of alkaline phosphatase enzyme. A developing solution was made by mixing 10ml alkaline phosphatase buffer with 66µl NBT and 33µl BCIP chromogenic reagents. The salt solution was decanted and replaced with developing solution, and agitated gently until filter development was optimal. This took 30 seconds to 2 minutes. Development was stopped by adding 50ml EDTA 2mM pH8.0 and agitating for 20 minutes. The filter was

Figure 2.2.5



CHAPTER 3

SbcCD IS A HAIRPIN NUCLEASE

Introduction

E. coli strains mutant in *recBC* are viable but recombination deficient (Chapter 1 Section 1). Homologous recombination can be activated by further mutation of *sbcB* (Chapter 1 Section 1). *recBCsbcB* cells are recombination proficient but have very low viability, and consequently mutate spontaneously to *recBCsbcB(C/D)* (Chapter 1 Section 2). The *recBCsbcB(C/D)* cells have wildtype level homologous recombination. Long palindromes generally confer inviability, but *sbcC/D* mutants survive when a long palindrome is introduced into them (Chapter 1 Section 2). SbcCD was therefore recognised as the primary determinant of long palindrome inviability. The effect was seen at replication. In addition plasmid replication may occur by the sigma mode in *recBCsbcB(C/D)* cells, suggesting that SbcB, SbcC and SbcD may have roles in inhibiting rolling-circle replication (Chapter 1 Section 2). Homologues of SbcCD were discovered in a number of other organisms with functions in genetic recombination and in the formation or processing of DSBs (Chapter 1 Sections 2 and 4). Work by John Connelly showed that SbcCD has both an ATP independent single-strand endonuclease activity and an ATP dependent double-strand exonuclease activity, and that the exonuclease is catalytic and processive (Chapter 1 Section 3).

A model for the regulation of palindrome maintenance by SbcCD was proposed to reconcile its multifarious phenotypes (Leach 1994; Chapter 1 Section 3, figure 1.3.2). The model postulates that SbcCD cleaves hairpin structures which arise in the lagging strand at a replication fork and the resulting collapsed replication fork is repaired by homologous recombination. This has been tested genetically (Leach *et al* 1997). Consistent with the model, a long (240bp) palindrome is lethal in *sbcCD*$^{+}$*recA*$^{-}$, *sbcCD*$^{+}$*recB*$^{-}$ and *sbcCD*$^{+}$*recC*$^{-}$ *E. coli* strains. Biochemical experiments on hairpin DNA substrates were undertaken to demonstrate the validity of the model. These have been described in this chapter. Much of the work was carried out in association with John Connelly (University of Edinburgh), whose prior work also provided the basis for it, and he has been appropriately acknowleged. The contents of this chapter have been

Results

SbcCD Cleaves a DNA Hairpin

SbcCD Cleaves Hairpin Oligonucleotides

HP65 and HP81 (hairpins with 5' overhangs), and HP78 and HP94 (blunt hairpins) were used initially as hairpin substrates, figure 2.1.1. These were each 5' end-labelled, heated, snap-cooled and incubated with 6.9nM SbcCD under nuclease conditions in the presence or absence of ATP at 37°C for 30 minutes. Products were analysed by denaturing PAGE (data not shown). The substrates were degraded in the presence of ATP. A major position of cleavage was near the loop.

SbcCD Cleavage of HP78

HP78 was incubated with SbcCD under nuclease conditions for 0 to 30 minutes in the presence of ATP at 37°C. Products were analysed by denaturing PAGE, figure 3.1. Products were sized accurately using a Maxam-Gilbert A/G cleavage ladder.

Where the 3' end was labelled, short incubation times (0 to 5 minutes) generated 4 main cleavage products of sizes between 40 and 43 nucleotides, figure 3.1 lanes 10, 11 and 12. After longer incubation times (5 to 30 minutes), these products disappeared and shorter products between 25 and 9 nucleotides were generated, figure 3.1 lanes 12, 13 and 14. Where the 5' end was labelled, short incubation times generated products between 38 and 35 nucleotides long, figure 3.1 lanes 4, 5 and 6. Longer incubation resulted in the disappearance of these products, and the appearance of shorter products between 20 and 6 nucleotides, figure 3.1 lanes 6, 7 and 8.

The size of the smallest cleavage products cannot be determined from this gel since the dye front interferes with migration of fragments smaller than approximately 4 nucleotides. However these results indicate that SbcCD nuclease is active on this hairpin substrate. They suggest that initial cleavage occurs in the vicinity of the stem-loop junction on the 5' side of the loop and that the substrate and these products are subsequently degraded.

Degradative cleavage occurs on both strands of the stem, as would be expected if the known double-strand processive exonuclease activity of SbcCD is responsible. These results do not show whether degradation occurs by single-strand or double-strand cleavage events. Nor do they show unambiguously the polarity of degradation (below).

Hairpin Degradation Occurs in Two Nucleolytic Steps

Initial Cleavage Products Can be Separated from Subsequent Products

Nucleolytic degradation of the hairpin substrates is extensive (above). Degradation generates a faint smear down the length of a gel track. In order to study the nature of hairpin cleavage by SbcCD, it was therefore necessary to identify conditions under which the processive double-strand exonuclease activity was inhibited, but the hairpin cleavage activity was retained.

Hairpin nuclease activity was abolished in the absence of metal ions (data not shown). The effects of other divalent cations than Mn^{2+} on the double-strand exonuclease activity have previously been described (Connelly *et al* 1997; Chapter 1 Section 3).

A degree of exonuclease inhibition was achieved as follows: HP78 substrate was incubated at 37°C for 30 minutes under nuclease conditions in the presence of various nucleotide cofactors. These nucleotide cofactors replaced ATP at the same concentration (1mM). Products were analysed by denaturing PAGE, figure 3.2. No SbcCD activity was observed in the absence of a nucleotide cofactor, figure 3.2a lane 2, nor in the presence of dATP, ADP or GDP, figure 3.2b lanes 3, 4 and 7 respectively. As previously noted, products were obtained in ATP which varied in size approximately from 40 to 9 nucleotides, figure 3.2a lane 3, figure 3.2b lane 2. In the presence of ATPγS and GTP, figure 3.2b lanes 5 and 6 respectively, a major product of approximately 40 nucleotides was obtained.

The desired exonuclease inhibition was achieved by John Connelly on the basis of these results. ATPγS or GTP was used as a nucleotide cofactor and

incubated with substrate under nuclease conditions at 16°C for 30 minutes. A major product of 41 nucleotides and two minor products of 40 and 42 nucleotides were obtained (data not shown; Connelly *et al* 1998). By phosphorimage analysis it was determined that cleavage in GTP totalled approximately 30% that in ATPγS. Thus ATPγS or GTP could be used at 16°C to uncouple the hairpin cleavage and double-strand exonuclease activities of SbcCD. Since ATPγS is a non-hydrolysable analogue of ATP, it is likely that this is achieved as a result of the enzyme not turning over; binding of ATPγS (or GTP) to SbcCD activates it only for initial cleavage, after which it is blocked. Hydrolysis of ATP may induce a conformational change in SbcCD which allows translocation. The lower activity of SbcCD in GTP than ATPγS may reflect a difference in binding affinity of the nucleotide cofactor.

Initial Cleavage of HP78 Occurs at the 5' Side of the Loop

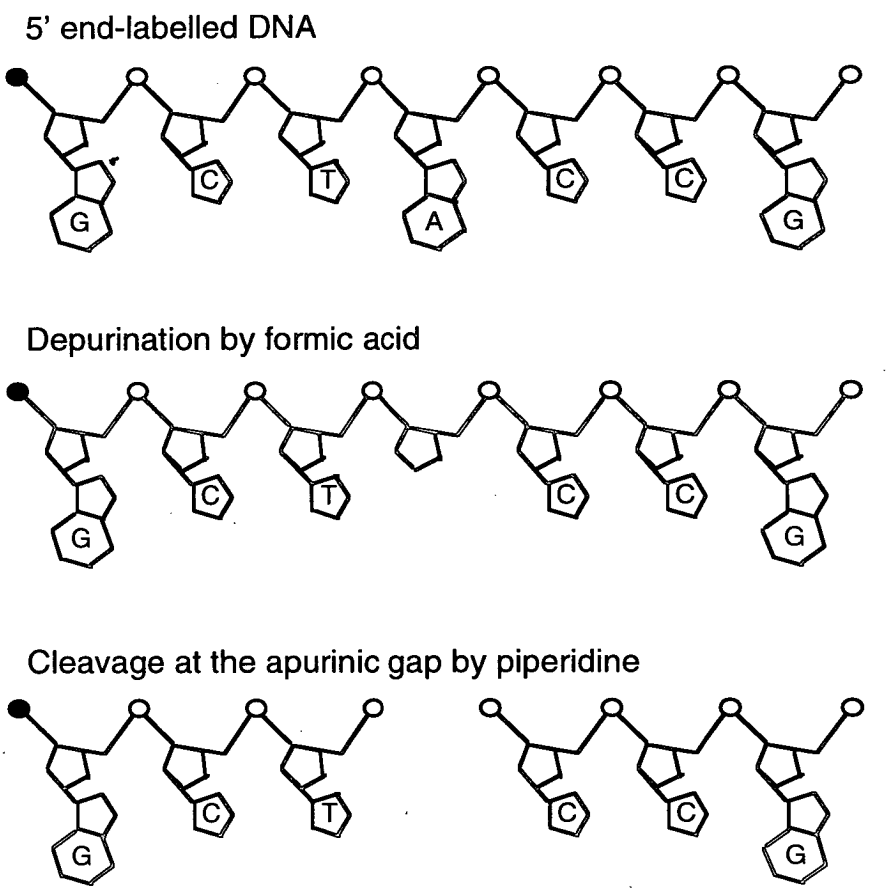
Cleavage of HP78 in ATP, ATPγS and GTP had indicated that the early cleavage event occurs 5' of the stem-loop junction (above). To confirm this the products of early cleavage were accurately sized. A Maxam-Gilbert A/G sequencing ladder was obtained both for 5' and 3' end-labelled HP78 (Maxam and Gilbert 1980, Papavassiliou 1994; Chapter 2). The products of cleavage of these substrates in ATPγS were run on a denaturing polyacrylamide gel alongside the respective ladder, figure 3.3. Differences in salt (or DNA) concentration between tracks on a gel can cause similar sized species to run divergently in different tracks. To acertain that banding was truly comparable between tracks, a mixture of the sequencing ladder and the products was also included, figure 3.3a lane 5, figure 3.3b lane 5.

In analysing the results it is necessary to consider the chemical nature of the Maxam-Gilbert products, figure 3.4. Formic acid treatment of DNA removes purines. Piperidine then cleaves the sugar-phosphate link either side of the apurinic gap to generate products carrying phosphate groups at both of the new termini. A 3' end-labelled DNA would therefore migrate alongside a similarly sized 3' end-labelled Maxam-Gilbert fragment, which has a 5' phosphate and a 3' hydroxyl group. And a 5' end-labelled DNA would migrate behind a similarly sized 5' end-labelled Maxam-Gilbert fragment, which has a 5' and a 3' phosphate group and thus an extra phosphate carrying negative charge. In the latter case, the 5' labelled Maxam-Gilbert

fragment actually migrates approximately 1.5 nucleotide positions ahead of the standard DNA on a 10% denaturing gel (Hyde *et al* 1994).

Assuming that SbcCD cleavage generates a 5' phosphate and a 3' hydroxyl group, the positions of early cleavage of both end-labelled HP78 substrates concur, figure 3.3: The major early cleavage event occurs between nucleotides 37 and 38 proximal to the 5' end of HP78, *ie* exactly at the stem-loop junction. The two minor early cleavage events occur between

Figure 3.4



Reactions which take place during Maxam-Gilbert piperidine cleavage of DNA at purines to generate a A/G sequencing ladder (Maxam and Gilbert 1980). Depurination is partial. Circles represent phosphate groups; filled circles represent ^{32}P; unlettered pentagons represent pentose sugars; lettered pentagons (C, T) and pentagon-hexagons (A, G) represent pyrimidines and purines respectively.

nucleotides 36 and 37, and 38 and 39 proximal to the 5' end of HP78. The minor cleavage products are each generated at approximately 10 to 25% of the frequency of the major one, as determined by phosphorimage analysis.

These results suggest that hairpin cleavage occurs with an initial cleavage event and a subsequent degradative step. Thus the progression of SbcCD digestion of a hairpin can be said to occur in two steps. However this work does not examine the chemical events which occur during a nucleolytic cleavage reaction.

Products of SbcCD Hairpin Nuclease have 5' Phosphate and 3' Hydroxyl Termini

The results of Maxam-Gilbert sizing of the early cleavage products are consistent with earlier, less accurate predictions, figures 3.1 and 3.2, when it is assumed that SbcCD hairpin cleavage generates 5' phosphate and 3' hydroxyl termini. The chemical nature of the cleaved termini of SbcCD products was confirmed by enzymatic tests.

The 3' termini of SbcCD cleavage products was analysed using calf intestinal terminal transferase which adds nucleotides onto a 3' hydroxyl terminus. Early products of cleavage of 5' end-labelled HP78 were obtained in ATPγS. These were denatured by heating, and incubated in terminal transferase buffers and ddATP in the presence or absence of terminal transferase enzyme. The products were analysed by denaturing PAGE, figure 3.5. After treatment with terminal transferase, all cleavage products migrate exactly 1 nucleotide position behind untreated products. This difference in mobility is attributed to an increase in size of the cleavage product by addition of a single nucleotide, ddAMP, onto the 3' hydroxyl terminus. SbcCD cleavage therefore generates a 3' terminus.

The 5' termini of SbcCD hairpin cleavage products were studied using shrimp alkaline phosphatase by John Connelly (Connelly *et al* 1998). This removes a 5' terminal phosphate group. Early cleavage products of 3' end-labelled HP78 were obtained using ATPγS. The products were denatured by heating, and incubated in shrimp alkaline phosphatase buffer with or without shrimp alkaline phosphatase enzyme (United States Biochemical).

The products were analysed by denaturing PAGE (data not shown). After treatment with alkaline phosphatase, all cleavage products are impeded by approximately 0.5 of a nucleotide position relative to untreated products. Such a difference in mobility is expected upon removal of the 5' terminal phosphate, since this decreases the negative charge of the treated product. It is consistent with the generation of products with 5' phosphate groups by SbcCD cleavage.

Generation of Dumbell DNA Substrates

The model proposed to explain SbcCD determination of palindrome-mediated inviability requires that SbcCD hairpin nuclease should recognise a hairpin structure. If this is so, SbcCD should recognise an aspect of the hairpin other than a 3′ or 5′ end. The model therefore predicts that it should cleave a hairpin structure which lacks termini. Based on the end-joining reactions performed by Petra Pfeiffer's group (Beyert *et al* 1994), two methods were devised to make such a substrate.

Preparation of DB146

Design of the Oligonucleotides

HP65 and HP81 have 13 nucleotide 5' overhangs complementary to one another (figure 2.1.1). These were to be ligated to make a dumbell substrate of 146 nucleotides, DB146. The sequences of these hairpins were based on those designed by Beyert *et al.* A 5' terminal dCMP is inefficiently kinased by T4 PNK, so dGMP was chosen instead. 5' overhangs were chosen because they are more efficiently kinased than 5' recessed termini. And to increase the stability of the hairpin conformation, a G-C pair was positioned at the last pair of the duplex before the single-strand region. To increase the efficiency of ligation with T4 DNA ligase, the oligonucleotides were designed with a 5' G and a 3' G which increases the stacking energy of this pair relative to any other pair in the vicinity. A compromise was sought for the length of the stem. A length greater than 27bp increases the stability of the hairpin conformation for labelling and ligation. Too long a stem prevents complete denaturation of the ligated form at the gel purification stage (below).

Preparation of DB146 from Oligonucleotides

1µl 200nmol/µl solution (in SMQ) each 65mer and 81mer were mixed with 3µl SMQ, and the solution was boiled and snap-cooled. A 10µl reaction volume was set up in SMQ adding 1µl 10x T4 DNA ligase buffer, 1µl 10mM ATP and 1µl 1mg/ml BSA. Dilution buffer was made with 3µl BSA and 3µl 10x ligase buffer and made up to 30µl with SMQ. 2µl T4 DNA ligase were diluted 10-fold in 18µl dilution buffer (in a pre-cooled tube) and 2µl of this (80U) were added to the reaction (preincubated at 37°C). The reaction was incubated at 37°C for 2 hours. Oligonucleotides in the absence of ligase were incubated alongside as a control.

Denaturing loading buffer was added to the samples, which were heated and separated by 10% formamide denaturing PAGE. Other percentage gels were tried, but the best separation of the unligated (nicked circle, ncDB) and the ligated (closed covalent circle, ccDB) forms was obtained with a 10% gel. When these forms are completely denatured, they run as a linear and a circle molecule at single-strand marker positions of 146 nucleotides and approximately 500 nucleotides respectively. It was extremely important to preheat the gel to 60°C before loading of the samples (Chapter 2, Section 2). It was also extremely important to ensure that the samples did not cool between heating and loading, that loading was completed quickly and the gel temperature did not drop more than 5°C before electrophoresis started, and that the temperature of the gel quickly returned to 60°C after loading and was maintained at 60°C throughout electrophoresis. If these precautions were not taken and the gel temperature dropped to 55°C, ncDB and ccDB forms migrated close or together and they could not be distinguished for excision. After running, the ccDB146 was excised, gel purified by eluting overnight at 56°C in 10mM TrisCl pH8.0, ethanol precipitated and resuspended in TrisCl.

An aliquot of the preparation was checked by analytical restriction digestion with *Alu*I and *Sma*I. *Alu*I restriction generated a single labelled fragment of 110 nucleotides. *Sma*I restriction generated a single labelled fragment of 126 nucleotides.

Inadequacies of the Method

A major problem with this method is that two ligation events are required to generate the dumbell substrate. Approximately 60% of the hairpin substrate mixture was singly ligated to form ncDB146, and approximately 30% was twice ligated to form ccDB146. Therefore recovery of ccDB146 was low. Attempts to increase the frequency of double ligation included using two rounds of ligation, rephosphorylating ncDB146 between successive rounds of ligation, and using various combinations of labelled and unlabelled oligonucleotide preparations (in case the terminus of one preparation was damaged or faulty). These methods were not successful.

The second major problem was that the existence of two labels complicated the analysis of cleavage products. This would even have affected studies of early cleavage products alone.

The third problem was in maintaining the denatured conformation of the ligated species during electrophoresis. Resolution on an alkaline agarose gel was insufficient. This difficulty could have been lessened if a shorter stem had been utilised. It was initially anticipated that multimerisation would be a problem, but in practice multimers only formed if the DNA was not snap-cooled. They were easily distinguished on the gel and could have been distinguished by restriction analysis.

Preparation of DB98

To circumvent some of the problems associated with DB146, a third oligonucleotide was designed by John Connelly. This was a 98mer and based on the sequence of DB146, figure 2.1.1. It has the 24 nucleotides adjacent to the 81mer loop and the 24 nucleotides adjacent to the 65mer loop. Its centre is composed of the overhang sequences plus the adjacent 10 nucleotides from the 81mer and the adjacent 8 nucleotides from the 65mer. It was self-ligated to generate ccDB98.

2pmol boiled and snap-cooled DB98 were added to 10µl 10mM ATP, 10µl 10x ligase buffer, 20µl 1mg/ml BSA and 10µl 10-fold diluted (400U) ligase enzyme in a 100µl reaction volume made up with SMQ. The dilution buffer

for the ligase consisted of 10µl BSA and 5µl 10x ligase buffer made up to 50µl with SMQ. 5µl ligase enzyme (in a pre-cooled tube) was diluted by the addition of 45µl dilution buffer before being added to the reaction (preincubated at 37°C) to inititate it. The ligation was incubated at 37°C for 1 hour and stopped by the addition of EDTA pH8.0 to 20mM.

DNA was concentrated by passing through a micoconcentrator (cut off size 20bp/30 nucleotides). Denaturing loading buffer was added to it. The DNA was heated to denature it and then separated by denaturing PAGE. A 10% formamide denaturing gel was used and the same precautions taken for maintaining denaturation as in the method for making DB146, figure 3.6a. An RF value is calculated as the ratio of the distance migrated by the species of interest : the distance migrated by a standard. Where migration was measured against that of ncDB98, the RF value of ccDB98 was 0.48. This band was excised, eluted and recovered as for DB146.

Advantages of the Method

The use of a single oligonucleotide circumvented the problems associated with the double ligation and double label of the DB146 substrate. It was necessary to separate the ligation products under highly denaturing conditions, as for DB146, but this was achievable. Inclusion of 10% formamide in the denaturing gel aided maintenance of highly denaturing conditions. Approximately 80% of DB98 oligonucleotide was ligated to ccDB98 as determined by phosphorimage analysis.

SbcCD Hairpin Nuclease does Not Require DNA Termini

A covalently closed, internally labelled dumbell substrate, ccDB98, was generated for use in testing whether SbcCD hairpin nuclease has a requirement for free DNA ends. The proposed model of palindrome maintenance supposes that it does not. The ccDB98 substrate was purified from a gel as a single species, figure 3.6a, however some ncDB98 was unavoidable as the product of radiolysis, figure 3.6b. ccDB98 was incubated with SbcCD under nuclease conditions for 30 minutes at 16°C in the presence of ATPγS. Cleavage products were analysed by denaturing PAGE. Products

of approximately 98 and 49 nucleotides in length were obtained, figure 3.6b. These are consistent with cleavage at the 5' stem-loop junctions, occuring at either one or both ends of the molecule. However some of the 98 nucleotide species will have been generated by random cleavage by radiolysis. These results indicate that SbcCD hairpin cleavage does not require DNA termini and that the model, which implies that SbcCD recognises a specific aspect of a hairpin structure, may be correct.

John Connelly similarly incubated ncDB98 with SbcCD. The expected 98 and 46 nucleotide products were obtained. Furthermore a novel nucleotide product was also obtained of a size corresponding to cleavage opposite the nick (John Connelly unpublished).

Incubation of ccDB146 or ccDB98 with SbcCD under nuclease conditions in the presence of ATP resulted in such extensive degradation that no products were visible on the gel (data not shown). The same result was obtained when 0.35nM SbcCD was incubated for only 1 minute (data not shown), and when DE81 paper was used for gel transfer before fixing and drying, indicating that low molecular weight species were not lost during transfer (data not shown). This is consistent with the processive exonuclease activity of SbcCD being responsible for hairpin degradation.

Discussion

Summary: Biochemical Evidence for the Proposed Model

SbcCD nuclease is active on DNA hairpin structures, including short synthetic and long substrates. Hairpin degradation occurs in two steps; initial cleavage occurs on the 5' side of the stem-loop junction of the HP78 substrate tested, followed by subsequent degradation of the double-strand stem. Initial cleavage occurs in ATPγS (or to a lesser extent in GTP), and ATP is required for the subseqent events. This suggests that initial cleavage requires binding but not hydrolysis of a nucleotide cofactor, whereas degradation requires ATP binding and may involve hydrolysis. The prefered site of initial cleavage of the HP78 test substrate is exactly between the stem and the loop. The minor cleavage sites are located one nucleotide away from this on either side. An initial cleavage product has a 5' phosphate and a 3' hydroxyl terminal group. It is expected that all products generated by SbcCD nuclease have identical terminal chemistry. An efficient method was devised for making a dumbell substrate. Upon exposure of the dumbell to SbcCD, early cleavage products were obtained corresponding to cleavage at both stem-loop junctions.

Conclusion: SbcCD Cleaves a DNA Hairpin Substrate

These results demonstrate that SbcCD recognises and cleaves a DNA hairpin substrate and does not require DNA termini. This supports the proposed model for palindrome maintanance (Leach 1994). However it is emphasised that SbcCD is not a "hairpin-specific" nuclease. The hairpin nuclease activity is distinct from the known SbcCD ATP independent single-strand endonuclease, since it requires a nucleoside triphosphate. It is also likely to be distinct from the ATP dependent double-strand exonuclease, since it appears to initiate at the internal site of the loop. Initial events also include cleavage of the stem at the three nucleotide bonds adjacent to the double-strand terminus (Connelly *et al* 1999; Chapter 1 Section 3). This raises the

possibility that double-strand degradation results from concerted single-strand cleavage events (Chapter 1 Section 3).

These results confirm those of the genetic studies which have demonstrated that SbcCD can generate DSBs at DNA hairpins formed during replication, and that it therefore might cleave DNA hairpins (Leach *et al* 1997). *E. coli* RecBCD is the only other *E coli* protein to have an ATP independent single-strand endonuclease activity and an ATP dependent double-strand exonuclease activity. However the SbcCD activity is distinct from RecBCD which is not active on dumbell substrates (Taylor and Smith 1995).

Speculation

Nuclease Polarity

After a 5 minute incubation most of the early cleavage products have been degraded to smaller products. This applies both for 5'HP78 and 3'HP78, figure 3.1. The result therefore does not show unambiguously whether degradation occurs with 5' to 3' or 3' to 5' polarity. The model of half-length measurement proposes that degradation occurs with 3' to 5' polarity (Connelly *et al* 1999; Chapter 1 Section 3). This would be consistent with the 3' to 5' exonuclease activity of Mre11 (Chapter 1 Section 4).

In support of the 3' to 5' exonuclease, the 5' end-labelled substrate early products are rapidly degraded. The 3' end-labelled early products persist rather longer. However they are degraded, and this indicates that an additional endonuclease or 5' to 3' exonuclease is also active. If the 3' to 5' exonuclease is the major nuclease, the majority of cleavage products of the 3'HP78 will not be visible due to removal of the 3' label.

Nuclease Activities on the Hairpin Substrate

Hairpin cleavage appears to involve at least three nuclease activities: An endonuclease activity is responsible for initial cleavage 5' of the loop. Degradation occurs, in part, by the activity of a 3' to 5' exonuclease (above). And an additional nuclease activity is also implicated (above). At least in part, the SbcCD hairpin nuclease may derive from the previously identified

single-strand ATP independent endonuclease and double-strand ATP dependent exonuclease. The studies of homologues indicate that the phosphodiester catalytic centre is most likely in the SbcD subunit (Chapter 1 Section 2). Therefore all the nuclease activities are derived from this same fundamental centre. However it is likely that several SbcD subunits are present in the intact SbcCD complex (Chapter 1 Section 3). Diversity in the nuclease activity might be generated by multiple interactions between the SbcD subunits, interactions with SbcC, and interactions with nucleotide cofactors.

Half-Length Cleavage

Few cleavage products are visible corresponding to cleavage occuring between the loop and the half-duplex position (approximately 20 nucleotides) on either substrate, figure 3.1. This may indicate that a cleavage event is particularly favoured when the two concerted nuclease domains meet, consistent with the half-length measurement model.

Past the half-length position, degradation is stimulated on both substrates; between nucleotides 20 and 12 on 5'HP78, and 20 and 15 on 3'HP78. It is then reduced between nucleotides 12 and 6 on 5'HP78, and 15 and 10 on 3'HP78. This pattern is consistently obtained. Perhaps it relates to the two 3' single-strand tails which exist at this stage of degradation.

De-protection of DSBs by SbcCD

The loop protects a hairpin stem from recombination nucleases. Hairpin cleavage by SbcCD can therefore be described as de-protection of a sealed double-strand end, and the role of SbcCD can be analysed in terms of facilitating the entry of DNA into a recombination pathway. Consistent with this model, T4 gp46/47 is required for the recombinational repair of a covalent *m*-AMSA/topoisomerase/DNA complex (Chapter 1 Section 3). In addition, the yeast Rad50/Mre11 complex acts during meiotic recombination to remove Spo11 protein from the ends of DSBs (Chapter 1 Section 4).

Other Substrates

The majority of this work was performed on a single model hairpin substrate, HP78, and hairpins based on it. Nine other synthetic hairpins with different sequences, sizes and loops have since been studied (Connelly *et al* 1999). All were cleaved with approximately equal efficiency on the 5' side of the loop, which confirms that the nuclease recognises the structure of the hairpin rather than its sequence.

SbcCD cleaved an unligated dumbell ncDB98 opposite the nick. This result, in conjunction with experiments on other candidate substrates for SbcCD nuclease, may help elucidate the aspect of the stem-loop junction which SbcCD hairpin nuclease recognises.

Figure 3.1

Time course degradation of HP78 by SbcCD. HP78 was 5' (lanes 3 to 7) or 3' (lanes 8 to 14) end-labelled. It was heated, snap-cooled and incubated with 6.9nM SbcCD under nuclease conditions in the presence of 1mM ATP at 37°C over a time course from 0 to 30 minutes: C represents the control lacking SbcCD which was incubated for 30 minutes. 0', 1', 5', 10' and 30' represent time in minutes. (The 0' time point was taken by the standard method of stopping the reaction immediately after adding protein). The products were separated by 20% denaturing PAGE at 60°C. An A/G ladder of the 3' end-labelled 78mer (lane 15), marker V (lane 1) and an 8 to 32 ladder (lane 2) have been included as size standards.

Figure 3.1

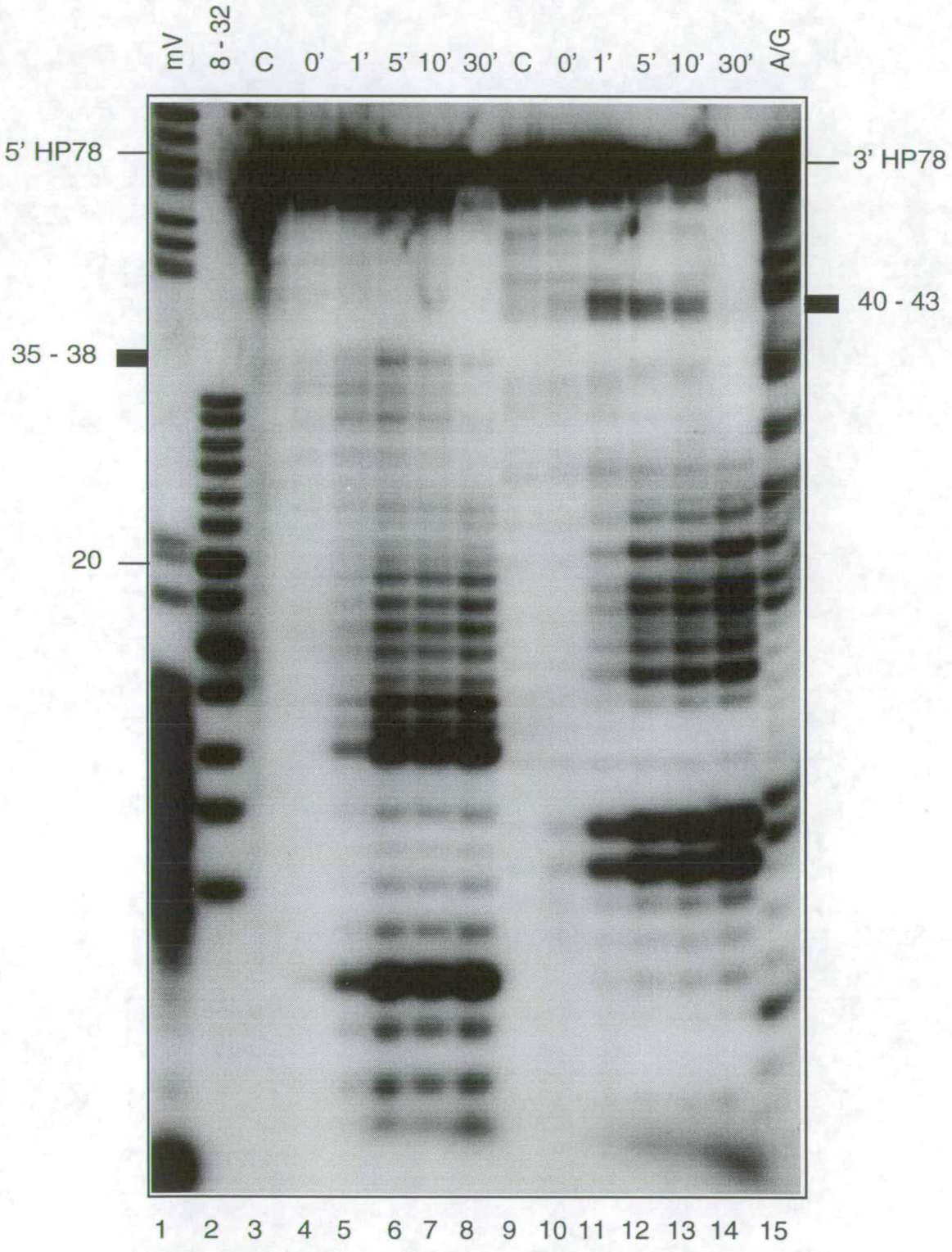

mV
8 - 32
C
0'
1'
5'
10'
30'
C
0'
1'
5'
10'
30'
A/G
5' HP78
3' HP78
40 - 43
35 - 38
20
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Figure 3.2

Trapping the products of initial cleavage of HP78 by preventing subsequent degradation. HP78 was 3' end-labelled, heated, snap-cooled and incubated with 6.9nM SbcCD under nuclease conditions at 37°C for 30 minutes. 1mM nucleotide cofactor was included in the nuclease assays. Products were separated by 20% denaturing PAGE at 60°C.

a

Control reactions were performed minus SbcCD plus ATP (lane 1), plus SbcCD minus ATP (lane 2) and plus SbcCD plus ATP (lane 3).

b

Reactions were performed minus SbcCD plus ATP (the control C, lane 1), plus SbcCD plus ATP (lane 2), plus SbcCD plus dATP (lane 3), plus SbcCD plus ADP (lane 4), plus SbcCD plus ATPγS (lane 5), plus SbcCD plus GTP (lane 6) and plus SbcCD plus GDP (lane 7). Sizes were estimated using marker V and the 8 to 32 ladder as size standards.

Figure 3.2

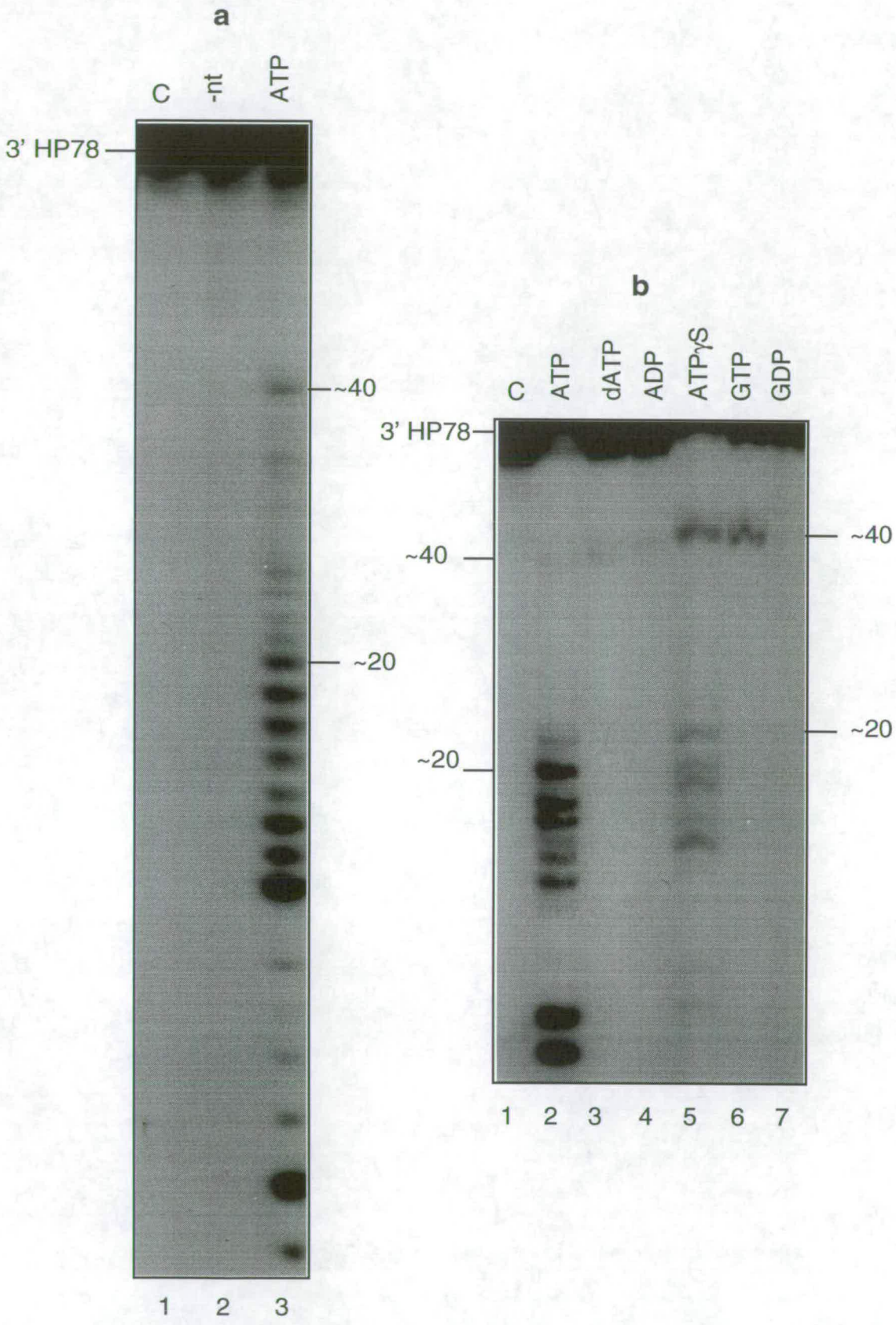

a
C
-nt
ATP
3' HP78
~40
~20
1
2
3
b
C
ATP
dATP
ADP
ATPγS
GTP
GDP
3' HP78
~40
~20
~40
~20
1
2
3
4
5
6
7

Figure 3.3

Determining the chemical nature of the cleavage termini. HP78 was end-labelled, heated, snap-cooled and incubated with 6.9nM SbcCD under nuclease conditions in 1mM ATPγS at 16°C for 30 minutes. Products were separated by 12% denaturing PAGE at 60°C.

a

3' end-labelled HP78 (5' phosphate, 3' hydroxyl) was digested and electrophoresed alongside an A/G ladder generated from 3' end-labelled 78mer (5' phosphate, 3' hydroxyl): 3' HP78 minus SbcCD plus ATPγS (control C, lane 1), 3' A/G ladder (lane 2), 3' HP78 plus SbcCD plus ATPγS (lane 3), 3' A/G ladder (lane 4), mixture of 3' HP78 plus SbcCD plus ATPγS with 3' A/G ladder (lane 5).

b

5' end-labelled HP78 (5' phosphate, 3' hydroxyl) was digested and electrophoresed alongside an A/G ladder generated from 5' end-labelled 78mer (5' phosphate, 3' phosphate): 5' HP78 minus SbcCD plus ATPγS (control C, lane 1), 5' A/G ladder (lane 2), 5' HP78 plus SbcCD plus ATPγS (lane 3), 5' A/G ladder (lane 4), mixture of 5' HP78 plus SbcCD plus ATPγS with 5' A/G ladder (lane 5).

Figure 3.3

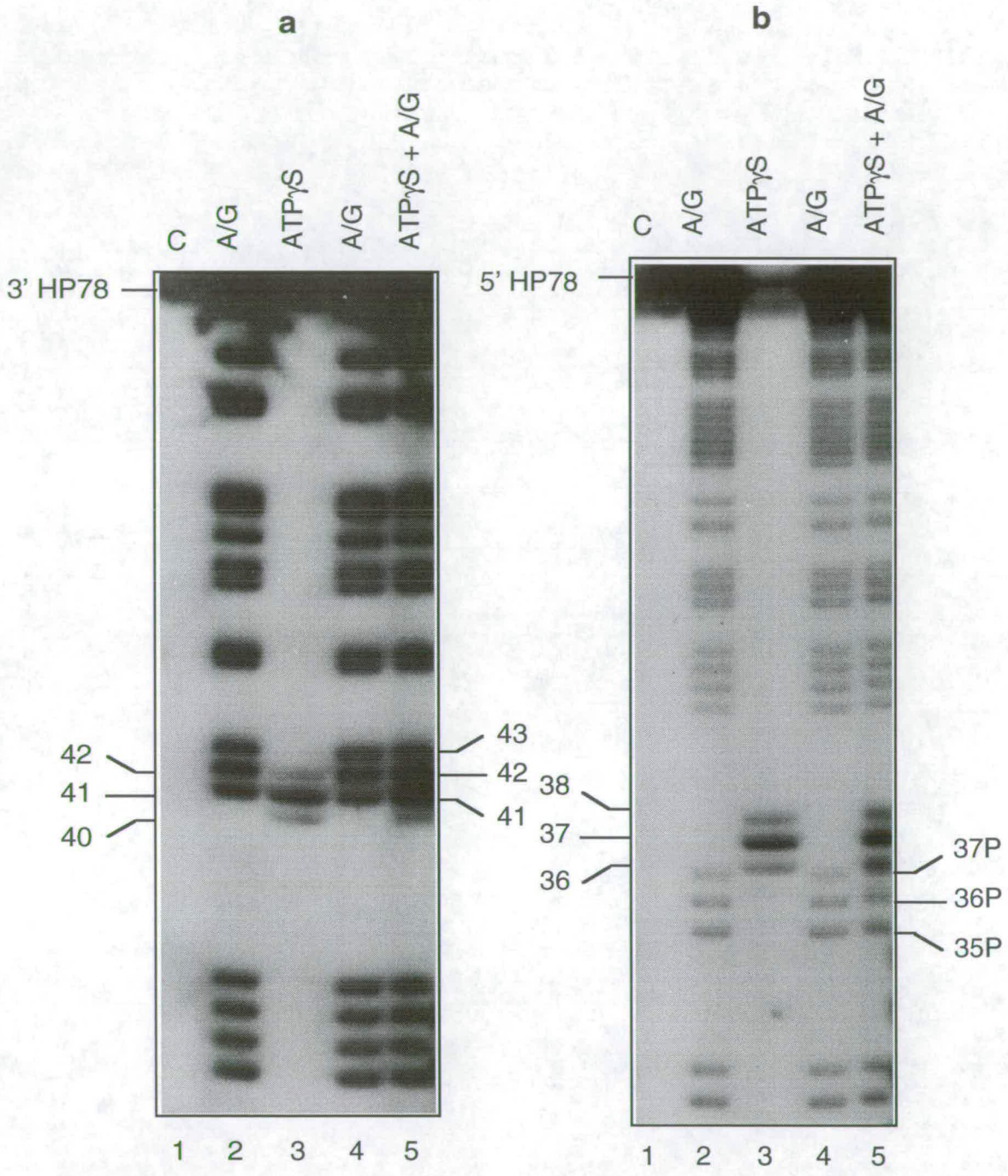
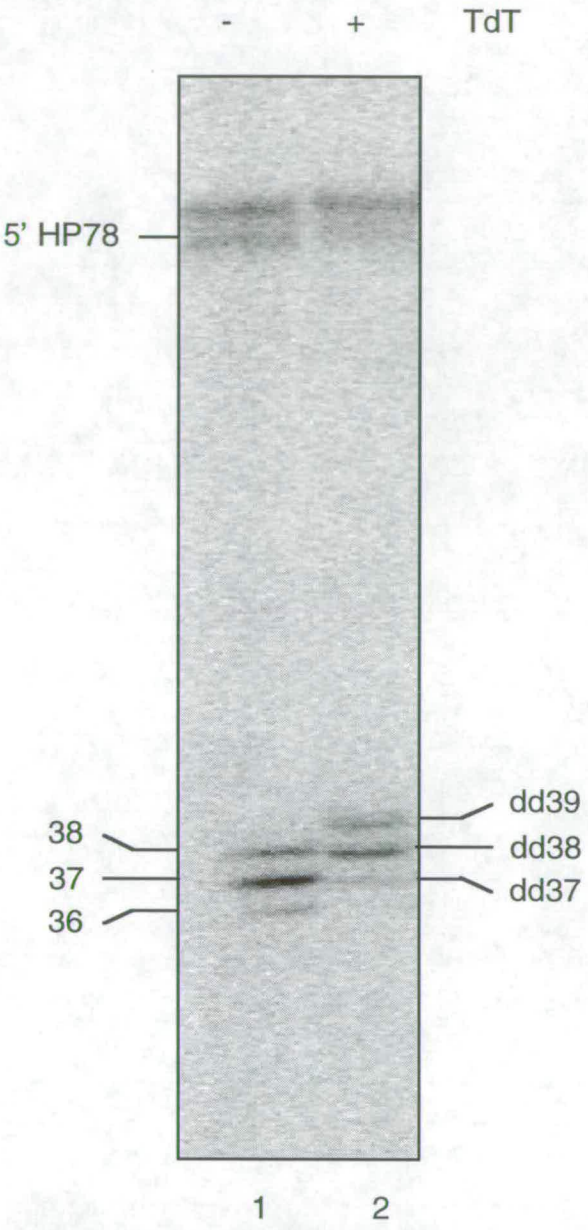
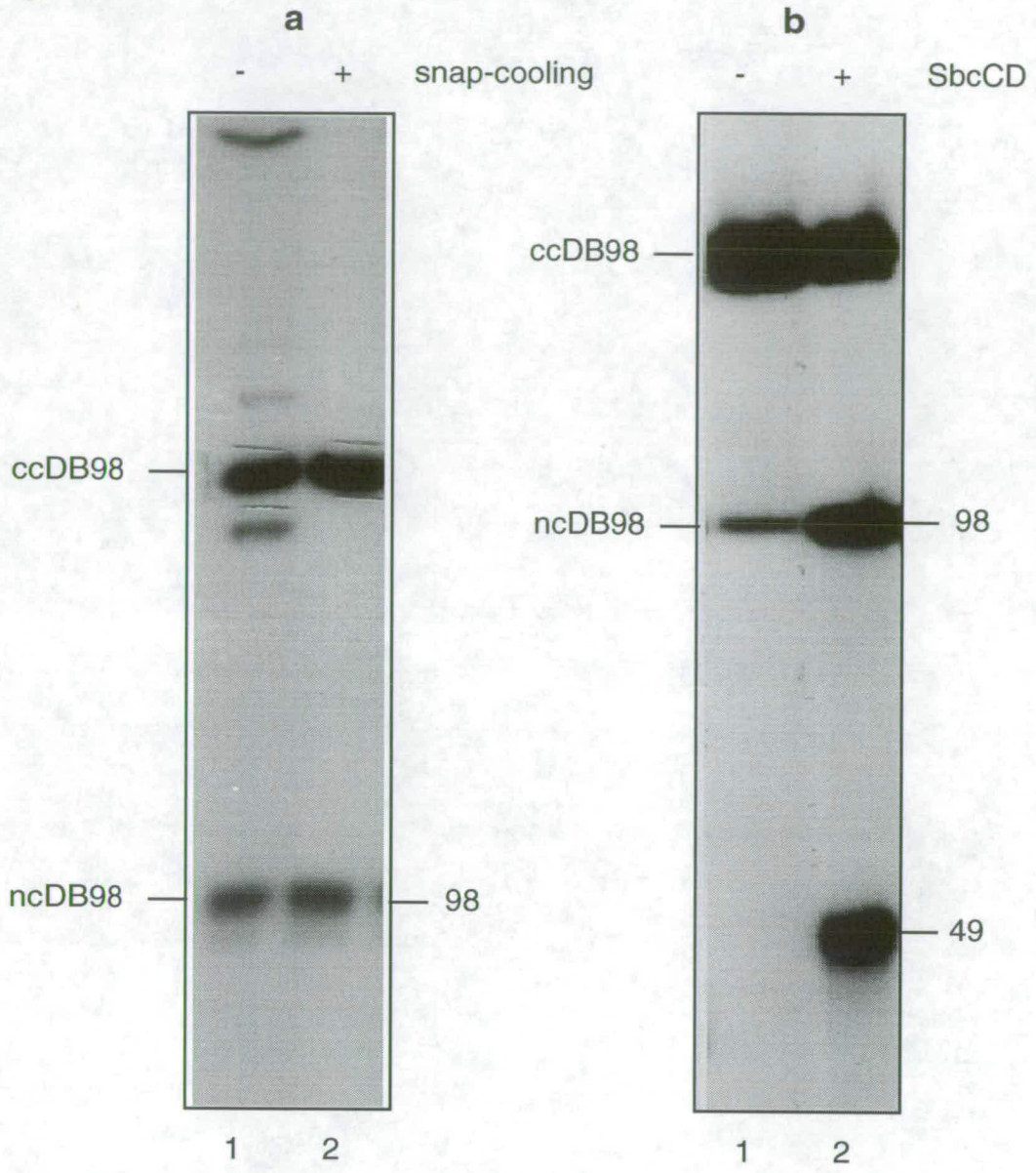

a
b
C
A/G
ATPγS
A/G
ATPγS + A/G
3' HP78
5' HP78
42
41
40
43
42
41
38
37
36
37P
36P
35P
1
2
3
4
5

Figure 3.5



Chemical nature of the 3′ end of SbcCD nuclease products. HP78 was 5′ end-labelled, heated, snap-cooled and incubated with 6.9nM SbcCD under nuclease conditions in 1mM ATPγS at 16°C for 30 minutes. Products were purified. They were then denatured by heating and incubated with 0.5μM ddATP under conditions recommended by the manufacturer at 37°C for 15 minutes in the presence (+) or absence (-) of calf intestinal terminal transferase. The end-modified products were separated by 12% denaturing PAGE at 60°C.

Figure 3.6



Gel purification of ccDB98 (a). 5′ end-labelled DB98 was ligated directly (lane1) or after heating and snap-cooling (lane 2). ccDB98 was separated from ncDB98 by denaturing PAGE at 60°C (detailed in the text). Snap-cooling appears to improve denaturing and reduce the formation of multimers. The ccDB98 product was excised for elution and purification.

Degradation of ccDB98 by SbcCD (b). ccDB98 was incubated with (+) or without (-) 6.9nM SbcCD under nuclease conditions in 1mM ATPγS at 16°C for 30 minutes. Products were separated by 10% denaturing PAGE at 60°C using 10% formamide in the gel mixture. Some ncDB98 is present in the ccDB98 preparation on account of radiolysis.

CHAPTER 4

SbcCD DOES NOT POSSESS HELICASE ACTIVITY ON A SHORT HAIRPIN SUBSTRATE

Introduction

A nuclease activity is generally coupled to a DNA helicase activity in providing the single-stranded substrate for pairing in homologous recombination (reviewed in Mendonca *et al* 1995). In the cases of RecBCD (Roman and Kowalczykowski 1989) and UvrABC (Gorbalenya and Koonin 1990) both activities are contained in the same large complex. In other instances separate proteins act in conjunction, *eg* RecQ helicase, helicase II and helicase IV (reviewed Matson *et al* 1994). It was of interest to determine whether SbcCD has a helicase activity.

SbcC and SbcD belong to subclasses of their respective protein families which are involved in genetic recombination (Chapter 1 Section 2). SbcCD and its homologues are believed to generate and process DSBs, and DSBs stimulate recombination (Chapter 1 Sections 3 and 4). The hairpin nuclease of SbcCD is postulated to generate a DSB in a replication fork, which is then repaired by homologous recombination (Chapter 3). It is proposed that the RecBCD nuclease-helicase complex may act at the DSB generated by SbcCD to perform the homologous recombination required to repair the replication fork. If SbcCD were to have a DNA helicase activity, it would be another example of a complex with both helicase and nuclease activities, and RecBCD might be redundant in this pathway.

Helicases are typically large, hexameric, ring complexes capable of encircling a DNA helix (Alexander Alexandrov, University of California Berkeley, personal communication), *eg* RuvAB (Yu *et al* 1997, Mitchell and West 1994), *eg* DnaB (Jezewska *et al* 1998). The large size of the putative SbcCD complex is therefore consistent with helicase activity.

There are also some indications that the eukaryotic homologues of SbcCD might possess helicase activity. The exonuclease conundrum associated with the Rad50/Mre11 complex (Chapter 1 Section 4) might be explained if the complex were to associate with a helicase (Haber 1998). A helicase might unwind DNA in a 5' to 3 direction, enabling the Rad50/Mre11 endonuclease to cleave the 5' ended strand. A 5' to 3' helicase activity could be envisaged to act on the 5' end-bound Spo11 protein at yeast meiotic DSBs. In addition,

the SMC family of condensins possesses a helicase-like activity involved in supercoiling DNA (Kimura and Hirano 1997).

In this chapter, the characterisation of a small band-shift is described. For the reasons given above, it was possible that this represented a helicase activity of SbcCD. However the sequence of SbcCD does not include the DEAD/H motif (Naom *et al* 1989), which is common to helicases (Saitoh *et al* 1995). That SbcCD does not require ATP for initial cleavage (Chapter 3) also suggests that it may not possess helicase activity. However its requirement for ATP in substrate degradation suggests that it may possess translocase activity. The band-shift might alternatively be explained by DNA binding by protein (either SbcCD, a part of it, or a contaminant) or migration of the nuclease product separately from substrate. Each of these possibilities is investigated.

Results

Products of SbcCD Hairpin Cleavage Migrate Anomalously on Native Gels

A Small Band-Shift is Observed on a Native Agarose Gel

During investigations of DNA binding activity by SbcCD, end-labelled, heated and snap-cooled HP78 substrate was incubated with SbcCD under standard binding conditions. Protein was preincubated with 1mM Mn^{2+} for 10 minutes at 0°C. Next this protein at 46nM was incubated with 1.5nM DNA, 1.25mM DTT, 7% glycerol, 100µg/ml BSA, 25mM TrisCl pH7.5, 5mM Mn^{2+} and 1mM ATPγS in a 30µl reaction volume made up with SMQ, at 16°C for 20 minutes. The samples were loaded immediately onto a pre-cooled native 0.8% agarose gel in 1xTAE and electrophoresed at 80V, 4°C. The migration of protein treated substrate was very slightly impeded relative to untreated substrate (data not shown).

Optimisation of Conditions for the Band-Shift

Various concentrations and types of native gel matrix were used to magnify the observed band-shift. Resolution was increasingly improved in a 3% NuSieve® GTG® Midi agarose gel run at 80V, and 3.5% to 7% Mini-Protean II polyacrylamide gels run at 50V. Resolution was not further improved in 10% or 12% polyacrylamide gels run at 4°C in pre-cooled 1x TAE buffer. Longer gels provided a greater distance for separation and better resolution than shorter ones. A clear band-shift was obtained using a 6.8% polyacrylamide Hoefer gel run in 1x TAE at 160V at 4°C (figure 4.5 below).

Resolution of DNA varies with the concentration of the matrix in a manner dependent on the dimensions of the DNA (De Wachter *et al* 1990). It is greater in polyacrylamide than agarose. However polyacrylamide is a harsher environment for unstable species than agarose and is increasingly harsh with concentration. This is a second reason why high percentage polyacrylamide gels were not used.

The Band-Shift Results from Protein Activity

Possible Protein Activities

The band-shift could have arisen from several possible protein activities. It could have represented gel retardation of substrate or product DNA by protein binding. This might occur either by SbcCD or contaminant protein binding. It may have resulted from a novel helicase activity of SbcCD which denatured the nicked hairpin product of the nuclease. Alternatively it could correspond to a difference in running position between uncleaved and cleaved hairpin DNA.

To determine whether the band-shift resulted from a preparation-specific contaminant, a different protein preparation was tested (Connelly *et al* 1997). An identical band-shift was obtained, indicating that the activity was present in other SbcCD preparations.

Protein Titration

SbcCD was titrated against substrate DNA to determine the effect of diluting protein to a concentration below that of the DNA. SbcCD dilutions were made to protein:DNA molar ratios of 30:1, 7.7:1, 3:1, 1.5:1, 0.8:1 and 0.5:1 and incubated with DNA. Products were separated by 6.8% native PAGE, figure 4.1. A decrease in the proportion of DNA shifted was observed with decreasing protein concentration, indicating that the shift was effected by protein. The same effect was observed minus BSA (data not shown).

The Small Band-Shift does Not Represent Protein Binding

Possible Protein Binding

Although too small to represent a typical gel retardation produced by SbcCD-DNA binding, the band-shift could still result from polypeptide binding (*ie* binding by SbcC, SbcD or a part of either). This is especially true because the concentration of purified SbcCD used to obtain the band-shift was very high. Low level contaminant protein in the SbcCD preparation

might have significant activity at this high concentration, which would not be observed if the SbcCD was further diluted.

It is unlikely that the complete 1.2MDa complex would even enter the polyacrylamide gel, still less cause such a small band-shift (Alexander Alexandrov personal communication). However it is conceivable that a fragment of SbcCD was slightly retarding the substrate DNA. This could be true if the polypeptide responsible was very small, if it had an unusual morphology, if it was unstable, or if its net charge at pH7.5 to pH8.0 (1x TAE is pH8.0) was highly negative. The latter hypothesis would imply that the pI of the polypeptide, *ie* the pH at which the polypeptide has zero net charge, is more acidic than pH7.5 to pH8.0. The pI of SbcC is pH5.52 and that of SbcD is pH5.80 (John Connelly personal communication). SbcCD becomes insoluble at pH5.5 (John Connelly personal communication). This is consistent with a pI between pH7.0 and pH5.0 because proteins are insoluble at pHs equal to their pI. However this insolubility might alternatively result from unfolding of the protein at this unphysiological pH. A small band-shift might result if the polypeptide responsible bound the substrate unstably and the complex dissociated in the well or during electrophoresis. However this generally causes the retarded band to appear as a characteristic smear, tailing towards the well. In contrast, the observed shift was only slightly diffuse.

The band-shift does not represent SbcCD-DNA binding

The running position of SbcC and SbcD on the gel was compared to that of the band-shift by Western Blotting. Products of incubation of SbcCD with substrate DNA under binding conditions were separated on two fully loaded 6.8% native polyacrylamide gels. Consecutive regions of the gels, from the well down to the band-shift, were excised. Eluted DNA from the different regions was loaded into different wells on two ATTO SDS gels, using one SDS gel as a duplicate of the other. The gels were blotted and analysed using anti-SbcC and anti-SbcD polyclonal immunoglobins, figure 4.2.

The anti-SbcC blot shows a large amount of immunoreactive material high in the gel, figure 4.2a. The anti-SbcD blot shows immunoreactive material lower in the gel, figure 4.2b. This indicates that the SbcC polypeptide migrates behind the smaller SbcD polypeptide in the denaturing SDS gel, as

expected. The majority of anti-SbcC immunoreactive material was present in the track corresponding to the native gel well band, indicating that SbcC hardly migrates into the gel, if at all. It may be that SbcC, which has a coiled-coil motif, was self-associating under these conditions. Anti-SbcD immunoreactive material was present in a non-adjacent track, indicating that SbcD migrates a little way into the native gel. No immunoreactive material was observed in the track corresponding to the band-shift, indicating that neither SbcC nor SbcD polypeptides are likely to be present there, and that the band-shift is not due to binding of SbcCD nor a component of it. No associated complex was observed on the native gel.

The Band-Shift Does Not Represent Contaminant Protein-DNA Binding

Several methods were employed to demonstrate that the shift was not due to retardation of any type of DNA-protein complex.

Protein Staining

During Western Blotting, the nitrocellulose filter was stained with Ponceau S. No protein was observed in the track corresponding to the band-shift (data not shown). Similarly no protein was observed in this track when the gel was stained with Coomassie Blue (data not shown). This indicates that no protein was present in the band-shift.

Silver Staining

SbcCD and HP78 substrate DNA were incubated under binding conditions in the presence of Mn^{2+} or Ca^{2+} ions (in both the preincubation and reaction mixtures) and using Tris nuclease buffer lacking BSA. It was necessary to exclude BSA on gels to be silver stained, to avoid interference by stained BSA. The reactions were separated on a 6.8% native polyacrylamide gel, which was silver stained and dried, figure 4.3. Material (presumably protein) is visible in the wells and a little below them. This may correspond to SbcC and SbcD. Unshifted material (presumably DNA) is visible in tracks corresponding to reactions in Mn^{2+} minus protein and in Ca^{2+}. The diffuse staining at the position of the band-shift in Mn^{2+} plus protein may correspond to protein or DNA. Therefore the experiment does not exclude

the possibility that the band-shift arises from a protein-DNA complex. This staining is not heavy, which may suggest that little (if any) protein is present at the band-shift. Silver staining does not detect protein at less than approximately 5ng (BioRad), *ie* 0.3% of the total protein present per reaction.

Glutaraldehyde Fixation

Glutaraldehyde crosslinking was used to try to increase the yield of band-shift material. 1% glutaraldehyde was immediately added after completion of the 20 minute binding reaction which was then incubated at 37°C for a further 10 minutes. Products of crosslinking were separated on a 6.8% native polyacrylamide Mini Protean II gel at 50V. However the quantity of material in the small shift was not affected by cross-linking and no additional retarded bands were observed (data not shown), suggesting that the shift does not represent a bound complex.

Column Purification

The reactions were treated with a QIAquick™ Nucleotide Removal column which purifies DNA from protein and salts. Four sets of binding reactions were set up, with and without protein. One set was not column purified. The second and third sets were column purified; they were made up in a double volume to allow for 50% recovery. Sets 1 and 2 were maintained on ice whilst the third set was subjected to binding conditions for another 20 minutes, alongside a fourth set. Set 1 controlled for the presence or absence of a normal shift after the reaction had sat on ice for up to 1 hour, which would be important if the cause of the shift was unstable under these conditions. Set 2 tested whether column purification eliminated the shift. Set 3 tested whether purified material was still capable of shifting, interesting if column purification eliminated the shift. Set 4 provided a control for the normal shift as originally observed by electrophoresis immediately after reaction.

Products were separated on a 12% native sequencing sized polyacrylamide gel, figure 4.4. A shift was not clearly visible in track 5 (the critical one) suggesting that the band-shift may result from protein-DNA binding. However later results support an alternative explanation not ruled out by this experiment.

The Band-Shift does Not Represent a Helicase Activity

Helicase Activity

The band-shift could have been an indication that SbcCD possesses a helicase activity. Cleavage 5' of the loop is expected to occur under these conditions for binding. SbcCD might separate the two single-strand products, which might migrate at the retarded position observed. The migration of a single-strand species relative to double-strand species on a native gel is difficult to predict. This behaviour would additionally suggest that SbcCD was able to prevent reannealing of the single-strands. If the nuclease and hypothetical helicase activities had similar kinetics, this would be consistent with the two acting concomitantly.

An Oligonucleotide Helicase Assay

A helicase assay was designed which utilised single-strand synthetic oligonucleotides equivalent to the expected products of SbcCD nuclease activity on a 5' end-labelled HP78 substrate. A 5' end-labelled 37mer oligonucleotide and an unlabelled 41mer oligonucleotide, figure 2.1.1, were annealed. These correspond to the 5' and 3' arms of the cleaved HP78 substrate respectively. The resulting double-strand species represented the expected early cleavage product. 5' end-labelled 37mer alone represented the expected single-strand labelled product of the hypothetical nuclease-helicase activity. These were electrophoresed on a 6.8% native polyacrylamide gel and their mobilities compared to that of the band-shift, figure 4.5. The single-strand species migrated ahead of the double-strand or the band-shift under these conditions. This clearly indicates that the band-shift does not simply represent the product of an unreversed helicase activity, *ie* it is not a cleavage product which has undergone denaturation by a helicase. This is pertinent since a helicase activity might only act after substrate cleavage. The double-strand species migrated very similarly to the band-shift, indicating that the band-shift is likely to represent a product of the hairpin nuclease. The band-shift is generally slightly smaller than the difference in mobility between the substrate and the annealed oligonucleotide duplex.

Synthetic oligonucleotides were used for this assay, instead of products generated by SbcCD nuclease, because there is no absolute evidence that the unlabelled arm of the cleaved hairpin is not degraded under early cleavage conditions. A 12% gel provided better resolution of the double-strand from the single-strand than the 6.8% gel, but resolved the band-shift from the substrate less well. A 6.8% Hoefer gel provided good all-round resolution.

Metal and Nucleotide Requirements for the Band-Shift

Helicases use the energy from NTP hydrolysis to unwind duplex DNA (reviewed Matson *et al* 1994). They are generally ATPases. They do not act in ATPγS. It would therefore be surprising if the band-shift observed in ATPγS were due to a helicase activity. It is conceivable that a helicase might act in Mg^{2+}, or in a mixture of Mn^{2+} (for protein stability) and Mg^{2+}, because ATPases generally require Mg^{2+}. Some proteins have been reported to have helicase activity which actually only denature short duplexes (Alexander Alexandrov personal communication). Alternatively therefore, a protein might be effective as a helicase by a different mechanism which does not require ATP hydrolysis, particularly on a short duplex substrate such as HP78 and *in vitro.*

HP78 was incubated under binding conditions with SbcCD in a range of metal ions and nucleotide cofactors. The preincubation metal ion was also varied. The products were separated on 6.8% native polyacrylamide gels, figure 4.6. The small band-shift was observed when Mn^{2+} ions were present in the reaction, irrespective of whether Mg^{2+} was also present, but it was not observed in the presence of Mg^{2+} or Ca^{2+} alone, figure 4.6a and figure 4.6b. It was observed when ATP or ATPγS was present in the reaction, but not in the absence of a nucleotide cofactor, figure 4.6c. In any case, no evidence was obtained of the production of single-stranded 37mer by helicase activity under the conditions tested. Ca^{2+} might stimulate formation of a stable cleavage DNA-protein bound complex (Chapter 5). The lack of band-shift in the presence of Ca^{2+} ions is therefore consistent with the above data which indicates that the shift is not due to protein binding.

A SSB Helicase Assay

If SbcCD did not prevent reannealing of the helicase separated single-strand products in the native reaction solution, a helicase activity would not be detected in the oligonucleotide assay. The following additional helicase assay was devised to clarify this.

Nuclease-free single-strand binding protein SSB (Sigma) binds single-strand DNA but not duplex (reviewed Chase and Williams 1986, Meyer and Laine 1990, Sambrook *et al* 1989). SSB was used in a helicase assay based on a previously published method (Roman and Kowalczykowski 1989). Binding reactions were set up with 5' end-labelled HP78 and ATP or ATPγS in the presence or absence of preincubated SbcCD and the presence or absence of SSB. SSB is predicted to prevent renaturation of separated single-strand DNA. The products were analysed on a 5% native polyacrylamide gel, figure 4.7.

The small shift of hairpin DNA is visible, although not well resolved on this gel. It appears in the presence of SbcCD and ATPγS, and aligns with the double-strand oligonucleotide control. The hairpin substrate is degraded in ATP. This degradation is visible as a smear ahead of the substrate DNA. A large retardation of single-strand oligonucleotide DNA is visible in the presence of SSB, indicating that SSB is able to bind single-strand substrate under these conditions. SSB retardation is not visible in any other track, indicating that SSB does not bind a double-strand substrate, as expected, and that it does not bind a hairpin substrate or a cleaved product with short single-strand overhangs. It can be concluded that the cleaved hairpin does not adopt a single-strand structure at any time prior to loading on the gel, and therefore that SbcCD does not denature the product and has no helicase activity. These results also indicate that the products are not denatured by any other means prior to entering the gel. It cannot be excluded that the single-strand generated may be too small to bind SSB.

The Band-Shift Corresponds to Cleaved Hairpin DNA

Nuclease Activity

The shift may correspond to a difference in running position between the nuclease substrate and product DNA. This is the simplest explanation.

The Band-Shift Corresponds to a Product of the Hairpin Nuclease

The shifted DNA was studied in denatured form by denaturing PAGE, figure 4.8. It can be seen that the band-shift, lane 5, included a product which migrated with the 37mer single-strand oligonucleotide model product, lanes 7 and 8, and with the products of early hairpin nuclease activity, lane 2. This indicates that the shifted DNA consists of hairpin nuclease products.

A sequencing sized denaturing gel was used to confirm the identity of the DNA present in the band-shift reactions, figure 4.9. The band-shift reaction DNA migrated as a mixture of substrate, 78 nucleotides long, and the 5' end-labelled products expected from early cleavage, 36, 37 and 38 nucleotides in length.

The Band-Shift is Sensitive to Running Conditions

The band-shift was not obtained in 1x TBE native gels using the gel systems described here (data not shown). TBE is known to be a harsher buffer than TAE with respect to maintaining the stability of transient species (above). It was also essential to run the native gels at relatively high voltages, as indicated. Furthermore, one of the non-denatured palindromic species observed on incompletely denaturing gels (Chapter 3) migrated in a position analogous to the band-shift observed on these native gels (data not shown).

Together the results suggest that these "native" gels were not completely native, despite being run at low temperatures to decrease denaturation. The band-shift may result when the DNA is not completely denatured and not

completely native. This suggests that the band-shift corresponds to a conformation which differs from the uncleaved form.

Discussion

Summary

A small band-shift was observed on native PAGE of the products of a binding reaction between SbcCD and a hairpin substrate. An optimal shift required a 6.8% polyacylamide Hoefer gel to be run at 160V at 4°C and in 1x TAE. A protein titration demonstrated that the shift was effected by protein, and Western Blotting demonstrated that it was extremely unlikely to correspond to a bound complex of SbcCD-DNA. The shift was unaffected by column purification of the protein-DNA mixture and by glutaraldehyde crosslinking. No protein was detected at the band-shift position by Ponceau S, Coomassie Blue or silver staining. Together these results indicated that the shift was not due to protein-DNA binding. An oligonucleotide helicase assay was devised. The results of this assay suggested that the shift was not due to helicase activity. This inference was supported by the metal ion and nucleotide cofactor requirements for the shift; Mn^{2+} was required and the shift occured in ATPγS as well as ATP. A SSB helicase assay was also devised. The results indicated that the shift was not due to helicase activity, and that the strands of the nicked hairpin product were not separated at any stage prior to entering the gel. Native and denaturing gels indicated that the shift consists of hairpin nuclease products. The metal ion and nucleotide cofactor requirements supported this. The migration of these nuclease products is dependent on gel running conditions.

Conclusions

SbcCD does Not Possess Helicase Activity on a Short Hairpin Substrate

The oligonucleotide helicase assay suggested that SbcCD is not a helicase on the hairpin substrate tested. This was supported by the metal and nucleotide cofactor requirements for the band-shift. The banding pattern on a denaturing gel raised the possibility that SbcCD is effective as a helicase by a mechanism other than ATP hydrolysis if it is also unable to prevent reannealing of the single-strands it generates. The observation which

highlighted this possibility, of increased denaturation in the SbcCD tracks relative to an untreated control, could be convincingly explained in other ways. The SSB helicase assay was devised to clarify this. It demonstrated that the cleavage products were not single-stranded at any stage prior to entering the gel. This indicated that SbcCD is not a helicase on the hairpin substrate and the nicked hairpin products are not separated by other means.

Consistent with these results, the closest homologues of SbcCD do not possess helicase activity: T5 gp*D12/D13* is likely to act with T5 gp*D10* helicase (Blinov *et al* 1989). T4 gp*46/47* acts with the T4 gp41/61 helicase-primase complex (reviewed Kreuzer *et al* 1995). Additionally it is plausible that SbcCD nuclease might act in conjunction with DnaB helicase at replication forks, or with UvrB during recombinational repair. Helicases use ATP, but the Rad50/Mre11 nuclease is ATP independent (Chapter 1 Section 4). It is therefore unlikely that it possesses helicase activity. It may associate with a helicase, *eg* Rad3 in yeast. Perhaps Xrs2 is a helicase.

These results do not exclude the possibility that SbcCD is a translocase. Indeed the degradation of a hairpin substrate (Chapter 3) could be explained by such an activity.

Products of SbcCD Hairpin Cleavage Can Adopt a Non-Duplex Conformation on Native PAGE

The band-shift consists of hairpin nuclease products and their migration is dependent on the running conditions of the gel. These results suggested that the band-shift occurs when the nicked hairpin adopts a conformation which differs sufficiently from the uncleaved hairpin that it migrates separately.

Speculation

The Products of Hairpin Cleavage May Develop "Splayed" Ends

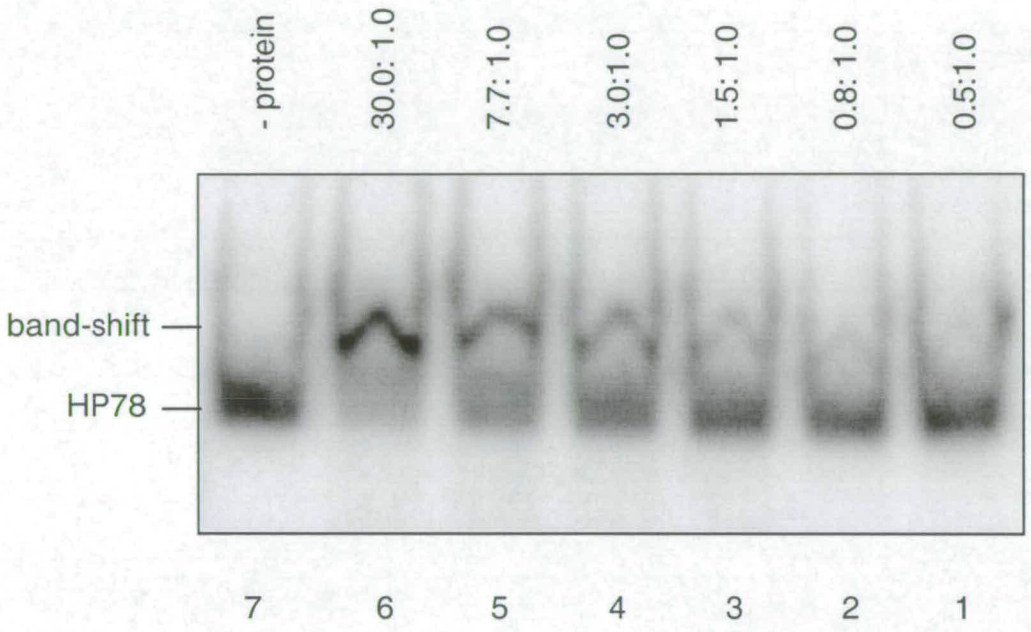
Studies on the hairpin nuclease acitivity indicate that the nicked hairpin would form a duplex with a single 5' overhang of between 3 and 5 nucleotides (Chapter 3). Recent studies indicate that the 3' terminal 3 nucleotides of a hairpin are also cleaved (Connelly *et al* 1999). Therefore the product might have 5' overhangs at both ends of the duplex. This has two effects. The cleaved hairpin product is slightly smaller than the model oligonucleotide product and this explains the smaller size of the shift compared to the difference in migration between the product and the oligonucleotide duplex. Furthermore a proportion of the cleaved hairpin duplex may have two single-strand ends. Structures like this promote partial denaturation of the duplex region, called "splaying". Splaying is likely to significantly affect the mobility of a species on a gel.

SbcCD is Unlikely to be a Helicase

This work does not exclude the possibility that SbcCD has helicase activity on substrates other than the short hairpin oligonucleotide tested. This is unlikely because the hairpin has a relatively short duplex region and short duplexes are more easily denatured than long ones (above).

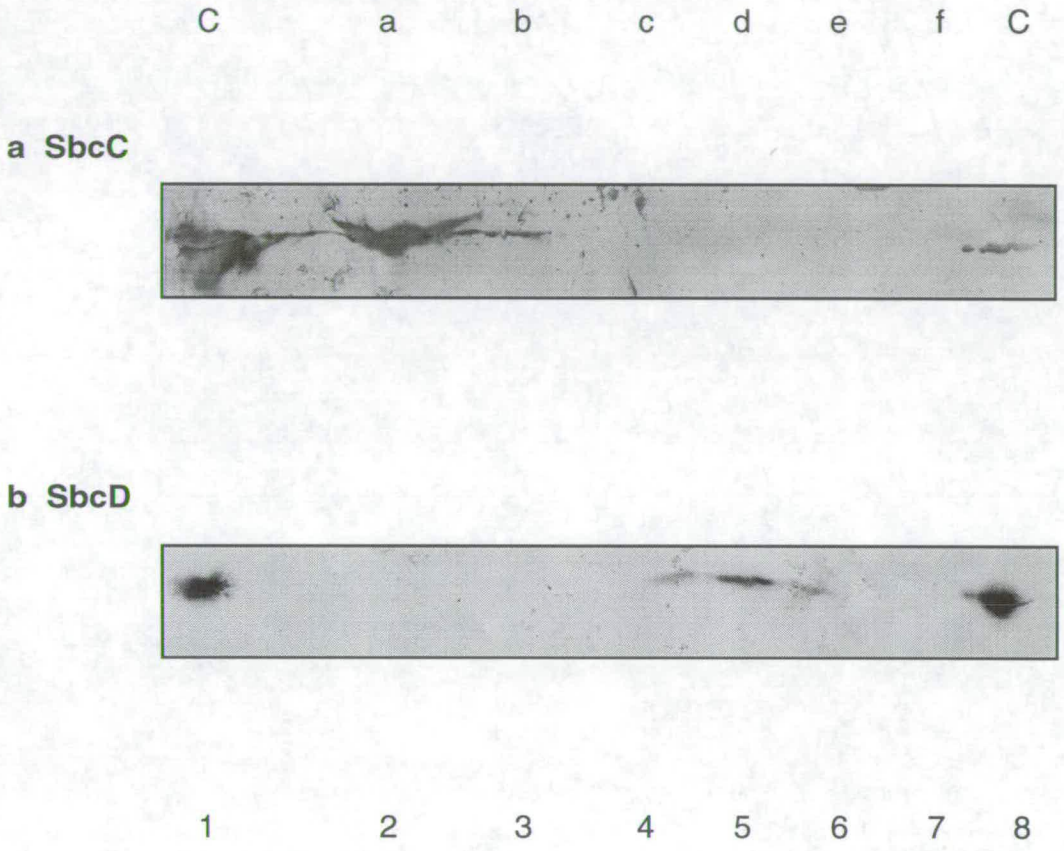
Helicases may have either 5' to 3' or 3' to 5' polarity (reviewed Matson *et al* 1994). Some require a single-strand region adjacent to the duplex for initial binding. Initial cleavage of the hairpin substrate can generate products with two 5' overhangs, but no 3' overhangs (Chapter 3, Connelly *et al* 1999). Therefore it is still possible that SbcCD possesses a 3' to 5' helicase which has a strict binding requirement for 3' single-strand overhangs. The potential helicase activity of SbcCD could be tested more definitively by using various specifically designed substrates, *eg* partially double-stranded circles and linear DNAs (Matson *et al* 1994).

Figure 4.1



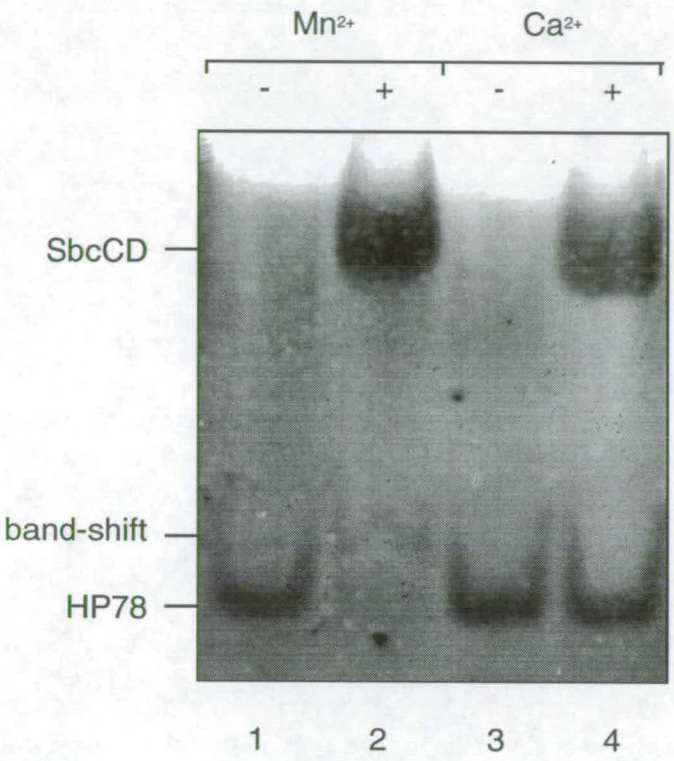
Band analysis by protein titration. HP78 was 5′ end-labelled, heated and snap-cooled. Protein was preincubated with 1mM Mn^{2+} for 10 minutes at 60 °C. This protein was then incubated under standard binding conditions, *ie* with 1.5nM DNA, 1.25mM DTT, 7% glycerol, 100µg/ml BSA, 25mM TrisCl pH7.5, 5mM Mn^{2+} and 1mM ATPγS in a 30µl reaction volume made up with SMQ, at 16°C for 20 minutes. A protein titration was prepared by performing this reaction using preincubated SbcCD which was diluted in BSA/Buffer B/Mn^{2+}. Products were separated on a 6.8% native polyacrylamide Mini Protean II gel in 1x TAE at 50V, 40°C. Shown are SbcCD:DNA molar ratios of 0.5:1.0 (lane 1), 0.8:1.0 (lane 2), 1.5:1.0 (lane 3), 3.0:1.0 (lane 4), 7.7:1.0 (lane 5) and 30.0:1.0 (lane 6). A control reaction minus SbcCD is shown (lane 7).

Figure 4.2



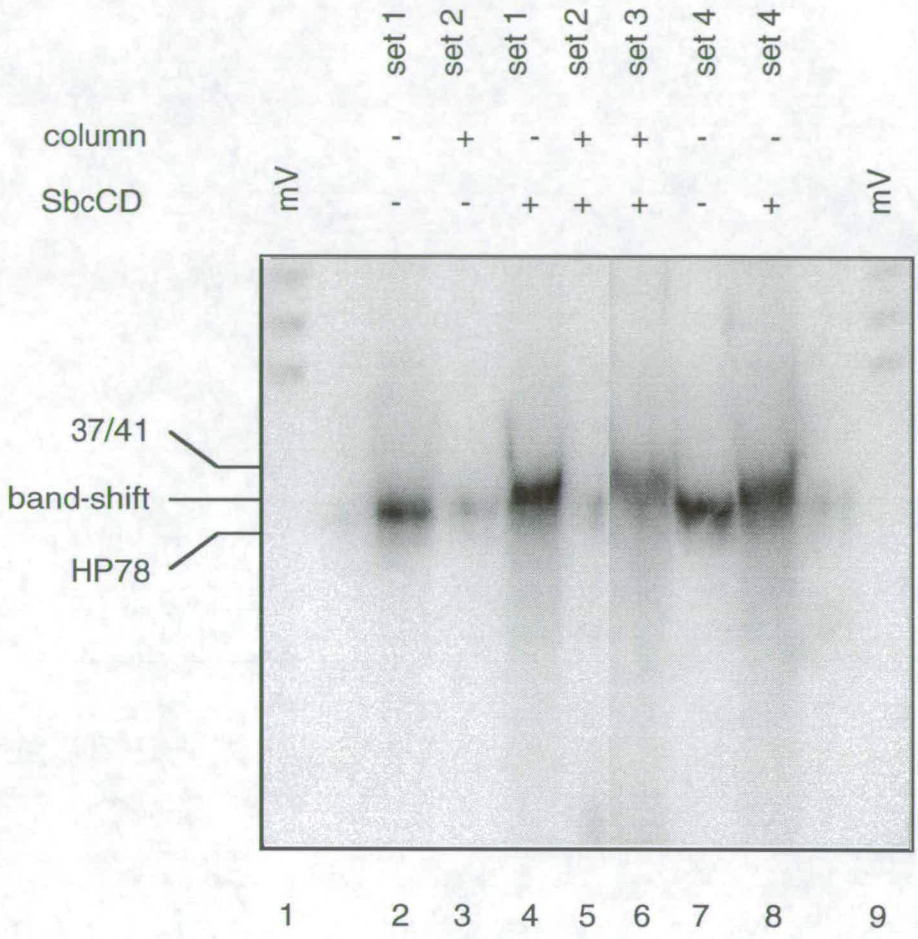
Band analysis by Western Blotting. The undiluted SbcCD standard binding reaction was scaled up. The products were loaded across two 6.8% native polyacrylamide Mini Protean II gels in 1x TAE and run at 50V, 40°C. Consecutive regions of the gels from the well (a) to the band-shift position (f) were excised, eluted, purified and concentrated by ethanol precipitation. Eluted material from these regions was loaded in duplicate onto two ATTO SDS gels. The gels were blotted in parallel onto nitrocellulose filters using a BioRad Semi Dry Cell, and analysed using anti-SbcC (a) and anti-SbcD (b) rabbit immunoglobulins. C indicates the control tracks in which untreated SbcCD protein was loaded as a size standard.

Figure 4.3



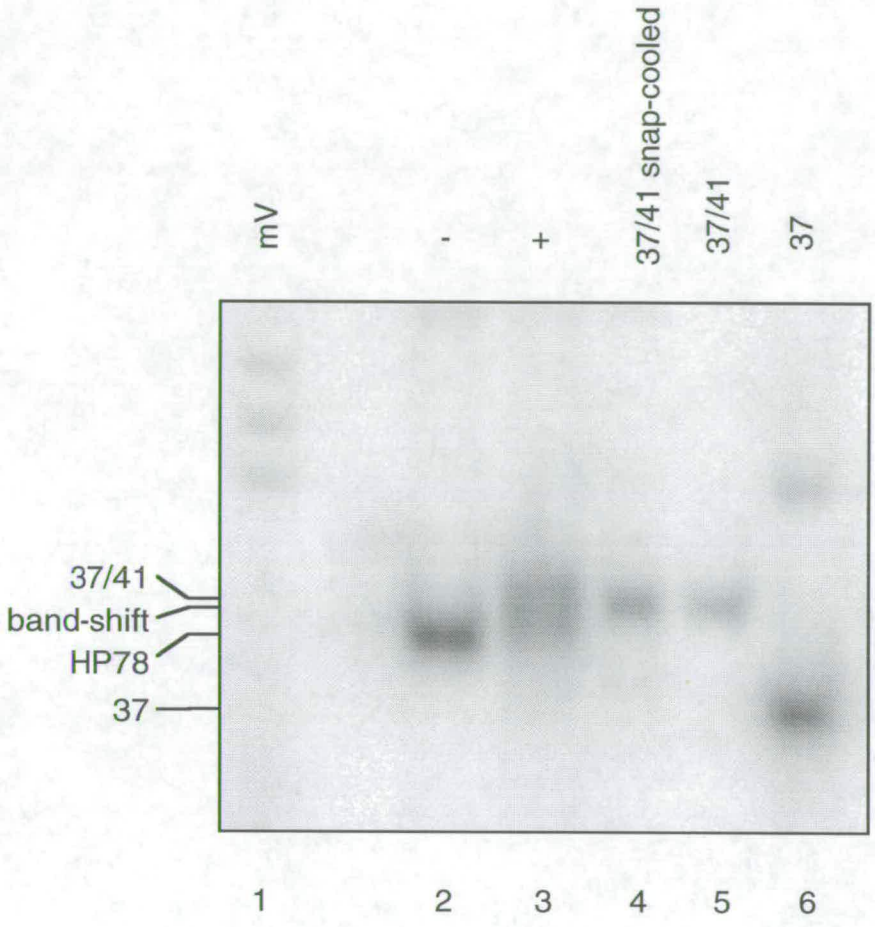
Band analysis by silver staining. The binding reaction was performed in the presence (lanes 2 and 4) and absence (lanes 1 and 3) of 69nM SbcCD, using 5mM either Mn^{2+} (lanes 1 and 2) or Ca^{2+} (lanes 3 and 4) and under conditions lacking BSA. Products were separated on a 6.8% native polyacrylamide Mini Protean II gel in 1x TAE at 50V, 40°C, silver stained and dried.

Figure 4.4



Band analysis by purification from protein. Four sets of standard binding reactions were performed with HP78 and 69nM SbcCD: Set 1 reactions were incubated under binding conditions in the presence (lane 4) and absence (lane 2) of SbcCD and then left on ice for 1 hour. Set 2 reactions were incubated under binding conditions in the presence (lane 5) and absence (lane 3) of SbcCD and purified from protein using QIAquick™ Nucleotide Removal columns. Set 3 reactions were incubated under binding conditions in the presence (lane 6) and absence (data not shown) of SbcCD, column purified and then subjected to a second standard binding incubation. Set 4 reactions were incubated in the presence (lane 8) and absence (lane 7) of SbcCD. All reactions were then immediately electrophoresed on a 12% native sequencing-sized polyacrylamide gel at 30W, 35°C in precooled buffer 1x TAE at 4°C.

Figure 4.5



The oligonucleotide helicase assay. The HP78 standard binding reaction was performed in the presence (lane 3) and absence (lane 2) of 69nM SbcCD and electrophoresed alongside the annealed 37/41 oligonucleotide duplex (lane 5) and the 37mer single-strand oligonucleotide (lane 6) as size standards. As a control, a sample of 37/41 duplex was heated and snap-cooled to promote denaturation. However the duplex ran in its native form (lane 4). Electrophoresis was conducted using a 6.8% Hoefer native polyacrylamide gel in 1x TAE at 30mA (120-180V) 4°C for 5 hours.

Figure 4.6

The effects of metal and nucleotide cofactor requirements. Standard HP78 binding reactions were set up in the presence (+) or absence (-) of 6.9nM SbcCD. The first metal indicates the preincubation metal at a concentration of 1mM. The second indicates the reaction metal at a concentration of 5mM. Figures 4.6a and 4.6b show metal cofactor tests which were carried out in 1mM ATPγS. Figure 4.6c shows nucleotide cofactor tests which were carried out in 1mM Mn^{2+} for preincubation and 5mM Mn^{2+} for reaction. 37mer single-strand (37), 37/41 duplex (37/41) and marker V (mV) were included as size standards. DNA was separated on 6.8% polyacrylamide gels.

Figure 4.6

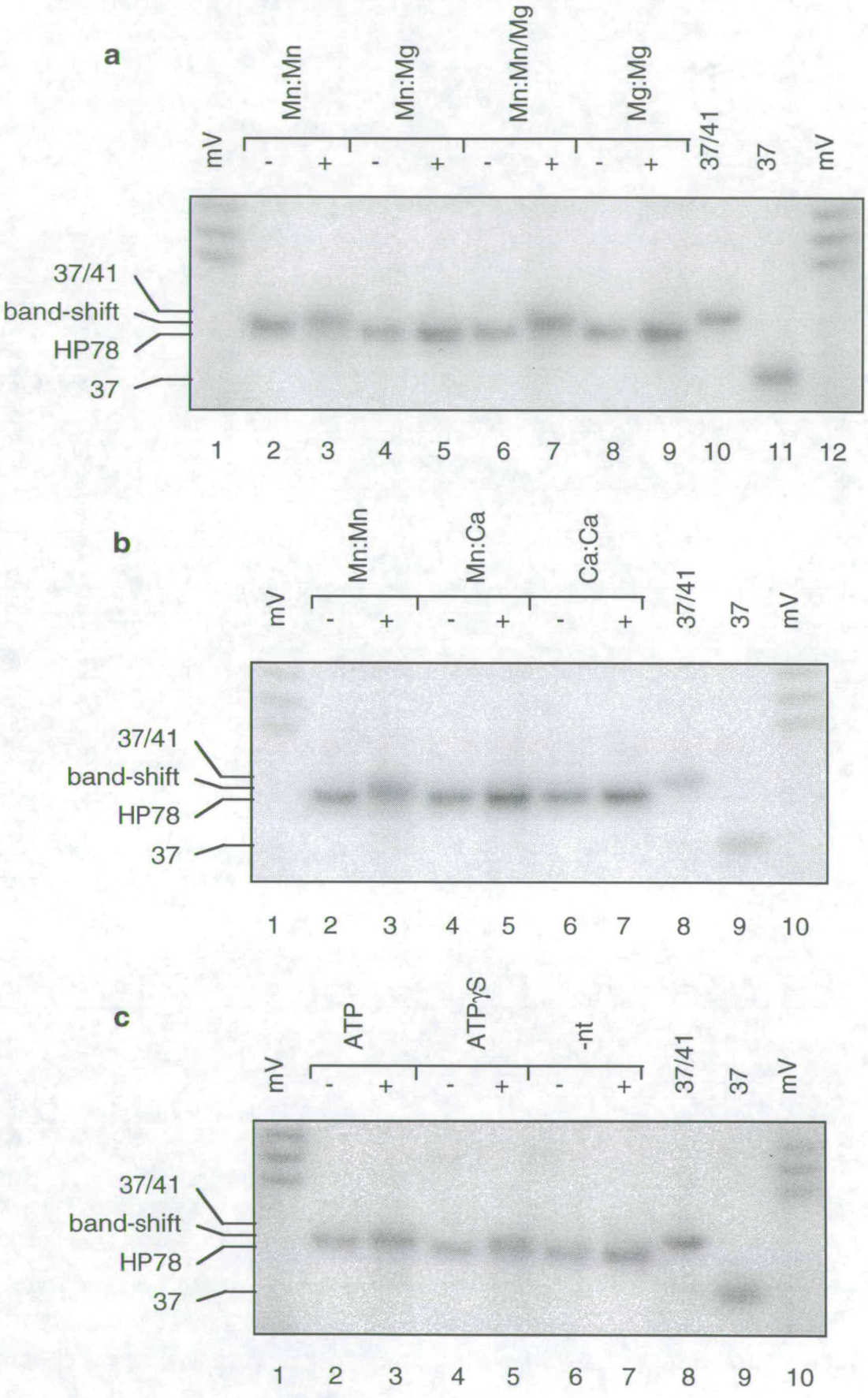
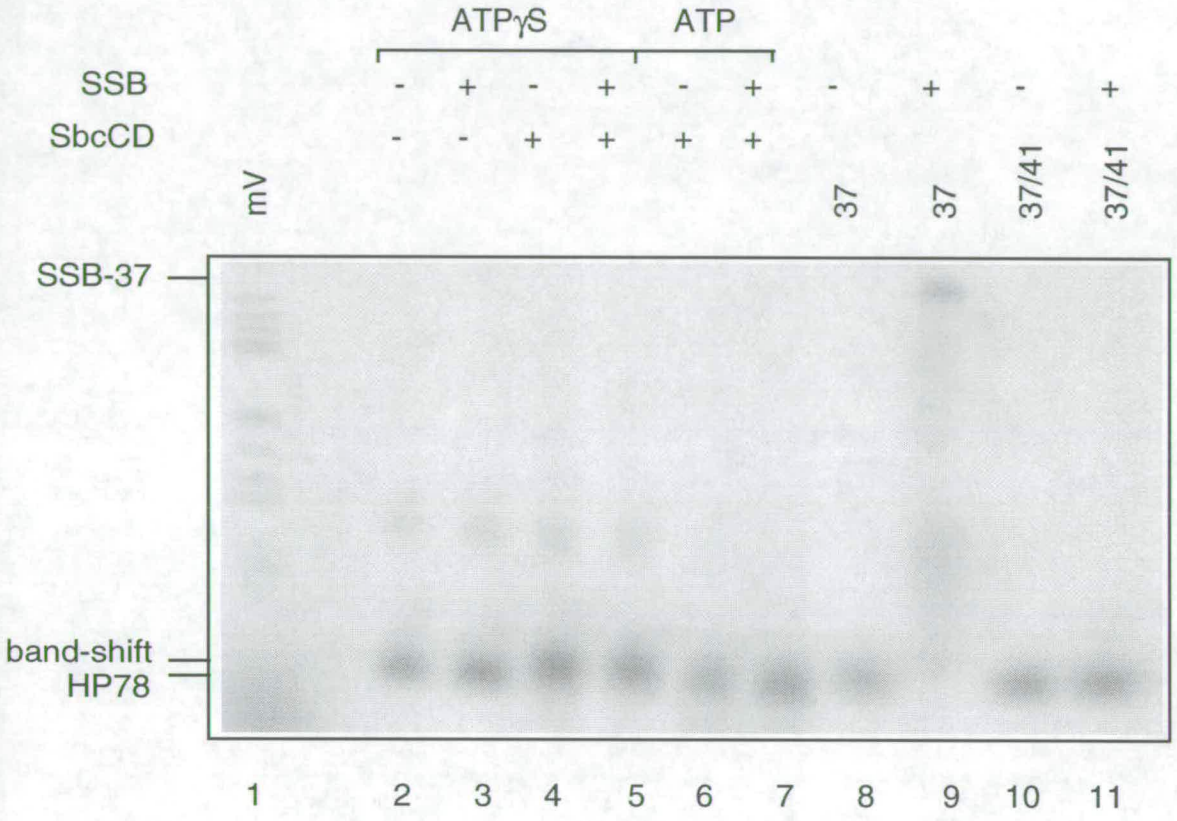
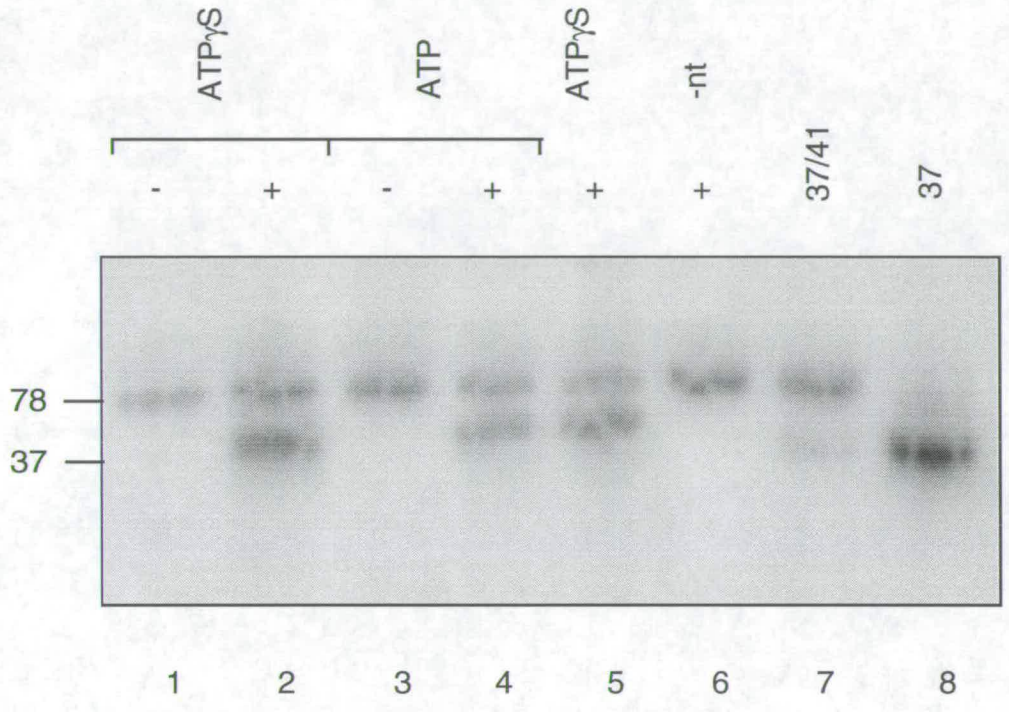

a
Mn:Mn
Mn:Mg
Mn:Mn/Mg
Mg:Mg
mV
- + - + - + - +
37/41
37
mV
37/41
band-shift
HP78
37
1 2 3 4 5 6 7 8 9 10 11 12
b
Mn:Mn
Mn:Ca
Ca:Ca
mV
- + - + - +
37/41
37
mV
37/41
band-shift
HP78
37
1 2 3 4 5 6 7 8 9 10
c
ATP
ATPγS
-nt
mV
- + - + - +
37/41
37
mV
37/41
band-shift
HP78
37
1 2 3 4 5 6 7 8 9 10

Figure 4.7



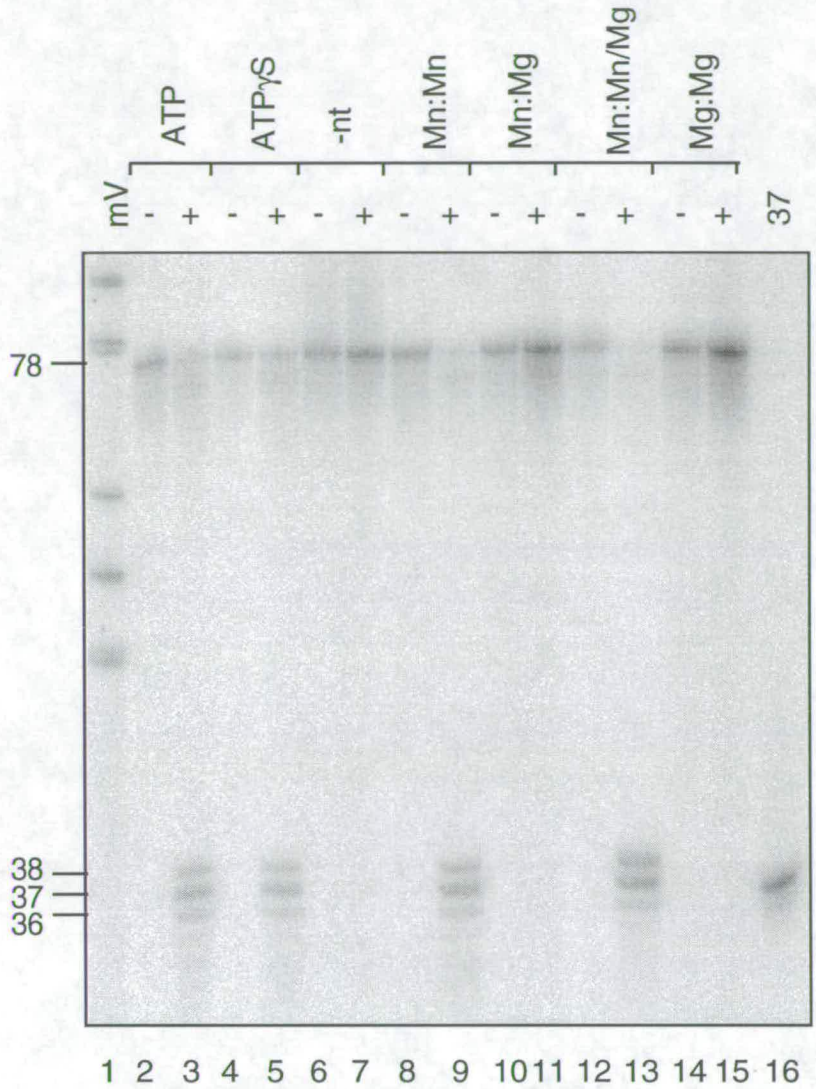
The SSB helicase assay. Binding reactions were set up in the presence (lanes 4, 5, 6 and 7) and absence (lanes 2 and 3) of 69nM SbcCD, and in the presence of ATP (lanes 6 and 7) or ATPγS (lanes 2, 3, 4 and 5). 1.4µg SSB (per 30µl reaction) were included in the reactions (lanes 3, 5 and 7) or not (lanes 2, 4 and 6). As controls, 37mer single-strand was incubated alongside in the presence (lane 9) or absence (lane 8) of SSB, and 37/41 duplex was similarly incubated in the presence (lane 11) or absence (lane 10) of SSB. The products of these reactions were separated on a prerun 5% native polyacrylamide Hoefer gel in 1x TAE at 30mA (120-180V) 4°C for 3 hours.

Figure 4.8



Band analysis by denaturing gel electrophoresis. Nuclease reactions were set up in the presence (lane 2) or absence (lane 1) of 6.9nM SbcCD in ATPγS. Binding reactions were set up in the presence (lanes 4, 5 and 6) or absence (lane 3) of 69nM SbcCD using ATP (lanes 3 and 4), ATPγS (lane 5) or no nucleotide cofactor (lane 6). Products of the binding reactions were added to denaturing gel loading buffer, heated for 4 minutes and loaded onto a 15% Mini Protean II urea polyacrylamide denaturing gel in 1x TAE at room temperature. The samples were run at 50V, 40°C. 37/41 duplex (lane 7) and 37mer single-strand (lane 8) were similarly treated and loaded alongside.

Figure 4.9



Analysis of the metal and nucleotide cofactor requirements for the band shift. Binding reactions were carried out in the presence of various nucleotide and metal cofactors in the presence (+) and absence (-) of undiluted SbcCD. The first metal indicates the preincubation metal, at a concentration of 1mM, and the second, the reaction metal at 5mM. Metal tests were carried out in the presence of 1mM ATPγS, and nucleotide tests in 5mM Mn^{2+}. 37mer single-strand and marker V were added as size standards. Products were resolved on a 10% urea polyacrylamide sequencing sized denaturing gel at 50W, 60°C in 1x TAE.

CHAPTER 5

DNA BINDING BY SbcCD

Introduction

The SbcCD nuclease activity acts on hairpin substrates in the presence of ATP, ATPγS or GTP (Chapter 2). To do so, it must bind DNA.

Binding studies were undertaken *in vitro* with the intention of biochemically dissecting the nuclease reaction to provide details of the interaction between SbcCD and its substrates. The metal ion and substrate requirements of SbcCD for binding were to be determined.

Moreover, DNA binding studies were intended to define alternative cellular roles of SbcCD and possibly its eukaryotic homologues. The model for the regulation of palindrome maintenance by SbcCD hairpin nuclease (Chapter 2) proposes one hypothetical role for SbcCD in *E. coli.* However the mutant phenotypes of the homologues of SbcCD suggest that the homologues have additional cellular roles. Studies using various substrates might provide information pertaining to these roles. If SbcCD were found to bind an alternative substrate but not to cleave it, this might implicate SbcCD in a different pathway, for which additional cofactors might be required.

This chapter describes the course of experiments undertaken to investigate the DNA binding requirements of SbcCD. Section 1 describes initial gel retardation studies. Section 2 describes attempts made to stabilize potentially weak binding. Section 3 describes attempts made to trap potentially transient binding, including the use of mutant SbcCD protein.

Section 1

DNA Binding by SbcCD is Weak or Transient

Attempting to Observe DNA Binding by Gel Retardation

DNA Binding was not Initially Detected by Gel Retardation

These days DNA binding is normally studied by gel retardation (reviewed Taylor *et al* 1994, Carey 1991). This is a preferred method because it allows the protein-DNA complex to be visualised and the size of the complex to be estimated, and because it can be used in conjunction with Western Blotting to identify the bound polypeptides (Sambrook *et al* 1989).

Protein-DNA complexes between SbcCD and end-labelled HP78 were initially studied using gel retardation conditions described by Powell *et al* 1993: A 20µl binding reaction was set up. It contained Tris nuclease buffer pH7.5 (1.25mM DTT, 100µg/ml BSA, 25mM Tris-Cl pH7.5, 2% glycerol), 5mM Mn^{2+}, 1mM ATPγS, 100mM NaCl, 1.5nM snap-cooled hairpin DNA, and 6.9nM protein, diluted in BufferB/BSA/Mn^{2+} and not preincubated. This was incubated at 16°C for 20 minutes whilst a native 10% polyacrylamide ATTO gel in TBE buffer (naturally pH7.0) was prerun at 4°C, 40mA for 20 minutes. The wells were rinsed. An equal volume of dyeless glycerol loading buffer (20% glycerol in 1x Buffer B, naturally pH7.5) was added to each reaction, the samples were immediately loaded onto the gel and the components separated by electrophoresis at 4°C, 40mA. Dye was not used because this might affect the stability of the complexes. A marker DNA was loaded alongside in native gel loading buffer (containing bromophenol blue dye) which allowed the gel running speed to be observed. No trace of any retarded bands was visible on the gel.

A protein titration was performed using concentrations of 69nM (undiluted), 6.9nM and 0.69nM, diluting in BufferB/BSA/Mn^{2+}, and no retardation was observed, figure 5.1.1a. However products of nuclease cleavage could be visualised on the native gel by boiling the binding reactions immediately before loading, indicating that the nuclease was active under these binding conditions, and therefore that binding had occured, figure 5.1.2a. Degradation is not apparent in ATP because the processive nuclease has relatively low activity at 16°C. The material which was retained in the wells when the samples were boiled is likely to correspond to DNA networks.

Possible Reasons for the Lack of Protein-DNA Candidates on Gel Retardation

The size of the active SbcCD nuclease complex has been estimated by sedimentation equilibrium analysis to be 1.2MDa (Connelly *et al* 1997, Chapter 1 Section 3). This corresponds to a very large, multisubunit complex, believed to require Mn^{2+} for assembly. The stoichiometry is unknown, but the complex is currently estimated to consist of eight polypeptides each of SbcC and SbcD (Chapter 1 Section 3). It is unlikely that such a large complex would enter a polyacrylamide gel matrix. Inspite of this, gel retardation would not be prevented if the complex eventually dissociated in the well during electrophoresis. Under these circumstances, the released DNA enters the gel a variable time after the unbound, producing a characteristic smear. Neither would the large size of the active complex prevent gel retardation if a proportion of the protein-DNA complexes were not intact. Complexes might be small enough to enter the gel if they included only some of the polypeptides of the active complex.

A band-shift would not have been observed if the complex were unstable under the reaction or gel conditions. A relatively stable complex might be unstable under suboptimal conditions. An unstable complex might dissociate during the binding reaction, or during transition from the reaction to the gel environment.

Alternatively, the complex might be intrinsically unstable. This could result from weak or transient binding.

Optimising Gel Retardation

Gel Retarded Bands are Obtained in the Absence of Crosslinking

Measures were taken to create reaction and gel conditions more conducive to stable complex formation. Glycerol was added to the reaction to a total of 7%. Glycerol provided the high density required to keep the reactions in the wells. This avoidance of the use of gel loading buffer reduced "dilution shock," the sudden change in conditions which occurs on loading of a DNA

binding reaction to a gel and can destabilize the protein-DNA complexes (Taylor *et al* 1994). In order to reduce the "electrophoretic dead time," the time taken for protein-DNA to migrate from the binding reaction into the gel, which should be as short as possible, the initial three minutes of electrophoresis were carried out at a higher current of 100mA. The current was returned to 40mA for the remainder of the run. A 6.8% native polyacrylamide gel was used instead of the 10%. This provided a less harsh environment for protein-DNA complexes, and it was hoped that the lower concentration of the matrix might allow entry of fragments of the active complex. 69nM protein was added per reaction instead of 6.9nM.

When these modifications were used in combination, two gel retarded bands were observed, figure 5.1.2b lanes 1-7. Using native HP78 as a standard, the first had an RF value of between 0.29 and 0.32, modal RF 0.30. It migrated at approximately 200bp. However this band constituted an extremely low proportion of the total DNA, equal to approximately 0.05% by phosphorimage ImageQuant™ analysis. Also present was a band migrating at approximately 80bp (RF 0.70). This probably corresponds to a duplex of the unfolded HP78 substrate.

Candidate Complexes are Obtained by Crosslinking

It was intended that the yield of candidate complexes be increased by glutararaldehyde crosslinking. The binding reactions were glutaraldehyde crosslinked by the method used to visualise RAG1-RAG2-DNA complexes (Hiom and Gellert 1997). A similar method was used to visualise RuvAB-Holliday Junction complexes (Parsons and West 1993, Hiom and West 1995b). Glutaraldehyde forms covalent crosslinks between amine groups when incubated at 37°C (Korn *et al* 1972). In practice the crosslinks form between amino acids, particularly lysine residues, and effectively trap protein-DNA complexes by locking protein around DNA.

Glutaraldehyde (Sigma) was added to the binding reactions, after incubation at 16°C, to a final concentration of 0.1%, and the reactions were transferred to 37°C for 10 minutes further incubation before immediately loading on the gel. The fixation reactions were not quenched. Crosslinking did not increase the yield of the previously observed band, RF 0.30, figure 5.1.2b lanes 8-12.

However two additional gel retarded bands were observed in the presence of Mn^{2+} and ATPγS, figure 5.1.2b lane 9. Using native HP78 as a standard, the first had an RF value of 0, corresponding to material trapped in the well. The second had an RF value between 0.38 and 0.54, modal RF 0.41. These bands corresponded to approximately 5% and 2% respectively of the total DNA loaded, estimated from ImageQuant™ analysis. The band RF 0.41 was not obtained consistently and its yield was highly variable. It migrated alongside DNA of approximately 170bp.

Cleavage products were visible (not shown), probably because the nicked hairpin dissociated during the glutaraldehyde incubation.

Glutaraldehyde crosslinking has limited use; the method is infamous for producing non-specific protein-DNA networks and complexes. The results obtained are therefore commonly artefactual. Caution must be exercised in interpreting such data.

Optimising Glutaraldehyde Crosslinking

The efficiency of glutaraldehyde fixation is reduced by high salt conditions. Salt was therefore removed from the binding reaction and this increased fixation 3-fold, figure 5.1.3a lanes 5-8.

4 moles glutaraldehyde crosslinks 1 mole amines (Korn *et al* 1972). Each mole Tris contains 1 mole amines. Since glutaraldehyde crosslinking occurs between amine side groups, use of the Tris buffer might have decreased the efficiency of fixation. Fixation in Tris was therefore compared to that in two non-amine buffers, MOPS and HEPES, figure 5.1.3a. Fixation was not much affected by the buffer, indicating that the buffer was not competing with SbcCD for glutaraldehyde. The greatest level of fixation occured in HEPES, as estimated by phosphorimage analysis.

Glutaraldehyde incubation time was varied between 1 and 60 minutes, and fixation temperatures of 0°C, 16°C and 37°C were tested. All three candidate complexes were visible using incubation at 16°C for 10 minutes, and the maximum yield of the faint bands was obtained under these conditions (data not shown). This temperature may be preferable to the standard 37°C on

account of the reduced disruption to binding conditions. The optimum glutaraldehyde concentration was determined by varying between 0.001% and 5%, figure 5.1.3b. The optimal yield of the three candidates was obtained using a concentration of 0.1%, and no additional complexes were observed at the intermediate concentrations. Higher glutaraldehyde concentrations resulted in DNA crosslinking in the absence of protein (data not shown). It was possible that use of a higher glutaraldehyde concentration in conjunction with a lower protein concentration might reduce networking and increase the yield of the candidate complexes. However the banding pattern obtained with 1% glutaraldehyde and 69nM protein was unchanged when the protein concentration was reduced to 13.8nM or 6.9nM (data not shown).

Optimising Binding Reaction Conditions

The glycerol concentration in the binding reaction was varied from 2% to 15% with no observable effect on the banding pattern; the bands RF 0 and RF 0.41 were obtained in all tracks (data not shown). 5% glycerol is commonly used for DNA binding studies.

The time of incubation of the binding reaction was varied between 0 minutes and 120 minutes. The candidate complex bands RF 0 and RF 0.41 were obtained after incubations between 2 minutes and 120 minutes. Binding candidates accumulated gradually over the 2 hours, but the increase beyond 20 minutes was slight. The same banding pattern was obtained whether the binding reactions were incubated at 0°C, 16°C or 37°C (data not shown).

After the importance of Mn^{2+} for the stability of SbcCD had been demonstrated (Connelly *et al* 1997, Chapter 1 Section 3), protein was preincubated in 1mM Mn^{2+} to increase its stability prior to nuclease assays. The time of preincubation, from 0 minutes to 30 minutes, had no effect on the banding pattern obtained under gel retardation conditions. The temperature of preincubation, 0°C, 16°C or 37°C, also had no effect. The preincubation metal ion had no effect unless Mn^{2+} in the binding reaction was removed, in which case Mn^{2+} was required in the preincubation. The banding pattern was also not affected by preincubation in Mn^{2+} with ATPγS, or preincubation

in 1mM ATPγS or 5mM metal. The sequence of binding reactant addition was varied with no effect on the banding pattern (data not shown).

Maximal binding in a DNA binding assay requires an excess of either protein or DNA, so that one saturates the other. A low concentration of DNA is preferred when optimising for gel retardation (*cf* Western Blotting). Protein titrations were performed to determine the effect of protein concentration on gel retardation, figure 5.1.1. ImageQuant™ analysis showed that the band RF 0.41 was more efficiently obtained using diluted DNA with undiluted protein, figure 5.1.1b. However this candidate was more clearly visible on the phosphorimage when a higher concentration of DNA was used, figure 5.1.1b and 5.1.1c. Even at the higher concentrations of DNA, the molarity of protein still exceeded that of DNA. When protein was titrated against DNA at a moderate to low concentration (below that of the protein) the banding pattern changed, figure 5.1.1d. RF 0.41 disappeared at the point where DNA and protein molarities were equal, and another, more diffuse, less retarded band appeared of RF 0.67. This change may correspond to a difference in stoichiometry of the candidate complex. RF 0.41 was retained at a protein dilution of up to 5-fold.

To test the effect of ATPγS on DNA binding, a titration of ATPγS from 0.01mM to 10mM was performed, figure 5.1.4a and 5.1.4b. A graph was plotted of ATPγS concentration against the percentage of the total loaded DNA in the candidate band (potentially equal to the percentage bound). This was done for the bands RF 0 and RF 0.41, figure 5.1.4c. The optimal concentration of ATPγS for retardation is between 0.1mM and 1mM. The maximum yield of the RF 0.41 candidate was 3% of total DNA, which was obtained at the 1mM ATPγS point. The maximum yield of RF 0 was approximately 8%, also at 1mM. The curve appears not to be unlike that for nuclease activity (Connelly *et al* 1997). No additional bands were observed.

A titration of Mn^{2+} from 0.1mM to 100mM was performed, figure 5.1.5a. A graph was plotted of Mn^{2+} concentration against the percentage of DNA in the band for bands RF 0 and RF 0.41, figure 5.1.5b. The optimal Mn^{2+} concentration for candidate retardation is between 15mM and 50mM. The maximum yield of RF 0.41 was 4% and of RF 0 was 18%, both at 50mM Mn^{2+}. No additional bands were observed.

The effects of different metal ions and nucleotide cofactors were also examined with the intention of optimising the yield of candidate complexes by gel retardation. These studies were also relevant to the project of uncoupling SbcCD binding activity from nuclease activity, and are discussed in Section 3.

Optimising Gel Conditions

The speed of electrophoresis was varied between 5mA and 40mA without affecting the banding pattern obtained. Resolution was slightly improved by running the gel in a 4°C environment, compared to room temperature (data not shown).

Summary of the Glutaraldehyde Binding Assay Conditions

In subsequent glutaraldehyde studies, binding reactions were performed in buffer containing 5% glycerol, and 1mM ATPγS, 5mM Mn^{2+}, 25mM Tris-Cl pH7.5, 1.25mM DTT and 100μg/ml BSA. 1.5nM substrate DNA was used, and 69nM protein, preincubated in 1mM Mn^{2+} at 0°C for 10 minutes, was added. The reactions were incubated at 16°C for 20 minutes. Glutaraldehyde fixation was then conducted by adding 0.1% glutaraldehyde and incubating at 16°C for 10 minutes. 6.8% gels in 1 x TBE were prerun 10 to 20 minutes at 4°C 40mA, loaded, the samples run in at 100mA for 3 minutes, and then the gels run 4°C at 40mA.

Gel Retardation using Agarose Gels

As detailed above, use of a lower percentage polyacrylamide matrix was one of several methods employed in conjunction to visualise candidate complexes. The 6.8% gel was still too concentrated to allow entry of a large amount of glutaraldehyde crosslinked material, which was retarded on these gels as the well band RF 0.

Complexes of 1MDa in size are known to enter 0.8% agarose gels (Alexander Alexandrov personal communication). Therefore SbcCD would be expected to enter such gels. Attempts were made to resolve the material trapped in the wells of the polyacrylamide gels by running 0.8% native agarose gels in 1x TAE buffer at 80mA, 4°C. In addition to a well-trapped band RF 0, two

candidate complex bands were observed. The first migrates on different agarose gels with an RF value between 0.67 and 0.72, modal RF 0.71. The second has an RF between 0.35 and 0.42, modal RF 0.38, figure 5.1.4b. Both bands were observed only in the presence of protein, and appeared under the same conditions as polyacrylamide (p) RF 0.41. The dimer band observed on 6.8% polyacrylamide gels was not visible, and presumably runs with substrate on these gels. It is likely that the smaller of the candidates, agarose (a) RF 0.71, corresponds to the candidate pRF 0.41, and that the larger, aRF 0.38, corresponds to part of pRF 0. The well material, aRF 0, did not enter the agarose gel even after 3 hours of electrophoresis (data not shown) which suggests it is considerably larger than 1MDa in size. It is likely to represent a non physiological artefact of glutaraldehyde crosslinking. The other agarose bands were more diffuse than polyacrylamide bands.

Identification of the Candidate Complexes

A total of four retarded bands was observed on 6.8% polyacrylamide gels and four on agarose gels. The variation in RF values of a species between similar gels presumably results from differences in running conditions. By examining multiple gels, the following deductions concerning the identity of these bands can be made.

The aRF 0 component of the pRF 0 band was insensitive to reaction conditions and usually obtained on glutaraldehyde fixation, irrespective of other conditions. It was frequently observed in the absence of protein, suggesting that it consists, at least in part, of non-specifically glutaraldehyde crosslinked DNA networks. It is unlikely to include protein-DNA complexes of less than 1MDa since it is not resolved in agarose gels.

The pRF 0.30 band, migrating at approximately 200bp, was inconsistently obtained. Its yield was very low. Where visible on a gel, it generally appeared faintly in the absence of protein, indicating that it does not correspond to a protein-DNA complex. (It was apparently absent minus protein in figure 5.1.2b, which led to its initial consideration as a candidate complex. But it is visible on other gels in the absence of protein. It is therefore likely that this band was actually present faintly in figure 5.1.2b.) It

was insensitive to variations in conditions. The identity of this band is unknown. It may represent a concatemer formed of unconventional base pairs.

This leaves three true candidate complexes. The first is represented by bands pRF 0.41/aRF 0.71 and was obtained inconsistently in the presence of protein, ATPγS and Mn^{2+}. Glutaraldehyde fixation was required. This species may correspond to glutaraldehyde fixed protein bound substrate or product DNA. Its yield was variable. These two possibilities could be distinguished by heating and snap-cooling after the reaction to favour hairpin conformation. Although glutaraldehyde crosslinks are covalent, they form between protein sites, and it had already been observed on other gels that boiling disrupts a glutaraldehyde crosslinked protein-DNA complex. If the band corresponded to protein bound DNA substrate, no single-strand products would be visible on snap-cooling. If it corresponded to protein bound product, snap-cooling would be expected to generate a 37mer single-strand product (discussed Chapter 4). This was visible, figure 5.1.6 (however see figure 5.3.4c, discussed Section 3). A protein bound product would be expected to be unstable and therefore obtained at variable low levels, and would require protein, ATPγS and Mn^{2+}, as does this band. (The duplex band was also lost on snap-cooling, consistent with its identification).

When protein molarity was reduced below that of DNA, the pRF 0.41 species was replaced with a pRF 0.67 species. This second complex suggests a different stoichiometry of protein and substrate molecules in the reaction under these conditions.

The other true candidate band was aRF 0.38. Like pRF 0/aRF 0 and pRF 0.41/aRF 0.71, it was obtained on glutaraldehyde treatment only. Like pRF 0.41/aRF 0.71, it was obtained inconsistently in the presence of protein, ATPγS and Mn^{2+}, and its yield was variable. It did not enter a 6.8% polyacrylamide gel indicating that it corresponds to a large protein-DNA complex. The dependence of this band on protein, ATPγS (ATP or GTP, detailed in Section 3) and Mn^{2+} suggests that, like pRF 0.41/aRF 0.71, it may be a protein-product or protein-substrate complex.

Discussion

Summary

No candidate complexes were observed under the gel retardation conditions of Powell *et al* 1993. Two gel retarded bands were obtained on a 6.8% polyacrylamide gel, one of which was duplex 78. Glutaraldehyde crosslinking was performed to increase the yield of candidate complexes. Four gel retarded species were obtained. One was too large to enter even a 0.8% agarose gel, pRF 0/aRF 0. This was likely to be a glutaraldehyde artefact possibly consisting of protein-DNA networks and non physiological protein-DNA complexes. Another did not enter a 6.8% polyacrylamide gel but migrated into a 0.8% agarose gel with RF 0.38. A third migrated into a 6.8% polyacrylamide gel with RF 0.41 and in a 0.8% agarose gel with RF 0.71. This species was replaced with a fourth, pRF 0.67, when protein molarity was decreased below DNA molarity. It was deduced that the latter three species might be true candidate complexes.

The variability in binding under controlled conditions may have been due misfolding or aggregation of the protein on repeated freeze-thaw cycles (Akhmedov *et al* 1998), however freeze-thaw was minimised to reduce this.

Conclusion: DNA Binding by SbcCD is Weak or Transient

The variable low yield and inconsistency of recovery of candidate complexes indicates that the active complex is not stable under any of the conditions of gel retardation employed. Many protein-DNA complexes, other than SbcCD-DNA, would have been observed under the wide range of gel conditions tested. It is therefore probable that the SbcCD bound substrate complex is relatively unstable.

Unstable DNA Binding

Weak Binding

Intrinsic instability of a protein-DNA complex can result from weak binding. In this case the protein, *P*, constantly associates with and dissociates from its substrate, D_s, and the equilibrium between bound and unbound protein favours unbound. High affinity binding is characterised by a low dissociation constant K_d or K_m (~pM to nM), *eg*SSB, *eg Eco*K methyltransferase with AdoMet cofactor has a K_d of 1.6nM (Powell *et al* 1993). Low affinity binding is characterised by a high K_d (~nM to μM), *eg* Taq*I* restriction endonuclease at low temperature has a K_d of 2.5μM (Zebala *et al* 1992). RecBCD binds DNA with a K_d between 0.1nM (Taylor and Smith 1995) and 3.0nM (Gabbidon *et al* 1998) depending on the nature of the double-strand end. During Mu transposition, MuA protein initially binds DNA reversibly. After donor cleavage MuA binding becomes essentially irreversible (low K_d) (Mizuuchi *et al* 1991).

$$D_s + P <=> D_sP$$

Transient Binding

Intrinsic instability can also result if the protein has a high turnover rate. In this case, binding is almost concomitant with activity, after which the protein rapidly dissociates from the product DNA. Binding occurs transiently, and its K_d may be high or low. Where D_p is product DNA,

$$D_s + P <=> D_sP -> D_pP -> D_p + P$$

RuvAB binding to Holliday Junctions is unstable (Hiom and West 1995b).

Attempting to Stabilize DNA Binding

Methods were designed specifically to increase the stability of the potentially weak or transient DNA binding. These are detailed in Sections 2 and 3 respectively. Gel retardation studies were mostly carried out using polyacrylamide gels, despite the inability of such gels to resolve an intact complex. This was because bands visible in agarose gels were diffuse and weak bands were therefore difficult to see. Any candidate material observed in the well could be analysed subsequently using agarose gels.

Figure 5.1.1

Protein titrations. HP78 substrate was heated, snap-cooled, 5' end-labelled and incubated at 16°C for 20 minutes under DNA binding conditions with SbcCD in the presence of ATPγS and Mn^{2+} ions. The *crescendo* symbol represents the ratio protein: DNA. Products were separated on 6.8% native polyacrylamide ATTO gels, except in figure a where a 10% gel was used. Precooled gels were prerun at 4°C, 40mA for 20 minutes and then run at 4°C 40mA.

a

Binding reactions were heated and snap-cooled immediately prior to loading on the gel. Reactions included 0mM (lane 1), 0.69nM (lane 2), 6.9nM (lane 3) and 69nM (lane 4) SbcCD and a constant concentration of 1.5nM DNA. This represents approximate SbcCD: DNA ratios of 0 (lane 1), 1:2 (lane 2), 9:2 (lane 3) and 46:1 (lane 4).

b

Binding reactions were set up using 1.5nM DNA and 6.9nM SbcCD (lane 2), 1.5nM DNA and 69nM SbcCD (lane 3), 0.15nM DNA and 69nM SbcCD (lane 4), 0.015nM DNA and 69nM SbcCD (lane 5), 0.15nM DNA and 6.9nM SbcCD (lane 6) and 0.015nM DNA and 6.9nM SbcCD (lane 7). All reactions were then incubated with 0.1% glutaraldehyde at 37°C for 10 minutes. Lane 1 represents the same reaction as lane 2, but products were heated and snap-cooled prior to loading. This represents approximate SbcCD: DNA ratios of 9:2 (lanes 1 and 2), 46:1 (lane 3), 460:1 (lane 4), 4600:1 (lane 5), 46:1 (lane 6) and 4600:1 (lane 7).

c

Binding reactions were set up with 1.5nM DNA (lane 1), 7.0nM DNA (lane 2) and 15nM DNA (lane 3) with a constant concentration of 69nM SbcCD. Reactions were then incubated with 0.1% glutaraldehyde at 16°C for 10 minutes before loading. This represents approximate SbcCD:DNA ratios of 46:1 (lane 1), 10:1 (lane 2) and 9:2 (lane 3).

d

Binding reactions were set up with 0nM (lane 1), 69nM (lane 2), 34.5nM (lane 3), 13.8nM (lane 4), 6.9nM (lane 5), 2.76nM (lane 6) and 1.38nM (lane 7) SbcCD and a constant concentration of 7nM DNA. Reactions were then incubated with 0.1% glutaraldehyde at 16°C for 10 minutes before loading. This represents approximate SbcCD:DNA ratios of 0 (lane 1), 10:1 (lane 2), 5:1 (lane 3), 2:1 (lane 4), 1:1 (lane 5), 2:5 (lane 6) and 5:1 (lane 7).

Figure 5.1.1

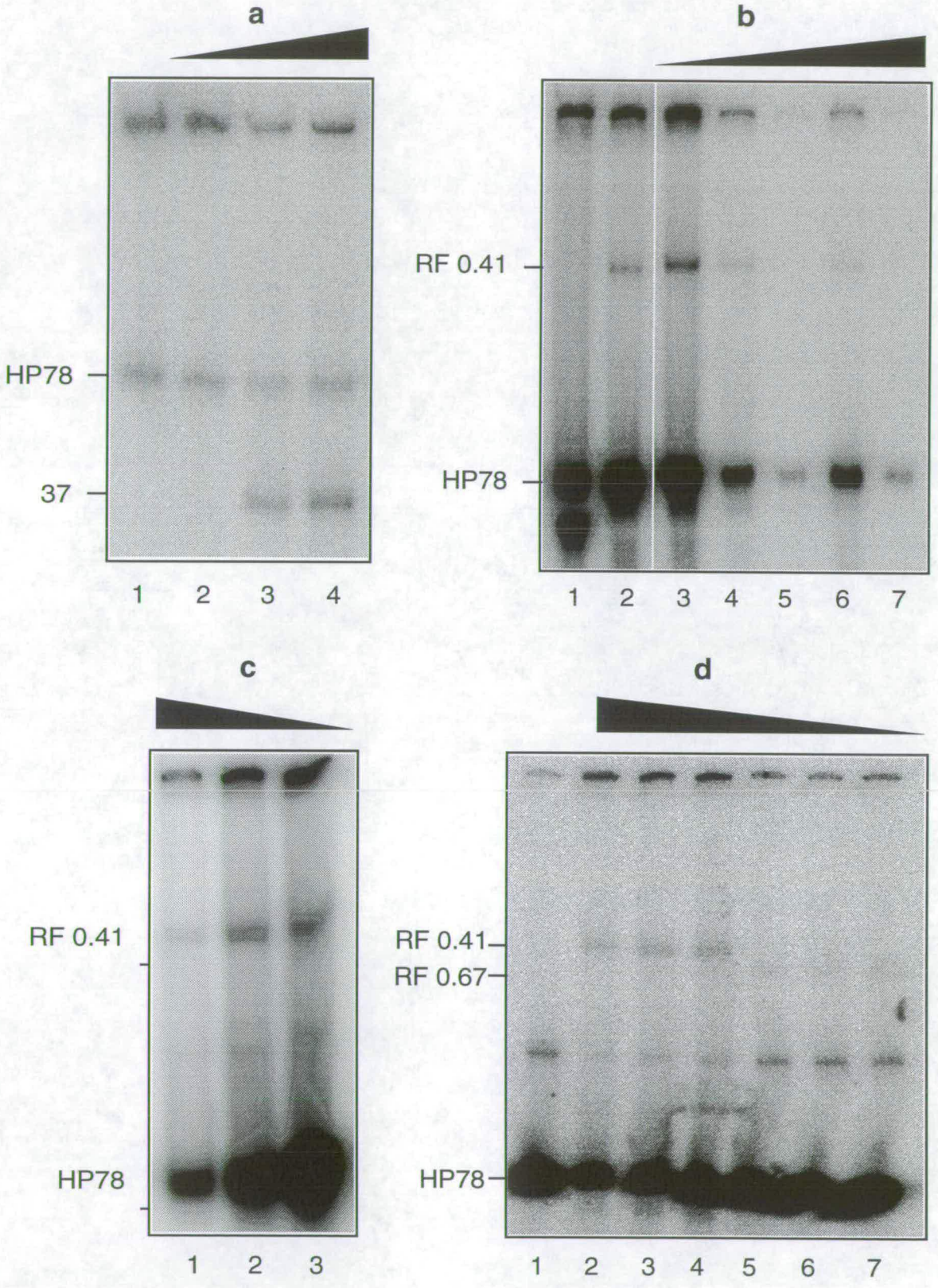

a
HP78
37
1 2 3 4
b
RF 0.41
HP78
1 2 3 4 5 6 7
c
RF 0.41
HP78
1 2 3
d
RF 0.41
RF 0.67
HP78
1 2 3 4 5 6 7

Figure 5.1.2

Optimisation of gel retardation conditions. HP78 substrate was heated, snap-cooled, 5' end-labelled and incubated at 16°C for 20 minutes under DNA binding conditions with SbcCD. Reactions were separated on 6.8% native polyacrylamide ATTO gels. The gels were prerun at 4°C for 20 minutes before electrophoresis of the samples at 4°C 40mA.

a

Reactions were set up in the absence of nucleotide (lanes 1 and 4), in ATP (lanes 2 and 5) and in ATPγS (lanes 3 and 6), and in the presence of Mn^{2+} ions. Reactions in lanes 1, 2 and 3 were heated and snap-cooled prior to loading.

b

Reactions were set up in the presence of ATPγS (lanes 1, 2, 3, 4, 8 and 9) or in the absence of nucleotide (lanes 5, 6, 7, 10, 11 and 12). SbcCD was absent in reactions 1 and 8. Mn^{2+}, Mg^{2+} and no metal ion was added to reactions 2, 3 and 4, 5, 6 and 7 and 10, 11 and 12 respectively. Mn^{2+} was also added to reaction 9. Reactions 8 to 12 were treated with 0.1% glutaraldehyde at 37°C for 10 minutes prior to loading.

Figure 5.1.2

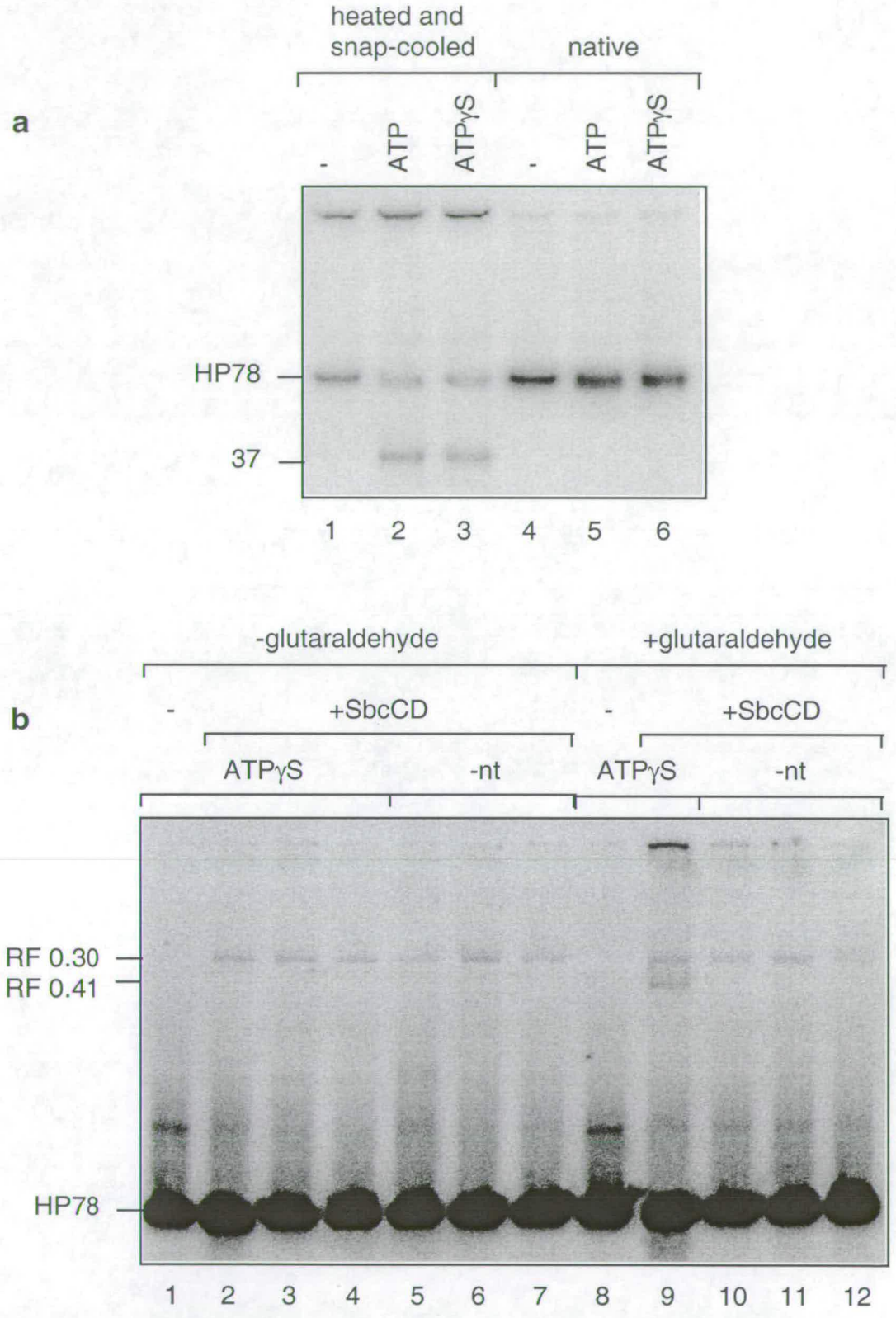

a
heated and
snap-cooled
native
-
ATP
ATPγS
-
ATP
ATPγS
HP78
37
1 2 3 4 5 6
b
-glutaraldehyde
+glutaraldehyde
-
+SbcCD
-
+SbcCD
ATPγS
-nt
ATPγS
-nt
RF 0.30
RF 0.41
HP78
1 2 3 4 5 6 7 8 9 10 11 12

Figure 5.1.3

Glutaraldehyde fixation. HP78 substrate was heated, snap-cooled, 5' end-labelled and incubated at 16°C for 20 minutes under DNA binding conditions with SbcCD. Reactions were separated on 6.8% native polyacrylamide ATTO gels. The gels were prerun at 4°C for 20 minutes before electrophoresis of the samples at 4°C 40mA.

a

Binding reactions were set up in 25mM HEPES nuclease buffer containing 2% glycerol and lacking NaCl (lanes 1 and 2), 25mM MOPS nuclease buffer containing 2% glycerol and lacking NaCl (lanes 3 and 4), 25mM Tris nuclease buffer containing 2% glycerol and lacking NaCl (lanes 5 and 6) and 25mM Tris nuclease buffer including 7% glycerol and 100mM NaCl (lanes 7 and 8). SbcCD was present (+) in reactions 1, 3, 5 and 7. Reactions were incubated with 0.1% glutaraldehyde at 16°C for 10 minutes prior to loading.

b

Binding reactions were incubated with 5% (lane 2), 2.5% (lane 3), 1% (lane 4), 0.5% (lane 5), 0.1% (lanes 1 and 6), 0.05% (lane 7), 0.01 (lane 8) and 0.001% (lane 9) glutaraldehyde at 16°C for 10 minutes prior to loading. SbcCD was absent (-) from reaction 1.

Figure 5.1.3

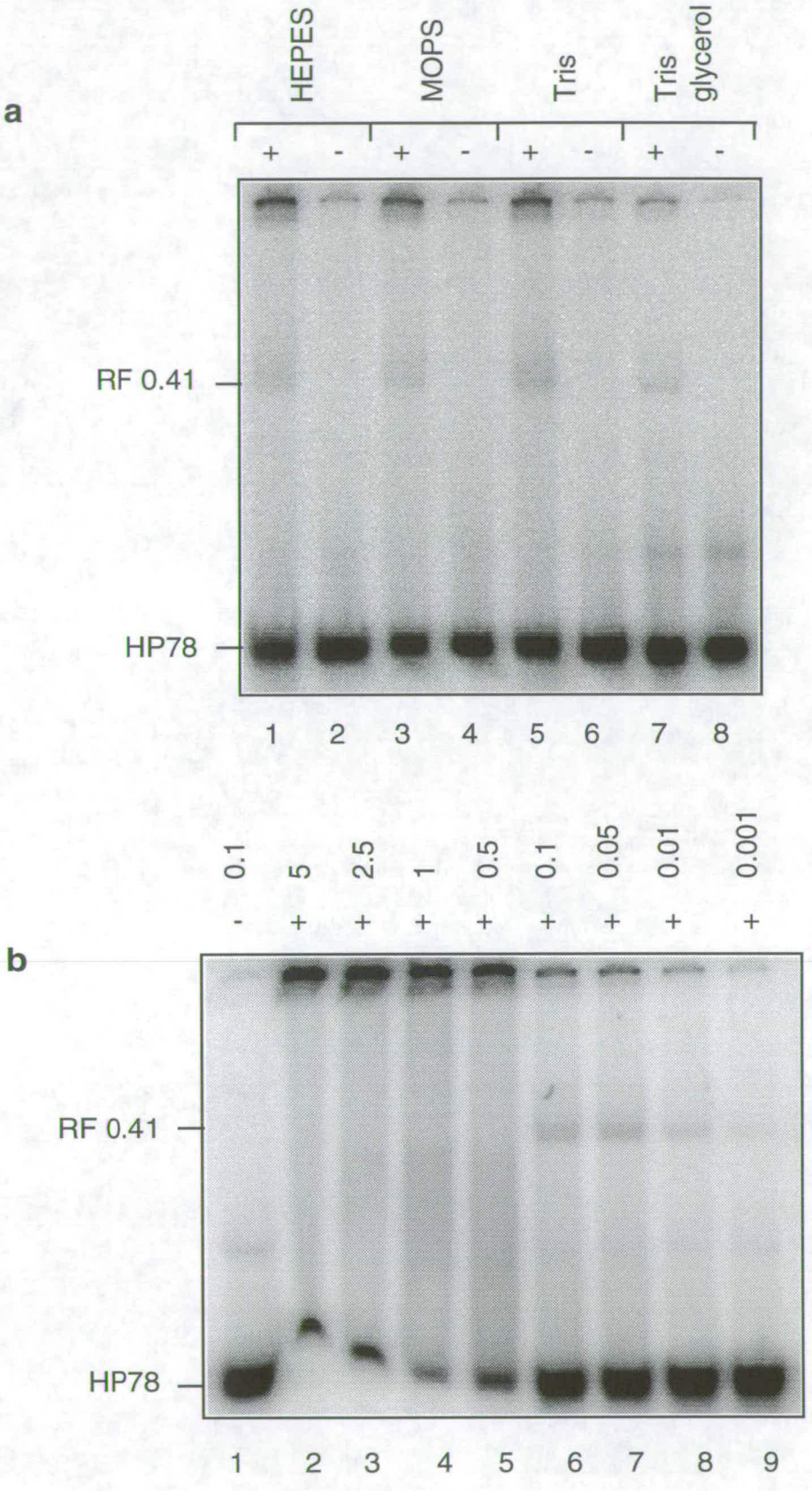

a
HEPES
MOPS
Tris
Tris glycerol
+ - + - + - + -
RF 0.41
HP78
1 2 3 4 5 6 7 8
b
0.1 5 2.5 1 0.5 0.1 0.05 0.01 0.001
- + + + + + + + +
RF 0.41
HP78
1 2 3 4 5 6 7 8 9

Figure 5.1.4

Preliminary ATPγS titration. HP78 substrate was heated, snap-cooled, 5' end-labelled and incubated at 16°C for 20 minutes under DNA binding conditions with SbcCD and in 5mM Mn^{2+}. In figures **a** (6.8% native polyacrylamide gel) and **b** (0.8% agarose gel), binding reactions were set up using ATPγS concentrations from 0.01mM (lanes 1 and 2) to 0.1mM (lanes 3 and 4), 1mM (lanes 5 and 6), 5mM (lanes 7 and 8) and 10mM (lanes 9 and 10). The presence (+) and absence (-) of SbcCD is indicated. RF values for the bands observed are also indicated. Reactions were treated with glutaraldehyde prior to loading.

Figure **c** shows a graph of the recovery of the candidate complex bands visible in a. The results are preliminary. Values were obtained by phosphorimage analysis and have been corrected for background present in the absence of SbcCD.

Figure 5.1.4

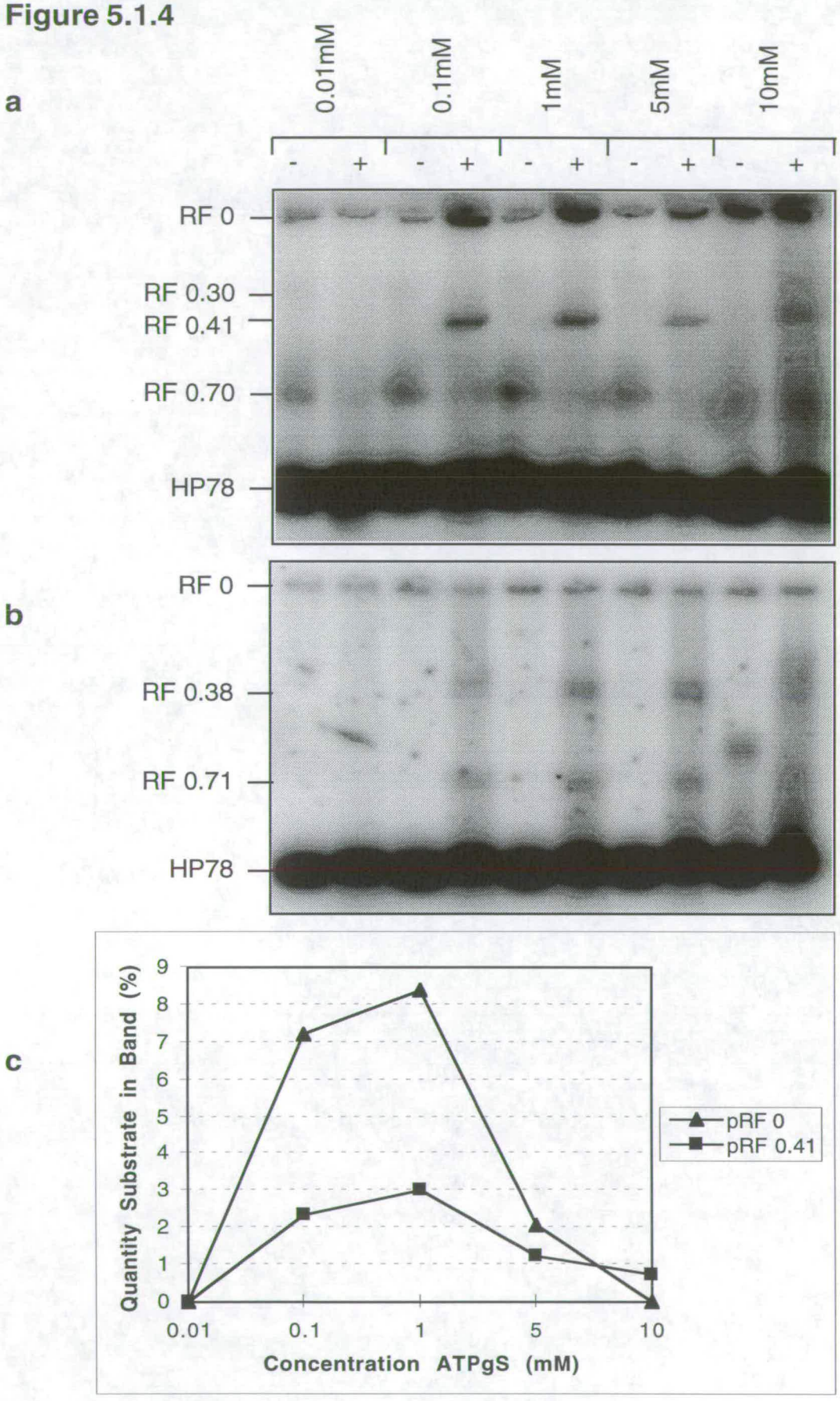

a
0.01mM
0.1mM
1mM
5mM
10mM
- + - + - + - + - +
RF 0
RF 0.30
RF 0.41
RF 0.70
HP78
b
RF 0
RF 0.38
RF 0.71
HP78
c
Quantity Substrate in Band (%)
0
1
2
3
4
5
6
7
8
9
0.01
0.1
1
5
10
Concentration ATPgS (mM)
pRF 0
pRF 0.41

Figure 5.1.5

Preliminary Mn^{2+} titration. HP78 substrate was heated, snap-cooled, 5' end-labelled and incubated at 16°C for 20 minutes under DNA binding conditions with SbcCD and in 1mM ATPγS.

a

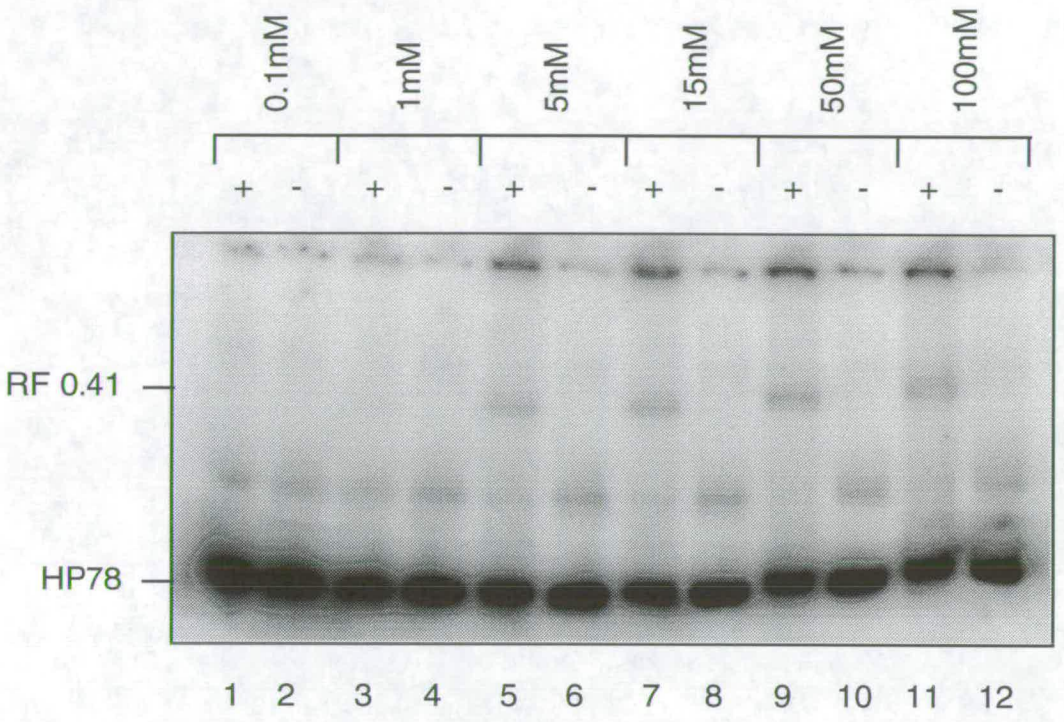
Binding reactions were set up in Mn^{2+} concentrations from 0.1mM (lanes 1 and 2) to 1mM (lanes 3 and 4), 5mM (lanes 5 and 6), 15mM (lanes 7 and 8), 50mM (lanes 9 and 10) and 100mM (lanes 11 and 12. The presence (+) and absence (-) of SbcCD is indicated. Reactions were treated with glutaraldehyde before separation on a 6.8% native polyacrylamide ATTO gel.

b

This graph shows the recovery of candidate complex bands visible in figure a. The results are preliminary. Values were obtained by phosphorimage analysis and have been corrected for background present in the absence of SbcCD.

Figure 5.1.5

a



b

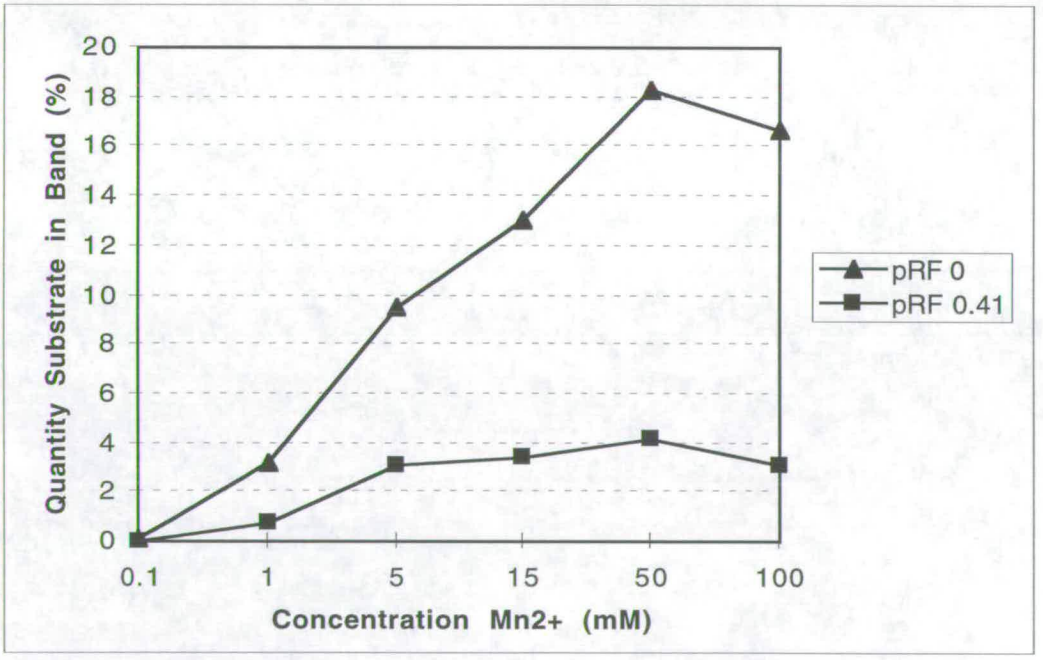
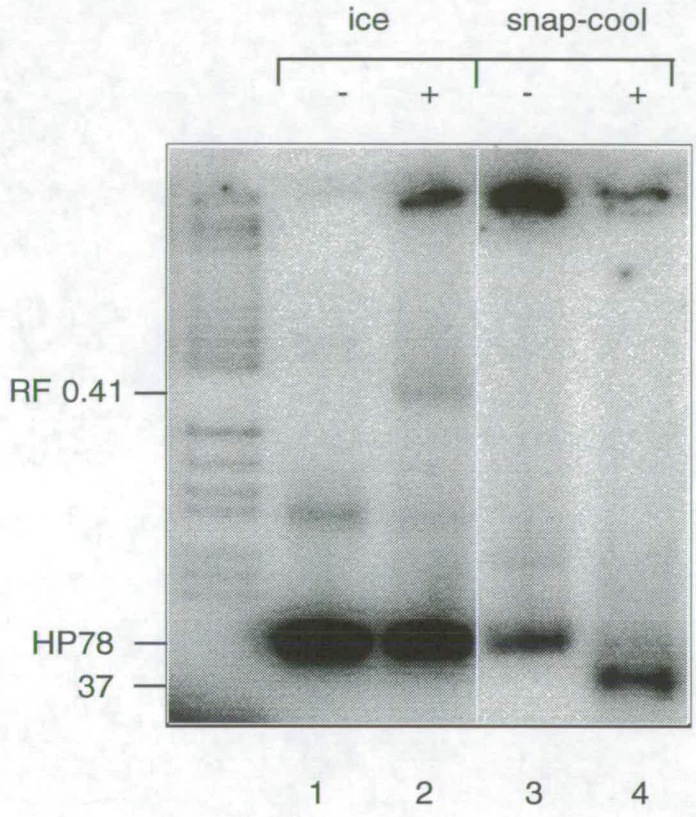


Figure 5.1.6



Identification of nuclease product. HP78 substrate was heated, snap-cooled and 5′ end-labelled. It was incubated at 16°C for 20 minutes under DNA binding conditions in the presence (+) or absence (-) of SbcCD and in the presence of Mn^{2+} and ATPγS. Reactions were then glutaraldehyde treated and either heated, snap-cooled (lanes 3 and 4) or not (lanes 1 and 2) and separated on a 6.8% native polyacrylamide ATTO gel.

Section 2

Testing for Weak DNA Binding

Attempting to Observe Gel Retardation of a Weak Complex

Substrates and gel conditions were both analysed together in this section.

Crosslinking

If DNA binding were weak, a low proportion of the total protein would be expected to participate in DNA binding at any one time. This low proportion might be visualised if trapped by covalent crosslinking. Three candidate complexes were observed on gel retardation of glutaraldehyde crosslinked binding reactions (Section 1). Glutaraldehyde crosslinking also generated artefacts. No retardation was observed by *p*-APB crosslinking of a thiophosphate substrate (Section 3).

Gentle Buffer Systems

Low Ionic Strength TAE

If the ionic strength of the gel buffer is too high, binding can be destabilized. It was claimed that DNA binding by SbcCD could be observed using a 4% low ionic strength TAE polyacrylamide gel (Matthew Whitby, University of Nottingham, personal communication). However the binding activity had not been confirmed by Western blotting. These candidate complexes had been obtained using a fraction of SbcCD different from that used in these studies but from the same purification (provided by John Connelly). Binding had been detected between SbcCD and a 50bp duplex substrate. Better binding had been obtained using a sequence-related synthetic Holliday Junction substrate, X-12. The two complexes had caused band retardations of equal size. These conditions were imitated to try to reproduce binding.

A 4% low ionic strength TAE polyacrylamide gel of 1.5mm thickness was poured and allowed to mature overnight at 4°C. Protein was preincubated for 10 minutes at 0°C in 1mM Mn^{2+}. Binding reactions were then carried out using 1.5nM substrate DNA, 6.9nM protein, 25mM Tris-HCl pH8.0, 1mM DTT, 6% glycerol, 100µg/ml BSA and 5mM Mn^{2+}, and incubating at 0°C for 10 minutes. The precooled gel was set up in precooled low ionic strength

TAE buffer at room temperature in a Hoefer SE 600 Vertical Slab Unit, with continual buffer recirculation by a peristaltic pump (Pharmacia) at a rate of 230ml per hour at room temperature. The binding reactions were immediately loaded onto the gel and electrophoresed at a constant voltage of 160V. The gel was dried and imaged by phosphorimage analysis. No candidate complexes were observed using snap-cooled HP78, a 26 nucleotide single-strand (26B), or a 26bp annealed double-strand (26T/26B) DNA as substrate, figure 5.2.1a.

TGE "Magic Buffer"

Dave Sherratt (University of Oxford) studied the DNA binding of site-specific recombinases XerC and XerD to their recombination site *dif*, and of resolvases Tn*3* and Tn*552* to their resolution sites using a Tris-glycine EDTA pH9.4 buffer (Rowland *et al* 1995, Bednarz *et al* 1990, Blake *et al* 1997, Blakely *et al* 1993). A method based on this was used to study SbcCD DNA binding.

1x TGE running buffer contains 50mM Tris, 10mM glycine and 1mM EDTA. The 10x stock is naturally pH9.8, changing to pH9.4 on dilution. 1x TGE binding buffer contains 10mM Tris and 2mM glycine. The 100x stock is naturally pH10, and changes to pH9.4 on dilution. A 7% ATTO polyacrylamide gel, made up in 1x TGE running buffer, was poured and allowed to mature overnight at room temperature. Protein was preincubated for 10 minutes at 0°C in 1mM Mn^{2+}. Binding reactions were then carried out in TGE binding buffer, using 1.5nM HP78 DNA, 69nM protein, 10mM Tris and 2mM glycine pH9.4, 50mM NaCl, 1.25mM DTT, 5% glycerol, 100μg/ml BSA, 5mM Mn^{2+} and 1mM ATPγS. These were incubated at 16°C for 20 minutes. The precooled gel was set up in precooled TGE running buffer and prerun for 10 minutes at 100V at 4°C. Binding reactions were immediately loaded and electrophoresed at 100V, figure 5.2.1b. No candidates for DNA bound protein complexes were observed.

pH

During protein purification, it was observed that SbcCD binds a hydroxyapatite column (which mimicks DNA) better at pH6 than other pHs (John Connelly personal communication). The nuclease is active between

pH7.0 and pH9.5 (Chapter 1 Section 3). At pH5, the protein precipitates, suggesting that it unfolds at this acidity. Acidic pH was extremely inhibitory for nuclease activity, whereas alkaline was less inhibitory (Connelly *et al* 1997). Perhaps SbcCD binding stability is increased at acid pH. This might result in inhibition of the nuclease activity. In addition pH affects gel retardation (Carey 1988).

The effect of pH on binding was initially investigated using glutaraldehyde crosslinking. A 0.5-step range of binding buffers was made from pH5.5 to pH10: NaAc (pH5.5), Tris-maleate (pH5.5, pH6.0, pH6.5, pH7.0, pH7.5, pH8.0), MOPS (pH7.0, pH7.5), HEPES (pH7.5, pH8.0), Tris-Cl (pH7.4, pH8.0, pH8.5), Glycine (pH8.5, pH9.0, pH9.5, pH10.0). A glutaraldehyde binding assay (Section 1) was performed using these buffers in place of TrisCl in the binding buffer, and Pal569 DNA. No candidates complexes were observed, figure 5.2.1d.

The effect of pH was then tested using pH-controlled gels in conjunction with pH-controlled binding reactions in the absence of crosslinking. pH buffers NaH_2PO_4/KOAc (pH6.2, pH6.75), Tris/EDTA (pH7.0, pH7.5), and TBE pH8.3 were used to pour 7% ATTO polyacrylamide gels. Protein was preincubated for 10 minutes at 0°C in 1mM Mn^{2+}. pH gels were prerun in appropriate pH buffers at 40mA for 20 minutres. Binding reactions were set up containing 1x pH buffer, 1.5nM HP78 DNA, 69nM preincubated protein, 1.25mM DTT, 5% glycerol, 100μg/ml BSA, 5mM Mn^{2+} and 1mM ATPγS. The reactions were incubated for 20 minutes at 16°C and immediately loaded onto the appropriate pH gel, run in at 100mA, and then run at 40mA. No candidate complexes were observed, figure 5.2.2. Bands of unknown origin were obtained in both the presence and absence of protein at pH6.75 and pH7.0. No candidate bands were obtained.

Additives

Certain additives have been employed to improve observations of DNA binding by proteins, such as dimethylsulfonate DMSO, which alleviates the cofactor requirements for the assembly of Mu transpososome (Baker and Mizuuchi 1992, Savilahti *et al* 1995), potassium or sodium glutamate (Menetski *et al* 1992, Gottlieb *et al* 1992), and glycerol (Sidorova and Rau

1996). The mechanisms by which these additives affect binding are not fully understood. It may be relevant that DMSO coordinates metals or that it alters solubility in water by affecting hydrogen bonding (Budavari *et al* 1996). Glycerol is known to increase the stability of complexes (Sambrook *et al* 1989).

Glutaraldehyde binding assays were carried out with the addition of 15% DMSO, 20% glycerol, or 30mM to 150mM potassium glutamate, figure 5.2.1c. Bands of unknown origin are visible in all tracks at this image intensity. A smear is additionally visible above the substrate band in the presence of DMSO. The experiment was repeated, this time adding DMSO to the reactions in both the presence and absence of protein, and the smear was observed in both tracks (data not shown). This indicates that DMSO may be interacting with DNA to partially retard it, but does not allow candidate complexes to be observed by this method.

Substrates

Although HP78 is cleaved by SbcCD hairpin nuclease, it might be that binding to HP78 is relatively unstable in comparison with that to other substrates. Dumbells may be preferred over short hairpins by the hairpin nuclease activity (Chapter 3). Also hairpins with stems up to 40bp are preserved in wildtype *E.coli* (Chapter 1 Section 1). This suggests that HP78, with a stem of 37bp, may not be an optimal substrate for SbcCD.

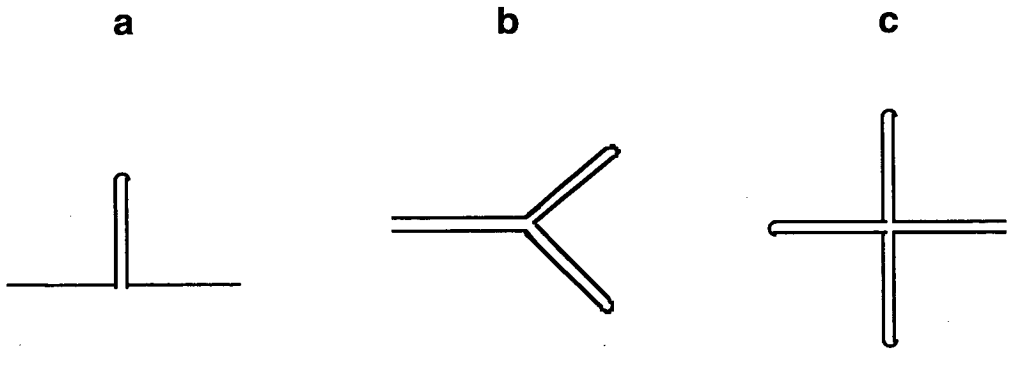
The DB98 substrate was preferred by SbcCD hairpin nuclease over HP78 (Chapter 3). Binding was compared using a glutaraldehyde binding assay, figure 5.2.3a. For HP78, the yield of pRF 0.41 was approximately 1% and of the well band pRF 0 was approximately 10%. These values are higher than obtained on any other occasion. For DB98, the yield of the well band was approximately 50% but no band RF 0.41 was observed. The reactions were also run on an agarose gel (data not shown). Much of the DB98 well band entered the gel. Bands aRF 0.71 and aRF 0.38 were observed at yields of approximately 1% and 10% respectively. This suggests that the DB98 bound complexes were of similar stoichiometry to the HP78 bound complexes previously observed (Section 1). The track was very smeared. This suggests that binding to DB98 was no more stable than to HP78.

The HP78 substrate might be rather smaller than optimal for DNA binding, especially considering the large estimated size for the active protein complex. To test this binding was compared using a 2.7kb duplex of linearised pUC19 as a substrate for the double-strand exonuclease. The samples were then visualised on a 0.8% agarose gel in TAE, figure 5.2.3b. No candidate complexes were visible. This suggests that the difficulty in observing binding to HP78 is not simply due to substrate size.

To determine the effect on gel retardation of using a longer hairpin substrate, a 569bp palindrome mixture was used, figure 2.1.1. This hairpin mixture, figure 5.2.3d, and HP56, figure 5.2.3c, were used in glutaraldehyde binding assays. No candidate complexes were observed for either substrate. However the "bubble" substrates (duplex bands) were observed to be efficiently digested in ATPγS and especially ATP. Perhaps this is not surprising, since these substrates include a loop-like structure.

Other substrates could also be tried. It may be that SbcCD recognises a duplex junction, such as that present in a cruciform or Holliday Junction, as implicated by Matthew Whitby's results (above), figure 5.2.4c. Perhaps the preferred substrate consists of a branched hairpin, figure 5.2.4a and 5.2.4b. John Connelly compared SbcCD nucleolytic cleavage of substrates a and b with that of HP56, and found cleavage was approximately equal (Connelly *et al* 1999). Cleavage was stimulated by DNA ends. However it is possible that DNA binding is stimulated by different substrates to the nuclease.

Figure 5.2.4



TAE Binding Assay

TAE is a gentler buffer than TBE (Chapter 4). A complex too weak to be observed in TBE might be visible in TAE. Binding reactions were set up using 5% glycerol, 1mM ATPγS, 5mM Mn^{2+}, 25mM TrisCl pH7.5, 1.25mM DTT, 100μg/ml BSA and 1.5nM end-labelled, freshly snap-cooled HP78 DNA. 69nM protein, preincubated in 1mM Mn^{2+} 0°C for 10 minutes, was added. Reactions were incubated at 16°C for 20 minutes. A 7% polyacrylamide TAE ATTO gel was matured for at least 2 hours. The gel and TAE buffer were precooled to 4°C. The gel was then prerun for 10 to 20 minutes at 4°C 40mA in TAE. Wells were rinsed and, after incubation, reactions immediately loaded onto the gel. The samples were run in 3 minutes 100mA and then electrophoresed 4°C 40mA.

No candidate complexes and no retarded bands were observed (Section 3). This result was similar to that using TBE. Nevertheless TAE was generally used in preference to TBE. Except where indicated, this TAE Binding Assay was used for further gel retardations. Exceptions include crosslinking investigations (for which the gel system was less important) and other buffer systems.

Salt Concentration

"Specific" binding generally occurs with very high affinity. "Non-specific" binding generally occurs with lower affinity (although theoretically could occur with high affinity). Weak, non-specific binding may be enhanced at certain high NaCl concentrations, whereas salt concentration has little effect on strong, specific binding (Engler *et al* 1997).

A TAE binding assay was carried out, including in the binding buffer a range of salt concentrations from 50mM to 200mM NaCl. No candidate complexes were visible at any of the tested salt concentrations, figure 5.2.5a. However nuclease activity was obtained at NaCl concentrations up to 100mM, was enhanced relative to 0mM NaCl at 50mM to 75mM, and was maximal at 50mM, figure 5.2.5b. John Connelly also previously found that nuclease activity was maximal at a NaCl concentration of 50mM, and further increases in NaCl were increasingly inhibitory relative to this activity

(Connelly *et al* 1997). The enhanced cleavage may result from an increase in nuclease activity without an increase in binding. Alternatively it may result from enhanced binding if this was not observable using the gel system.

Observing DNA Binding by Microspin Gel Filtration

Microspin Gel Filtration for DNA Binding Analysis

The Principal

Gel filtration is a gentle method of analysing a protein or a protein-DNA complex because the binding buffer is used in the column. The complexes are therefore maintained in the same buffer throughout binding and analysis. This minimises any changes in conditions between binding and analysis. It was anticipated that this method might stabilize a weak complex.

Small gel filtration columns packed with Sephacryl 400 beads (Pharmacia) were used for analysis of DNA binding. These beads were chosen because they are large enough to be penetrated by DNA up to 100bp in length. However the anticipated protein-DNA complex, or fragments of it, would not enter the beads. On centrifugation of the sample in the column, candidate complexes should pass the beads and enter the collected fraction. Some unbound DNA background is expected in the collected fraction since not all DNA capable of penetrating the beads will do so.

Obtaining and Optimising DNA Binding by Microspin Gel Filtration

Radioactive counts were obtained in the collected fraction under a range of conditions, figure 5.2.6 (and Section 3). This material constitutes candidate complexes. A low level of material was also obtained in the absence of protein, indicating that there is a background trickle of DNA which is not absorbed by the column.

Most efficient purification is obtained when the smallest purified product is twenty times larger than the largest impurity (Pharmacia). If the protein-DNA complex is negligably larger than the active protein complex, 1.2MDa,

the substrate DNA therefore should be less than 50kDa in size. HP78 (25kDa) was therefore chosen in preference to the Pal569 substrate (190kDa).

The manufacturers also recommend that the volume to be analysed is between 25µl and 100µl, and that the DNA concentration used be of the order of ng. Reaction volumes between 30µl and 90µl were tested and the optimum for recovery of collected counts relative to column-retained counts was found to be 60µl, figure 5.2.6a. Recovery was unaffected by DNA concentration, figure 5.2.6b, indicating that DNA concentration is proportional to recovery. This would result if DNA is limiting in the binding reaction, which would be expected since a large molar excess of protein over DNA is used. The optimum DNA concentration was around 2.5nM.

A protein titration was performed, figure 5.2.6c. There was little effect of protein concentration on recovery. Despite using a 100-fold range of molarities in vast excess of the DNA molarity, recovery remained only 10 to 20%. If detectable SbcCD-DNA binding were strong, these protein concentrations would be well above the K_d, and 100% bound DNA might be recovered. If binding were weak, recovered bound DNA might be expected to increase with protein concentration. Perhaps the detected DNA-binding represents that of a minor component of the protein preparation, which is saturated (see Discussion).

Similar to observations on gel retardation, binding for 20 minutes provided better yield than overnight incubation, figure 5.2.6d, and DMSO had no effect on yield, figure 5.2.6e. Metal and nucleotide requirements were also analysed (Section 3). These studies determined, amongst other things, that the bound fraction included product DNA as well as substrate DNA. Since background substrate DNA was collected in the absence of protein, it is possible that the recovered bound fraction is exclusively product DNA, figure 5.3.3.

Since this method of analysis maintains the complexes in their binding buffer, many of the conditions optimised for gel filtration would also apply to any other non-invasive method of analysis of this complex.

Limitations of the Gel Filtration Method

Once binding has been obtained by gel filtration, the DNA species bound can be identified using a denaturing polyacrylamide gel, figure 5.3.3. However further study is limited. Unlike the gel retardation method (or certain other methods; refer to the final discussion of the chapter), binding cannot be visualised, and the nature of the bound protein cannot be investigated.

Discussion

Summary

The gel retardation method was adapted to enhance weak DNA binding. Methods used included crosslinking, gentle buffers (including low-ionic strength TAE, TGE), pH-controlled buffers, additives (including potassium glutamate, DMSO and high glycerol), different substrates (including a 2.7kb duplex, DB98 and 56 and 569 nucleotide hairpin substrates) and high salt concentration. None of these methods generated a candidate complex.

Microspin gel filtration columns can be used to observe weak binding because the binding buffer is retained during analysis. Candidates for bound protein-DNA complexes were obtained. Their recovery was unaffected by DNA concentration. DMSO did not increase recovery. The candidate complex included protein bound product DNA. It may (or may not) have consisted entirely of protein bound product DNA.

Conclusion: DNA Binding by Nucleolytically Active SbcCD Is Not Weak

A low yield of candidate DNA-protein bound complexes was initially obtained (Section 1). One hypothesis proposed to explain this was that binding between SbcCD and its substrate was weak. In this section, the use of conditions designed to enhance weak binding did not increase the yield of candidate complexes. This suggests that the reason for the difficulty in obtaining binding is not low affinity of interaction. In support of this, the apparent K_m of SbcCD binding to a duplex substrate was 0.5nM to 3.5nM between protein concentrations of 0.25nM and 1.0nM (Connelly *et al* 1997). Although this figure increased with protein concentration and therefore is not definitive, it is not indicative of a weak interaction. The measurement was repeated in the presence of Mn^{2+}, predicted to be necessary for protein stability, and a consistent value of approximately 4nM was obtained (John Connelly and Erica de Leau, unpublished). This value is middle range in comparison to K_m for other protein-DNA interactions (Section 1).

Candidate complexes were obtained by gel filtration. Binding too weak to be observed by any type of gel retardation system might be observed by gel filtration because this system is innately gentler. However the bound complex obtained consisted largely, and perhaps entirely, of protein bound product DNA. The interaction between product and protein would be expected to be weak. If this were not so, the enzyme would be blocked from further reaction, inhibited by its own products.

The protein titration gel filtration study indicated that the protein-DNA binding detected may have corresponded to saturated binding of a minor component of the protein preparation. The nucleolytically active form of SbcCD (thought to be $SbcC_8SbcD_8$, Chapter 1 Section 3) is assumed not to be a minor component. A nucleolytically inactive complex of different stoichiometry, or SbcC or SbcD alone, might constitute a minor component of the preparation which might bind non-specifically. If any of these were responsible for the dectected binding, this would suggest that the (specific) binding of active protein is transient. It is possible that the minor binding protein component detected recognises both substrate and product DNA.

Figure 5.2.1

Buffers and additives for DNA binding.

a

Low-ionic strength buffer. HP78 was heated, snap-cooled and 5' end-labelled. A duplex was prepared by annealing 5' end-labelled 27B and unlabelled 27T single-strand oligonucleotides. 27B was used as a single-strand substrate. Binding reactions were set up using low ionic strength buffer and incubating HP78 (lanes 1 to 4), duplex (lanes 5 to 8) and single-strand DNA (lanes 9 to 12) with SbcCD (lanes 2, 4, 6, 8, 10 and 12) in Mn^{2+}. ATPγS (lanes 3, 4, 7, 8, 11 and 12) or no nucleotide (lanes 1, 2, 5, 6, 9, and 10) was included in the reactions. Reactions were separated on a low ionic strength 4% native polyacrylamide gel.

b

TGE buffer. HP78 was heated, snap-cooled and 5' end-labelled. Binding reactions were set up using TGE buffer in the presence (lanes 4 and 5) or absence (lanes 2 and 3) of ATPγS, and the presence (lanes 3 and 5) or absence (lanes 2 and 4) of SbcCD. They were separated on a TGE native polyacrylamide gel. Marker V is in lane 1.

c

Additives. Pal569 was heated and snap-cooled (to generate HP569 which has a duplex stem of 277bp) and 5' end-labelled. Binding reactions were set up including 150mM (lane 1), 90mM (lane 2), 30mM (lane 3) potassium glutamate, 15% DMSO (lane 4) or 20% glycerol (lane 5). The reactions were performed in ATPγS and Mn^{2+}. No additive was included in reaction 6, and SbcCD was not added to reaction 7. The reactions were separated on a 6.8% native polyacrylamide gel.

d

pH tests. Pal569 was heated and snap-cooled (to generate HP569 which has a duplex stem of 277bp) and 5' end-labelled. Binding reactions were set up in Tris buffer pH7.5 (lanes 1 and 2), NaAc buffer pH5.5 (lane3), Tris-maleate buffer pH5.5 (lane 4), pH6.0 (lane 5), pH6.5 (lane 6), pH7.0 (lane 7), pH7.5 (lane 8) and pH8.0 (lane 9), MOPS buffer pH7.0 (lane 10) and pH7.5 (lane 11), HEPES buffer pH7.5 (lane 12) and pH8.0 (lane 13), TrisCl buffer pH7.5 (lane 14), pH8.0 (lane 15) and pH8.5 (lane 16) and Glycine buffer pH8.5 (lane 17), pH9.0 (lane 18), pH9.5 (lane 19) and pH10.0 (lane 20). Reactions included ATPγS and Mn^{2+}. They were glutaraldehyde treated and separated on a 6.8% native polyacrylamide gel in 1x TBE.

Figure 5.2.1

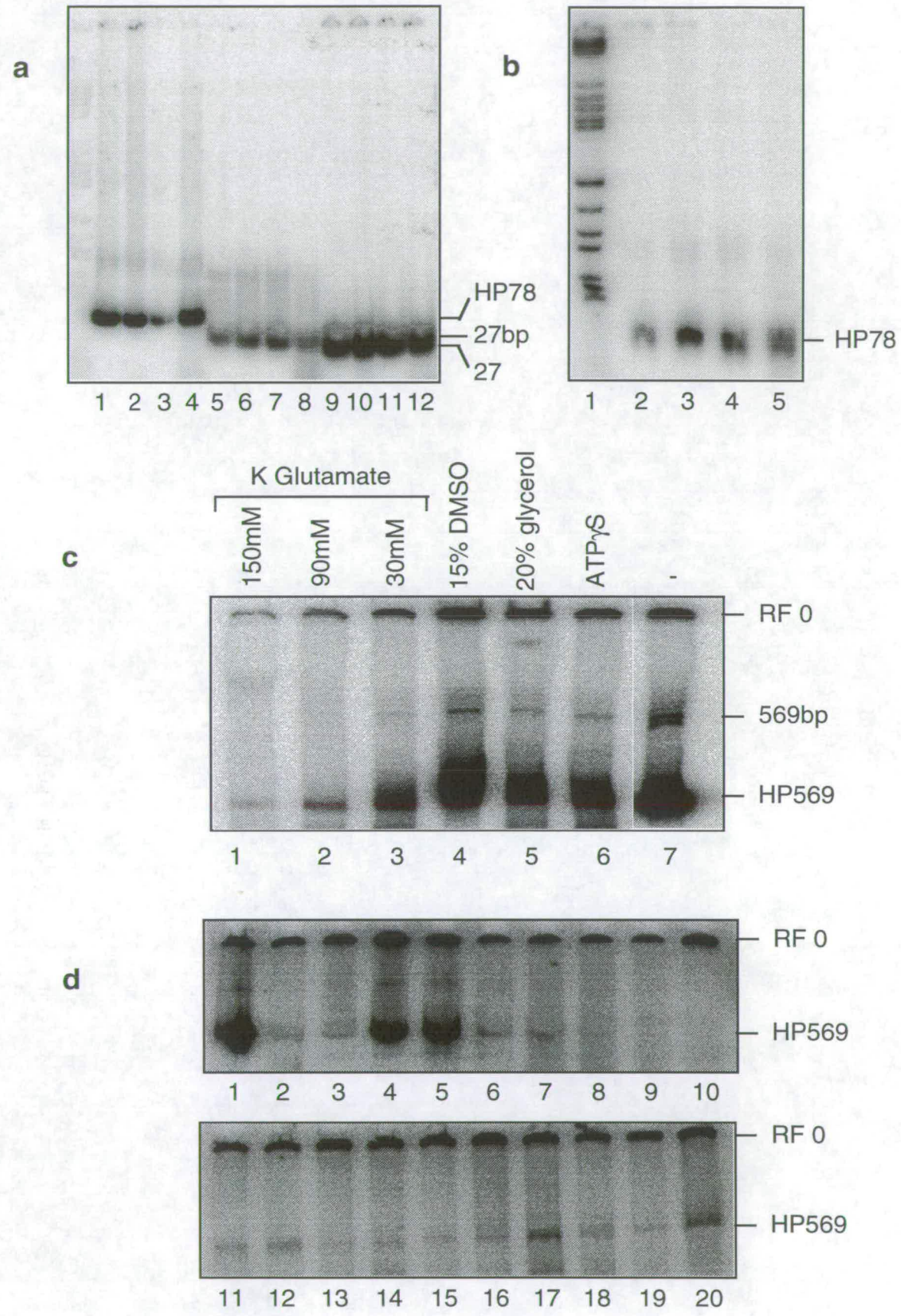
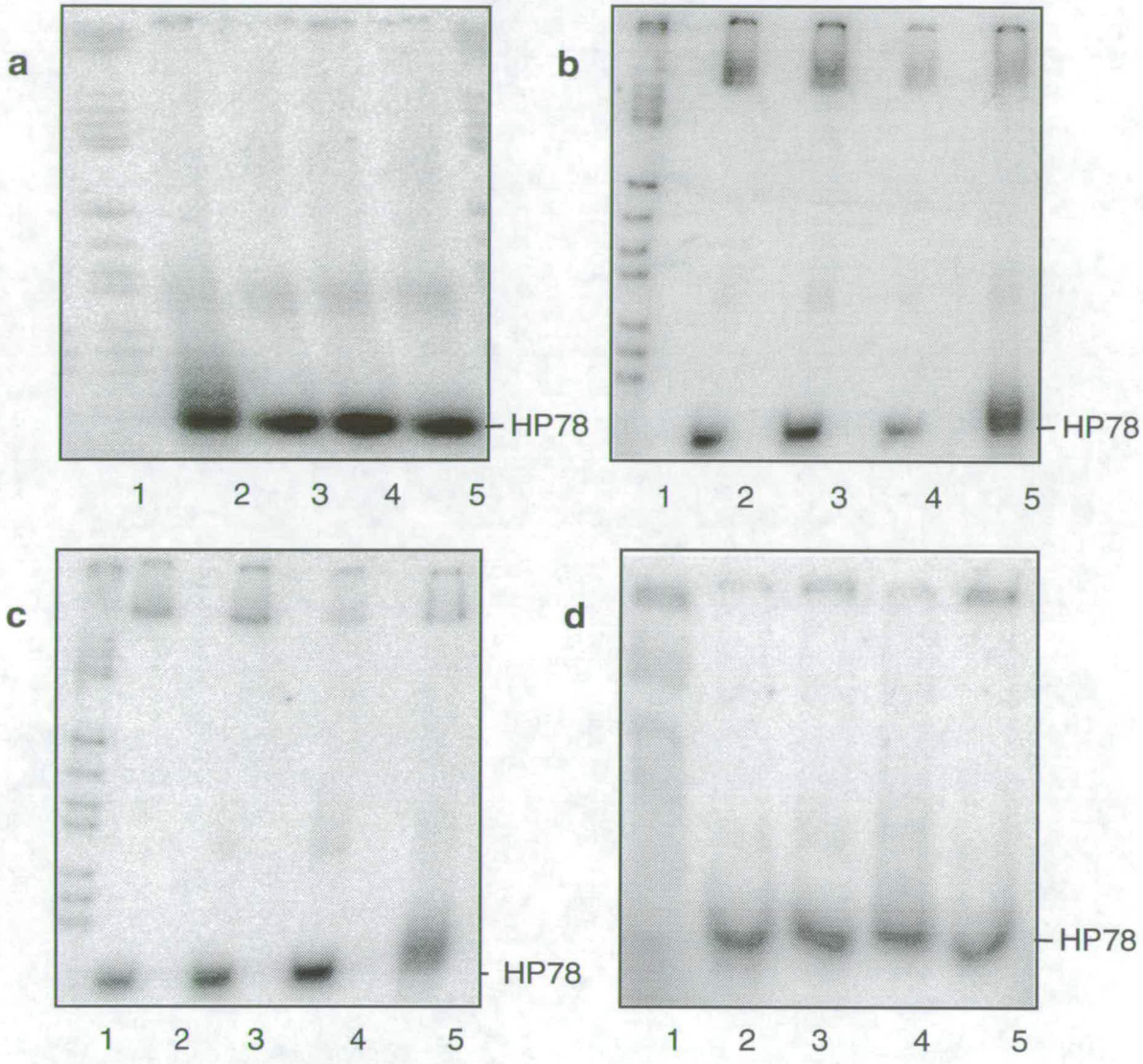

a
HP78
27bp
27
1 2 3 4 5 6 7 8 9 10 11 12
b
HP78
1 2 3 4 5
c
K Glutamate
150mM
90mM
30mM
15% DMSO
20% glycerol
ATPγS
-
RF 0
569bp
HP569
1 2 3 4 5 6 7
d
RF 0
HP569
1 2 3 4 5 6 7 8 9 10
RF 0
HP569
11 12 13 14 15 16 17 18 19 20

Figure 5.2.2



pH tests. HP78 was heated, snap-cooled and 5′ end-labelled. Binding reactions were set up in NaH_2PO_4 buffer pH6.25 (**a**) and pH6.75 (**b**), Tris EDTA buffer pH7.0 (**c**) and pH7.5 (**d**), and Tris Borate EDTA pH8.3 (**e**). Marker V (lane 1) and reactions minus SbcCD minus nucleotide (lane 2), plus SbcCD minus nucleotide (lane 3), minus SbcCD plus ATPγS (lane 4) and plus SbcCD plus ATPγS (lane 5) are shown. The reactions were separated on 6.8% native polyacrylamide gels using the same buffers as in the reactions.

Figure 5.2.3

Substrates for DNA binding.

a

HP78 (lanes 1 and 2) was heated, snap-cooled and 5' end-labelled. DB98 (lanes 3 and 4) was heated, snap-cooled, 5' end-labelled, ligated and purified (Chapter 3). These substrates were incubated with (+) or without (-) SbcCD in the presence of ATPγS and Mn^{2+} under binding conditions. The reactions were treated with glutaraldehyde and separated on a 6.8% native polyacrylamide gel.

b

pUC19 DNA was linearised, 5' end-labelled and used to set up binding reactions in the presence of ATPγS and Mn^{2+} and the presence (+) or absence (-) of SbcCD. Reaction products were treated with glutaraldehyde and separated by electrophoresis on a 0.8% native agarose gel.

c

HP56 was heated, snap-cooled and 5' end-labelled. Binding reactions were set up in the presence of ATP and SbcCD (lane 1), the presence of ATPγS and SbcCD (lane 2), the absence of nucleotide and the presence of SbcCD (lane 3) and the absence of nucleotide and SbcCD (lane 4). Marker V is shown in lane 5. Reactions were treated with glutaraldehyde and electrophoresed on a 6.8% native polyacrylamide gel.

d

Pal569 was heated and snap-cooled (to generate HP569 which has a duplex stem of 277bp) and 5' end-labelled. Reactions were set up using 1.5nM (lane 1) and 0.15nM (lanes 2 and 3) DNA in the presence (lanes 1 and 2) or absence (lane 3) of SbcCD. Marker V is included in lane 4. ATPγS and Mn^{2+} were present in each reaction. Reactions were treated with glutaraldehyde and separated by 6.8% native polyacrylamide gel electrophoresis.

Figure 5.2.3

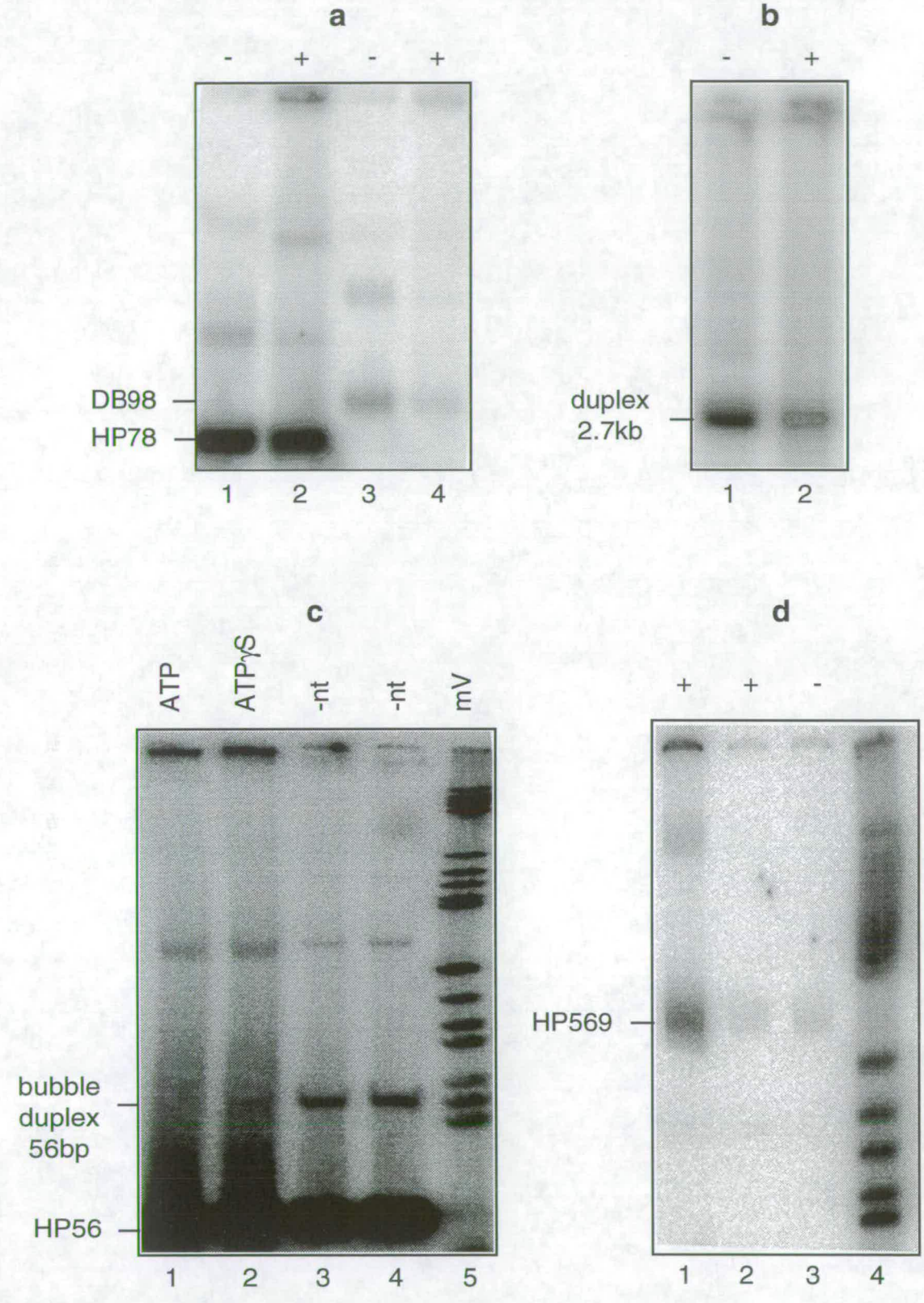
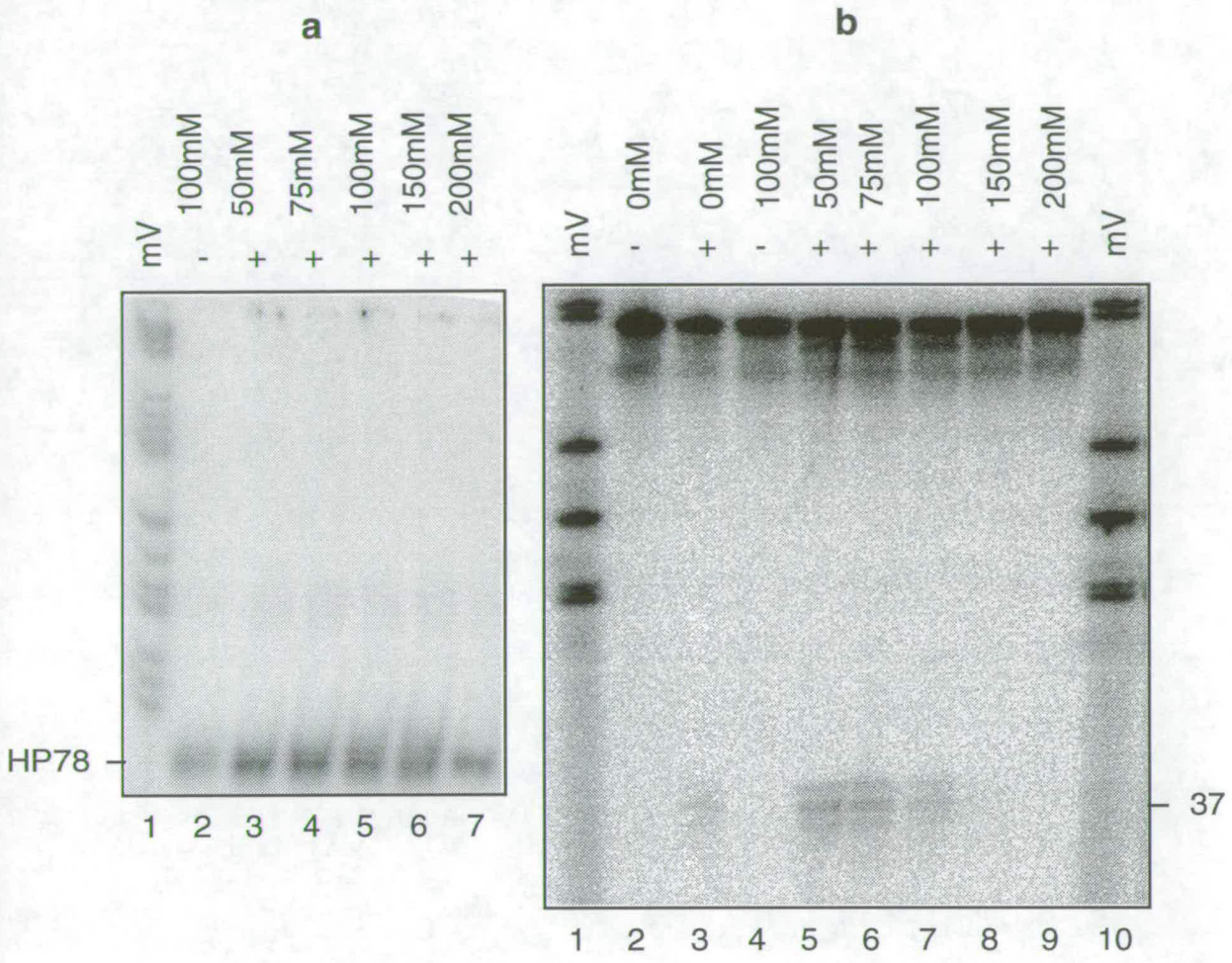
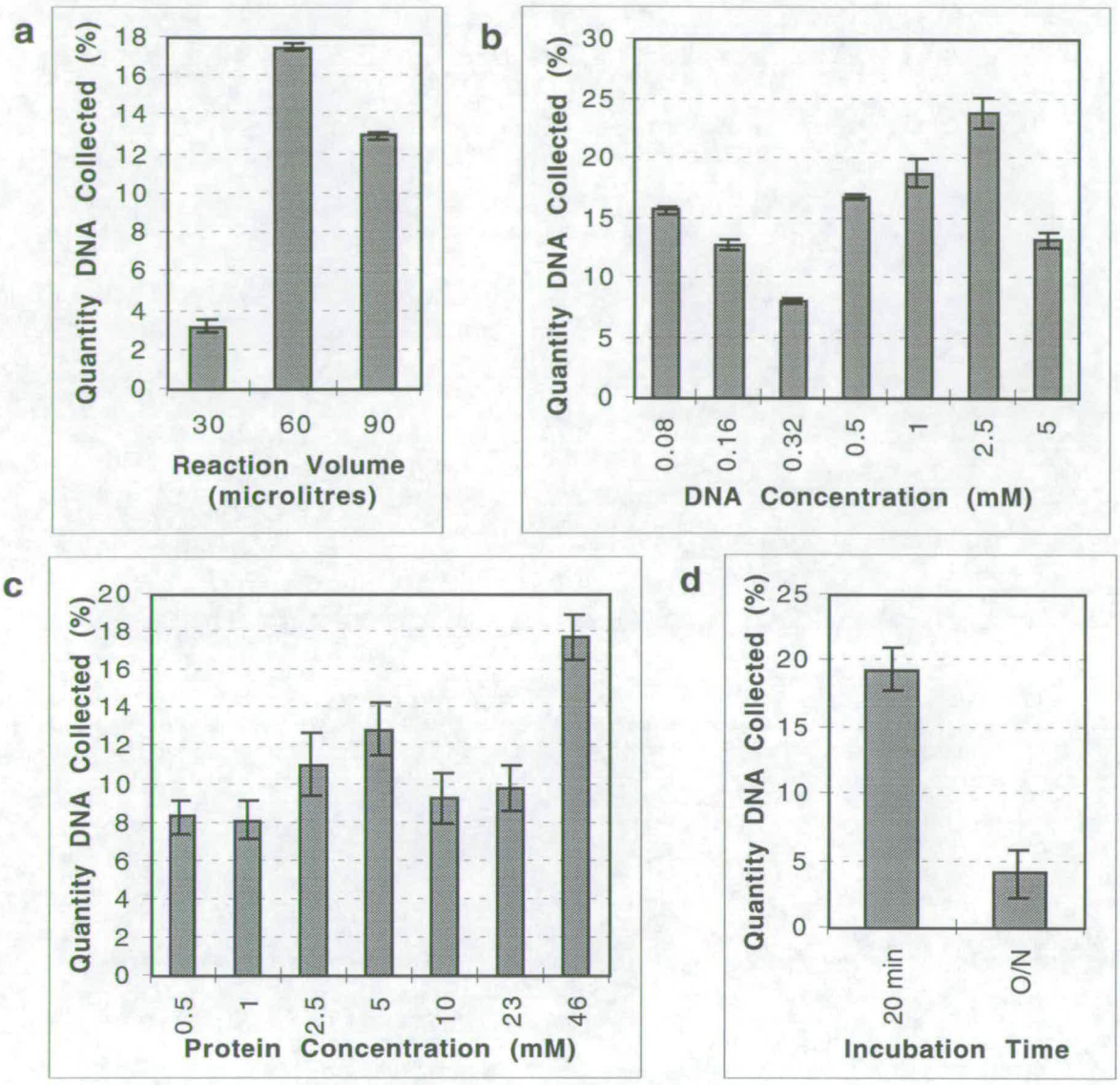

a
- + - +
DB98
HP78
1 2 3 4
b
- +
duplex
2.7kb
1 2
c
ATP
ATPγS
-nt
-nt
mV
bubble
duplex
56bp
HP56
1 2 3 4 5
d
+ + -
HP569
1 2 3 4

Figure 5.2.5



The effect of salt on binding and nuclease activity. Binding reactions (**a**) were set up using heated, snap-cooled 5′ end-labelled HP78 substrate in the presence of 50mM (lane 3), 75mM (lane 4), 100mM (lanes 2 and 5), 150mM (lane 6) and 200mM (lane 7) NaCl and the presence (+) or absence (-) SbcCD. Marker V is present in lane 1. Reactions were separated on a 6.8% native polyacrylamide gel in TAE. Nuclease reactions (**b**) were set up using heated, snap-cooled 5′ end-labelled HP78 substrate in the presence of 0mM (lanes 2 and 3), 50mM (lane 5), 75mM (lane 6), 100mM (lanes 4 and 7), 150mM (lane 8) and 200mM (lane 9) NaCl in the presence (+) or absence (-) SbcCD. Marker V is present in lanes 1 and 10. Reactions were separated on a 10% denaturing sequencing sized gel.

Figure 5.2.6



Optimising gel filtration. HP78 was heated, snap-cooled and 5′ end-labelled. It was used for microspin gel filtration assays of DNA binding, where DNA collected represents the bound fraction. Histograms show the quantity of the total DNA collected as a percentage of that used in the reaction. Reaction volume (**a**), DNA concentration in the reaction (**b**), SbcCD concentration in the reaction (**c**) and incubation time of the reaction (**d**) were varied, and DMSO was added to the reaction (**e**). The data shown represent the mean values obtained on performing each experiment in duplicate. Error bars represent twice the standard error, which is equal to the statistical range. Values shown are corrected for background collected in the absence of SbcCD.

Section 3

Trapping a Stable Cleavage Complex

Attempting to Trap Transient DNA Binding

DNA binding by SbcCD is unstable, and this is likely to be either because binding is weak, or because it is transient (Section 1). It was concluded that DNA binding is probably not weak (Section 2). There remains the possibility that binding is not observed because it is quickly followed by cleavage of the substrate and dissociation. In order to investigate this, conditions were sought to uncouple binding from nuclease activity and thus to chemically trap the uncleaved substrate bound protein complex, as a stable cleavage complex.

Altering Binding Reaction Conditions

Salt Concentration

A salt concentration was sought at which the nuclease is inactive but binding occurs. Nuclease activity was maximal in a salt concentration of 50mM NaCl (Section 2). In 75mM and 100mM NaCl it was above that in the absence of NaCl. It was abolished in salt concentrations of 150mM and above. However no binding was observed despite the published observation that high salt enhances binding of certain protein-DNA complexes.

Nucleotide Cofactors

SbcCD nuclease is active in ATP, ATPγS and GTP (Chapter 3). If binding could be obtained in the presence of a different nucleotide cofactor, the bound complex might be trapped for study.

Glutaraldehyde binding assays were set up using a range of nucleotide cofactors at 1mM, either ATPγS, ATP, GTP, ADP or no nucleotide. These were separated by 6.8% native gel electrophoresis, figure 5.3.1a or a 0.8% TAE agarose gel, figure 5.3.1b. The pRF 0.41/aRF 0.71 and the aRF 0.38 candidate complexes were obtained in ATPγS and ATP.

Similar binding reactions were set up with an additional dATP nucleotide cofactor test. These were not glutaraldehyde treated but subjected to analysis by S-400 microspin column filtration. The results are represented in

figure 5.3.1c. Binding is clearly apparent in ATP and ATPγS, and there is a much lower level of binding (hardly above background) in GTP.

Results obtained with both gel retardation and gel filtration methods showed approximately equal candidate binding in ATP and ATPγS.. However cleavage in ATPγS is approximately 30% that in ATP (Chapter 3). This supports the hypothesis that the binding observed includes that between protein and product. Its requirement for ATP or ATPγS suggests that the detected bound protein species includes SbcC.

Metal Ion Cofactors

Enzymes which require divalent cations for their activities have been modulated by substitution of the metal ion. For example, a stable cleavage complex between RAG1 and RAG2 proteins and their substrate was obtained by substituting Mn^{2+}, required for the nuclease activity, for Ca^{2+} (Hiom and Gellert 1997). Specific binding by *Eco*RV to its recognition site is stimulated approximately 700-fold in Ca^{2+}, relative to binding in the absence of metal ions, and cleavage does not occur in Ca^{2+} (Engler *et al* 1997).

A range of metal ions was tested in place of 5mM Mn^{2+} in the SbcCD-HP78 binding reaction, and the reactions were glutaraldehyde crosslinked and run on 6.8% TBE polyacrylamide gels, figures 5.3.2a and 5.3.2c, and a 0.8% agarose gel, figure 5.3.2b. Where Mn^{2+} is present in the binding reaction or the preincubation, the pRF 0.41/aRF 0.71 and aRF 0.38 binding candidates are observed. Where Mn^{2+} is only present in the preincubation (diluted) solution, the yield is correspondingly lower than where it is present in the binding reaction. This binding is inhibited by Zn^{2+} and the high 15mM concentration of Mg^{2+}, figure 5.3.2a. Analysis of metal requirements by gel filtration showed binding in 5mM Mn^{2+}, Ca^{2+}, and less strongly Mg^{2+}. No binding was obtained in 5mM Fe^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+} or Zn^{2+}, using either a HP78 or a HP56 substrate, figure 5.3.2d.

To confirm that bound products had been obtained on gel filtration, protein was preincubated in Mn^{2+}, binding reactions were carried out in 5mM Mn^{2+} and the collected material was run on a 12% denaturing sequencing polyacrylamide gel, figure 5.3.3. The bound fraction included product DNA

and substrate DNA, confirming that binding and nuclease activity had been obtained under these conditions. A low level of substrate DNA was obtained in the absence of protein, corresponding to a small proportion passing through the column as background. This was unaffected by the presence or absence of protein.

To define in greater detail the metal requirements for binding and nuclease activity, protein was preincubated in Mn^{2+} or Ca^{2+}, binding reactions were carried out in 5mM metal in the presence or absence of 0.05mM EDTA, and the collected material was run on a 12% denaturing sequencing polyacrylamide gel, figure 5.3.3. Preincubation Mn^{2+} was diluted to 0.03mM on addition of the protein to the reaction. 0.05mM EDTA was therefore sufficient to chelate any Mn^{2+} present in the binding reaction after equilibration of the protein, *ie* any Mn^{2+} not co-ordinated to the protein to stabilize it. Nuclease activity was obtained only where Mn^{2+} was present in the preincubation, and was obtained even when EDTA was then added to the reaction. Binding was obtained where Mn^{2+} or Ca^{2+} was present in the reaction. Together these results suggested that Mn^{2+} is required for nuclease activity, probably in stabilizing the protein structure for catalysis, and that binding can occur either in Mn^{2+} or Ca^{2+}. This in turn suggested that Ca^{2+} could substitute for Mn^{2+} in the protein structure but the resulting protein would be able only to bind, not to cleave, its substrate. If these results were to be believed, reaction in Ca^{2+} would uncouple binding and nuclease activity as desired. However these results contrast with those shown in figure 5.3.2c. Additionally, if the above logic is correct, it is surprising that protein preincubated in Mn^{2+} does not support cleavage and binding. This may indicate that these results are anomalous.

On the basis of these results, it was intended to test a range of Ca^{2+} concentrations for binding activity. However gel retardation or gel filtration binding in Ca^{2+} was never again obtained, despite changing stocks and repeating on multiple occasions. Therefore it must be concluded that Mn^{2+} is required both for the nuclease activity and for the binding activity. Mn^{2+} probably has a unique role in maintaining the structure of the SbcCD complex, which is essential for protein stability. No other tested metal can substitute for this function, not even for binding. The requirement of

Mn^{2+} for binding suggests that the detected bound protein species include SbcD.

Thiophosphate Substrate

Modified Substrates

The effect of a DNA substrate modification on protein activity depends on the individual chemical behaviour of a protein with a substrate. Two types of effect would be useful in trapping transient binding. A modification of the hairpin substrate might cause it to be bound but not cleaved by the protein. To be useful in these studies, the intermediate bound complex so produced would need to be stable. A modification might alternatively allow the substrate to be bound and cleaved but disallow dissociation after cleavage.

Binding affinity of the bacteriophage R17 coat protein to its RNA hairpin substrate can be increased or decreased up to 10-fold by substituting certain binding site nucleotides with 5'-O-(1-thiophosphate), which creates 5'-bridging phosphorothioate linkages in place of phosphodiester links (Milligan and Uhlenbeck 1989). If such a substrate were to similarly increase the affinity of SbcCD binding to the extent that the bound substrate or bound product was stable, it would be useful in studying binding. If binding is transient, a bound product would be the more likely of these two complexes.

*Eco*RV restriction endonuclease cleavage of its DNA recognition site occurs in the presence of a Mg^{2+} cofactor. A phosphorothioate substrate for *Eco*RV can be generated from 3'-phosphorothiolate or 4'-thiodeoxyribose units. This substrate does not affect protein binding, and is not cleaved in the presence of the Mg^{2+} cofactor (Engler *et al* 1997). If a phosphorothioate substituted substrate were to similarly affect the SbcCD reaction, and the trapped intermediate bound complex were stable, it would be useful in studying binding.

A similarly modified substrate affects DNA topoisomerases and site-specific recombinases differently. These families of enzymes both catalyse two transesterification reactions. The first breaks the substrate and generates a 3'-

DNA covalent intermediate and a 5' hydroxyl terminus; the second rejoins the ends by strand ligation. Since the second event occurs much faster than the first, the intermediate is transient and mechanistic studies of the reaction were difficult. Substitution of substrate phosphodiesters with 5'-bridging phosphorothioate linkages was found to generate a suicide substrate in which the breakage step occurred but rejoining was prevented. The 5' sulfhydryl terminus, formed from breakage, was not competent for ligation (Burgin *et al* 1995).

SbcCD Activity on a Thiophosphate Hairpin Substrate

Design of a Thiophosphate Substrate

A thiophosphate hairpin substrate based on HP78 was designed. The phosphodiester links 5' of the nucleotides 37, 38 and 39 proximal to the 5' end of HP78 were substituted with 5' phosphorothioate links. This generated the substrate thioHP78. These positions were chosen for substitution since they correspond to the phosphodiester links which are initially cleaved by the hairpin nuclease. Substitution was carried out during synthesis (by Oswel DNA Service).

In the case of thiophosphate modifications, the effects of the DNA modification on protein activities arise because the phosphate-sulphur bond is much more polar than the corresponding phosphate-oxygen bond. Although a sulphur atom is larger than an oxygen atom, the phosphate-sulphur bond length is comparable with the phosphate-oxygen bond length (Milligan and Uhlenbeck 1989).

Synthesis of phosphorothioate DNA is not completely efficient and the resulting DNA contains a low level of phosphodiester links at the positions studied. There is also a slow conversion of phosphorothioates to phosphodiesters resulting from the exchange of sulphur with oxygen. It has been estimated that, for each intended phosphorothioate link, 1% or 2% of DNAs in the preparation will instead have a phosphodiester link (Tom Brown, University of Southhampton, personal communication). Therefore for thioHP78, approximately (1.5% x 3 positions =) 5% of intended phosphorothioates will be phosphodiesters. An additional source of

variation within a phosphorothioate DNA preparation results from the fact that thiophosphates, unlike phosphates, are chiral centres. The R*d* and R*p* isomers, which are randomly incorporated into DNA during synthesis, may have differing activities.

DNA Binding

A TAE binding assay was performed using 1.5nM either HP78 or thioHP78 5' end-labelled substrate in the presence or absence of 1mM ATPγS. No bound candidate complexes were observed for either substrate by gel retardation, figure 5.3.4a. Analysis of binding in 1mM ATPγS by S-400 microspin gel filtration showed an equal level of binding to both substrates, figure 5.3.4b. Therefore the protein bound thiophosphate substrate was not trapped.

Hairpin Nuclease

The protein bound thiophosphate substrate may not have been trapped if the bound intermediate was unstable or if it was cleaved. A hairpin nuclease assay was performed to test this at 37°C, figure 5.3.5a, and 16°C, figure 5.3.5b. Products were visualised on denaturing polyacrylamide sequencing gels. The experiment was repeated at 16°C because smaller products were initially observed on thioHP78 cleavage at 37°C, which suggested that the reaction on thioHP78 was different to that on HP78, figure 5.3.5a.

Taking into consideration the estimated content of phophodiesters at intended phosphorothioate positions (5%), and the background obtained in the absence of protein, the relative quantities of HP78 and thioHP78 cleavage products obtained at 16°C was determined by phosphorimage analysis, Table 5.3.1. There was considerably more 3' terminal cleavage of thioHP78 than HP78 in ATPγS, but these products were presumably degraded in ATP. Cleavage at the preferred positions 5' of the loop was reduced in thioHP78 below HP78 in both ATPγS and ATP, and the effect was more pronounced in ATPγS. Cleavage further towards the 5' terminus was increased in thioHP78 above HP78 in both ATPγS and ATP, and the effect was more pronounced in ATPγS. The effects are probably enhanced in ATPγS as a result of degradation in ATP.

The cleavage of the phosphorothioate substrate indicated that it could not be used to trap transiently bound complexes. It appears that the phosphorothioate substrate is cleaved at least as efficiently as HP78, however the pattern of cleavage differs. The slight increase in cleavage of the 5' proximal arm apparently occurs at the expense of cleavage immediately 5' of the loop, which is slightly reduced. Whereas the preferred sites of cleavage of the unmodified hairpin are 36, 37 and 38 nucleotides proximal to the 5' end, the preferred sites of cleavage of the modified hairpin are at positions 33, 34 and 35. However cleavages at 36, 37 and 38 are still pronounced.

Secondary Structure of a Thiophosphate Substrate

Certain phosphorothioate substitutions may affect the structure of DNA. Where thiophosphates are present on both strands of a duplex, the T_m is drastically reduced because the sulphur atom is larger than the oxygen and does not hydrogen bond so tightly. However duplex thiophosphate DNA is of B form (Cruse *et al* 1986). Where phosphorothioates are present on one strand only, as in thioHP78, T_m is only slightly reduced. T_m is also affected by the sequence and length of the oligonucleotide. It has been estimated that the reduction in T_m for thioHP78, as compared to HP78, would be approximately 3°C or 4°C, which is small (Tom Brown personal communication).

Table 5.3.1

Percentage recovery of products obtained on degradation of HP78 and thioHP78 by SbcCD. These values were derived by phosphorimage analysis of figure 5.3.5b. They have been corrected for background obtained in the absence of SbcCD.

	ATP (%)		ATPγS (%)	
	HP78	thioHP78	HP78	thioHP78
uncleaved 78	55	46	65	57
77, 76, 75	5	5	2	6
38, 37, 36	6	4	7	4
35 and less	17	27	8	16

It is possible that the differing cleavage pattern of thioHP78, relative to HP78, results from a different secondary structure, perhaps one with a larger unpaired region. In order to test this, a nuclease assay was performed in the presence of a range of NaCl concentrations. Apart from its effects on binding affinity (Engler *et al* 1997) and nuclease activity (Connelly *et al* 1997), high salt concentration may alter the conformation of DNA. High salt did not affect the cleavage pattern obtained for either HP78, figure 5.3.5c, or thioHP78, figure 5.3.5d, suggesting that thioHP78 adopts the predicted 4T loop structure under all the conditions investigated. This is consistent with the advice of Tom Brown (above). It is therefore suggested that the differing cleavage pattern results from a direct stimulation of nuclease activity by the phosphorothioate linkages.

p-APB Crosslinking of a Thiophosphate Hairpin Substrate

A photoreactive *p*-APB (*p*-azidophenacyl bromide) moiety has been used to crosslink a specifically positioned phosphorothioate linkage in a substrate DNA to λ integration host factor (IHF) to determine which parts of the protein contact the DNA (Yang and Nash 1994). This method can also be used to stabilize and detect a protein-DNA complex. The reaction occurs in two steps: Nucleophilic sulphur in the phosphorothioate attacks bromide in *p*-APB and displaces it, with the result that *p*-APB is chemically coupled to the phosphorothioate DNA through the sulphur. *p*-APB may then act as a crosslinkable arm, and if strategically positioned, may contact the binding site of the protein. On exposure to long-wavelength light (UV 302nm), the *p*-APB moiety becomes covalently bound to the protein contact site. *p*-APB is effective as a crosslinker because of its reactivity with sulphur and its short length (which increases accuracy). *p*-ABA (*p*-azidobromoacetanilide) and *p*-AIA (*p*-azidoiodoacetanilide) moieties can be similarly used, and *p*-AIA has been reported to have greater crosslinking efficiency than *p*-APB or *p*-ABA, crosslinking at approximately twice the level of *p*-APB (Zhang *et al* 1995). *p*-APB, *p*-ABA and *p*-AIA can also form protein-protein and protein-RNA crosslinks if the reactive sulphur is that of a cysteine residue in the protein (and *p*-AIA also incorporates with greater specificity into cysteine residues) (Zhang *et al* 1995). Despite the greater efficiency of crosslinking by *p*-AIA, *p*-APB was chosen for this work due to its availability.

A method based on that described by Yang and Nash was used to try to trap a transient complex by crosslinking *p*-APB coupled phosphorothioate hairpin DNA to SbcCD. Since the phosphorothioate links of thioHP78 are positioned at the preferred site of HP78 cleavage, they are suitably placed for specific interaction with the binding site.

During crosslinking, a 7% TAE ATTO polyacrylamide gel was prerun at 180V for 10 minutes. The crosslinked samples were subjected to native PAGE, figure 5.3.4c. As expected, no retarded bands were observed in tracks corresponding to untreated DNA. A heavy background of DNA is visible in the *p*-APB coupled DNA tracks both in the presence and absence of protein. Crosslinking of thioHP78 generated gel retarded bands of RF 0.30 and RF 0.41. However these were obtained both in the presence and absence of SbcCD.

Mutant SbcCD

Design of SbcCD Mutants

Three mutants in SbcD were designed (David Pinder, John Connelly and David Leach, University of Edinburgh). Mutant SbcD(H184Y) (DL1127) is altered in phosphodiesterase motif IV and might have a similar phenotype as Mre11-58S (Chapter 1 Section 4). Mutant SbcD(D8A) (DL1128) is altered in motif I. Since aspartic acid has a carboxyl side chain, whereas alanine is a short aliphatic amino acid, this change might severely affect the phosphodiesterase activity or abolish it as a null. Mutant SbcD(D8E) (DL1129) is a conservative change which retains the carboxyl side chain and therefore would be expected to have a less severe effect than SbcD(D8A). Mutant SbcCD was prepared by QuikChange™ site-directed mutagenesis using the primers shown in figure 5.3.6.

DNA binding and hairpin nuclease assays were performed using these proteins. It was expected that Mre11-58S-like SbcD(H184Y) might have partial activity. It might bind but lack nuclease activity, or display alternative partial activity. The non conservative nature of SbcD(D8A) might affect, or prevent, binding or nuclease activity. The conservative mutant SbcD(D8E) was unlikely to display a mutant binding or nuclease phenotype.

Gel Retardation

A TAE binding assay was performed using 35nM protein (preincubated at 0°C for 10 minutes in 1mM Mn^{2+}). The three mutant proteins and wildtype protein were tested for binding activity to HP78 and thioHP78 substrates. No candidate complexes were observed in any track, figure 5.3.7a.

The binding assay was repeated using TGE buffer for the gel and the running buffer and in the reaction buffer. No candidate complexes were observed, figure 5.3.7b. The binding assay was repeated using the original TAE buffers but including NaCl at concentrations of 75mM and 150mM. Optimal nuclease activity had been obtained at 75mM, whereas nuclease activity had not been obtained at 150mM NaCl (Section 2). The HP78 substrate only was analysed. No candidate complexes were obtained, figure 5.3.7e. The binding assay was repeated using pH controlled buffers for the gel and running buffer and in the reactions. Binding was tested at pH6.75, figure 5.3.7c, and pH7.0, figure 5.3.7d. No candidate complexes were obtained.

Figure 5.3.6

DL1127 primers (31 nucleotides)

5' CCCATCATCGCCACGGGA**TAT**TTAACGACCG 3'
3' GGGTAGTAGCGGTGCCCT**ATA**AATTGCTGGC 5'

DL1128 primers (29 nucleotides)

5' CCTTCACACCTCA**GCC**TGGCATCTCGGCC 3'
3' GGAAGTGTGGAGT**CGG**ACCGTAGAGCCGG 5'

DL1129 primers (29 nucleotides)

5' CCTTCACACCTCA**GAA**TGGCATCTCGGCC 3'
3' GGAAGTGTGGAGT**CTT**ACCGTAGAGCCGG 5'

Microspin Gel Filtration

A microspin gel filtration assay was performed using each of the mutants and wildtype, and including a control lacking protein. A low level of binding was detected in all reactions except in the presence of mutant SbcCD(D8A), figure 5.3.8a. This is consistent with the expected mutant phenotypes. It is surprising that wildtype protein displayed only a low level of binding. This may indicate that protein activity is low following purification.

Nuclease Activity

A nuclease assay was conducted using each of the mutants and wildtype protein preparations supplied by David Pinder (Fraction 71) and John Connelly (Fraction V) with HP78 substrate in 1mM ATPγS or 1mM ATP, figure 5.3.8b.

Products corresponding to cleavage 5' of the loop (in ATPγS), and to degradation (in ATP), were obtained in the presence of both wildtype protein fractions.

Neither cleavage at the loop nor degradation was obtained in the presence of any of the mutants. This is unexpected for SbcD(D8E). Dissociation of SbcC from the mutant SbcD polypeptide in the recovered fractions could result in poor binding and/or poor nuclease activity. It is therefore possible that nuclease activity was not observed for the mutants because insufficient intact mutant protein was present. Alternatively, of course, nuclease activity might be absent because the mutant proteins bind the HP78 substrate but cannot cleave, or cannot bind (less likely in the light of the microspin gel filtration result).

Products of less than 12 nucleotides in length were obtained with all the QuikChange™ purified proteins. The exact size of the small products cannot be deduced since their migration is affected by the dye front, but they may correspond to single radiophosphates generated from a contaminant phosphatase activity in these preparations. This is likely since preliminary purifications by John Connelly also included a phosphatase activity, and the

QuikChange™ purified proteins were considerably less pure than John Connelly's later preparations (John Connelly and David Pinder personal communication).

Discussion

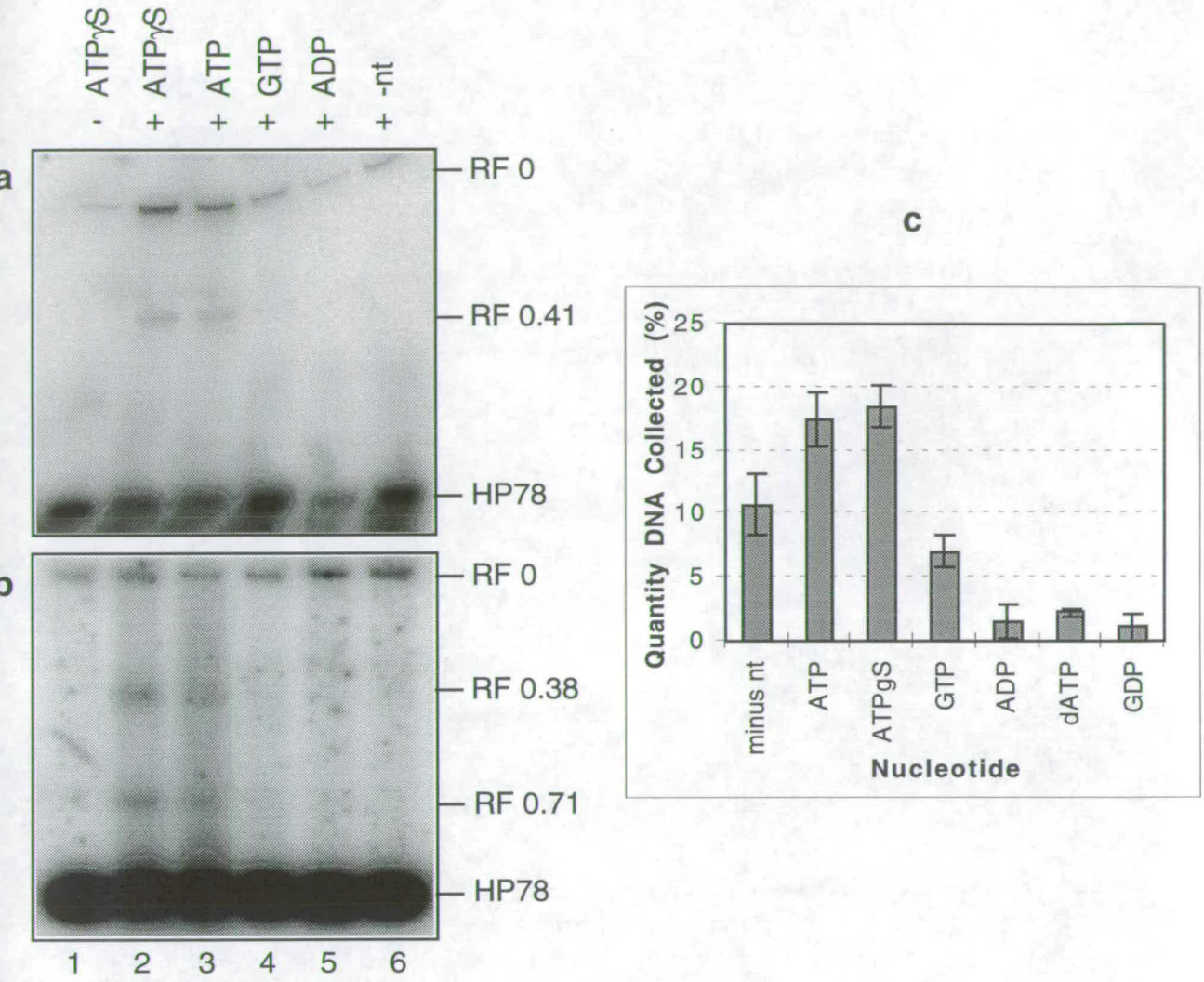
Summary

Attempts have been made to uncouple DNA binding from the nuclease activity of SbcCD on its HP78 substrate, in order to trap transient bound protein. There was no evidence of binding in the high salt concentrations at which the nuclease is inactive. Binding was not obtained in the presence of nucleotide cofactors unable to support nuclease activity. Initial experiments suggested that binding, but not cleavage, was obtained in the presence of Ca^{2+} ions, however this result was not later reproducible. A phosphorothioate substrate, thioHP78, which has thiophosphate linkages at the preferred cleavage sites 5' of the hairpin loop, was designed and tested for binding and nuclease activity. Binding at a comparable level to HP78 was observed by microspin gel filtration. However bound complex was not observed on gel retardation and the substrate was cleaved. The pattern of thioHP78 cleavage differed from that of HP78, corresponding to an overall enhancement of nuclease activity. Crosslinking of thioHP78 to protein was attempted using the sulphur reactive, photoreactive moiety, *p*-APB, but no significant bound complexes were obtained on gel retardation. Three mutant SbcD polypeptides were tested. These had single amino acid changes; SbcD(H184Y) is mutant in phosphoesterase motif IV and was expected to be homologous to Mre11-58S, SbcD(D8A) is a non conservative mutant in motif I and is expected to have a null phenotype, and SbcD(D8E) is a conservative mutant in motif I, and is expected to have a less severe phenotype. A low level of binding was detected for mutants SbcD(H184Y) and SbcD(D8E) by microspin gel filtration, which could be consistent with the expected phenotypes. However binding by gel retardation was not obtained using the TAE assay, high salt concentrations, TGE buffer or buffers controlled at pH6.75 or pH7.0. Hairpin cleavage was also not obtained. The mutant SbcD polypeptides may interacted poorly. Microspin gel filtration results alone provide rather weak evidence for DNA binding. It therefore cannot be concluded with certainty whether these mutant proteins have DNA binding activity.

Conclusion: DNA Binding by SbcCD May Be Transient

None of the attempts to trap transient binding were successful. It is therefore still possible that SbcCD binds its substrate only transiently before cleaving and dissociating. Transience is offered tentatively as a possible explanation of the unstable binding.

Figure 5.3.1



Nucleotide tests. HP78 was heated, snap-cooled and 5′ end-labelled and then incubated under DNA binding conditions with (+) or without (-) SbcCD in the presence of 5mM Mn^{2+} and 1mM ATPγS (lanes 1 and 2), ATP (lane 3), GTP (lane 4), ADP (lane 5) and in the absence of nucleotide (lane 6). The reactions were glutaraldehyde treated and run on a 6.8% native polyacrylamide ATTO gel (**a**) and a 0.8% agarose gel (**b**). The histogram (**c**) shows the mean percentage of the DNA which was collected as the candidate bound fraction on microspin gel filtration. The reactions were performed in duplicate; the error bar represents the range which equals twice the standard error. Values are corrected for backgound.

Figure 5.3.2

Metal cofactor tests. HP78 was heated, snap-cooled and 5' end-labelled and then incubated under DNA binding conditions with (+) or without (-) SbcCD in the presence of 1mM ATPγS.

a and b

Included in the reaction was 5mM Mn^{2+} (lanes 1 and 2), 5mM Mg^{2+} (lanes 3 and 4), 15mM Mg^{2+} (lane 5), 5mM Zn^{2+} (lanes 6 and 7) and 15mM Zn^{2+} (lane 8), and in the absence of nucleotide (lanes 9 and 10). Reactions were glutaraldehyde treated and separated on a 6.8% native polyacrylamide gel (**a**) and a 0.8% agarose gel (**b**).

c

SbcCD was preincubated in 5mM Ca^{2+} (lanes 1, 2 and 3), 1mM Ca^{2+} (lanes 4, 5 and 6) or 1mM Mn^{2+} (lanes 7, 8 and 9), or was not preincubated (lanes 10, 11 and 12). Reactions were set up in 5mM Mn^{2+} and 5mM Mg^{2+} (lanes 1, 4, 7 and 10), 5mM Ca^{2+} (lanes 2, 5, 8 and 11) or 5mM Mn^{2+} (lanes 3, 6, 9 and 12). The control lacking SbcCD was incubated in 5mM Mn^{2+} (lane 13). Reactions were glutaraldehyde treated and separated on a 6.8% native polyacrylamide gel.

d

The histogram shows the mean percentage DNA collected as candidate complex from gel filtration DNA binding analysis performed in duplicate using 1mM Mn^{2+} for preincubation and 5mM Mn^{2+}, Mg^{2+}, Ca^{2+}, Fe^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+} or Zn^{2+} for the reaction. The error bar represents the range which equals twice the standard error. Values are corrected for background.

Figure 5.3.2

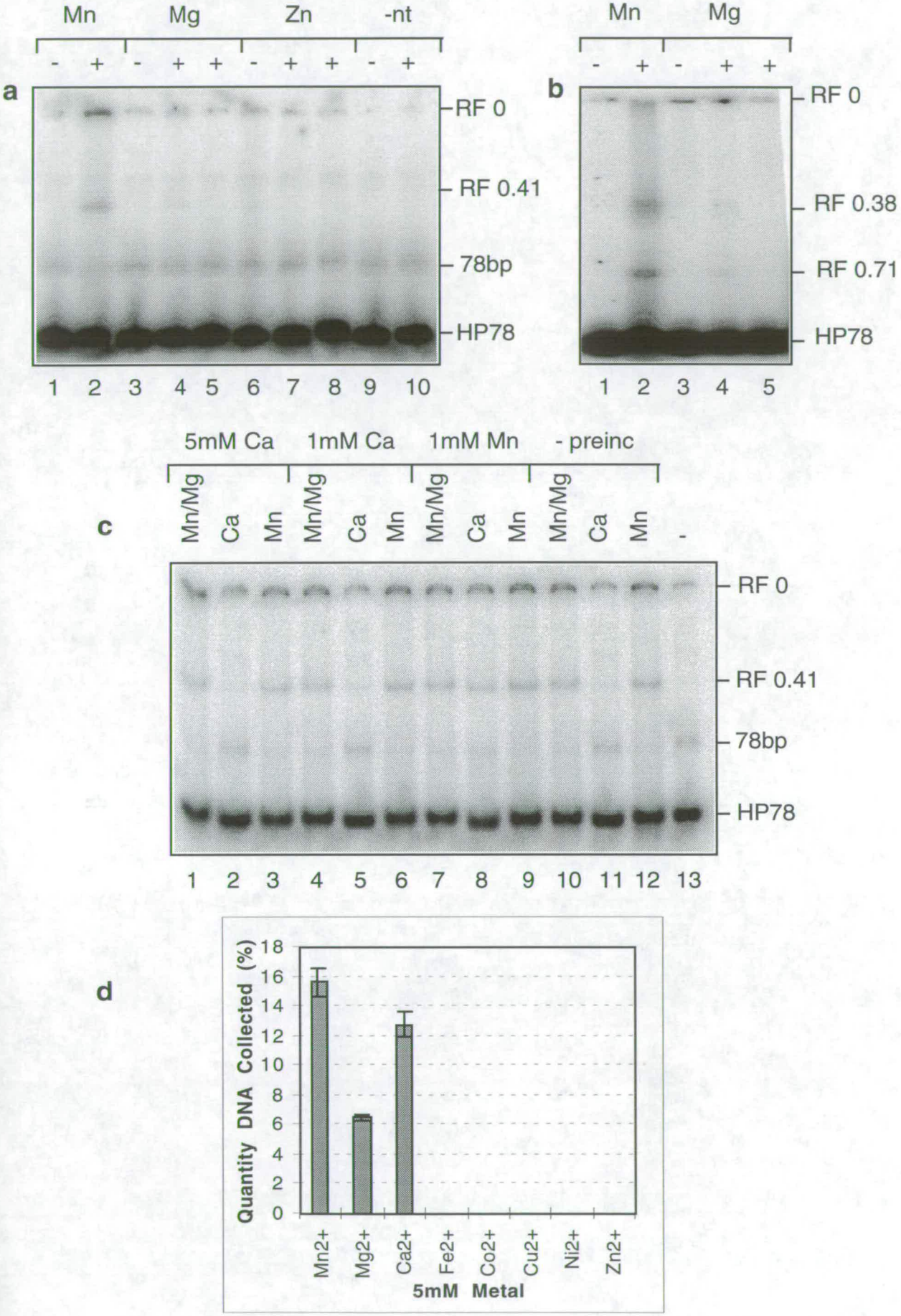
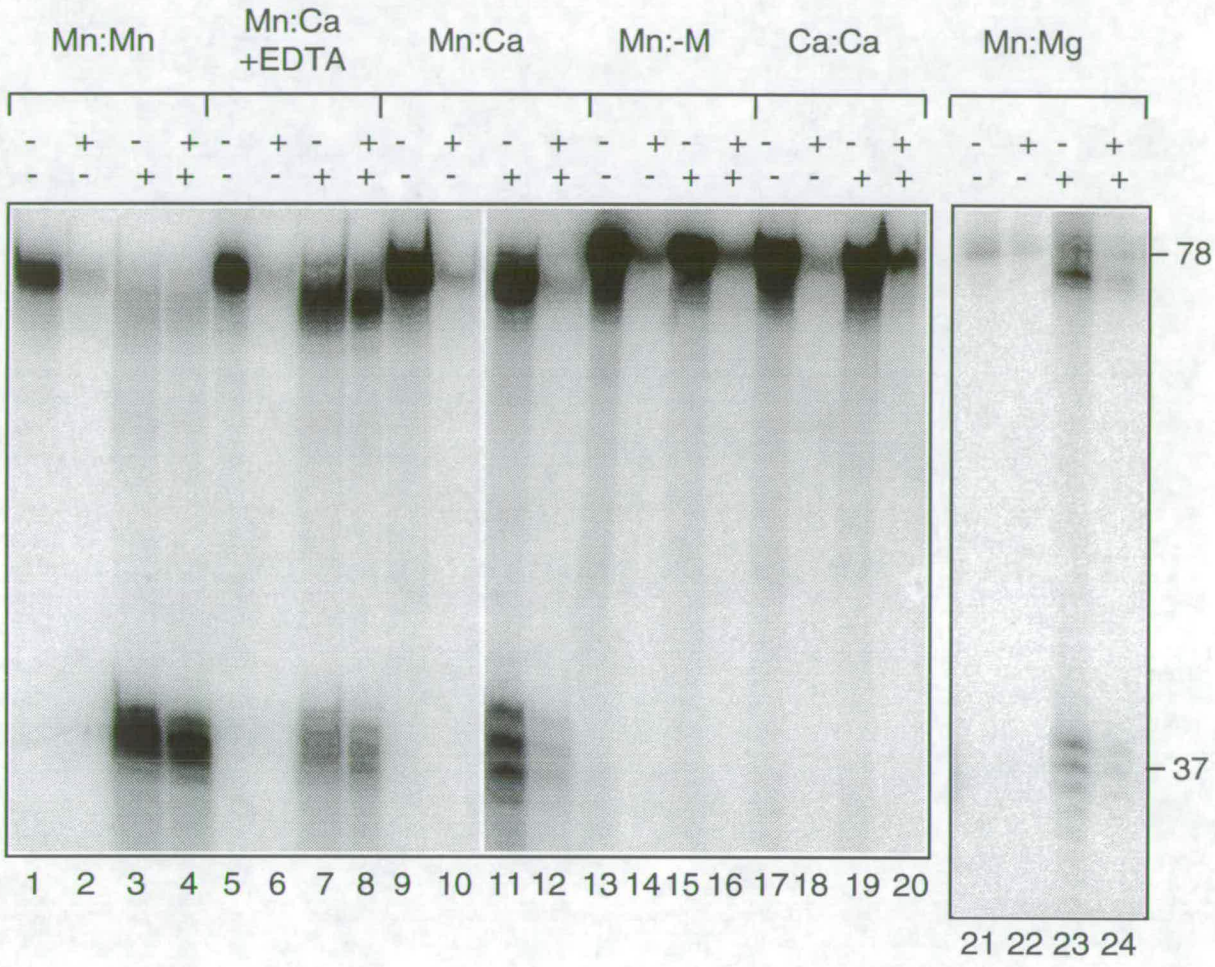

a
Mn
Mg
Zn
-nt
- + - + + - + + - +
RF 0
RF 0.41
78bp
HP78
1 2 3 4 5 6 7 8 9 10
b
Mn
Mg
- + - + +
RF 0
RF 0.38
RF 0.71
HP78
1 2 3 4 5
c
5mM Ca
1mM Ca
1mM Mn
- preinc
Mn/Mg
Ca
Mn
-
RF 0
RF 0.41
78bp
HP78
1 2 3 4 5 6 7 8 9 10 11 12 13
d
Quantity DNA Collected (%)
0 2 4 6 8 10 12 14 16 18
Mn2+
Mg2+
Ca2+
Fe2+
Co2+
Cu2+
Ni2+
Zn2+
5mM Metal

Figure 5.3.3



Metal tests using the gel filtration assay. HP78 was heated, snap-cooled and 5′ end-labelled and then incubated under DNA binding conditions. The lower line of (+) and (-) represents the presence or absence of SbcCD in the binding reaction. The upper line indicates whether (+) or not (-) the reaction was passed through a microspin gel filtration column. The first metal indicates the SbcCD preincubation metal at a concentration of 1mM, and the second indicates the reaction metal at a concentration of 5mM. Reactions 5, 6, 7 and 8 were performed in the presence of 0.05mM EDTA, and 13, 14, 15 and 16 were performed in the absence of metal ion. Reactions not treated by gel filtration were maintained on ice during this procedure. Reactions were separated on 10% denaturing polyacrylamide sequencing sized gels.

Figure 5.3.4

DNA binding activity on the thiophosphate substrate. HP78 and thioHP78 were heated, snap-cooled and 5' end-labelled.

a

HP78 was incubated under conditions of the TAE binding assay in the presence (lanes 4 and 5) and absence (lanes 2 and 3) of ATPγS, and thioHP78 similarly in the presence (lanes 8 and 9) and absence (lanes 6 and 7) of ATPγS. The presence (+) and absence (-) of SbcCD is indicated. Reactions were separated on a 6.8% native polyacrylamide gel. Marker V is loaded in lane 1.

b

HP78 and thioHP78 binding reactions were subjected to microspin gel filtration analysis in duplicate and the mean percentage candidate bound DNA recovered is indicated on the histogram. The error bars represent the range which is equal to twice the standard error. Values are corrected for background.

c

HP78 (lanes 2, 3, 4 and 5) and thioHP78 (lanes 6, 7, 8 and 9) were incubated under conditions of the TAE binding assay in the presence (+) or absence (-) of SbcCD, and in the presence of 1mM ATPγS and 5mM Mn^{2+}. The reactions marked X were then UV crosslinked as described in the text. Reactions were separated on a 6.8% native polyacrylamide gel at 180V, 4°C. Marker V was included in lane 1.

Figure 5.3.4

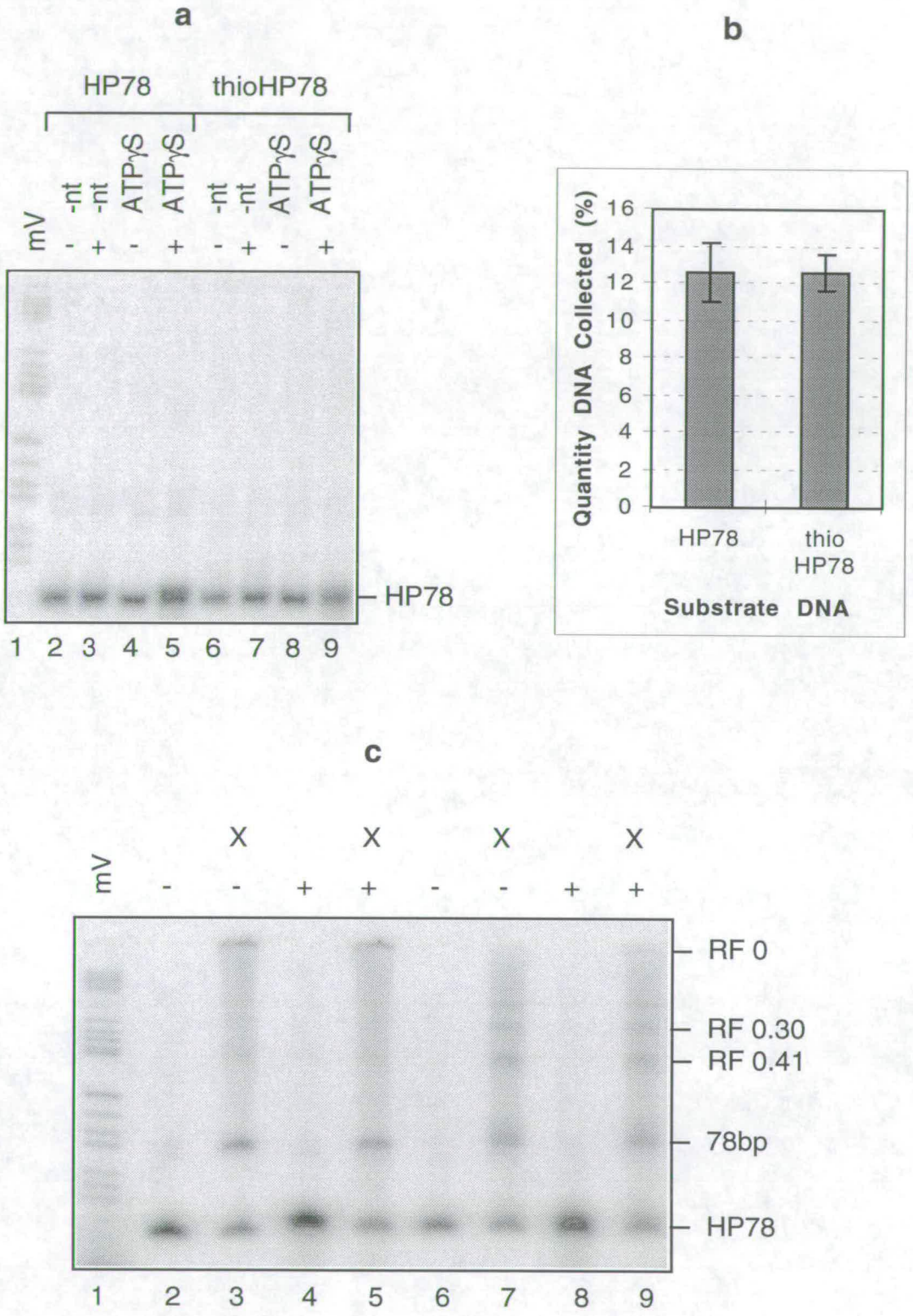

a
HP78
thioHP78
mV
-nt
-nt
ATPγS
ATPγS
- + - + - + - +
HP78
1 2 3 4 5 6 7 8 9
b
Quantity DNA Collected (%)
16
14
12
10
8
6
4
2
0
HP78
thio HP78
Substrate DNA
c
X X X X
mV
- - + + - - + +
RF 0
RF 0.30
RF 0.41
78bp
HP78
1 2 3 4 5 6 7 8 9

Figure 5.3.5

SbcCD nuclease activity on the thiophosphate substrate. HP78 and thioHP78 were heated, snap-cooled and 5' end-labelled, and subjected to nuclease assays. The products were separated on 10% denaturing sequencing sized polyacrylamide gels.

a and b

HP78 (reactions 2, 3 and 4) and thioHP78 (reactions 5, 6 and 7) were incubated at 37°C (**a**) and 16°C (**b**) in the presence of 1mM ATP (lanes 2, 3, 5 and 6) or ATPγS (lanes 4 and 7) and the presence (+) or absence (-) of SbcCD. Similar reactions of HP78 and thioHP78 have been loaded adjacent in (**b**) for closer comparison. Marker V was loaded in lanes 1 and 8.

c and d

HP78 (**c**) and thioHP78 (**d**) were incubated at 16°C in the presence (+) or absence (-) of SbcCD in 1mM ATPγS and in 0mM (lanes 2 and 3), 50mM (lane 5), 75mM (lane 6), 100mM (lanes 4 and 7), 150mM (lane 8) or 200mM (lane 9) NaCl. Marker V was loaded in lanes 1 and 10.

Figure 5.3.5

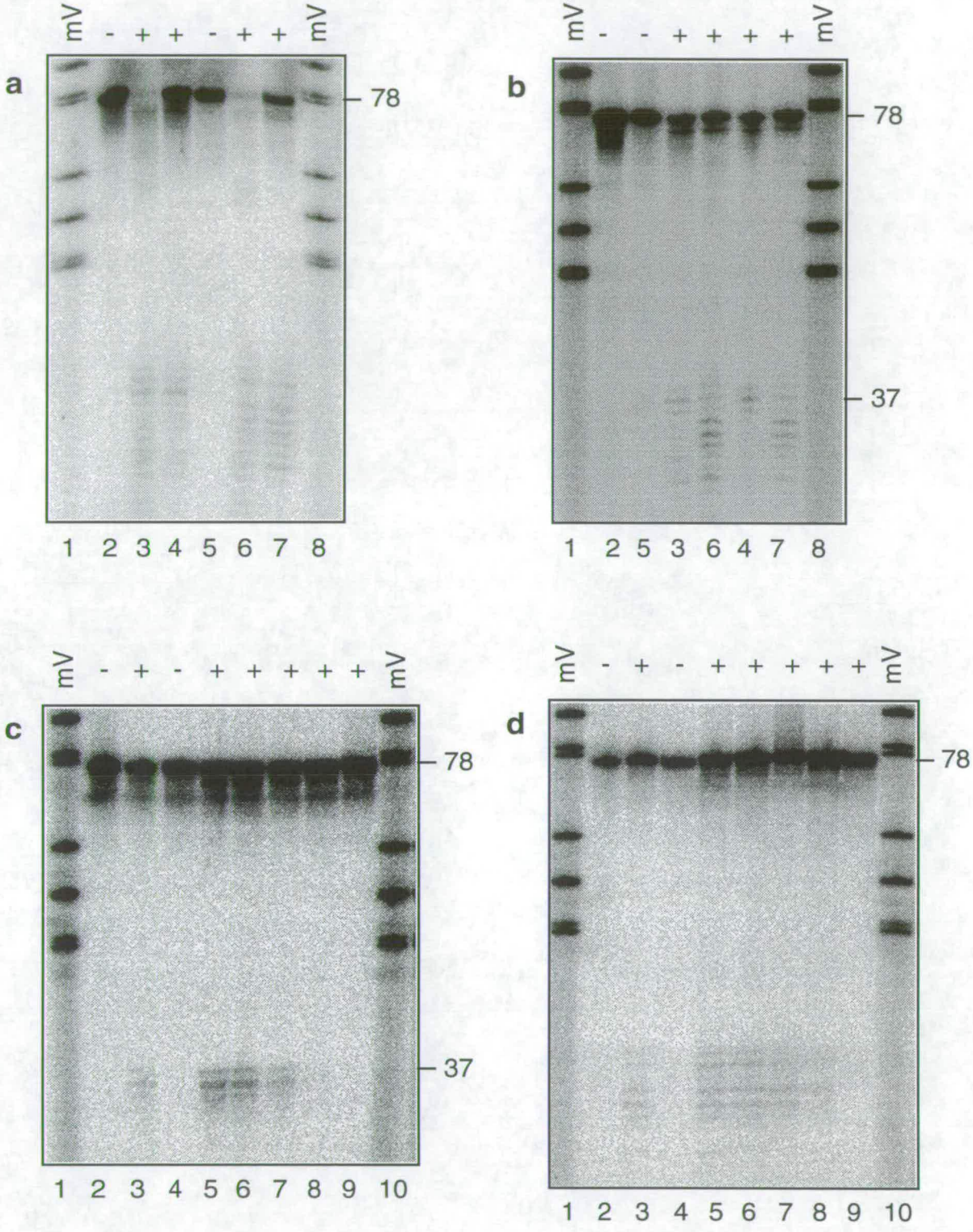

a
mV - + + - + + mV
78
1 2 3 4 5 6 7 8
b
mV - - + + + + mV
78
37
1 2 5 3 6 4 7 8
c
mV - + - + + + + + mV
78
37
1 2 3 4 5 6 7 8 9 10
d
mV - + - + + + + + mV
78
1 2 3 4 5 6 7 8 9 10

Figure 5.3.7

DNA binding by mutant SbcCD. HP78 and thioHP78 were heated, snap-cooled and 5' end-labelled.

a and b

HP78 (lanes 1, 2, 5, 6, 9, 10, 13 and 14) and thioHP78 (lanes 3, 4, 7, 8, 11 and 12) were incubated with mutant SbcD(H184Y) (lanes 2 and 4), SbcD(D8A) (lanes 6 and 8), SbcD(D8E) (lanes 10 and 12) or wildtype (lane 14) SbcCD, or without (lanes 1, 3, 5, 7, 9, 11 and 13) SbcCD under conditions of the TAE (**a**) or TGE (**b**) binding assays. ATPγS and Mn^{2+} were included in the reactions. The reactions were run on TAE (**a**) or TGE (**b**) 6.8% native polyacrylamide gels.

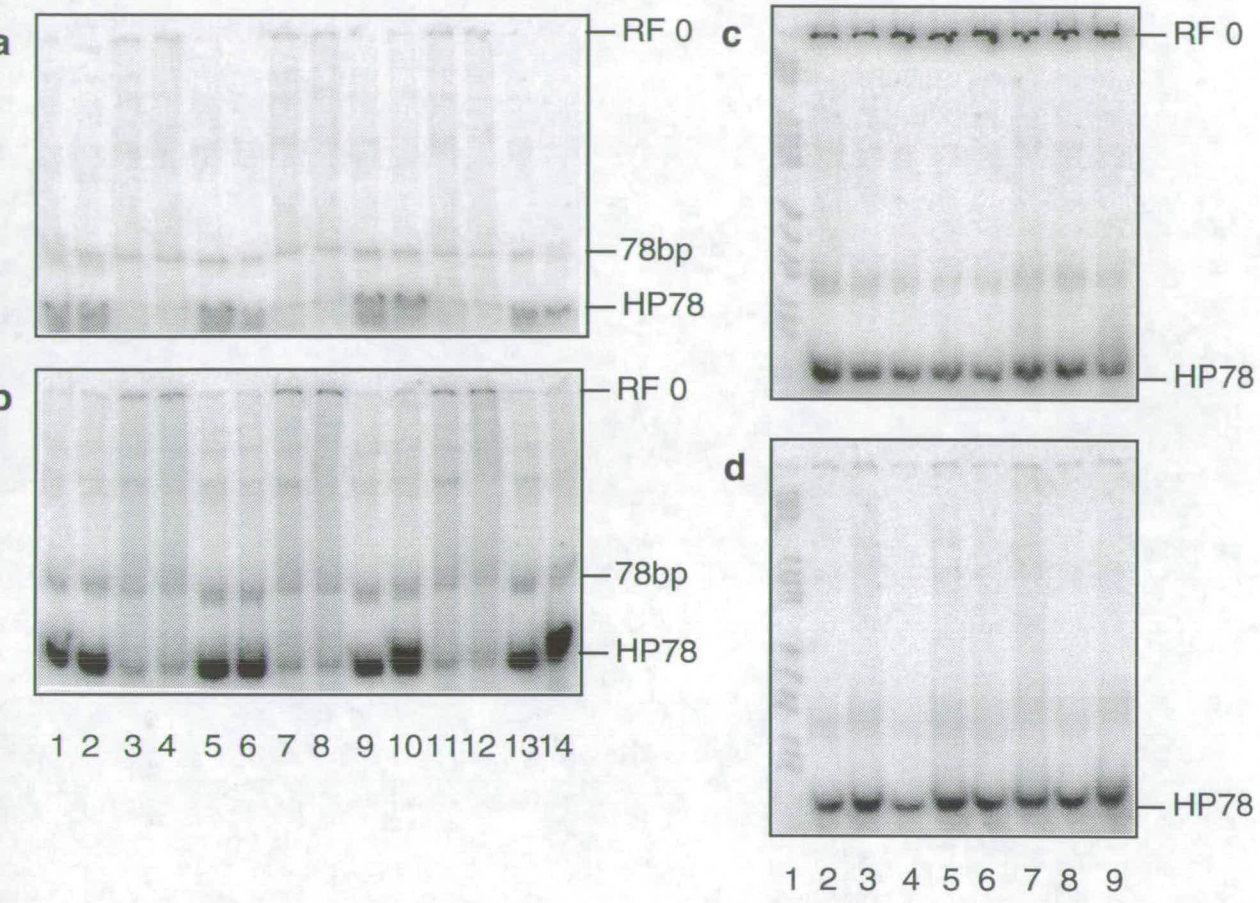
c and d

HP78 was incubated with mutant SbcD(H184Y) (lane 2), SbcD(D8A) (lane 4), SbcD(D8E) (lane 6) or wildtype (lane 8) SbcCD, or without (lanes 1, 3, 5 and 7) SbcCD in buffer pH6.75 (**c**) or pH7.0 (**d**) under binding conditions. ATPγS and Mn^{2+} were included in the reactions. The reactions were run on 6.8% native polyacrylamide gels at the respective pHs. Marker V was loaded in lane 1.

e

HP78 was incubated with mutant SbcD(H184Y) (lanes 2 and 3), SbcD(D8A) (lanes 5 and 6), SbcD(D8E) (lanes 8 and 9) or wildtype (lanes 11 and 12) SbcCD, or without (lanes 1, 4, 7 and 10) SbcCD. 75mM (lanes 1, 2, 4, 5, 7, 8, 10 and 11) or 150mM (lanes 3, 6, 9 or 12) NaCl was included in the reactions, in addition to ATPγS and Mn^{2+}. The reactions were run on a 6.8% native polyacrylamide gel.

Figure 5.3.7


a
RF 0
78bp
HP78
b
RF 0
78bp
HP78
1 2 3 4 5 6 7 8 9 10 11 12 13 14
c
RF 0
HP78
d
HP78
1 2 3 4 5 6 7 8 9

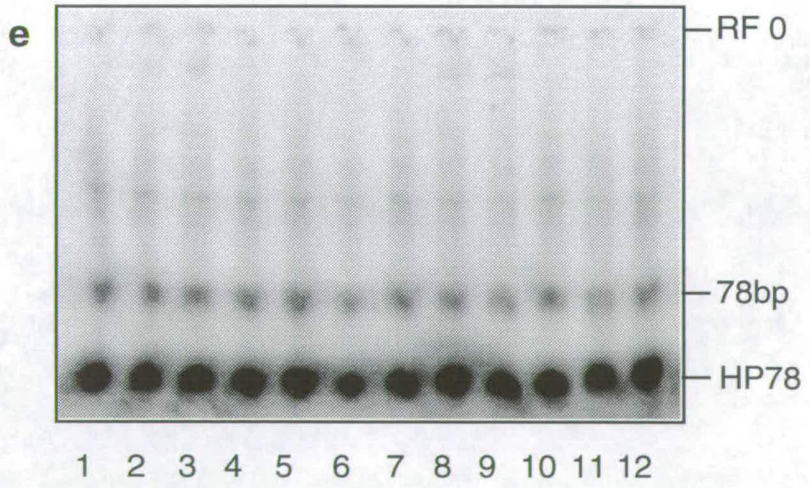

e
RF 0
78bp
HP78
1 2 3 4 5 6 7 8 9 10 11 12

Figure 5.3.8

Further tests on the mutant SbcCD protein. HP78 was heated, snap-cooled and 5' end-labelled.

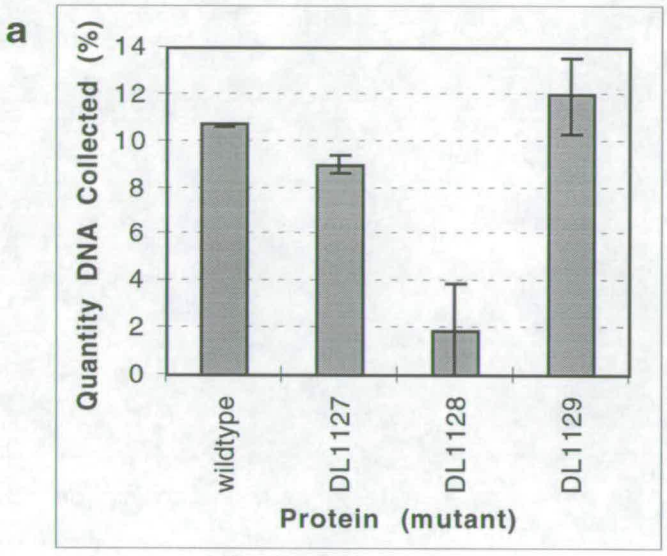
a

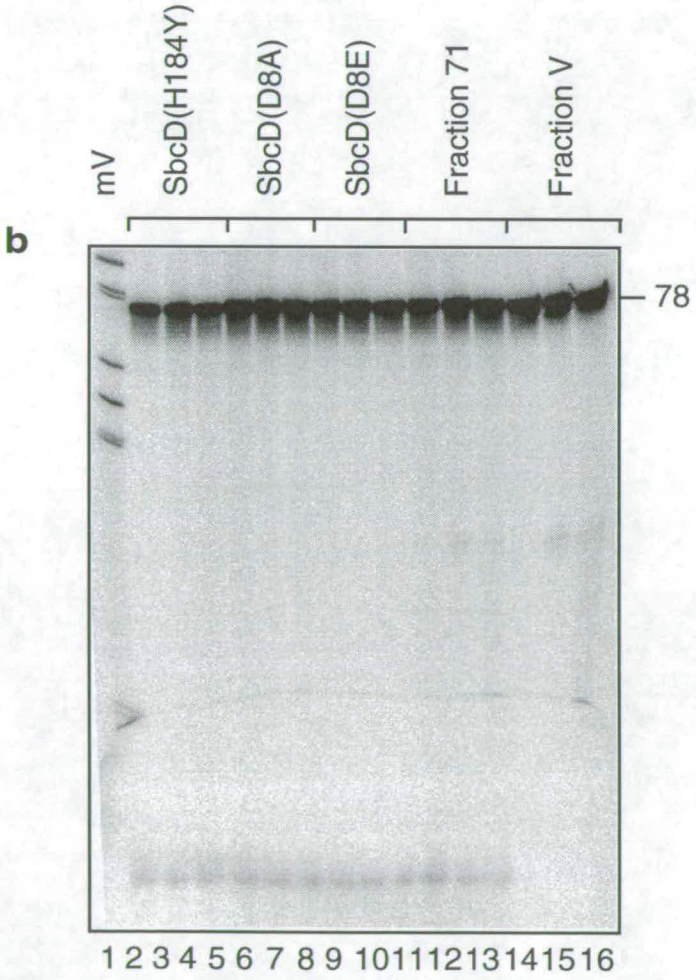
Gel filtration analysis of DNA binding by mutant SbcCD. HP78 was incubated in the presence of ATPγS and Mn^{2+} with wildtype and mutants SbcD(H184Y), SbcD(D8A) and SbcD(D8E) under binding conditions and the reactions subjected to microspin gel filtration in duplicate. Mean percentages are given for the quantity of DNA collected as candidate bound complex, and the error bars represent the range, which equals twice the standard error. Values are corrected for background.

b

Nuclease activity of mutant SbcCD. HP78 was incubated with mutant SbcD(H184Y) (lanes 3 and 4), SbcD(D8A) (lanes 6 and 7), SbcD(D8E) (lanes 9 and 10), Fraction 71 wildtype (lanes 12 and 13) and Fraction V wildtype (lanes 15 and 16) SbcCD, or without SbcCD (lanes 2, 5, 8, 11 and 14). Reactions were performed in the presence of Mn^{2+} and ATPγS (lanes 2, 3, 5, 6, 8, 9, 11, 12, 14 and 15) or ATP (lanes 4, 7, 10, 13 and 16) and run on a 6.8% native polyacrylamide gel. Marker V was loaded in lane 1.

Figure 5.3.8


a
Quantity DNA Collected (%)
14
12
10
8
6
4
2
0
wildtype
DL1127
DL1128
DL1129
Protein (mutant)


b
mV
SbcD(H184Y)
SbcD(D8A)
SbcD(D8E)
Fraction 71
Fraction V
78
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

DNA Binding by SbcCD

SbcCD May Bind DNA Transiently

Candidate protein-DNA complexes were obtained under certain conditions using glutaraldehyde crosslinking. However these species were recovered at very low yield and were not obtained without crosslinking, a method infamous for artefactual results. Otherwise, no binding was detected by gel retardation under any conditions, including those that enhanced weak binding. DNA binding by SbcCD was detected by gel filtration. That the bound DNA largely consisted of product suggests that substrate DNA is rapidly converted to product on protein binding, and this is consistent with the tentative model for transient binding. That the product-protein complexes were obtained by gel filtration and not gel retardation suggests that they bound weakly, also consistent with the model.

The lack of detected binding is unlikely to be due to the absence of a required cofactor. A nucleotide cofactor and a divalent metal ion cofactor are required for SbcCD nuclease activity (Chapter 1). Binding activity has been investigated alongside these requirements. In addition, there is no evidence for binding sites for other cofactors in the SbcC or SbcD sequences.

It is concluded that DNA binding by SbcCD is difficult to detect. Indirect evidence is presented that this may result from transient, rather than weak, interaction between protein and substrate.

SbcD May Lack a Eukaryotic-Specific C-Terminal DNA Binding Domain

The C-terminal deletion mutant Mre11ΔC49 exhibits very weak DNA binding compared to wildtype yeast Mre11 (Furuse *et al* 1998) and two DNA binding sites have been mapped to the C-terminus (Usui *et al* 1998, Chapter 1 Section 4). This acidic C-terminal region is eukaryotic-specific, *ie* absent in SbcD (Chapter 1 Section 4). Its absence in SbcD may account for the low stability of DNA binding by SbcCD.

It has been suggested that the acidic C-terminus acts as a protein-protein interaction site, recruiting meiosis-specific proteins which facilitate DNA binding (Usui *et al* 1998, Chapter 1 Section 4). However DNA binding by wildtype Mre11 has been demonstrated in the absence of other proteins. Therefore protein interactions are not essential. It may be that Mre11 has differing DNA binding modes in mitosis and meiosis.

The double-strand exonuclease activity of Mre11ΔC49 is significantly increased above that of wildtype Mre11 (Kunihiro Ohta and Takehiko Shibata unpublished). It is possible that the 3' terminal region of Mre11 increases the stability of the protein-substrate complex, and in turn decreases the nuclease activity. Nuclease activity would be reduced if the intermediate complex was more stable, since further reaction would be delayed and turnover would be lower. Therefore this is consistent with transient binding by SbcCD. It suggests that the 3' terminal region of Mre11 reduces the transience of Mre11 DNA binding sufficiently for complexes to be clearly observed by gel retardation. Perhaps there was an evolutionary advantage to reducing the Rad50/Mre11 nuclease activity in yeast, and an evolutionary advantage to maintaining a high SbcCD nuclease activity in *E.coli*. These differences may reflect the different cellular roles of these proteins.

SbcCD as a Processive Nuclease

In further support of these ideas, glycerol gradient centrifugation was required to demonstrate binding between RecBCD and its double-strand substrate (Goldmark and Linn 1972). Values for the K_d vary according to the stoichiometry of the complex (Taylor and Smith 1995). Like SbcCD, RecBCD is a processive nuclease. It may be that these proteins remain bound whilst translocating (in ATP), and when translocation is inhibited (in ATPγS in the case of SbcCD) or finishes (perhaps at the end of the substrate molecule) the protein undergoes a conformational change which prevents it from binding.

Alternative DNA Binding Assays

Certain methods of detecting DNA binding have not been used because they are unlikely to shed new light on this study: Other types of gel filtration and glycerol gradient sedimentation (Hiom and West 1995a) might produce results similar to those obtained from microspin gel filtration, but they do not identify the polypeptides bound. Other types of crosslinking, such as formaldehyde PCR crosslinking (Alan Grossman personal communication), bromodeoxyU, UV laser, and biotin-streptavidin pull down methods (reviewed Kneale *et al* 1994) are vulnerable to the same non-specific artefacts as glutaraldehyde and *p*-APB crosslinking. Filter binding (Coombs and Pearson 1978, Strauss *et al* 1981) offers no benefits over gel retardation, especially since agarose gels can be used to analyse large complexes.

Methods which utilise the anisotropy (polarisation) of a beam of light resulting from the movement of a protein in solution include measurement of circular dichroism and intrinsic fluorimetry (reviewed Kneal *et al* 1994). Although they do not enable a transient complex to be trapped, they would allow transient binding to be observed *in situ* which might provide useful information. However fluorimetry requires approximately 100µl 1µM protein per experiment and the SbcCD protein currently available is not concentrated enough nor sufficient for a series of such experiments to be conducted.

It would be interesting to study the SbcC and SbcD polypeptides separately, since they may bind more stably alone. However SbcC and SbcD have not yet been individually purified. Binding to alternative substrates might be fruitfully studied if SbcCD has another cellular role. A Holliday Junction, cruciform or branched hairpin might be analysed. Nuclease activity on a branched hairpin was obtained (Section 2, Connelly *et al* 1999). Alternatively DNA binding might be observed *in situ* by electron microscopy.

CHAPTER 6

DISCUSSION

Summary

SbcCD cleaves a DNA hairpin substrate (Chapter 3). Initial cleavage occurs immediately 5' of the loop and additional early cleavage events occur at the 5' end of the molecule. These products are subsequently degraded in the presence of ATP. Early products can be trapped using ATPγS or GTP at a lower incubation temperature. Cleavage products have 5' phosphate and 3' hydroxyl termini, like those of most nucleases. A dumbell substrate was generated to test the structure requirements for the hairpin nuclease. This has a hairpin structure but lacks termini. It was cleaved by SbcCD, indicating that an aspect of the hairpin other than its termini is recognised by the nuclease. This observation supports the model for the regulation of palindrome maintenance by SbcCD proposed (Leach 1994) and is consistent with genetic data (Pinder *et al* 1997). SbcCD may therefore be involved in the formation and processing of DSBs, like its eukaryotic homologues.

A potential DNA helicase activity of SbcCD has been investigated (Chapter 4). SbcCD does not possess helicase activity on the hairpin substrate tested. It is proposed that SbcCD is unlikely to be a helicase. The nicked hairpin product of SbcCD activity in ATPγS migrates anomalously on native gels under certain conditions and this is likely to result from splaying of the cleaved double-strand termini.

DNA binding by SbcCD was studied with a view to testing different substrates to characterise alternative cellular roles (Chapter 5). Retarded bands but no candidate complexes were observed by gel retardation in the absence of fixation (Chapter 5, Section 1). By glutaraldehyde crosslinking the binding reactions and using a combination of agarose and polyacrylamide gel retardation, three candidates for SbcCD-DNA complexes were obtained. They were obtained inconsistently and at very low level. They were therefore not studied further by Western Blotting. Gel retardation of glutaraldehyde fixed material is subject to artefacts, and these candidates may or may not correspond to true reaction intermediates.

Candidate complexes were obtained by microspin gel filtration and largely consisted of protein bound product DNA (Chapter 5, Section 2). However it

is not possible to identify the bound polypeptides by this method. Attempts to improve gel retardation conditions for potentially weak DNA binding were unsuccessful. Together with other experimental and circumstantial evidence this suggests that SbcCD does not bind its substrate weakly.

Attempts were made to uncouple binding from the nuclease activity to trap potentially transient binding (Chapter 5, Section 3). None of the methods employed were successful in inhibiting the nuclease. It was concluded that the difficulties in obtaining SbcCD-DNA binding may result from transient binding, but this hypothesis has yet to be tested.

Post-Modern SbcCD

Mn^{2+} Dependence And Other Metals

Mn^{2+} dependent protein activities *in vitro* are often diagnostic of Mg^{2+} activities *in vivo*. They are often less specific than those in Mg^{2+}. It is possible that the Mn^{2+} dependent nuclease activities which form the basis of this work may therefore correspond to activities less specific than those which might be detected in other conditions. However it is feasible that Mn^{2+} is the metal cofactor utilised in vivo, since the phosphoesterase family proteins use Mn^{2+} and transition metals but not Mg^{2+} (Chapter 1).

DNA Binding by SbcCD

It is interesting that SMCs bind cruciform DNA in preference to duplex, in preference to single-strand DNA, and long DNA in preference to short DNA. It is not know how well they bind hairpin substrates. Here it has been shown that SbcCD binds a hairpin substrate poorly. It is unknown how well SbcCD binds duplex and single-strand DNA. It would be interesting to test whether SbcCD binds cruciform DNA, or other DNA substrates with a "corner" structure. This merits detailed investigation.

The studies on SbcCD DNA binding were based on the premise that binding to a known nuclease substrate should be optimised before alternative substrates were analysed. If it is true that binding to the nuclease substrate is transient, hairpin DNA may not be the optimal substrate on which to

demonstrate DNA binding. Other substrates, such as cruciforms, may have provided clearer results. The preferred DNA binding substrates of hMre11 differed from the preferred double-strand exonuclease substrates (Paull and Gellert 1998, Chapter 1 Section 4).

Cellular Roles of SbcCD, Rad50/Mre11 and SMC Proteins

SbcCD is a hairpin nuclease, and Rad50/Mre11 is a candidate for hairpin opening and V(D)J recombination in mammals. It is not yet known whether hairpins are similarly intermediates in non homologous end-joining. In addition, SbcCD is implicated in stimulating recombination and DNA repair. However a direct involvement of SbcCD in recombination is not implicated, in contrast to Rad50/Mre11.

Rad50 is involved in recombination and in sister chromatid cohesion at sites of synapsis. These are roles expected for a homologue of the SMC proteins. It will be interesting to find out whether Rad50 forms SMC heterodimers in complexes with different roles to the Rad50/Mre11 complex. It is possible that SbcCD interacts with other proteins.

SbcCD and TREDs

In this thesis, evidence is presented to indicate that SbcCD recognises and cleaves hairpin structures in regulating the maintenance of palindromes (Chapter 3). Pseudo-hairpin formation and strand-slippage is believed to be responsible for the expansion of trinucleotide repeat arrays observed in TREDs (Chapter 1, Section 1). It might be expected that SbcCD should recognise and remove these pseudo-hairpins. It is therefore interesting that SbcCD cleaves a trinucleotide repeat oligonucleotide (Connelly *et al* 1999). This suggests that the homologues of SbcCD may have a role in determining trinucleotide repeat expansion.

Consistent with this, amplification of TRED tracts in the bimodal *E. coli* model is prevented in the presence of functional SbcC(D) (Sarkar *et al* 1998). However human MSH2 binds a $d(CTG)_n$ pseudo-hairpin with an affinity which increases with the length of the array (Pearson *et al* 1997), presumably by recognising the mismatches present in the stem. It has been suggested

that hMSH2 may block access of SbcCD to the pseudo-hairpins associated with trinucleotide repeat expansion (Schmidt *et al* 1999). The human homolgue of SbcCD might perform a similar function in human disease. If it were possible to modify the human homologue of SbcCD (Rad50/Mre11/NBS1/x_1/x_2) so that it were to recognise the hMSH2 bound pseudo-hairpin, a therapy for all TREDs might be developed.

Last Word

During the past six months or so, research into Mre11 and SMC proteins has boomed. The field of study of SbcCD is also new. Tantalising glimpses of the effects of the eukaryotic homologues on human disease and the apparent complexity of the cellular roles of these homologues demonstrate the importance of these studies.

BIBLIOGRAPHY

Akgün E, Zahn J, Baumes S, Brown G, Liang F, Romanienko P J, Lewis S and Jasin M (1997) Palindrome resolution and recombination in the mammalian germ line. *Mol Cell Biol* **17** 5559-5570

Ajimura M, Leem S and Ogawa H (1992) Identification of new genes required for meiotic recombination in *Saccharomyces cerevisiae*. *Genetics* **133** 51-66

Akhmedov A T, Frei C, Tsai-Pflugfelder M, Kemper B, Gasser S M and Jessberger R (1998) Structural maintenance of chromosomes protein C-terminal domains bind preferentially to DNA with secondary structure. *J Biol Chem* **273** 24088-24094

Alani E, Padmore R and Kleckner N (1990) Analysis of wild-type and *rad50* mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. *Cell* **61** 419-436

Alani E, Subbiah S and Kleckner N (1989) The yeast *RAD50* gene encodes a predicted 153-kD protein containing a purine nucleotide-binding domain and two large heptad-repeat regions. *Genetics* **122** 47-57

Albertini A M, Hofer M, Calos M P and Miller J H (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* **29** 319-328

Allen J B, Zhou Z, Siede W, Friedberg E C and Elledge S J (1994) The SAD1/RAD3 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev* **8** (20) 2401-2415

Allers T and Leach D R (1995) DNA palindromes adopt a methylation-resistant conformation that is consistent with DNA cruciform or hairpin formation in vivo. *J Mol Biol* **252** 70-85

Allgood N D and Silhavy T J (1991) *Escherichia coli xonA (sbcB)* mutants enhance illegitimate recombination. *Genetics* **127** 671-680

Asai T and Kogoma T (1994) D-loops and R-loops: Alternative mechanisms for the initiation of chromosome replication in *Escherichia coli*. *J Bacteriol* **176** (7) 1807-1812

Baker T A and Mizuuchi K (1992) DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev* **6** 2221-2232

Bednarz A L, Boocock M R and Sherratt D J (1990) Determinants of correct res site alignment in site-specific recombination by Tn3 resolvase. *Genes Dev* **4** 2366-2375

Beese L S and Steitz T A (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J* **10** 25-33

Bennett R J and West S C (1995) RuvC protein resolves Holliday junctions via cleavage of the continuous (noncrossover) strands. *Proc Natl Acad Sci U S A* **92** 5635-5639

Bennett R J, Dunderdale H J and West S C (1993) Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. *Cell* **74** 1021-1031

Benson F, Bauman P and West S (1998) Synergistic interactions of Rad51 and Rad52 in recombination and DNA repair. *Nature* **391** 401-404

Bergerat A, de Massy B, Gadelle D, Varoutas P C, Nicolas A and Forterre P (1997) An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386** 414-417

Berstein H (1968)Repair and recombination in phage T4. I: Genes affecting recombination. *Cold Spring Harbor Symp Quant Biol* **33** 325-331

Beyert N, Reichenberger S, Peters M, Hartung M, Gottlich B, Goedecke W, Vielmetter W and Pfeiffer P (1994) Nonhomologous DNA end joining of synthetic hairpin substrates in *Xenopus laevis* egg extracts. *Nucleic Acids Res* **22** 1643-1650

Bhattacharyya A, Murchie A I H, von Kitzing E, Diekmann S, Kemper B and Lilley D M J (1991) Model for the interaction of DNA junctions and resolving enzymes. *J Mol Biol* **221** 1191-1207

Bianchi M E (1994) Prokaryotic HU and eukaryotic HMG1: a kinked relationship. *Mol Microbiol* **14** 1-5

Bierne H and Michel B (1994) When replication forks stop. *Mol Microbiol* **13** 17-23

Birkenbihl R P and Subramani S (1992) Cloning and characterization of *rad21* an essential gene of *Schizosaccharomyces pombe* involved in double-strand break repair. *Nucl Acids Res* **20** (24) 6605-6611

Blake J A, Ganguly N and Sherratt D J (1997) DNA sequence of recombinase-binding sites can determine Xer site- specific recombination outcome. *Mol Microbiol* **23** 387-398

Blakely G, May G, McCulloch R, Arciszewska L K, Burke M, Lovett S T and Sherratt D J (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* **75** 351-361

Blinov V M, Koonin E V, Gorbalenya A E, Kaliman A V and Kryukov V M (1989) Two early genes of T5 encode proteins containing an NTP-binding sequence motif and probably involved in DNA replication, recombination and repair. *FEBS Lett* **252** 47-52

Boubnov N V, Hall K T, Wills Z, Lee S E, He D M, Benjamin D M, Pulaski C R, Band H, Reeves W, Hendrickson E A and et al. (1995) Complementation of the ionizing radiation sensitivity, DNA end binding, and V(D)J recombination defects of double-strand break repair mutants by the p86 Ku autoantigen. *Proc Natl Acad Sci U S A* **92** 890-894

Boubrik F and Rouviere-Yaniv J (1995) Increased sensitivity to gamma irradiation in bacteria lacking protein HU. *Proc Natl Acad Sci U S A* **92** 3958-3962

Boulton S J and Jackson S P (1998) Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J* **17** 1819-1828

Bressan D A, Olivares H A, Nelms B E and Petrini J H (1998) Alteration of N-terminal phosphoesterase signature motifs inactivates *Saccharomyces cerevisiae* Mre11. *Genetics* **150** 591-600

Bridges B A (1995) Sexual potency and adaptive mutation in bacteria. *Trends Microbiol* **3** 291-292; discussion 292-294

Budavari S, O'Neil M J, Smith A, Heckelman P E and Kinneary J F (Eds) (1996) *Merck Index: An encyclopaedia of chemicals, drugs and biologicals* (12th edition) Merck Research Laboratories New Jersey

Burgin A B, Jr., Huizenga B N and Nash H A (1995) A novel suicide substrate for DNA topoisomerases and site-specific recombinases. *Nucleic Acids Res* **23** 2973-2979

Cao L, Alani E and Kleckner N (1990) A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. *Cell* **61** 1089-1101

Carey J (1988) Gel retardation at low pH resolves trp repressor-DNA complexes for quantitative study. *Proc Natl Acad Sci U S A* **85** 975-979

Carney J P, Maser R S, Olivares H, Davis E M, Le Beau M, Yates J R r, Hays L, Morgan W F and Petrini J H (1998) The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93** 477-486

Cary R B, Peterson S R, Wang J, Bear D G, Bradbury E M and Chen D J (1997) DNA looping by Ku and the DNA-dependent protein kinase. *Proc Natl Acad Sci U S A* **94** 4267-4272

Chalker A F, Leach D R and Lloyd R G (1988) *Escherichia coli sbcC* mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene* **71** 201-205

Chalker A F, Okely E A, Davison A and Leach D R (1993) The effects of central asymmetry on the propagation of palindromic DNA in bacteriophage lambda are consistent with cruciform extrusion *in vivo*. *Genetics* **133** 143-148

Chamankhah M and Xiao W (1998) Molecular cloning and genetic characterization of the *Saccharomyces cerevisiae NGS1/MRE11* gene. *Curr Genet* **34** 368-374

Chase J W and Williams K R (1986) Single-stranded DNA binding proteins required for DNA replication. *Annu Rev Biochem* **55** 103-136

Chuang P T, Albertson D G and Meyer B J (1994) DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell***79** 459-474

Chuang P T, Lieb J D and Meyer B J (1996) Sex specific assembly of a dosage compensation complex on the nematode X chromosome. *Science* **274** 1736-1739

Claassen L A and Grossman L (1991) Deletion mutagenesis of the *Escherichia coli* UvrA protein localizes domains for DNA binding, damage recognition, and protein-protein interactions. *J Biol Chem* **266** 11388-11394

Cohen A and Clark A J (1986) Synthesis of linear plasmid multimers in *Escherichia coli* K-12. *J Bacteriol* **167** 327-335

Cohen P T and Cohen P (1989) Discovery of a protein phosphatase activity encoded in the genome of bacteriophage lambda. Probable identity with open reading frame 221. *Biochem J* **260** (3) 931-934

Collick A, Drew J, Penberth J, Bois P, Luckett J, Scaerou F, Jeffreys A and Reik W (1996) Instability of long inverted repeats within mouse transgenes. *EMBO J* **15** (5) 1163-1171

Connelly J C and Leach D R (1996) The sbcC and sbcD genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells* 1 285-291

Connelly J C, de Leau E S, Okely E A and Leach D R (1997) Overexpression, purification, and characterization of the SbcCD protein from *Escherichia coli*. *J Biol Chem* **272** 19819-19826

Connelly J C, deLeau E S and Leach D R F (1999) DNA cleavage and degradation by the SbcCD protein complex from *Escherichia coli*. *Nucleic Acids Res* 27 (4) 1039-1046

Connelly J C, Kirkham L A and Leach D R (1998) The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc Natl Acad Sci U S A* **95** 7969-7974

Coombs D H and Pearson G D (1978) Filter-binding assay for covalent DNA-protein complexes: adenovirus DNA- terminal protein complex. *Proc Natl Acad Sci U S A* **75** 5291-5295

Cox T M and Sinclair J (Eds) (1997) *Molecular Biology in Medicine*, Blackwell Science Ltd

Cruse W B, Salisbury S A, Brown T, Cosstick R, Eckstein F and Kennard O (1986) Chiral phosphorothioate analogues of B-DNA. The crystal structure of Rp- d[Gp(S)CpGp(S)CpGp(S)C]. *J Mol Biol* **192** 891-905

Darlow J M and Leach D R (1995) The effects of trinucleotide repeats found in human inherited disorders on palindrome inviability in *Escherichia coli* suggest hairpin folding preferences *in vivo*. *Genetics* 141 825-832

Das Gupta U, Weston-Hafer K and Berg D E (1987) Local DNA sequence control of deletion formation in *Escherichia coli* plasmid pBR322. *Genetics* 115 41-49

Davison A and Leach D R (1994a) The effects of nucleotide sequence changes on DNA secondary structure formation in *Escherichia coli* are consistent with cruciform extrusion *in vivo*. *Genetics* **137** 361-368

Davison A and Leach D R (1994b) Two-base DNA hairpin-loop structures in vivo. *Nucleic Acids Res* **22** 4361-4363

De Wachter R, Maniloff J and Fiers W (1990) Two-dimensional gel electrophoresis of nucleic acids, In *Practical Approach IRL: Gel Electrophoresis of Nucleic Acids*. Eds Rickwood D and Hames B D, Oxford University Press New York

de Massy B, Weisberg R A and Studier F W (1987) Gene 3 endonuclease of bacteriophage T7 resolves conformationally branched structures in double-stranded DNA. *J Mol Biol* **193** 359-376

Dolganov G M, Maser R S, Novikov A, Tosto L, Chong S, Bressan D A and Petrini J H (1996) Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Mol Cell Biol* **16** 4832-4841

Dunderdale H J, Sharples G J, Lloyd R G and West S C (1994) Cloning, overexpression, purification, and characterization of the *Escherichia coli* RuvC Holliday junction resolvase. *J Biol Chem* **269** 5187-5194

Earnshaw W C, Halligan B, Cooke C A, Heck M M S and Liu L F (1985) Topoisomerase II is a structural component of mitotic chromosome scaffolds. *J Cell Biol* **100** 1706-1715

Ehrlich S D, Bierne H, d'Alencon E, Vilette D, Petranovic M, Noirot P and Michel B (1993) Mechanisms of illegitimate recombination. *Gene* **135** 161-166

Engler L E, Welch K K and Jen-Jacobson L (1997) Specific binding by *Eco*RV endonuclease to its DNA recognition site GATATC. *J Mol Biol* **269** 82-101

Espeli O and Boccard F (1997) In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. *Mol Microbiol* **26** 767-777

Fairman J, Wang R Y, Liang H, Zhao L, Saltman D, Liang J C and Nagarajan L (1996) Translocations and deletions of 5q13.1 in myelodysplasia and acute myelogenous leukemia: evidence for a novel critical locus. *Blood* **88** 2259-2266

Featherstone C and Jackson S P (1998) DNA repair: The Nijmegan Breakage Syndrome protein. *Curr Biol* 8 R622-R625

Fraser M J (1994) Endo-exonucleases: enzymes involved in DNA repair and cell death? *Bioessays* **16** 761-766

Furuse M, Nagase Y, Tsubouchi H, Murakami-Murofushi K, Shibata T and Ohta K (1998) Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J* **17** (21) 6412-6425

Gabbidon M R, Rampersaud V E and Julin D A (1998) Salt-stable complexes of the *Escherichia coli* RecBCD enzyme bound to double-strand DNA. *Arch Biochem Biophys* **350** (2) 266-272

Gacy A M, Goellner G, Juranic N, Macura S and McMurray C T (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell* **81** 533-540

Gasser S M (1995) Chromosome structure. Coiling up chromosomes. *Curr Biol* **5** 357-360

Gibson F P, Leach D R and Lloyd R G (1992) Identification of *sbcD* mutations as cosuppressors of *recBC* that allow propagation of DNA palindromes in *Escherichia coli K-12*. *J Bacteriol* 174 1222-1228

Gillen J R, Willis D K and Clark A J (1981) Genetic analysis of the *recE* pathway of genetic recombination in *Escherichia coli* K-12. *J Bacteriol* 145 521-532

Gilson E, Clément J M, Brutlag D and Hofnung M (1984) A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J* 3 1417-1421

Goldberg J, Huang H B, Kwon Y G, Greengard P, Nairn A C and Kuriyan J (1995) Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* **376** 745-753

Goldmark P J and Linn S (1972) Purification and properties of the recBC DNase of *Escherichia coli* K-12. *J Biol Chem* **247** (6) 1849-1860

Gorbalenya A E and Koonin E V (1989) Viral proteins containing the purine NTP-binding sequence pattern. *Nucl Acids Res* **17** (21) 8413-8440

Gorbalenya A E and Koonin E V (1990) Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. *J Mol Biol* **213** 583-591

Gordenin D A, Malkova A L, Peterzen A, Kulikov V N, Pavlov Y I, Perkins E and Resnick M A (1992) Transposon Tn5 excision in yeast: influence of DNA polymerases alpha, delta, and epsilon and repair genes. *Proc Natl Acad Sci U S A* **89** 3785-3789

Gottlieb P A, Wu S, Zhang X, Tecklenburg M, Kuempel P and Hill T M (1992) Equilibrium, kinetic, and footprinting studies of the Tus-Ter protein-DNA interaction. *J Biol Chem* **267** 7434-7443

Gram H and Rüger W (1985) Genes *55, agt, 47 and 46* of bacteriophage T4: the genomic organization as deduced by sequence analysis. *EMBO J* **4** (1) 257-264

Haber J E (1998) The many interfaces of Mre11. *Cell* **95** 583-586

Harlow E and Lane D (1988) *Antibodies: A Laboratory Manual.* Cold Spring Harbor Laboratory Press

Hendrix R W, Roberts J W, Stahl F W and Weisberg R A (1983) Appendix II, In *Lambda II.* Cold Spring Harbor Laboratory Press

Higgins C F (1986) The regulation of gene expression in *Escherichia coli* and its bacteriophage. In *Genetic Engineering* 5 Ed Rigby P W J, Academic Press London

Hill K E, Dasouki M, Phillips J A r and Burk R F (1996) Human selenoprotein P gene maps to 5q31. *Genomics* **36** 550-551

Hiom K and Gellert M (1997) A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. *Cell* **88** 65-72

Hiom K and West S C (1995a) Branch migration during homologous recombination: assembly of a RuvAB- Holliday junction complex *in vitro. Cell* **80** 787-793

Hiom K and West S C (1995b) Characterisation of RuvAB-Holliday junction complexes by glycerol gradient sedimentation. *Nucleic Acids Res* **23** 3621-3626

Hiom K, Tsaneva I R and West S C (1996) The directionality of RuvAB-mediated branch migration: *in vitro* studies with three-armed junctions. *Genes Cells* **1** 443-451

Hirano T (1995) Biochemical and genetic dissection of mitotic chromosome condensation. *TIBS* **20** 357-361

Hirano T and Mitchison T J (1994) A heterodimeric coiled-coil protein required for mitotic chromosome condensation *in vitro. Cell* **79** 449-458

Hirano T, Mitchison T J and Swedlow J R (1995) The SMC family: from chromosome condensation to dosage compensation. *Curr Op Cell Biol* **7** 329-336

Horii Z I and Clark A J (1973) Genetic Analysis of the *recF* pathway to genetic recombination in *Escherichia coli* K-12: isolation and characterization of mutants. *J Molec Biol* **80** 327-344

Hosoda J, Mathew E and Jansen B (1971) Role of genes 46 and 47 in bacteriophage T4 replication: I *In vivo* deoxyribonucleic acid replication. *J Virol* **8** (4) 372-387

Hsu T and Karam J D (1990) Transcriptional mapping of a DNA replication gene cluster in bacteriophage T4: Sites for initiation, termination and mRNA processing. *J Biol Chem* **265** (9) 5303-5316

Hubbard M J and Cohen P (1993) On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem* **18** (5) 172-177

Hulton C S J, Seirafi A, Hinton J C D, Sidebotham J M, Waddell L, Pavitt G D, Owen-Hughes T, Spassky A, Buc H and Higgins C F (1990) Histone-like protein H1 (H-NS), DNA supercoiling and gene expression in bacteria. *Cell* **63** 631-642

Hulton C S, Higgins C F and Sharp P M (1991) ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. *Mol Microbiol* **5** 825-834

Huxley H E (1998) Getting to grips with contraction: the interplay of structure and biochemistry. *Trends Biochem* 23 (2) 84-87

Hyde H, Davies A A, Benson F E and West S C (1994) Resolution of recombination intermediates by a mammalian activity functionally analogous to *Escherichia coli* RuvC resolvase. *J Biol Chem* **269** 5202-5209

Ivanov E L, Korolev V G and Fabre F (1992) *XRS2*, a DNA repair gene of *Saccharomyces cerevisiae*, is needed for meiotic recombination. *Genetics* **132** 651-664

Ivanov E L, Sugawara N, Fishman-Lobell J and Haber J E (1996) Genetic requirements for the single-strand annealing pathway of double- strand break repair in *Saccharomyces cerevisiae. Genetics* **142** 693-704

Ivanov E L, Sugawara N, White C I, Fabre F and Haber J E (1994) Mutations in *XRS2* and *RAD50* delay but do not prevent mating-type switching in *Saccharomyces cerevisiae. Mol Cell Biol* **14** 3414-3425

Jenny T F, Gerloff D L, Cohen M A and Benner S A (1995) Predicted secondary and supersecondary structure for the serine- threonine-specific protein phosphatase family. *Proteins* **21** 1-10

Jessberger R, Frei C and Gasser S (1998) Chromosome dynamics: the SMC protein family. *Curr Op Genet Dev* **8** 254-259

Jessberger R, Riwar B, Baechtold H and Akhmedov A T (1996) SMC proteins constitute two subunits of the mammalian recombination complex RC-1. *EMBO J* **15** 4061-4068

Jezewska M J, Rajendran S, Bujalowska D and Bujalowski W (1998) Does single-stranded DNA pass through the inner channel of the protein hexamer in the complex with the *Escherichia coli* DnaB helicase? Fluoresence energy transfer studies. *J Biol Chem* **273** (17) 10515-10529

Johzuka K and Ogawa H (1995) Interaction of Mre11 and Rad50: two proteins required for DNA repair and meiosis-specific double-strand break formation in *Saccharomyces cerevisiae. Genetics* **139** 1521-1532

Kabotyanski E B, Zhu C, Kallick D A and Roth D B (1995) Hairpin opening by single-strand-specific nucleases. *Nucleic Acids Res* **23** 3872-3881

Kaliman A V, Kryukov V M and Bayev A A (1988) The nucleotide sequence of the region of bacteriophage T5 early genes D10-D15. *Nucl Acids Res* **16** (21) 10353-10354

Karu A E, Sakaki Y, Echols H and Linn S (1975) The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of *Escherichia coli*. *J Biol Chem* **250** 7377-7387

Kashina A S, Baskin R J, Cole D G, Wedaman K P, Saxton W M and Scholey J M (1996) A bipolar kinesin. *Nature* **379** 270-272

Keeney S and Kleckner N (1995) Covalent protein-DNA complexes at the 5' strand termini of meiosis- specific double-strand breaks in yeast. *Proc Natl Acad Sci U S A* **92** 11274-11278

Keeney S, Giroux C N and Kleckner N (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88** 375-384

Kennedy A K, Guhathakurta A, Kleckner N and Haniford D B (1998) Tn*10* transposition via a DNA hairpin intermediate. *Cell* **95** 125-134

Kim K K, Daud A I, Wong S C, Pajak L, Tsai S C, Wang H, Henzel W J and Field L J (1996) Mouse RAD50 has limited epitopic homology to p53 and is expressed in the adult myocardium. *J Biol Chem* **271** 29255-29264

Kimura K and Hirano T (1997) ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation. *Cell* **90** 625-634

Kironmai K M and Muniyappa K (1997) Alteration of telomeric sequences and senescence caused by mutations in *RAD50* of *Saccharomyces cerevisiae*. *Genes Cells* **2** 443-455

Kirtikar D M, Cathcart G R, White J G, Ukstins I and Goldthwait D A (1977) Mutations in *Escherichia coli* altering an apurinic endonuclease, endonuclease II, and exonuclease III and their effect on *in vivo* sensitivity to methylmethanesulfonate. *Biochemistry* **16** 5625-5631

Kleff S, Kemper B and Sternglanz R (1992) Identification and characterization of yeast mutants and the gene for a cruciform cutting endonuclease. *EMBO J* **11** 699-704

Kneale G G (1994) *Methods in Molecular Biology* **30** *DNA-Protein Interactions: Principles and Protocols.* Ed Kneale G G, Humana Press Inc Clifton New Jersey

Kogoma T and Lark K G (1975) Characterisation of the replication of *Escherichia coli* DNA in the absence of protein synthesis: Stable DNA replication. *J Mol Biol* **94** 243-256

Koonin E V (1994) Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Sci* **3** 356-358

Korn A H, Feairheller S H and Filachione E M (1972) Glutaraldehyde: nature of the reagent. *J Mol Biol* **65** 525-529

Koshland D and Strunnikov A (1996) Mitotic Chromosome Condensation. *Annu Rev Cell Dev Biol* **12** 305-333

Kramer P R, Stringer J R and Sinden R R (1996) Stability of an inverted repeat in a human fibrosarcoma cell. *Nucleic Acids Res* **24** 4234-4241

Kreuzer K N, Saunders M, Weislo L J and Kreuzer H W (1995) Recombination-dependent DNA replication stimulated by double-strand breaks in bacteriophage T4. *J Bacteriol* **177** 6844-6853

Kulkarni S K and Stahl F W (1989) Interaction between the *sbcC* gene of *Escherichia coli* and the *gam* gene of phage lambda. *Genetics* **123** 249-253

Kutter E M and Wiberg J S (1968) Degradation of cytosin-containing bacteria and bacteriophage DNA after infection of *Escherichia coli* B with bacteriophage T4D wildtype and with mutants defective in genes 46, 47 and 56. *J Mol Biol* **38** (3) 395-411

Kuzminov A (1995a) Collapse and repair of replication forks in *Escherichia coli*. *Mol Microbiol* **16** 373-384

Kuzminov A (1995b) Instability of inhibited replication forks in *E. coli*. *Bioessays* **17** 733-741

Kuzminov A, Schabtach E and Stahl F W (1994) Chi sites in combination with RecA protein increase the survival of linear DNA in *Escherichia coli* by inactivating exoV activity of RecBCD nuclease. *EMBO J* **13** 2764-2776

Larionov V L, Karpova T S, Kouprina N Y and Jouravlena G A (1985) A mutant of *Saccharomyces cerevisiae* with impaired maintenance of centromeric plasmids. *Curr Genet* **10** (1) 15-20

Le Beau M M, Albain K S, Larson R A, Vardiman J W, Davis E M, Blough R R, Golomb H M and Rowley J D (1986) Clinical and cytogenetic correlations in 63 patients with therapy-related myelodysplastic syndromes and acute non lymphocytic leukemias: further evidence for characteristic abnormalities of chromosomes 5 and 7. *J Clin Oncol* **4** (3) 325-345

Leach D and Lindsey J (1986) *In vivo* loss of supercoiled DNA carrying a palindromic sequence. *Mol Gen Genet* **204** 322-327

Leach D R (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* **16** 893-900

Leach D R and Stahl F W (1983) Viability of lambda phages carrying a perfect palindrome in the absence of recombination nucleases. *Nature* **305** 448-451

Leach D R F (1996a) *Genetic Recombination*. Blackwell Science Ltd

Leach D R F (1996b) Cloning and characterisation of DNAs with palindromic sequences. In *Genetic Engineering* Ed Setlow J, Plenum Press New York

Leach D R F (1999) Molecular Processing of DNA Folding Anomalies in *Escherichia coli*. In *DNA recombination and repair: A practical approach* Ed Smith P and Jones C, Oxford University Press (in press)

Leach D R, Lloyd R G and Coulson A F (1992) The SbcCD protein of *Escherichia coli* is related to two putative nucleases in the UvrA superfamily of nucleotide-binding proteins. *Genetica* **87** 95-100

Leach D R, Okely E A and Pinder D J (1997) Repair by recombination of DNA containing a palindromic sequence. *Mol Microbiol* **26** 597-606

Leach D, Lindsey J and Okely E (1987) Genome interactions which influence DNA palindrome mediated instability and inviability in *Escherichia coli*. *J Cell Sci Suppl* 7 33-40

Lee S E, Moore J K, Holmes A, Umezu K, Kolodner R D and Haber J E (1998) Saccharomyces Ku70, Mre11/Rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell* **94** 399-409

Lehman I R and Nussbaum A L (1964) The deoxyribonucleases of *Escherichia coli* V: On the specificity of exonuclease I (phosphodiesterase). *J Biol Chem* **239** (8) 2628-2636

Lehmann A R, Walicka M, Griffiths D J F, Murray J M, Watts F Z, McCready S and Carr A M (1995) The *rad18* gene of *Schizosaccharomyces pombe* defines a new subgroup of the SMC superfamily involved in DSB repair. *Mol Cell Biol* **15** 7067-7080

Lewis S M (1994) P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells. *Proc Natl Acad Sci U S A* **91** 1332-1336

Lilley D (1986) Bacterial chromatin. A new twist to an old story [news]. *Nature* **320** 14-15

Lilley D M J (1981) *In vivo* consequences of plasmid topology. *Nature* **292** (5821) 380-382

Lin F L, Sperle K and Steinberg N (1984) Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Mol Cell Biol* 4 (6) 1020-1034

Lindsey J C and Leach D R (1989) Slow replication of palindrome-containing DNA. *J Mol Biol* **206** 779-782

Lloyd R G (1991) Linkage distortion following conjugational transfer of *sbcC+* to *recBC sbcBC* strains of *Escherichia coli*. *J Bacteriol* **173** 5694-5698

Lloyd R G and Buckman C (1985) Identification and genetic analysis of *sbcC* mutations in commonly used *recBC sbcB* strains of *Escherichia coli K-12*. *J Bacteriol* **164** 836-844

Lloyd R G and Sharples G J (1992) Genetic analysis of recombination in prokaryotes. *Curr Opin Genet Dev* **2** 683-690

Lloyd R G and Sharples G J (1993) Processing of recombination intermediates by the RecG and RuvAB proteins of *Escherichia coli*. *Nucleic Acids Res* **21** 1719-1725

Loidl J, Klein F and Scherthan H (1994) Homologous pairing is reduced but not abolished in asynaptic mutants of yeast. *J Cell Biol* **125** 1191-1200

Lovett S T and Clark A J (1984) Genetic analysis of the *recJ* gene of *Escherichia coli* K-12. *J Bacteriol* **157** 190-196

Luder A and Mosig G (1982) Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: Priming by RNA polymerase and by recombination. *Proc Natl Acad Sci USA* **79** 1101-1105

Lupas A, Van Dyke M and Stock J (1991) Predicting coiled coils from protein sequences. *Science* **252** 1162-1164

Lynch H T and Smyrk T (1998) An update on Lynch Syndrome. *Curr Op Oncol* **10** 349-356

Malagón F and Aguilera A (1998) Genetic stability and DNA rearrangements associated with a 2 x 1.1kb perfect palindrome in *Escherichia coli*. *Mol Gen Genet* **259** 639-644

Malkova A, Ross L, Dawson D, Hoekstra M F and Haber J E (1996) Meiotic recombination initiated by a double-strand break in *rad50 delta* yeast cells otherwise unable to initiate meiotic recombination. *Genetics* **143** 741-754

Malone R E, Ward T, Lin S and Waring J (1990) The *RAD50* gene, a member of the double strand break repair epistasis group, is not required for spontaneous mitotic recombination in yeast. *Curr Genet* **18** 111-116

Marrione P E and Cox M M (1995) RuvB protein-mediated ATP hydrolysis: functional asymmetry in the RuvB hexamer. *Biochemistry* **34** 9809-9818

Marrione P E and Cox M M (1996) Allosteric effects of RuvA protein, ATP, and DNA on RuvB protein- mediated ATP hydrolysis. *Biochemistry* **35** 11228-11238

Maser R S, Monsen K J, Nelms B E and Petrini J H (1997) hMre11 and hRad50 nuclear foci are induced during the normal cellular response to DNA double-strand breaks. *Mol Cell Biol* **17** 6087-6096

Matson S W, Bean D W and George J W (1994) DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays* **16** 13-22

Maxam A M and Gilbert W (1980) Sequencing end-labelled DNA with base-specific chemical changes. *Methods Enzym* **65** 499-561

McBlane J F, van Gent DC, Ramsden D A, Romeo C, Cuomo C A, Gellert M and Oettinger M A (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83** 387-395

McKee A H and Kleckner N (1997a) A general method for identifying recessive diploid-specific mutations in *Saccharomyces cerevisiae*, its application to the isolation of mutants blocked at intermediate stages of meiotic prophase and characterization of a new gene *SAE2*. *Genetics* **146** 797-816

McKee A H and Kleckner N (1997b) Mutations in *Saccharomyces cerevisiae* that block meiotic prophase chromosome metabolism and confer cell cycle arrest at pachytene identify two new meiosis-specific genes *SAE1* and *SAE3*. *Genetics* **146** 817-834

Medina N, Vannier F, Roche B, Autret S, Levine A and Seror S J (1997) Sequencing of regions downstream of *addA* (98 degrees) and *citG* (289 degrees) in *Bacillus subtilis*. *Microbiology* **143** 3305-3308

Melby T E, Ciampaglio C N, Briscoe G and Erickson H P (1998) The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge [In Process Citation]. *J Cell Biol* **142** 1595-1604

Mendonca V M, Klepin H D and Matson S W (1995) DNA helicases in recombination and repair: construction of a *delta uvrD delta helD delta recQ* mutant deficient in recombination and repair. *J Bacteriol* **177** 1326-1335

Menetski J P, Varghese A and Kowalczykowski S C (1992) The physical and enzymatic properties of *Escherichia coli* RecA protein display anion-specific inhibition. *J Biol Chem* **267** 10400-10404

Meselson M S and Radding C M (1975) A general model for genetic recombination. *Proc Natl Acad Sci USA* **72** 385-361

Meyer R R and Laine P S (1990) The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev* **54** 342-380

Michaelis C, Ciosk R and Nasmyth K (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91** (1) 35-45

Michel B, Ehrlich S D and Uzest M (1997) DNA double-strand breaks caused by replication arrest. *EMBO J* **16** 430-438

Milligan J F and Uhlenbeck O C (1989) Determination of RNA-protein contacts using thiophosphate substitutions. *Biochemistry* **28** 2849-2855

Milne G T, Jin S, Shannon K B and Weaver D T (1996) Mutations in two Ku homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. *Mol Cell Biol* **16** 4189-4198

Mitas M, Yu A, Dill J and Haworth I S (1995) The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry* **34** 12803-12811

Mitchell A H and West S C (1994) Hexameric rings of *Escherichia coli* RuvB protein. Cooperative assembly, processivity and ATPase activity. *J Mol Biol* **243** 208-215

Mizuuchi K, Mizuuchi M and Gellert M (1982) Cruciform structures in palindromic DNA are favored by DNA supercoiling. *J Mol Biol* **156** 229-243

Mizuuchi M, Baker T A and Mizuuchi K (1991) DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proc Natl Acad Sci U S A* **88** 9031-9035

Modrich P (1995) Mismatch repair, genetic stability and tumour avoidance. *Philos Trans R Soc Lond B Biol Sci* **347** 89-95

Moore J K and Haber J E (1996) Cell cycle and genetic requirements of two pathways of non homologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* **16** 2164-2173

Moreau S, Ferguson J R and Symington L S (1999) The nuclease activity of mre11 is required for meiosis but not for mating type switching, end-joining or telomere maintenance. *Mol Cell Biol* **19** (1) 556-566

Mueller J E, Newton C J, Jensch F, Kemper B, Cunningham R P, Kallenbach N R and Seeman N C (1990) Resolution of Holliday junction analogs by T4 endonuclease VII can be directed by substrate structure. *J Biol Chem* **265** 13918-13924

Myers R S and Stahl F W (1994) Chi and the RecBCD enzyme of *Escherichia coli*. *Annu Rev Genet* **28** 49-70

Nag D K and Kurst A (1997) A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast *Saccharomyces cerevisiae*. *Genetics* **146** 835-847

Nairz K and Klein F (1997) *mre11S*--a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes Dev* **11** 2272-2290

Naom I S, Morton S J, Leach D R and Lloyd R G (1989) Molecular organization of *sbcC*, a gene that affects genetic recombination and the viability of DNA palindromes in *Escherichia coli K-12*. *Nucleic Acids Res* **17** 8033-8045

Neece S H, Carles-Kinch K, Tomso D J and Kreuzer K N (1996) Role of recombinational repair in sensitivity to an antitumour agent that inhibits bacteriophage T4 type II DNA topoisomerase. *Mol Microbiol* **20** 1145-1154

Niki H, Imamura R, Kitaoka M, Yamanaka K, Ogura T and Hiraga S (1992) *E. coli* MukB protein involved in chromosome partition forms a homodimer with a rod-and-hinge structure having DNA binding and ATP/GTP binding activities. *EMBO J* **11** 5101-5109

Nugent C I, Bosco G, Ross L O, Evans S K, Salinger A P, Moore J K, Haber J E and Lundblad V (1998) Telomere maintenance is dependent on activities required for end repair of double-strand breaks [In Process Citation]. *Curr Biol* **8** 657-660

O'Neill M C (1989) *Escherichia coli* promoters. I. Consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organization. *J Biol Chem* **264** 5522-5530

Ohta K, Nicolas A, Furuse M, Nabetani A, Ogawa H and Shibata T (1998) Mutations in the *MRE11, RAD50, XRS2,* and *MRE2* genes alter chromatin configuration at meiotic DNA double-stranded break sites in premeiotic and meiotic cells. *Proc Natl Acad Sci U S A* **95** 646-651

Ohta K, Shibata T and Nicolas A (1994) Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J* **13** (23) 5754-5763

Okazaki R, Arisawa M and Sugino A (1971) Slow joining of newly replicated DNA chains in DNA polymerase I- deficient *Escherichia coli* mutants. *Proc Natl Acad Sci U S A* **68** 2954-2957

Okazaki R, Okazaki T, Sakabe K, Sugimoto K and Sugino A (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A* **59** 598-605

Padmore R, Cao L and Kleckner N (1991) Temporal comparison of recombination and synaptonemal complex formation during meiosis in *S. cerevisiae*. *Cell* **66** 1239-1256

Papavassiliou A G (1994) Phenanthroline-copper footprinting, In *Methods in Molecular Biology* **30** *DNA-Protein Interactions: Principles and Protocols.* Ed Kneale G G, Humana Press Inc Clifton New Jersey

Parsons C A and West S C (1993) Formation of a RuvAB-Holliday Junction Complex *in vitro*. *J Mol Biol* **232** 397-405

Paull T T and Gellert M (1998) The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol Cell* **1** 969-979

Pearson C E, Ewel A, Acharya S, Fishel R A and Sinden R R (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum Mol Genet* **6** 1117-1123

Peterson C L (1994) The SMC family: Novel motor proteins for chromosome condensation? *Cell* **79** 389-392

Petrini J H, Walsh M E, DiMare C, Chen X N, Korenberg J R and Weaver D T (1995) Isolation and characterization of the human MRE11 homologue. *Genomics* **29** 80-86

Petruska J, Arnheim N and Goodman M F (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res* **24** 1992-1998

Picksley S M, Parsons C A, Kemper B and West S C (1990) Cleavage specificity of bacteriophage T4 endonuclease VII and bacteriophage T7 endonuclease I on synthetic branch migratable Holliday junctions. *J Mol Biol* **212** 723-735

Pinder D J (1996) *Illegitimate recombination in plasmids.* PhD thesis, University of Edinburgh

Pinder D J, Blake C E and Leach D R F (1996) DIR: A novel DNA rearrangement associated with IRs. *Nucl Acids Res* **25** (3) 523-529

Pottmeyer S and Kemper B (1992) T4 endonuclease VII resolves cruciform DNA with nick and counter-nick and its activity is directed by local nucleotide sequence. *J Mol Biol* **223** 607-615

Powell L M, Dryden D T F, Willcock D F, Pain R H and Murray N E (1993) DNA recognition by the *Eco*K methyltransferase: The influence of DNA methylation and the cofactor *S*-adenosyl-L-methionine. *J Mol Biol* **234** 60-71

Prashad N and Hosoda J (1972) Role of genes *46* and *47* in bacteriophage T4 reproduction. II. Formation of gaps on parental DNA of polynucleotide ligase defective mutants. *J Mol Biol* **70** 617-635

Radman M, Matic I, Halliday J A and Taddei F (1995) Editing DNA replication and recombination by mismatch repair: from bacterial genetics to mechanisms of predisposition to cancer in humans. *Philos Trans R Soc Lond B Biol Sci* **347** 97-103

Rao B J and Radding C M (1995) RecA protein mediates homologous recognition via non-Watson-Crick bonds in base triplets. *Philos Trans R Soc Lond B Biol Sci* **347** 5-12

Raymond W E and Kleckner N (1993a) RAD50 protein of *S.cerevisiae* exhibits ATP-dependent DNA binding. *Nucleic Acids Res* **21** 3851-3856

Raymond W E and Kleckner N (1993b) Expression of the *Saccharomyces cerevisiae RAD50* gene during meiosis: steady-state transcript levels rise and fall while steady-state protein levels remain constant. *Mol Gen Genet* **238** 390-400

Roman L J and Kowalczykowski S C (1989) Characterization of the helicase activity of the *Escherichia coli* RecBCD enzyme using a novel helicase assay. *Biochemistry* **28** 2863-2873

Rosche W A, Trinh T Q and Sinden R R (1995) Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. *J Bacteriol* **177** 4385-4391

Roth D B and Wilson J H (1985) Relative rates of homologous and non homologous recombination in transfected DNA. *Proc Natl Acad Sci* **82** (10) 3355-3359

Rowland S J, Sherratt D J, Stark W M and Boocock M R (1995) Tn*552* transposase purification and *in vitro* activities. *EMBO J* **14** 196-205

Ryder L, Sharples G J and Lloyd R G (1996) Recombination-dependent growth in exonuclease-depleted *recBC sbcBC* strains of *Escherichia coli* K-12. *Genetics* **143** 1101-1114

Saitoh N, Goldberg I and Earnshaw W C (1995) The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *BioEssays* **17** (9) 759-766

Saitoh N, Goldberg I, Wood E and Earnshaw W C (1994) ScII: an abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. *J Cell Biol* **127** 303-318

Saka Y, Sutani T, Yamashita Y, Saitoh S, Takeuchi M, Nakaseko Y and Yanagida M (1994) Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis. *EMBO J* **13** (20) 4938-4952

Sambrook J, Frisch E F and Maniatis T (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press

Sarkar P S, Chang H, Boudi F B and Reddy S (1998) CTG repeats show bimodal amplification in *E. coli. Cell* **95** 531-540

Savilahti H, Rice P A and Mizuuchi K (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J* **14** 4893-4903

Sawitzke J A and Stahl F W (1994) The phage lambda *orf* gene encodes a trans-acting factor that suppresses *Escherichia coli recO, recR,* and *recF* mutations for recombination of lambda but not of *E. coli. J Bacteriol* **176** 6730-6737

Schmidt K H, Abbott C and Leach D R F (1999) Opposing effects of the mismatch repair protein MutS on instability of trinucleotide repeats. (submitted for publication)

Seigneur M, Bidnenko V, Ehrlich S D and Michel B (1998) RuvAB acts at arrested replication forks. *Cell* **95** 419-430

Sharples G J and Leach D R (1995) Structural and functional similarities between the SbcCD proteins of Escherichia coli and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast [letter]. *Mol Microbiol* **17** 1215-1217

Sharples G J and Lloyd R G (1990) A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res* **18** 6503-6508

Sharples G J and Lloyd R G (1993) Location of the *Bacillus subtilis sbcD* gene downstream of *addAB*, the analogues of *E. coli recBC*. *Nucleic Acids Res* **21** 2010

Shinohara A and Ogawa T (1995) Homologous recombination and the roles of double-strand breaks. *Trends Biochem Sci* **20** 387-391

Shlyakhtenko L S, Potaman V N, Sinden R R and Lyubchenko Y L (1998) Structure and dynamics of supercoil-stabilized DNA cruciforms. *JMB* **280** 61-72

Shurvinton C E, Stahl M M and Stahl F W (1987) Large palindromes in the lambda phage genome are preserved in a *rec+* host by inhibiting lambda DNA replication. *Proc Natl Acad Sci U S A* **84** 1624-1628

Sidorova N Y and Rau D C (1996) Differences in water release for the binding of *Eco*RI to specific and nonspecific DNA sequences. *Proc Natl Acad Sci U S A* **93** 12272-12277

Silberstein Z and Cohen A (1987) Synthesis of linear multimers of OriC and pBR322 derivatives in *Escherichia coli* K-12: role of recombination and replication functions. *J Bacteriol* **169** 3131-3137

Sinden R R and Pettijohn D E (1984) Cruciform transitions in DNA. *J Biol Chem* **259** 6593-6600

Sinden R R, Broyles S S and Pettijohn D E (1983) Perfect palindromic *lac* operator DNA sequence exists as a stable cruciform structure in supercoiled DNA *in vitro* but not *in vivo*. *Proc Natl Acad Sci U S A* **80** 1797-1801

Smith G R, Amundsen S K, Dabert P and Taylor A F (1995) The initiation and control of homologous recombination in *Escherichia coli*. *Philos Trans R Soc Lond B Biol Sci* **347** 13-20

Solaro P C, Birkenkamp K, Pfeiffer P and Kemper B (1993) Endonuclease VII of phage T4 triggers mismatch correction in vitro. *J Mol Biol* **230** 868-877

Solaro P, Greger B and Kemper B (1995) Detection and partial purification of a cruciform-resolving activity (X- solvase) from nuclear extracts of mouse B-cells. *Eur J Biochem* **230** 926-933

Stern M J, Ames G F L, Smith N H, Robinson E C and Higgins C F (1984) Repetetive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37** (3) 1015-1026

Strauss H S, Boston R S, Record M T, Jr. and Burgess R R (1981) Variables affecting the selectivity and efficiency of retention of DNA fragments by *E. coli* RNA polymerase in the nitrocellulose-filter- binding assay. *Gene* **13** 75-87

Strunnikov A V, Hogan E and Koshland D (1995) SMC2, a *Saccharomyces cerevisiae* gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. *Genes Dev* **9** 587-599

Strunnikov A V, Larionov V L and Koshland D (1993) SMC1: An essential yeast gene encoding a putative Head-Rod-Tail protein is required for nuclear division and defines a new ubiquitous protein family. *J Cell Biol* **123** (6, ii) 1635-1648

Suen I, Rhodes J N, Christy M, McEwen B, Gray D M and Mitas M (1999) Structural properties of Friedreich's ataxia d(GAA) repeats. *Biochem Biophys Acta* **1444** 14-24

Sugawara N and Haber J E (1992) Characteristics of double-strand break-induced recombination: homology requirements and single-strand DNA formation. *Mol Cell Biol* **12** (2) 563-575

Sugawara N, Paques F, Colaiacovo M and Haber J E (1997) Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *Proc Natl Acad Sci U S A* **94** 9214-9219

Summers D K (1996) Plasmid Inheritance, In *The Biology of Plasmids*. Blackwell Science Ltd

Sutani T and Yanagida M (1997) DNA renaturation activity of the SMC complex implicated in chromosome condensation. *Nature* **388** 798-801

Szostak J W, Orr-Weaver T L, Rothstein R J and Stahl F W (1983) The double-strand break repair model for conversion and crossing over. *Cell* **33** 25-35

Tavassoli M, Shayeghi M, Nasim A and Watts F Z (1995) Cloning and characterisation of the *Schizosaccharomyces pombe rad32* gene: a gene required for repair of double-strand breaks and recombination. *Nucleic Acids Res* **23** 383-388

Taylor A F and Smith G R (1995) Monomeric RecBCD enzyme binds and unwinds DNA. *J Biol Chem* **270** 24451-24458

Taylor J D, Ackroyd A J and Halford S E (1994) The gel shift assay for the analysis of DNA-protein interactions, In *Methods in Molecular Biology* **30** *DNA-Protein Interactions: Principles and Protocols.* Ed Kneale G G, Humana Press Inc Clifton New Jersey

Thaler D S, Stahl M M and Stahl F W (1987) Tests of the double-strand break repair model for Red-mediated recombination of phage λ and plasmid λ dv. *Genetics* **116** 501-511

Thoms B and Wackernagel W (1998) Interaction of RecBCD enzyme with DNA at double-strand breaks produced in UV-irradiated *Escherichia coli*: Requirement for DNA end-processing. *J Bacteriol* **180** (21) 5639-5645

Tran H T, Degtyareva N P, Koloteva N N, Sugino A, Masumoto H, Gordenin D A and Resnick M A (1995) Replication slippage between distant short repeats in *Saccharomyces cerevisiae* depends on the direction of replication and the *RAD50* and *RAD52* genes. *Mol Cell Biol* **15** 5607-5617

Trinh T Q and Sinden R R (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352** 544-547

Trujillo K M, Yuan S S, Lee E Y and Sung P (1998) Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J Biol Chem* **273** 21447-21450

Tsubouchi H and Ogawa H (1998) A novel *mre11* mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. *Mol Cell Biol* **18** 260-268

Tsukamoto Y and Ikeda H (1998) Double-strand break repair mediated by DNA end-joining [In Process Citation]. *Genes Cells* **3** 135-144

Tsukamoto Y, Kato J and Ikeda H (1996a) Hdf1, a yeast Ku-protein homologue, is involved in illegitimate recombination, but not in homologous recombination. *Nucleic Acids Res* **24** 2067-2072

Tsukamoto Y, Kato J and Ikeda H (1996b) Effects of mutations of *RAD50, RAD51, RAD52,* and related genes on illegitimate recombination in *Saccharomyces cerevisiae. Genetics* **142** 383-391

Tsukamoto Y, Kato J and Ikeda H (1997) Budding yeast Rad50, Mre11, Xrs2, and Hdf1, but not Rad52, are involved in the formation of deletions on a dicentric plasmid. *Mol Gen Genet* **255** 543-547

Usui T, Ohta T, Oshiumi H, Tomizawa J, Ogawa H and Ogawa T (1998) Complex formation and functional versatility of Mre11 of budding yeast in recombination. *Cell* **95** 705-716

van Holde K and Zlatanova J (1994) Unusual DNA structures, chromatin and transcription. *Bioessays* **16** 59-68

Villarroya M, Pérez-Roger I, Macián F and Armengod M E (1998) Stationary phase induction of *dnaN* and *recF*, two genes of *Escherichia coli* involved in DNA replication and repair. *EMBO J* **17** (6) 1829-1837

Walker J E, Saraste M, Runswick M J and Gay N J (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1** 945-951

Walton K M and Dixon J E (1993) Protein tyrosine phosphatases. *Ann Rev Biochem* **62** 101-120

Wang T C and Chen S H (1994) Okazaki DNA fragments contain equal amounts of lagging-strand and leading-strand sequences. *Biochem Biophys Res Commun* **198** 844-849

Wang T C and Smith K C (1985) Mechanism of *sbcB*-suppression of the *recBC*-deficiency in postreplication repair in UV-irradiated *Escherichia coli* K-12. *Mol Gen Genet* **201** 186-191

Warren G J and Green R L (1985) Comparison of physical and genetic properties of palindromic DNA sequences. *J Bacteriol* **161** 1103-1111

Weaver D T (1995) What to do at an end: DNA double-strand-break repair. *Trends Genet* **11** 388-392

Weaver D T and DePamphilis M L (1984) The role of palindromic and non-palindromic sequences in arresting DNA synthesis *in vitro* and *in vivo*. *J Mol Biol* **180** (4) 961-986

Weiner B M and Kleckner N (1994) Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. *Cell* **77** 977-991

Whitby M C, Bolt E L, Chan S N and Lloyd R G (1996) Interactions between RuvA and RuvC at Holliday junctions: inhibition of junction cleavage and formation of a RuvA-RuvC-DNA complex. *J Mol Biol* **264** 878-890

Whitby M C, Ryder L and Lloyd R G (1993) Reverse branch migration of Holliday junctions by RecG protein: a new mechanism for resolution of intermediates in recombination and DNA repair. *Cell* **75** 341-350

Whitby M C, Vincent S D and Lloyd R G (1994) Branch migration of Holliday junctions: identification of RecG protein as a junction specific DNA helicase. *EMBO J* **13** 5220-5228

Woodworth D L and Kreuzer K N (1996) Bacteriophage T4 mutants hypersensitive to an antitumor agent that induces topoisomerase-DNA cleavage complexes. *Genetics* **143** 1081-1090

Wu T C and Lichten M (1995) Factors that affect the location and frequency of meiosis-induced double-strand breaks in *Saccharomyces cerevisiae*. *Genetics* **140** 55-66

Xiao W, Chow B L and Rathgeber L (1996) The repair of DNA methylation damage in *Saccharomyces cerevisiae*. *Curr Genet* **30** 461-468

Xiao Y and Weaver D T (1997) Conditional gene targeted deletion by Cre recombinase demonstrates the requirement for the double-strand break repair Mre11 protein in murine embryonic stem cells. *Nucleic Acids Res* **25** 2985-2991

Xodo L E, Manzini G, Quadrifoglio F, van der Marel G and van Boom J (1991) DNA hairpin loops in solution. Correlation between primary structure, thermostability and reactivity with single-strand-specific nuclease from mung bean. *Nucleic Acids Res* **19** 1505-1511

Xu L and Kleckner N (1995) Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. *EMBO J* **14** 5115-5128

Xu L, Weiner B M and Kleckner N (1997) Meiotic cells monitor the status of the interhomolog recombination complex. *Genes Dev* **11** 106-118

Yang S W and Nash H A (1994) Specific photocrosslinking of DNA-protein complexes: identification of contacts between integration host factor and its target DNA. *Proc Natl Acad Sci U S A* **91** 12183-12187

Yu X, West S C and Egelman E H (1997) Structure and subunit composition of the RuvAB-Holliday junction complex. *J Mol Biol* **266** 217-222

Zebala J A, Choi J and Barany F (1992) Characterization of steady-state, single-turnover and binding kinetics of the *Taq*I restriction endonuclease. *J Biol Chem* **267** (12) 8097-8105

Zhang J, Lee M H and Walker G C (1995) P-azidoiodoacetanilide, a new short photocrosslinker that has greater cysteine specificity than p-azidophenacyl bromide and p- azidobromoacetanilide. *Biochem Biophys Res Commun* **217** 1177-1184

Zhang L and Hermans J (1993) Molecular dynamics study of structure and stability of a model coiled coil. *Proteins* **16** 384-392

Zhou N E, Kay C M and Hodges R S (1992) Synthetic model proteins: the relative contribution of leucine residues at the nonequivalent positions of the 3-4 hydrophobic repeat to the stability of the two-stranded alpha-helical coiled-coil. *Biochemistry* **31** 5739-5746

Zhuo S, Clemens J C, Hakes D J, Barford D and Dixon J E (1993) Expression, purification, crystallization and biochemical characterization of a recombinant protein phosphatase. *J Biol Chem* **268** (24) 17754-17761

Zhuo S, Clemens J C, Stone R L and Dixon J E (1994) Mutational analysis of a Ser/Thr phosphatase. Identification of residues important in phosphoesterase substrate binding and catalysis. *J Biol Chem* **269** 26234-26238

APPENDIX 1

The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 7969–7974, July 1998
Biochemistry

The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA

(hairpin nuclease/palindromic DNA/Rad50-Mre11)

JOHN C. CONNELLY, LUCY A. KIRKHAM, AND DAVID R. F. LEACH*

Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JR, United Kingdom

Communicated by John R. Roth, University of Utah, Salt Lake City, UT, April 27, 1998 (received for review October 17, 1997)

ABSTRACT **Hairpin structures can inhibit DNA replication and are intermediates in certain recombination reactions. We have shown that the purified SbcCD protein of *Escherichia coli* cleaves a DNA hairpin. This cleavage does not require the presence of a free (3′ or 5′) DNA end and generates products with 3′-hydroxyl and 5′-phosphate termini. Electron microscopy of SbcCD has revealed the "head-rod-tail" structure predicted for the SMC (structural maintenance of chromosomes) family of proteins, of which SbcC is a member. This work provides evidence consistent with the proposal that SbcCD cleaves hairpin structures that halt the progress of the replication fork, allowing homologous recombination to restore DNA replication.**

Long DNA palindromes inhibit DNA replication and are unstable in the genomes of *Escherichia coli* (see ref. 1), *Bacillus* (2), *Streptococcus* (3), *Streptomyces* (4), *Saccharomyces cerevisiae* (5–8), mice (9, 10), and humans (see ref. 11). These effects result from the formation of hairpin or cruciform structures by intrastrand base pairing (1). In *E. coli*, the inhibition of DNA replication is overcome significantly in *sbcC* or *sbcD* mutants (12, 13).

The *sbcC* and *sbcD* genes have been cloned and sequenced (14). SbcC has the characteristic molecular organization predicted for a polypeptide belonging to the SMC (*s*tructural *m*aintenance of *c*hromosomes) family (15, 16). This family of polypeptides is implicated in chromosome condensation and segregation (17–21), transcriptional repression (22), and genetic recombination (15, 23, 24). Members possess "Walker-A" and "Walker-B" nucleotide binding motifs (25) separated by an α-helical region (several hundred amino acids long) predicted to form two coiled-coil domains interrupted by a short spacer region (reviewed in refs. 16, 19, and 26). SbcC belongs to the subclass of SMC proteins implicated in genetic recombination that show more extensive sequence similarity in and around their ATP binding motifs and possess a conserved CXXC motif within their putative coiled-coil spacer region (15). This subclass includes the Rad50 polypeptide of *S. cerevisiae* (scRad50) (23), its human (27) and mouse (28) homologues (hRad50 and mRad50), and the product of bacteriophage T4 gene *46* (T4 gp46) (29). All of these polypeptides are involved in generating and/or processing DNA double-strand breaks. *In vivo*, scRad50 and hRad50 interact with the (sc or h) Mre11 polypeptide that also is involved in double-strand break repair (27, 30). The murine MRE11 gene has been shown to be essential for normal cell proliferation in embryonic stem cells (31). T4 gp46 has been shown genetically to interact with the product of gene *47* (gp47) (32). Both genes *46* and *47* are essential for recombination and DNA replication in T4 (33). ScMre11, hMre11 (34), and T4 gp47 are all putative nucleases that share sequence similarity with SbcD (15). They belong to a family of phosphoesterases, including various serine/threonine phosphatases, that contain the conserved sequence $DXH(X_{25})GDXXD(X_{25})GNHD/E$. This coupling of an SMC-family polypeptide to a phosphoesterase seems to be a common feature of these proteins involved in recombination and double-strand break repair.

SbcCD protein forms a large (1.2-MDa) complex that functions as an ATP-dependent double-strand DNA exonuclease and an ATP-independent single-strand endonuclease (35, 36). Density transfer and de-methylation studies have indicated that DNA replication is required before SbcCD can recognize DNA palindromes (37, 38). These observations have led to the proposal that SbcCD collapses replication forks by attacking hairpin structures that arise on the lagging-strand template and that broken replication forks are repaired by homologous recombination (1, 35). This proposal has been tested genetically by observing that a 240-bp palindrome introduced into the chromosome of *E. coli* causes cell death in *sbcCD*$^+$ cells that are either *recA*, *recB*, or *recC* mutants (39).

Here, we test the prediction that SbcCD might recognize and cleave hairpin DNA. We demonstrate that SbcCD protein cleaves a hairpin DNA substrate at the 5′ side of the loop to yield products with 5′ phosphate and 3′ hydroxyl ends. This reaction does not require DNA termini. In addition, we provide physical evidence that SbcC(D) forms the head-rod-tail structure expected of an SMC protein.

MATERIALS AND METHODS

DNA Substrates. To remove short chain termination products, all oligonucleotides (synthesized by Oswel DNA Service, University of Southampton) were run on 10% denaturing polyacrylamide gels and were visualized by UV shadowing (according to standard protocols; see ref. 40). DNA was recovered from excised bands by eluting at 60°C for 16 h in TE (10 mM Tris·HCl, pH 7.5/1 mM EDTA, pH 8.0) buffer, followed by ethanol precipitation. After ^{32}P-end-labeling, oligonucleotides once again were gel purified and recovered as described above.

Hairpin substrate. 5′HP78 (5′-GTTTCTATTCAGCCCTTTGACGTAATCCAGCCCCGGGTTTTCCCGGGGCTGGATTACGTCAAAGGGCTGAATAGAAAC-3′) is a 78-nt oligomer capable of forming the hairpin structure, with a 37-bp stem and 4-nt loop, shown in Fig. 2. 3′HP78 is a 77-nt version of 5′HP78 that lacks the 3′-terminal cytosine nucleotide and contains a 5′-terminal phosphate group. After 3′ end-labeling, 3′HP78 possesses exactly the same sequence as 5′HP78 (see

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: SMC, structural maintenance of chromosomes.

*To whom reprint requests should be addressed. e-mail: D.Leach@ed.ac.uk.

Fig. 2*B*). 5′HP78 was labeled (according to manufacturers' instructions) at the 5′ end with [γ-^{32}P]ATP (Amersham, 3,000 Ci/mmol) by using T4 polynucleotide kinase (Boehringer Mannheim) to yield a substrate with a specific activity of $\approx 6.7 \times 10^6$ cpm/μg total nucleic acid. 3′ end-labeled substrate was generated by filling in the 1-nt gap of 3′HP78 with [α-^{32}P]dCTP (Amersham, 3,000 Ci/mmol) by using the Klenow fragment of DNA polymerase I (New England Biolabs) to yield a substrate with a specific activity of $\approx 7.4 \times 10^6$ cpm/μg total nucleic acid. The hairpin-forming oligonucleotides used here are based on those described by Beyert *et al.* (41). Restriction digestion and native PAGE were used to confirm the presence of hairpin secondary structure (data not shown).

Covalently closed dumbbell substrate. ncDB98 (5′-GCTGAATAGAAACCTATCTGTGTGCCCGGGTTTTCCCGGGCACACAGATAGGTTTCTATTCAGCCCTTCCAGCCCCGGGTTTTCCCGGGGCTGGAAGG-3′) is a 98-nt oligomer capable of forming a nicked-circular dumbbell (ncDB98). It was labeled (according to manufacturers' instructions) at the 5′ end with [γ-^{32}P]ATP by using T4 polynucleotide kinase to yield an oligonucleotide with a specific activity of $\approx 1.5 \times 10^7$ cpm/μg total nucleic acid. Covalently closed DB98 (ccDB98; see Fig. 4) was obtained by treating 2 pmol of boiled and snap-cooled ^{32}P-labeled nicked-DB98 (ncDB98) with 400 units of T4 DNA ligase (New England Biolabs) at 37°C for 1 h in a 100-μl reaction volume. ccDB98 was separated from ncDB98 by denaturing gel electrophoresis. When ccDB98 is totally denatured it behaves as a single band (see Fig. 4*B*), with the slowly migrating characteristic of an open circle. Restriction digestion, chemical sequencing, and native PAGE were used to confirm that ccDB98 exists as a monomer–dumbbell (data not shown).

Analysis of Hairpin Cleavage Termini. All reactions were performed according to manufacturers' recommendations.

5′-termini. 3′HP78 (3′-^{32}P-labeled) DNA ($\approx$1.5 nM) and SbcCD (2.5 nM) were incubated in SbcCD assay buffer (160 μl) containing 1 mM ATPγS for 30 min at 16°C. The reaction was stopped by the addition of EDTA to 10 mM, and DNA was recovered by ethanol precipitation. After resuspending in 10 μl of TE buffer, the DNA products were denatured by heating to 100°C for 3 min and were split into two 5-μl aliquots. The aliquots were added to two separate tubes containing alkaline phosphatase buffer (total volume 20 μl). After incubating for 60 min at 37°C, in the presence or absence of 4 units of shrimp alkaline phosphatase (United States Biochemical), the ^{32}P-labeled products were analyzed by denaturing electrophoresis.

3′ termini. 5′HP78 (5′-^{32}P-labeled) DNA ($\approx$1.5 nM) and SbcCD (2.5 nM) were incubated in SbcCD assay buffer (80 μl) containing 1 mM ATPγS for 30 min at 16°C. The reaction was stopped by the addition of EDTA to 10 mM, and DNA was recovered by using a QIAquick nucleotide removal kit (Qiagen, Chatsworth, CA). After resuspending in 10 μl of 10 mM Tris·HCl (pH 7.5), the DNA products were denatured by heating to 100°C for 3 min and were split into two 5-μl aliquots. The aliquots were added to two separate tubes containing terminal transferase solution 1 (Boehringer Mannheim), 2.5 mM $CoCl_2$, and 0.5 μM ddATP. After a 15-min incubation at 37°C in the presence or absence of 25 units of terminal transferase (Boehringer Mannheim), ^{32}P-labeled products were analyzed by denaturing electrophoresis.

DNA Size Standards. SbcCD cleavage sites were confirmed by comparing SbcCD products with an A/G sequencing ladder, produced by the chemical degradation of 3′-labeled 3′HP78 or 5′-labeled 5′HP78 (42). Chemical cleavage sites were confirmed by using an 8–32 base oligonucleotide sizing marker (Pharmacia) and DNA molecular weight marker V (Boehringer Mannheim).

Gel Electrophoresis. Reactions were stopped by adding an equal volume of DNA loading buffer (1 mM EDTA, pH 8.0/95% formamide/0.01% bromophenol blue). Samples were heated to 100°C for 4 min before electrophoresis and were loaded as quickly as possible thereafter. To ensure conditions were as denaturing as possible, 10% polyacrylamide gels (0.4 mm thick) containing 7 M urea and 10% formamide were run on a Sequi-Gen Nucleic Acid Sequencing Cell (Bio-Rad) at $\approx$60°C for 1.5 h at 60 W using a TBE (89 mM Tris borate, pH 8.3/2 mM EDTA, pH 8.0) buffer system. DNA was visualized by autoradiography.

SbcCD Protein and Nuclease Assay. SbcCD protein was prepared as described (36). Protein concentrations were estimated by using a protein assay kit (Bio-Rad) with BSA as a standard. The molar concentration of SbcCD was calculated assuming a stoichiometry of $SbcC_6$:$SbcD_{12}$. In the SbcCD nuclease assay, 2.5 nM protein and DNA ($\approx$1.5 nM) were incubated for 30 min at the temperature indicated in a 20-μl reaction volume. The reactions were carried out in the presence of: 5 mM Mn^{2+}, 1.25 mM DTT, 2% glycerol, 100 μg/ml BSA, 25 mM Tris (pH 7.5), and 1 mM ATPγS (or 1 mM ATP). Reactions were stopped and analyzed on a 10% denaturing polyacrylamide gel before drying and autoradiography.

Electron Microscopy. Low angle rotary shadowing and unidirectional shadowing were performed as a service at the National Centre for Macromolecular Hydrodynamics, University of Leicester.

RESULTS

SbcCD Recognizes and Degrades a Hairpin Substrate. To test whether SbcCD was active on a hairpin substrate, a 3′-^{32}P-end-labeled oligonucleotide (3′HP78) capable of forming a DNA hairpin with a 37-bp stem and a 4-nt loop (illustrated in Fig. 2*A*) was incubated with protein over a varying time period. After short incubation times (1 min or less) at 37°C, four main early cleavage products, $\approx$41–44 nt in length, were seen (Fig. 1*A*, lanes b and c). Longer incubation times (5–30 min) resulted in the disappearance of these fragments and the appearance of products ranging in size from $\approx$9–26 nt (Fig. 1*A*, lanes d–f). This experiment indicates that SbcCD nuclease is active on a hairpin substrate and suggests that initial cleavage occurs 41–44 nt distal to the 3′ end of the hairpin, in the vicinity of the stem-loop junction. It appears that SbcCD possesses a hairpin cleaving activity and that initial hairpin-cleavage products subsequently are degraded by the processive ATP-dependent double-strand exonuclease activity of the protein.

Initial Cleavage Products Can Be Separated from Subsequent Products. Conditions were sought that would reduce (or inhibit) the processive double-strand exonuclease activity of SbcCD but retain hairpin cleavage activity. Protein was incubated with 3′HP78 (3′-end-labeled) hairpin oligonucleotide at 16°C in the presence of various nucleotide cofactors. No SbcCD activity was observed in the absence of nucleotide (Fig. 1*B*, lane h) or in the presence of dATP, ADP, or GDP (Fig. 1*B*, lanes c, d, and g, respectively). In the presence of ATP, products ranged in size from $\approx$9–44 nt (Fig. 1*B*, lane b). In the presence of ATPγS (a nonhydrolyzable analogue of ATP) or GTP (Fig. 1*B*, lanes e and f, respectively) one major cleavage product of $\approx$41 nt and two minor products of $\approx$40 and 42 nt were observed. This result indicated that ATPγS or GTP could be used to uncouple hairpin cleavage from double-strand exonuclease activity. It is possible that, under these conditions, the protein may not be turning over.

Initial Cleavage Occurs at the 5′ Side of a Hairpin Loop. One major and two minor products were generated when hairpin DNA was incubated with SbcCD protein in the presence of ATPγS (Fig. 1*B*, lane e). To map these SbcCD cleavage sites more precisely, protein was incubated with hairpin substrates labeled at either their 3′ or 5′ termini. Cleavage sites were confirmed by comparing the SbcCD products with A/G sequencing ladders produced by the chemical degradation of

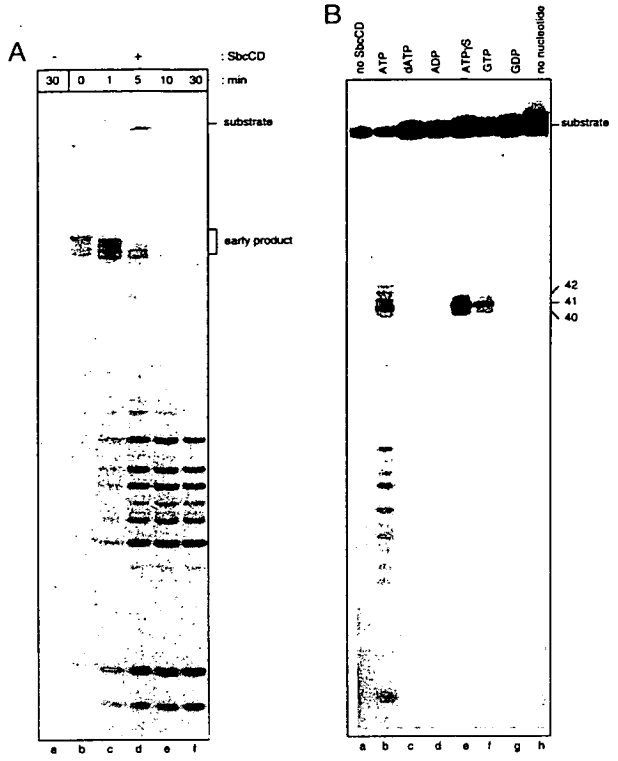


FIG. 1. Trapping the initial hairpin cleavage products generated by SbcCD. (*A*) An autoradiograph illustrating the DNA fragments produced by SbcCD over a 30-min time period. SbcCD was incubated with 3′-end-labeled hairpin substrate (3′HP78), in the presence of ATP at 37°C, for 0 min (lane b), 1 min (lane c), 5 min (lane d), 10 min (lane e), or 30 min (lane f). 3′HP78 also was incubated for 30 min in the absence of SbcCD (lane a). Early products produced by SbcCD are indicated. (B) Effect of various nucleotides on the SbcCD hairpin-nuclease activity. SbcCD was incubated with 3′-end-labeled hairpin substrate (3′HP78), at 16°C for 30 min, in the presence of various nucleotides. Reactions were performed in the presence of ATP (lane b), dATP (lane c), ADP (lane d), ATPγS (lane e), GTP (lane f), GDP (lane g), or no nucleotide (lane h). 3′HP78 also was incubated with ATP in the absence of SbcCD (lane a). Products were visualized by gel electrophoresis. The sizes of the fragments produced in the presence of ATPγS (lane e) and GTP (lane f) are indicated.

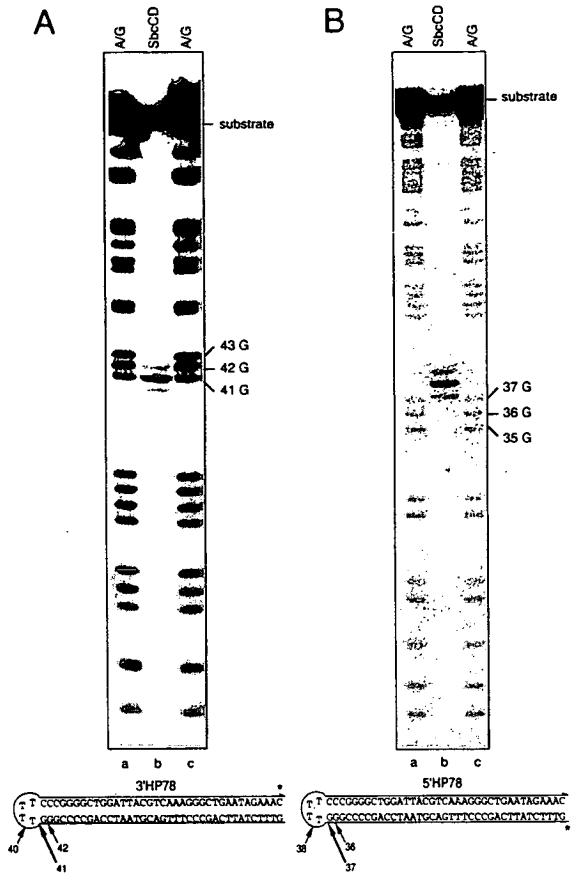


FIG. 2. 5′- and 3′-labeled hairpin substrates incubated with SbcCD in the presence of ATPγS. *, the position of the ^{32}P-label. The SbcCD product is flanked by A/G sequencing ladders. Fragments arising as a result of chemical cleavage at the three guanosines adjacent to the hairpin loop are indicated. (*A*) 3′-end-labeled hairpin (3′HP78) was incubated with SbcCD in the presence of ATPγS, at 16°C for 30 min, and reaction products were resolved on a polyacrylamide gel (lane b). A/G sequencing ladders of 3′HP78 are shown in lanes a and c. (*B*) 5′-end-labeled hairpin (5′HP78) was incubated with SbcCD in the presence of ATPγS, at 16°C for 30 min, and reaction products were resolved on a polyacrylamide gel (lane b). *, the position of the ^{32}P-label. A/G sequencing ladders of 5′HP78 are shown in lanes a and c. A one-and-one-half base allowance was made to compensate for the nucleoside eliminated in the sequencing reaction.

3′HP78 or 5′HP78 (both substrates are illustrated in Fig. 2 *A* and *B*). When 3′HP78 was incubated with SbcCD, the major product observed was a 41-nt DNA fragment. Two minor products of 40 and 42 nt also were seen (Fig. 2*A*, lane b). This result indicates that SbcCD preferentially cleaves 3′HP78 immediately adjacent to the loop/double-strand junction. To confirm that this was so, SbcCD was incubated with 5′-end-labeled hairpin 5′HP78. With this substrate, cleavage at the same junction should yield a major product of 37 nt and two minor products of 36 and 38 nt. All of these products can be seen in Fig. 2*B*, lane b. The migration of the SbcCD cleavage products relative to the A/G sequencing ladders was consistent with the presence of 5′-phosphate and 3′-hydroxyl termini at the cleavage site. To confirm this, enzymatic tests were performed (see below).

SbcCD Generates Products with 5′-Phosphate and 3′-Hydroxyl Termini. To investigate the nature of the 5′-terminal group created by SbcCD, 3′-end-labeled hairpin (3′HP78) was incubated with SbcCD, and then the products were treated with shrimp alkaline phosphatase. After phosphatase treatment, all of the SbcCD products showed an altered mobility (Fig. 3*A*, lane a). This result indicates that SbcCD leaves a 5′-phosphate group when it cleaves DNA.

To determine the nature of the 3′-terminal group created by SbcCD, 5′-end-labeled hairpin (5′HP78) was incubated with SbcCD, and then the resulting products were treated with terminal transferase in the presence of ddATP. SbcCD products that had been treated with terminal transferase were seen to have increased in size by 1 nt (Fig. 3*B*, lane b), which demonstrates that SbcCD generates a 3′-hydroxyl group when it cleaves DNA. Therefore, SbcCD cleaves a hairpin substrate to leave products with 5′-phosphate and 3′-hydroxyl termini.

SbcCD Does Not Require a 3′ or 5′ Terminus for Activity. To determine whether SbcCD was active on a molecule without free ends, protein was incubated with a covalently closed, internally labeled oligonucleotide (ccDB98; see Fig. 4). This substrate was capable of forming a dumbbell structure with a 45-bp duplex linked by two loop regions. ccDB98 was purified from a preparative gel as a single species (Fig. 4*B*). When ccDB98 was incubated at 16°C with ATPγS, in the presence of SbcCD, two major products of ≈98 and 49 nt (ncDB98) were observed (Fig. 4*A*, lane b). The size of both of these products strongly suggests that SbcCD is cleaving the covalently closed substrate at one or both junctions. In the absence of SbcCD, only a small amount of the band of ≈98 nt (ncDB98) was seen (Fig. 4*A*, lane a). This band is the expected product of radiolysis of ccDB98. A single nick anywhere in the

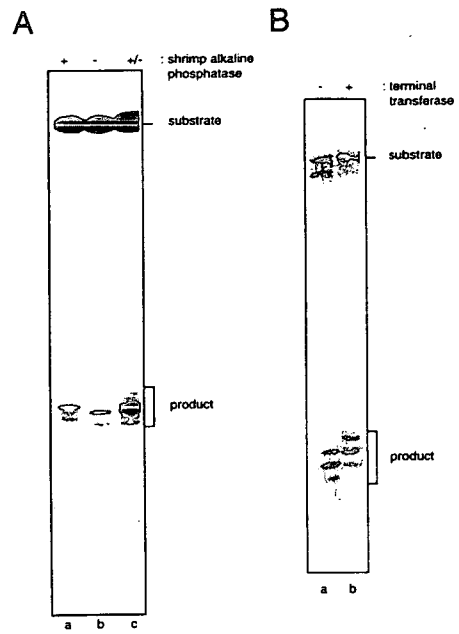


FIG. 3. Analysis of hairpin cleavage termini. (*A*) 5′ termini: 3′-end-labeled hairpin (3′HP78) was digested with SbcCD for 30 min at 16°C, in the presence of ATPγS, and the reaction products then were incubated with (lane a) or without (lane b) shrimp alkaline phosphatase. SbcCD cleavage products are indicated. A 1:1 mix of the samples loaded to lanes a and b were analyzed in lane c to demonstrate that the mobility shift observed was not a lane-specific artefact. (*B*) 3′-termini: 5′-end-labeled hairpin (5′HP78) was digested with SbcCD, in the presence of ATPγS, and the reaction products were incubated in the absence (lane a) or presence (lane b) of terminal transferase.

backbone of this circular molecule would result in the appearance of a 98-nt fragment.

SbcC(D) Has a Head-Rod-Tail Structure. SbcC is a polypeptide of 1,048 amino acids. Its sequence suggests that it contains ATP-binding motifs in globular regions at either end of the protein. These regions are separated by a 700-amino acid α-helical region likely to form a coiled-coil structure, broken by a conserved CXXC motif. This organization is typical of the head-rod-tail structure predicted for the SMC-family of proteins. Physical analysis revealed that SbcCD forms a 1.2-MDa complex (36). By using, EM we have demonstrated the existence of a structure consisting of two globular domains, linked by a long filamentous rod (Fig. 5). The globular domains at either end of the structure are ≈3.4 nm in diameter. The rod region between is ≈80 nm long and 1.5 nm wide (estimated by unidirectional shadowing). This region presumably comprises the amino acids of SbcC that are predicted to form an α-helical coiled-coil region. Additional experiments are required to determine the composition and stoichiometry of the structure visualized. Whether the D subunit remains associated to SbcC under the conditions of microscopy is unknown. For this reason, the D of SbcC(D) is indicated in parenthesis.

DISCUSSION

The SbcCD protein of *E. coli* has been shown to have exonuclease and endonuclease activities (35, 36), and genetic studies have suggested that it can act to generate double-strand breaks at DNA hairpins formed during DNA replication (39). These observations suggested that SbcCD might be able to recognize and cleave hairpin DNA. Here, we show that SbcCD does indeed cleave a hairpin substrate. Cleavage occurs on the 5′ side of the loop to yield products with 5′-phosphate and 3′-hydroxyl termini. In the presence of ATP, the initial products of cleavage are degraded by the processive double-strand

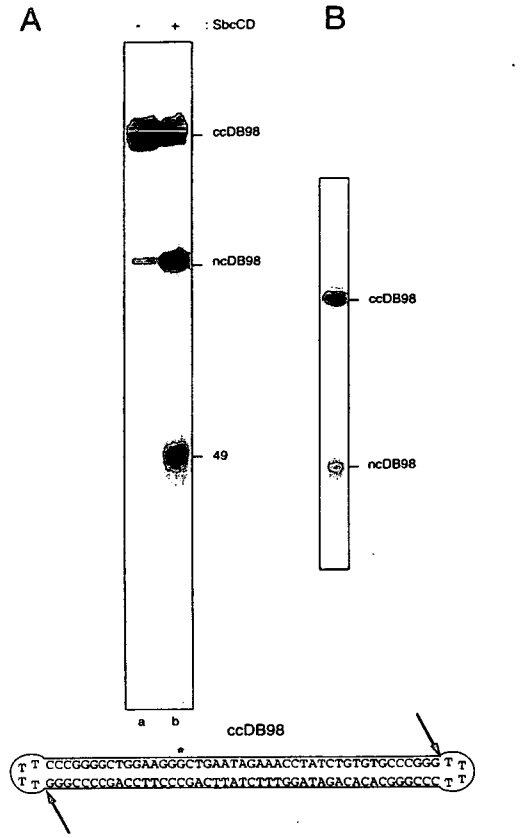


FIG. 4. SbcCD cleaves a covalently closed dumbbell substrate. (*A*) Internally labeled, covalently closed dumbbell substrate (ccDB98) was incubated without (lane a) or with (lane b) SbcCD in the presence of ATPγS, at 16°C for 30 min. *, the position of the ^{32}P-label. Reaction products then were resolved on a denaturing polyacrylamide gel. The 49- and 98-nt fragments produced are indicated (49 and ncDB98). (*B*) Preparative 10% polyacrylamide gel showing the single product obtained after the ligation of ncDB98. The upper band is covalently closed dumbbell (ccDB98), the lower band is nicked-circular dumbbell (ncDB98), which failed to ligate.

exonuclease activity of SbcCD to yield smaller oligonucleotide products. In the presence of ATPγS or GTP, initial products accumulate, suggesting that nucleotide binding, but not hydrolysis, is required for hairpin cleavage.

Although SbcCD possesses an ATP-independent, single-strand endonuclease activity (35), a purine nucleoside triphosphate is required for hairpin cleavage. This requirement suggests that hairpin cleavage is not simply a consequence of the single-strand endonuclease activity of SbcCD. Hairpin cleavage is also distinct from the ATP-dependent double-

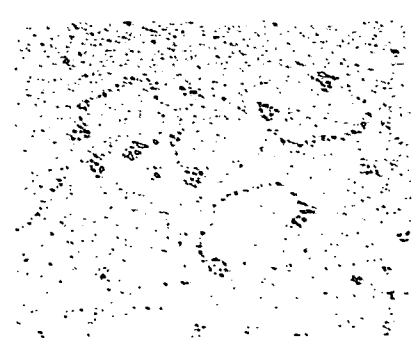


FIG. 5. Electron micrograph of SbcCD protein. SbcCD was visualized by using the low angle rotary shadowing method (Magnification, ×150,000.) Structures can be seen that consist of two globular domains linked by a long filamentous rod.

strand exonuclease activity of SbcCD because no fragments released from the blunt (nonloop) end of the hairpin substrate were detected. SbcCD is active not only on a hairpin substrate but also on a closed circular dumbbell substrate. This indicates that SbcCD does not require 5′ or 3′ termini to cleave a molecule containing hairpin loops and suggests that it is some feature of the loop, or stem-loop junction, that is initially recognized. In contrast, the RecBCD enzyme of *E. coli* is an ATP-dependent, double-strand exonuclease and ATP-independent, single-strand endonuclease that is not active on dumbbell substrates (43). The ability of SbcCD to cleave a hairpin in the absence of a double-strand end is consistent with the recognition of a hairpin formed on the template strand of DNA replication. Cleavage of such a hairpin would lead to the formation of a double-strand break, resulting in collapse of the replication fork. The latter then could be reconstituted via a homologous recombination repair pathway.

DNA Replication and the Repair of Double-Strand Breaks. Bacteriophage T4 and λ DNA replication can be initiated by homologous recombination at sites of DNA double-strand breaks (44–47), and it has been proposed that an important role of RecABCD-χ-mediated recombination in *E. coli* is the reconstitution of broken replication forks (48–50). Double-strand breaks may form during DNA replication because of the presence of a nick or gap in the template strand before its copying (50). Alternatively, replication arrest can lead to breakage of the fork (51–54), and palindromic sequences are known to cause replication arrest *in vivo* and *in vitro* (55, 56). It has been proposed that processing of hairpin secondary structures can result in the formation of double-strand breaks that then can be repaired by recombination (1, 35). For recombination to proceed, the loop at the apex of a hairpin must be removed because this structure can protect a DNA double-strand from recombination nucleases such as RecBCD enzyme (43). Homologous recombination indeed is required to repair the consequences of SbcCD attack of a 240-bp palindrome in the *E. coli* chromosome (39), and in *S. cerevisiae*, a 140-bp palindrome has been shown to act as a meiotic recombination hotspot by inducing a double-strand break (8).

Recombination Reactions That Require De-Protection of DNA Ends. A hairpin DNA end is protected from attack by nucleases such as RecBCD that recognize 3′ or 5′ termini. Cleavage by SbcCD leads to homologous recombination (39), suggesting that the protein has a role in deprotecting a substrate and allowing it to enter the homologous recombination pathway. In *S. cerevisiae*, the Rad50/Mre11 complex has been suggested to play a different deprotecting role (39). In meiotic recombination, the SPO11 protein has been shown to become covalently attached to the 5′ ends of the initiating double-strand breaks (39). Resection of these ends requires Rad50/Mre11 (23). Perhaps the role of Rad50/Mre11 is to deprotect the double-strand break site by removing the bound SPO11 and to generate recombinogenic 3′ single-strand overhangs (57). Another possible example of an SbcCD family member acting to deprotect a covalently protected end has been described in bacteriophage T4. When mutants of T4 were sought that showed increased sensitivity to *m*-AMSA [4′-(9-acridinylamino)methanesulfon-*m*-anisidide], most mutations were found in genes that encode T4 recombination proteins, suggesting that recombination is involved in the repair of *m*-AMSA-induced covalent topoisomerase-DNA cleavage complexes (58, 59). Of the nine mutations identified, one was in gene *46* and another in a gene designated *47.1* (located upstream of gene *47*). It was suggested that the mutation in gene *47.1* acts in a polar fashion on gene *47*. Consistent with this explanation, a nearly complete deletion of gene *47.1* was constructed and was found not to cause hypersensitivity to *m*-AMSA (59).

In *S. cerevisiae*, *RAD50* and *MRE11* also play a role in nonhomologous end joining (60–62). Again, Rad50/Mre11 is acting at DNA ends although its role in end-joining is unknown. However, an interesting observation is that the same end-joining phenotype is obtained in *hdf1* and *ku80* mutants (61, 63). Hdf1 and scKu80 are the *S. cerevisiae* homologues of the mammalian Ku70 and Ku80 proteins that have been implicated in binding the hairpin DNA ends generated as intermediates in V(D)J recombination (64, 65). This raises the question of whether the human homologue of SbcCD (hRad50/hMre11) is involved in V(D)J recombination.

The factor that cleaves the hairpin intermediates in V(D)J recombination has not been identified. It has been established that an activity capable of nicking hairpin structures several nucleotides from the hairpin tip is present in murine lymphoid and nonlymphoid cells (10, 66). This is reminiscent of the preference SbcCD has for cleaving at the side of a hairpin loop. The presence of such an activity in nonlymphoid cells suggests that the ability to cleave a hairpin is required in a wider context than V(D)J recombination (10, 66). Whether or not hRad50/hMre11 is implicated directly in hairpin cleavage, the activity of SbcCD closely resembles the reaction predicted to be required. It will be interesting to compare the action of the mammalian hairpin cleaving nuclease(s) with SbcCD when they have been identified and characterized.

SbcC(D) Has the Head-Rod-Tail Structure Predicted for the SMC Family of Proteins. EM has revealed that SbcC(D) has a filamentous structure with globular domains, of similar size, at either end of a rod. This tripartite head-rod-tail organization is typical of that predicted for the SMC-family of proteins and has been observed previously for two proteins implicated in chromosome segregation: the MukB protein of *E. coli* (67) and the BimC-related bipolar kinesin from *Drosophila melanogaster* (68). Neither of these proteins has been classified as belonging to the SMC family. In many SMC proteins, the potential coiled-coil domain is believed to be interrupted by a hinged spacer region. The putative coiled-coil of SbcC is broken by a conserved CXXC motif; yet, it is not obvious (from Fig. 5) that this sequence forms a hinged spacer within SbcC.

Despite their ubiquity, little is known of the biochemical activities of many SMC proteins. SbcC is an SMC protein that interacts with a non-SMC protein, SbcD (35, 36). Recently, several other SMC proteins also have been shown to exist in protein complexes with non-SMC proteins. The DPY27 polypeptide of *Caenorhabditis elegans* forms a dosage compensation complex with DPY26 and DPY28 (69, 70). The XCAP-C and XCAP-E polypeptides of *Xenopus laevis* form a chromosome condensation complex (condensin) with three other polypeptides (71). This complex exhibits a high affinity for structured DNA, such as cruciform DNA, and introduces positive supercoils into closed circular DNA in the presence of topoisomerase I (72). Two other SMC proteins, BSMC1 and BSMC2, exhibit DNA reannealing activity and form a recombination complex (RC-1) consisting of at least five polypeptides proposed to repair double-strand breaks in mammalian cells (24). This complex also recognizes and cleaves branched DNA substrates (73).

We have shown that SbcCD recognizes and cleaves a DNA hairpin substrate. However, the unusual protein structure would appear to be more complex than that required for such a relatively simple reaction. The SMC-family of proteins clearly are involved in many aspects of DNA metabolism in many different organisms. It is intriguing that several of these large SMC protein complexes, like SbcCD, recognize features of DNA secondary structure. Perhaps their common role is in targeting non-SMC proteins to these regions, where they perform varied roles important for maintaining genomic integrity.

We thank Bill Earnshaw and Dave Sherratt for critical reading of the manuscript; Karin Birkenkamp, Börries Kemper, and Petra Pfeiffer

for help in designing and handling hairpin oligonucleotides; Stefan Hyman and Arthur Rowe for EM. This research has been supported by a project grant from the Biotechnology and Biological Sciences Research Council. L.A.K. is supported by a Biotechnology and Biological Sciences Research Council studentship.

1. Leach, D. R. F. (1994) *BioEssays* **16,** 893–900.
2. Peeters, B. P. H., De Boer, J. H., Bron, S. & Venema, G. (1988) *Mol. Gen. Genet.* **212,** 450–458.
3. Behnke, D., Malke, H., Hartmann, M. & Walter, F. (1979) *Plasmid* **2,** 605–616.
4. Kieser, T. & Melton, R. (1988) *Gene* **65,** 83–91.
5. Henderson, S. T. & Petes, T. D. (1993) *Genetics* **113,** 57–62.
6. Gordenin, D. A., Lobachev, K. S., Degtyareva, N. P., Malkova, A. L., Perkins, E. & Resnick, M. A. (1993) *Mol. Cell. Biol.* **13,** 5315–5322.
7. Ruskin, B. & Fink, G. R. (1993) *Genetics* **133,** 43–56.
8. Nag, D. K. & Kurst, A. (1997) *Genetics* **146,** 835–847.
9. Collick, A., Drew, J., Penberth, J., Bois, P., Luckett, J., Scaerou, F., Jeffreys, A. & Reik, W. (1996) *EMBO J.* **15,** 1163–1171.
10. Akgun, E., Zahn, J., Baumes, S., Brown, G., Liang, F., Romanienko, P. J., Lewis, S. & Jasin, M. (1997) *Mol. Cell. Biol.* **17,** 5559–5570.
11. Kramer, P. R., Stringer, J. R. & Sinden, R. R. (1996) *Nucleic Acids Res.* **24,** 4234–4241.
12. Chalker, A. F., Leach, D. R. F. & Lloyd, R. G. (1988) *Gene* **71,** 201–205.
13. Gibson, F. P., Leach, D. R. F. & Lloyd, R. G. (1992) *J. Bacteriol.* **174,** 1222–1228.
14. Naom, I. S., Morton, S. J., Leach, D. R. F. & Lloyd, R. G. (1989) *Nucleic Acids Res.* **17,** 8033–8045.
15. Sharples, G. J. & Leach, D. R. F. (1995) *Mol. Microbiol.* **17,** 1215–1220.
16. Hirano, T., Mitchison, T. J. & Swedlow, J. R. (1995) *Curr. Opin. Cell Biol.* **7,** 329–336.
17. Strunikov, A. V., Larinov, V. & Koshland, D. (1993) *J. Cell Biol.* **123,** 1635–1648.
18. Hirano, T. & Mitchison, T. J. (1994) *Cell* **79,** 449–458.
19. Saitoh, N., Goldberg, I. G., Wood, E. R. & Earnshaw, W. C. (1994) *J. Cell. Biol.* **127,** 303–318.
20. Saka, Y., Sutani, T., Yamashita, Y., Saitoh, S., Takeuchi, M., Nakaseko, Y. & Yanagida, M. (1994) *EMBO J.* **13,** 4938–4952.
21. Strunikov, A. V., Hogan, E. & Koshland, D. (1995) *Genes Dev.* **9,** 587–599.
22. Chuang, P. T., Albertson, D. G. & Meyer, B. J. (1994) *Cell* **79,** 459–474.
23. Alani, E., Subbiah, S. & Kleckner, N. (1989) *Genetics* **122,** 47–57.
24. Jessberger, R., Riwar, B., Baechtold, H. & Akhmedov, A. T. (1996) *EMBO J.* **15,** 4061–4068.
25. Walker, J., Saraste, M., Runswick, M. & Gay, N. (1982) *EMBO J.* **1,** 945–951.
26. Gasser, S. M. (1995) *Curr. Biol.* **5,** 357–360.
27. Dolganov, G. M., Maser, R. S., Novikov, A., Tosto, L., Chong, S., Bressan, D. A. & Petrini, J. H. J. (1996) *Mol. Cell. Biol.* **16,** 4832–4841.
28. Kim, K. K., Daud, A. I., Wong, S. C., Pajak, L., Tsai, S. C., Wang, H., Henzel, W. J. & Field, L. J. (1996) *J. Biol. Chem.* **271,** 29255–29264.
29. Leach, D. R. F., Lloyd, R. G. & Coulson, A. F. C. (1992) *Genetica* **87,** 95–100.
30. Johzuka, K. & Ogawa, H. (1995) *Genetics* **139,** 1521–1532.
31. Xiao, Y. & Weaver, D. T. (1997) *Nucleic Acids Res.* **25,** 2985–2991.
32. Mickelson, C. & Wiberg, J. S. (1981) *J. Virol.* **40,** 65–77.
33. Kreuzer, K. N., Saunders, M., Weislo, L. J. & Kreuzer, H. W. E. (1995) *J. Bacteriol.* **177,** 6844–6853.
34. Petrini, H. J., Walsh, M. E., DiMare, C., Chen, X.-N., Kornberg, J. R. & Weaver, D. T. (1995) *Genomics* **29,** 80–86.
35. Connelly, J. C. & Leach, D. R. F. (1996) *Genes to Cells* **1,** 285–291.
36. Connelly, J. C., deLeau, E. S., Okely, E. A. & Leach, D. R. F. (1997) *J. Biol. Chem.* **272,** 19819–19826.
37. Shurvinton, C. E., Stahl, M. M. & Stahl, F. W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1624–1628.
38. Lindsey, J. C. & Leach, D. R. F. (1989) *J. Mol. Biol.* **206,** 7024–7027.
39. Leach, D. R. F., Okely, E. A. & Pinder, D. J. (1997) *Mol. Microbiol.* **26,** 597–606.
40. Sambrook, J., Frisch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
41. Beyert, N., Reichenberger, S., Peters, M., Hartung, M., Gottlich, B., Goedecke, W., Vielmetter, W. & Pfeiffer, P. (1994) *Nucleic Acids Res.* **22,** 1643–1650.
42. Papavassiliou, A. G. (1994) in *Methods in Molecular Biology*, ed. Kneale, G. G. (Humana, Clifton, NJ), Vol. 30, pp. 43–78.
43. Taylor, A. F. & Smith, G. R. (1995) *J. Biol. Chem.* **41,** 24459–24467.
44. Luder, A. & Mosig, G. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1101–1105.
45. Mosig, G. (1983) in *Relationship of T4 DNA replication and recombination.* In Bacteriophage T4, eds Mathews, C., Kutter, E., Mosig, G. & Berget, P. (Am. Soc. Microbiol., Washington DC), pp. 120–130.
46. Stahl, F. W., Kobayashi, I. & Stahl, M. (1984) *J. Mol. Biol.* **181,** 199–209.
47. Stahl, F. W. & Stahl, M. M. (1986) *Genetics* **113,** 1–12.
48. Asai, T., Bates, D. B. & Kogoma, T. (1994) *Cell* **78,** 1051–1061.
49. Kuzminov, A., Schabtach, E. & Stahl, F. W. (1994) *EMBO J.* **13,** 2764–2776.
50. Kuzminov, A. (1995) *Mol. Microbiol.* **16,** 373–384.
51. Bierne, H., Ehrlich, S. D. & Michel, B. (1991) *EMBO J.* **10,** 2699–2705.
52. Bierne, H. & Michel, B. (1994) *Mol. Microbiol.* **13,** 17–23.
53. Kuzminov, A. (1995) *BioEssays* **17,** 733–741.
54. Michel, B., Ehrlich, S. D. & Uzest, M. (1997) *EMBO J.* **16,** 430–438.
55. La Duca, R. J., Fay, P. J., Chuang, C., McHenry, C. S. & Banbara, R. A. (1983) *Biochemistry* **22,** 5177–5188.
56. Weaver, D. T. & DePamphlis, M. L. (1984) *J. Mol. Biol.* **180,** 961–986.
57. Keeney, S., Giroux, C. N. & Kleckner, N. (1997) *Cell* **88,** 375–384.
58. Neece, S. H., Carles-Kinch, K., Tomso, D. J. & Kreuzer, K. N. (1996) *Mol. Microbiol.* **20,** 1145–1154.
59. Woodworth, D. L. & Kreuzer, K. N. (1996) *Genetics* **143,** 1081–1090.
60. Moore, J. K. & Haber, J. E. (1996) *Mol. Cell. Biol.* **16,** 2164–2173.
61. Milne, G. T., Jin, S., Shannon, K. B. & Weaver, D. T. (1996) *Mol. Cell. Biol.* **16,** 4189–4198.
62. Tsukamoto, Y., Kato, J.-I. & Ikeda, H. (1996) *Genetics* **142,** 383–391.
63. Tsukamoto, Y., Kato, J.-I. & Ikeda, H. (1996) *Nucleic Acids Res.* **24,** 2067–2072.
64. Taccioli, G., Gottlieb, M. T., Blunt, T., Priestly, A., Demengeot, J., Mizuta, R., Lehman, A. R., Alt, F. W., Jackson, S. P. & Jeggo, P. A. (1994) *Science* **266,** 1442–1445.
65. Boubnov, N. V., Hall, K. T., Wills, Z., Lee, S. E., He, D. M., Benjamin, D. M., Pulaski, C. R., Band, H., Reeves, W., Hendrickson, E. A., *et al.* (1995) *Proc. Natl. Acad. Sci. USA* **92,** 890–894.
66. Lewis, S. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1332–1336.
67. Niki, H., Imamura, R., Kitaoka, M., Yamanaka, K., Ogura, T. & Hiraga, S. (1992) *EMBO J.* **13,** 5101–5109.
68. Kashina, A. S., Baskin, R. J., Cole, D. G., Wedaman, K. P., Saxton, W. M. & Scholey, J. M. (1996) *Nature (London)* **379,** 270–272.
69. Chuang, P. T., Lieb, J. D. & Meyer, B. J. (1996) *Science* **274,** 1736–1739.
70. Lieb, J., Capowski, E., Meneely, P. & Meyer, B. (1996) *Science* **274,** 1732–1736.
71. Hirano, T., Kobayashi, R. & Hirano, M. (1997) *Cell* **89,** 511–521.
72. Kimura, K. & Hirano, T. (1997) *Cell* **90,** 625–634.
73. Jessberger, R., Chui, G., Linn, S. & Kemper, B. (1996) *Mutation Res.* **350,** 217–227.

APPENDIX 2

SbcCD: A Hairpin Stylist

SbcCD: A Hairpin Stylist

DNA hairpins are fold-back structures which can potentially form from single-strand palindromic sequences. Pseudo-hairpins can similarly form from sequences with sufficient self-complementarity to fold-back (reviewed Holde and Zlatanova, 1994, Leach 1999). Hairpins are implicated in a wide range of DNA metabolic reactions in prokaryotes and eukaryotes. An activity which recognises hairpins and pseudo-hairpins is therefore of considerable interest. The SbcCD protein of *Escherichia coli* is a hairpin nuclease and may be key to understanding these reactions.

The Ubiquitous Hairpin

Much DNA metabolism occurs via hairpin intermediates. Certain bacterial transposons are postulated to transpose via a hairpin intermediate, *eg* Tn*10* and Tn*5* (Gordenin *et al* 1992), by a mechanism which may be conserved throughout a range of hydrolysis-transesterification reactions (Kennedy *et al* 1998). Short inverted repeats of a few nucleotides (palindromic "P" inserts) are generated on transposition of certain plant transposons and are also generated at sites of non homologous end-joining in murine fibroblastoid cells (Lewis 1994). These P inserts are proposed to be derived from a hairpin intermediate by nicking of the hairpin loop at a point a few nucleotides from the apex, where this distance determines the length of the P insert (Lewis 1994).

The DNA-PK complex functions in both the non homologous end-joining and strand-slippage (discussed below) pathways of illegitimate recombination. The DNA-PK is essential for non homologous end-joining in eukaryotes (Boubnov *et al* 1995) and the Ku component of DNA-PK displays avid sequence independent DNA binding to a range of substrates including hairpin loops (Cary *et al* 1997). This is consistent with a role of hairpins in non homologous end-joining.

Diversity in immunoglobulin and T-cell receptor genes is generated by DNA rearrangement by the mechanism of V(D)J recombination. V(D)J recombination in vertebrates (reviewed Leach 1996a) bears similarities with that of non homologous end-joining. During V(D)J recombination, Figure 1, RAG1 and RAG2 proteins cleave the signal sequence (McBlane *et al* 1995) and generate covalently closed hairpins from the coding ends (Hiom and Gellert 1997). P inserts are generated at the coding joint by hairpin cleavage by an unknown factor (Paull and Gellert 1998), nucleotide incorporation and end ligation. P insert products have been observed in murine lymphoid cells (Lewis 1994). The RAG1 and RAG2 proteins can also promote intermolecular transposition into a nonspecific target site *in vitro* (Hiom *et al* 1998) and the V(D)J recombination system bears global similarities with transposition by the HAT family of transposons (reviewed Roth and Craig 1998).

In addition to the formation of hairpins as intermediates in DNA recombination reactions, hairpins can form during replication (Trinh and Sinden 1991). Strand-slippage is a hypothetical type of illegitimate recombination (Albertini 1982) in which DNA polymerase penetrates a hairpin structure in the template strand only partially, pauses at a short direct repeat just within the structure called the donor, and then slips to a short direct repeat just beyond the structure called the target (Leach 1994). The main body of the hairpin is thus deleted. This was studied in *E. coli* using a cassette carrying a palindrome with asymmetric donor and target repeats (Trinh and Sinden 1991). In one orientation, strand-slippage could occur on the leading strand only, and on inversion of the cassette, it could occur on the lagging strand only. Strand-slippage, and therefore hairpin formation at a replication fork, was found to occur at a 20-fold (Trinh and Sinden 1991) to 25-fold (Rosche *et al* 1995) greater frequency in the lagging strand, presumably because single-strand regions are present which can form secondary structures.

The rate of hairpin extrusion of a palindromic sequence is dependent on temperature (according to melting temperature) and increases with the superhelical density of the carrier DNA (Mizuuchi *et al* 1982, Sinden *et al* 1983, Sinden and Pettijohn 1984, Warren and Green 1985). The thermodynamic stability of a hairpin decreases with loop size, increases

with stem length (Mizuuchi *et al* 1982, Xodo *et al* 1991) and is affected by sequence (Davison and Leach 1994a, Davison and Leach 1994b, Chalker *et al* 1993). Short hairpins (*eg* a 33bp stem) rarely extrude and may do so only in the presence of a nucleotide cofactor (Sinden *et al* 1983).

A growing number of hereditary neuromuscular disorders, called trinucleotide repeat expansion diseases (TREDs), are thought to result from erroneous replication of $(CAG)_n$ and $(CGG)_n$ repeat tracts by a similar strand-slippage mechanism (Mitas *et al* 1995, Leach 1996a, Leach 1998). Disease symptoms arise when the tract expands above a threshold length, where expansion starts to occur in large discrete saltatory steps rather than small incremental steps. These $(CXG)_n$ repeats can form pseudo-hairpins (Mitas *et al* 995, Marquis Gacy *et al* 1995, Darlow and Leach 1995), and it has been suggested that large discrete saltatory expansion occurs when the replication machinery slips across the base of pseudo-hairpins (Sarkar *et al* 1998).

Palindromes Are Mutagenic and Cause Replication Arrest

Perfect palindromes longer than approximately 100bp or 150bp present *E. coli* with a problem at replication. They cannot be propagated and either are unstable or confer inviability (Lilley 1981, Warren and Green 1985). Instability is the partial internal deletion of a palindrome during replication, probably by strand-slippage. Inviability is the loss of the carrier replicon from the population due to a palindrome-mediated block to replication. Kinetic and thermodynamic factors which increase hairpin extrusion increase instability and inviability (Warren and Green 1985). Both processes occur across the phyla, although there are different levels of palindrome tolerance in different organisms and at different locations (Das Gupta *et al* 1987, Akgun *et al* 1997, reviewed Leach 1996b).

Other palindromes up to 40bp in length are maintained in *E. coli* (Leach *et al* 1987). These are integral parts of the genome (distinct from the transient hairpin intermediates described above). They have important functions resulting from their hairpin forming ability (reviewed Higgins 1986) as barriers to degradation of mRNAs, bacteriophage replication

origins, *eg* M13, G4, ϕ29, and transcriptional and replication terminators. Short palindromes are conserved across a range of prokaryotic genomes. These include Repetitive Extragenic Palindromes REPs (Stern *et al* 1984, Gilson *et al* 1984), Intergenic Repeat Units IRUs (Sharples and Lloyd 1990, Hulton *et al* 1990), and repeat tracts of REPs, called Bacterial Interspersed Mosaic Elements BIMEs (Espeli and Boccard 1997). This suggests that palindrome maintenance is regulated.

How Are Palindromes Tolerated?
The Usual Suspects

A model has been proposed for the regulation of palindrome maintenance which cites homologous recombination as a major participant (Leach 1994). The main pathway of homologous recombination in wildtype prokaryotes is mediated by RecABCD, as originally proposed (Meselson and Radding 1975), Figure 2a. The potent RecBCD exonuclease V enters DNA via a blunt or nearly blunt double-strand end and degrades duplex DNA until it meets a properly oriented χ site (5'-GCTGGTGG-3') (reviewed Myers and Stahl 1994, Smith *et al* 1995, Fraser 1994). In combination with RecA and SSB, this temporarily inactivates the RecD subunit, and reduces the affinity of the RecBC heterodimer for RecD (Kuzminov *et al* 1994). RecD dissociates to activate RecBC(D$^-$) as a recombinogenic helicase which proceeds to generate single-strand DNA. RecBC loads RecA onto the 3' terminal single-strand at the χ site, although χ–specific loading requires RecD *in vitro* (Churchill *et al* 1999). Head-to-tail polymerisation of RecA generates a right-handed helical nucleoprotein filament (reviewed Rao and Radding 1995). This filament promotes pairing with a second DNA molecule in a homology search to form a DNA triplex or quadruplex of non Watson-Crick bonds. RecA then promotes strand exchange by single-strand invasion of a homologous duplex and formation of a displacement (D)-loop (Holliday Junction) intermediate. It is the 3' terminal RecA-coated strand which is preferentially utilised in homologous pairing (Churchill *et al* 1999). The Holliday Junction may undergo branch migration, catalysed by RuvAB in the forward direction (Yu *et al* 1997), and resolution by cleavage by RuvC Holliday Junction-specific endonuclease (Bennett *et al* 1993, Whitby *et al*

1996). RecG catalyses branch migration in the reverse direction (Whitby *et al* 1993).

In RecABCD-mediated homologous recombinational repair of a double-strand break (DSB), the resected DSB has two 3' termini and two Holliday Junctions are formed. These may either be resolved (Szostak *et al* 1983) or undergo further branch migration until they meet (Thaler *et al* 1987), Figure 2b. It has been proposed that the RecABCD system evolved primarily for the repair and reconstitution of replication forks which sustain DSBs (Kuzminov *et al* 1994). It was observed that 90% of χ sequences in *E. coli* are oriented towards *oriC*, *ie* the opposite orientation to replication, such that they would protect against degradation of DNA in the terminus to origin direction. This suggests that most DNA damage repaired by RecABCD takes the form of one double-strand end rather than the two present at normal DSBs, Figure 2c, and such a structure is predicted to occur at a broken replication fork. Homologous recombination would reconstitute the fork at a prior stage, figure 2d. Consistent with this model, the repair of broken forks requires RecABCD (Kuzminov 1995a, Kuzminov 1995b, Michel *et al* 1997) and homologous recombination is increased in the vicinity of arrested forks which are susceptible to breakage (Bierne and Michel 1994).

Several other pathways of homologous recombination exist in bacteria (reviewed Summers 1996). They rescue *recBC* mutants and are specified by null mutation of *sbc* "suppressor of *recBC*" genes. The RecF pathway is partially activated in *sbcB* mutants and fully activated for wildtype levels of recombination in *sbcBsbcC/D* mutants (Lloyd and Buckman 1985). Indeed *sbcC/D* mutations arise spontaneously in *sbcB* cells. *sbcB* encodes processive 3' to 5' single-strand exonuclease I (Lehman and Nussbaum 1964) which might degrade the putative 3' single-strand recombination intermediate (Horii and Clark 1973). The pathway utilises RecF, RecO and RecR which are postulated to act in strand pairing. It utilises RecJ, a 5' to 3' single-strand exonuclease, and RecQ, a DNA helicase which might act with it in a manner analogous to RecBC. RecN has a role, perhaps in protecting the intermediate (Leach 1996a). RdgC also has a role and is postulated to be an exonuclease with a role in replication fork repair (Ryder *et al* 1996). The RecE pathway is activated when mutation of *sbcA*

induces *recE* (Gillen *et al* 1981). *recE* is located in the chromosomal segment identified as the Rac prophage, a cryptic lambdoid prophage. It encodes ATP independent double-strand 5' to 3' exonuclease VIII. The pathway also utilises RecT, an analogue of the λ redβ gene which catalyses the annealing of complementary single-stranded DNA, RecF and RecJ. The key reactions in homologous recombination are conserved between the phyla (reviewed Kanaar and Hoeijmakers 1997).

New Suspects

It had been shown that palindrome inviability results from a block to replication (Leach and Lindsey 1986, Chalker *et al* 1988, Shurvinton *et al* 1987), and that long palindrome replication is slow (Lindsey and Leach 1989). Since the major role of RecABCD-mediated homologous recombination is the repair of DSBs at replication forks, and palindromes form hairpins in the lagging strand during replication, it was proposed that palindrome maintenance might be regulated by a mechanism involving DSB formation at replication fork hairpins followed by repair (Leach 1994). Replication pauses at sites of secondary structure (Weaver and DePamphilis 1984, Kuzminov 1995b). The determinant of palindrome inviability was postulated to be a nuclease which cleaves a hairpin formed at a replication fork during replication arrest (Leach 1994), Figure 2d. This activity would generate a DSB with a single double-strand end, of the type acted upon by the RecABCD enzymes. The replication fork would be repaired and reconstituted by homologous recombination at a prior stage. This would provide another opportunity for unhindered replication of the region. The frequency with which a palindrome forms a hairpin would determine whether it would be maintained or confer inviability. Long palindromes would always form hairpins, the replication fork would collapse and the strain would be inviable. Short palindromes would not form hairpins and would be replicated unhindered. Several cycles might be required for the replication of medium sized palindromes. Palindromes which form hairpins only occasionally would be potentially subject to instability by strand-slippage if the hairpin nuclease did not cleave absolutely all hairpins formed at the replication fork. This model of arm-directed repair depends on the leading arm of the replication fork remaining free of secondary structure

to act as a repository of correct genetic information. It also depends on the hairpin nuclease not requiring DNA ends.

The model predicts that RecBC is required for the repair of DSBs formed at a palindrome, and that the palindrome template sequence is preserved in a cell although it may be selectively lost from a population depending on its stability as a hairpin. These were tested using a 240bp palindrome which is sufficiently short to be viable in *E. coli* (Leach *et al* 1997). As expected, it exhibited 10^4- to 10^5-fold lower viability in a *rec*$^-$ strain, and the daughter cells mostly retained the palindrome. Consistent with the model, two copies of the palindrome inserted in tandem commonly underwent intramolecular homologous recombination to leave a single copy, indicating that homologous recombination occurs in the vicinity of the palindrome. Palindromes have independently been found to be hot spots for DSB formation and homologous recombination in *E. coli* (Warren and Green 1985, Kuzminov 1995b), yeast (Nag and Kurst 1997), and mouse (Collick *et al* 1996, Akgun *et al* 1997).

sbcC was initially identified as a locus involved in full activation of the RecF pathway (Lloyd and Buckman 1985). It was then observed that mutation of *sbcC* allowed growth of strains carrying long perfect palindromes (Leach and Stahl 1983, Chalker *et al* 1988). *sbcD* was then found to map adjacent to *sbcC* (Naom *et al* 1989) and to have an identical effect on recombination and palindrome maintenance (Gibson *et al* 1992). It was predicted that SbcC and SbcD would form a complex, a notion supported by the observation that the homologues in T4, gp46/47, T5, gpD13/D12 (Leach *et al* 1992), and yeast, Rad50Mre11 (Johzuka and Ogawa 1995) do so. As expected, SbcC and SbcD polypeptides copurify (Connelly *et al* 1997).

SbcC and SbcD also have eukaryotic homologues, called Rad50 and Mre11 respectively. The homologues of SbcCD are nucleases (Gram and Ruger 1985, Blinov *et al* 1989, Furuse *et al* 1998, Usui *et al* 1998, Trujillo *et al* 1998, Moreau *et al* 1999, Paull and Gellert 1998). SMC, structural maintenance of chromosomes, proteins form a superfamily of ATP and DNA binding proteins with conserved Head-Rod-Tail domains (Melby *et al* 1998). They fall into six groups, four of which show considerable

homology, and two of which show less homology and are called SMC-like proteins. Rad50 homologues exist throughout the phyla, and belong to the subfamily of SMC-like proteins which function in recombination (Sharples and Leach 1995, Melby *et al* 1998). The homologues of SbcD also exist through the phyla and share the conserved motifs characteristic of phosphoesterase proteins (Sharples and Leach 1995, Tsubouchi and Ogawa 1998). In addition, the λ Gam protein binds either RecBC nuclease or SbcCD, suggesting some similarity (Kulkarni and Stahl 1989). It was therefore not a surprise to discover that SbcCD is a nuclease (Connelly and Leach 1996). It possesses an ATP independent single-strand endonuclease activity, an ATP dependent double-strand exonuclease activity (Connelly and Leach 1996), and a hairpin nuclease activity (Connelly *et al* 1998, Connelly *et al* 1999), all of which require Mn^{2+}.

The SbcCD protein is therefore a candidate for the hairpin nuclease determinant of palindrome inviability in *E. coli*. In order to test this, the nuclease activity of SbcCD was studied on a hairpin substrate lacking ends, *ie* a dumbell DNA (Connelly *et al* 1998, Kirkham 1999). This was generated by ligation of an oligonucleotide. SbcCD was found to cleave the dumbell at the phosphodiester link immediately 5' of the loop, identically to its cleavage of hairpin oligonucleotide substrates. This supports a role of SbcCD as the hairpin nuclease determinant of palindrome maintenance described by the model. ATP binding was required for this initial cleavage, and ATP hydrolysis was required for degradation of the substrate. The protein was further characterised with a view to identifying other potential substrates and roles of SbcCD and its eukaryotic homologues (Kirkham 1999). Here SbcCD was shown not to have helicase activity on a short oligonucleotide hairpin substrate. DNA binding was difficult to detect, but a low level was observed by gel filtration and by gel retardation of glutaraldehyde crosslinked binding reactions. This suggested that DNA binding by SbcCD might be transient.

Unanswered Questions

In contrast to SbcCD, DNA binding by yeast Rad50 (Raymond and Kleckner 1993a) and Mre11 (Furuse *et al* 1998, Usui *et al* 1998) and human Mre11 (Paull and Gellert 1998) has been detected by gel retardation without the need for crosslinking. It is interesting that the C-terminus of eukaryotic Mre11 includes an acidic tail which is absent in prokaryotic SbcD (Sharples and Leach 1995), and it is in this region that the two DNA binding sites of yeast Mre11 are contained (Usui *et al* 1998). Furthermore, a C-terminal deletion mutant of yeast Mre11 exhibits very weak DNA binding compared to wildtype (Furuse *et al* 1998). It has been suggested that the acidic C-terminus acts as a protein-protein interaction site, recruiting meiosis-specific factors which facilitate DNA binding (Usui *et al* 1998). However these proteins are not essential for DNA binding (Usui *et al* 1998, Chamankhah and Xiao 1998). Why should the prokaryotic and eukaryotic homologues bear this fundamental biochemical dissimilarity? It may be that stronger binding slows down the nuclease reaction and transient binding stimulates rapid nuclease activity. Mre11 has similar nuclease activities to SbcCD (Furuse *et al* 1998, Usui *et al* 1998, Trujillo *et al* 1998, Moreau *et al* 1999, Paull and Gellert 1998) and regulation of palindrome maintenance appears to occur in eukaryotes. This suggests that Rad50Mre11 may have a similar function to SbcCD. The differences in DNA binding may indicate that Rad50Mre11 also has additional function(s) to SbcCD. In support of this, the Rad50Mre11 complex includes a third protein in both meiosis and mitosis, Xrs2 in yeast (Johzuka and Ogawa 1995, Raymond and Kleckner 1993a, Ivanov *et al* 1992) and NBS1 in human (Trujillo *et al* 1998, Carney *et al* 1998, Featherstone and Jackson 1998), and may (Dolganov *et al* 1996, Carney *et al* 1998) or may not (Trujillo *et al* 1998) also include two other proteins. The feature common to all the homologues may be catalysis of the de-protection of DNA ends. T4 gp46/47 may be involved in the recombinational repair of *m*-AMSA induced covalently crosslinked topoisomerase-DNA complexes (Woodworth and Kreuzer 1996, Neece *et al* 1996). SbcCD cleaves a hairpin adjacent to the loop, opening up the stem to degradative enzymes (Connelly *et al* 1998).. Similarly Rad50Mre11 liberates Spo11 covalently bound to DSB sites during DSB formation in

yeast meiotic recombination (Keeney and Kleckner 1995, Keeney *et al* 1997).

Although the processive nuclease activities of SbcCD (Connelly *et al* 1997, Connelly *et al* 1999) and Mre11 (Usui *et al* 1998, Trujillo *et al* 1998) have 3' to 5' polarity *in vitro*, Rad50Mre11 in yeast is implicated in 5' to 3' DSB resection *in vivo* (Nairz and Klein 1997). It is puzzling to reconcile the two different polarities. This paradox might be resolved by half-length measurement of a hairpin substrate by SbcCD (Connelly *et al* 1999) or by activity of the Mre11 nuclease in conjunction with a helicase (Usui *et al* 1998, reviewed Haber 1998) if both nucleases initially cleaved 5' of the DSB site.

Rad50Mre11 is implicated in genetic disease in mammals. Mouse Rad50 shares epitopic homology to the p53 tumour suppressor protein (Kim *et al* 1996). Human *RAD50* is located at 5q31 which is in a region associated with several genetic cancers (Dolganov *et al* 1996, Hill *et al* 1996, Fairman *et al* 1996). Human *MRE11* maps to 11q21, which is in a region associated with cancer-related chromosomal abnormalities (Petrini *et al* 1995). Mre11 is essential for embryonic stem cell proliferation (Xiao and Weaver 1997) and thus no genetic diseases arise from their mutation. However mutation of NBS1 confers Nijmegan Breakage Syndrome, a disease characterised by early onset cancers, immune system defects and developmental abnormalities (Featherstone and Jackson 1998). It is caused by the absence of the signal transduction pathway which regulates the G1/S DNA damage response checkpoint (Carney *et al* 1998, Trujillo *et al* 1998) by sensing the presence of DSBs and stimulating non homologous recombinational repair (Maser *et al* 1997). In addition to preventing carcinogenesis, Rad50Mre11 may have a role in preventing trinucleotide repeat expansion. A CTG trinucleotide repeat oligonucleotide was cleaved by SbcCD (Connelly *et al* 1999) and CTG amplification in a bacterial model system requires loss of SbcC (Sarkar *et al* 1998). This is consistent with SbcCD cleavage of pseudo-hairpins formed from trinucleotide repeat arrays and may be relevant in developing a therapy for TREDs.

The fields of study of SMC proteins, Mre11, and SbcCD have intensified over the past six months or so. Studies on SbcCD and its homologues are being rewarded with tantalising glimpses of their complex, fundamental roles.

Supplementary Bibliography

Churchill J J, Anderson D G, Kowalczykowski S C (1999) The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of χ, resulting in constitutive recombination activation. *Genes Dev* **13** 901-911

Hiom K, Melek M and Gellert M (1998) DNA transposition by the RAG1 and RAG2 proteins: A possible source of oncogenic translocations. *Cell* **94** 4 463-470

Kanaar R and Hoeijmakers H J (1997) Recombination and joining: Different means to the same ends. *Genes Function* **1** 165-174

Kirkham L A (1999) The SbcCD Protein of *Escherichi coli*. PhD thesis, The University of Edinburgh

Roth D B and Craig N L (1998) VDJ Recombination: A Transposase Goes to Work. *Cell* **94** 4 463-470

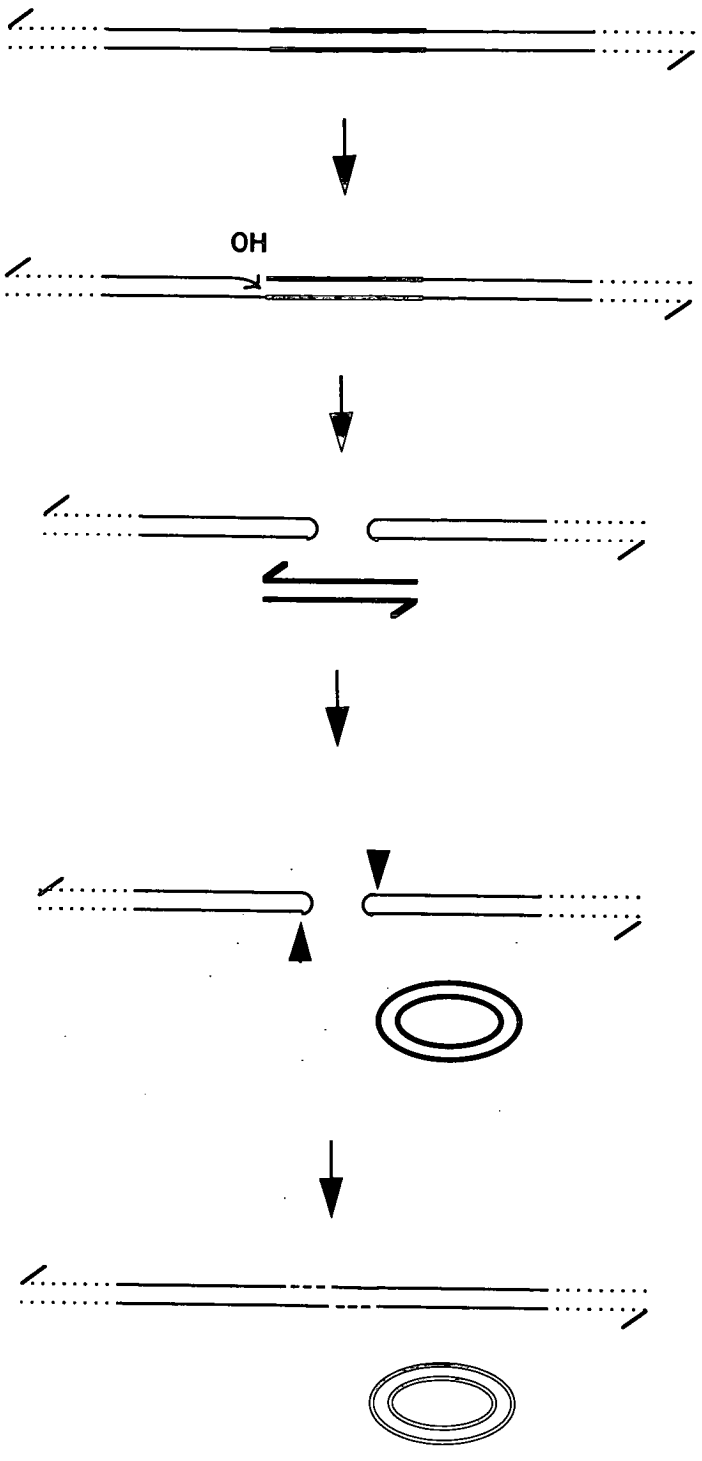


Figure 1 V(D)J Recombination. Coding sequence (thin lines) is separated by signal sequence (thick lines). DSB formation is mediated by RAG1 and RAG2 which catalyse the nucleophilic attack of the bottom strand by the coding sequence 3′ OH (Roth and Craig 1998). This transesterification reaction occurs at both coding sequences to be joined. A hairpin intermediate is generated by each of these reactions. Ku might also be used. An unidentified factor (which might include Mre11) cleaves the hairpins at a site adjacent to the loop in a manner reminiscent of SbcCD hairpin nuclease activity. Joining is completed by nucleotide incorporation and generates a coding joint. Joining of the signal sequence ends results in circularisation and generates a signal joint (Roth and Craig 1998).

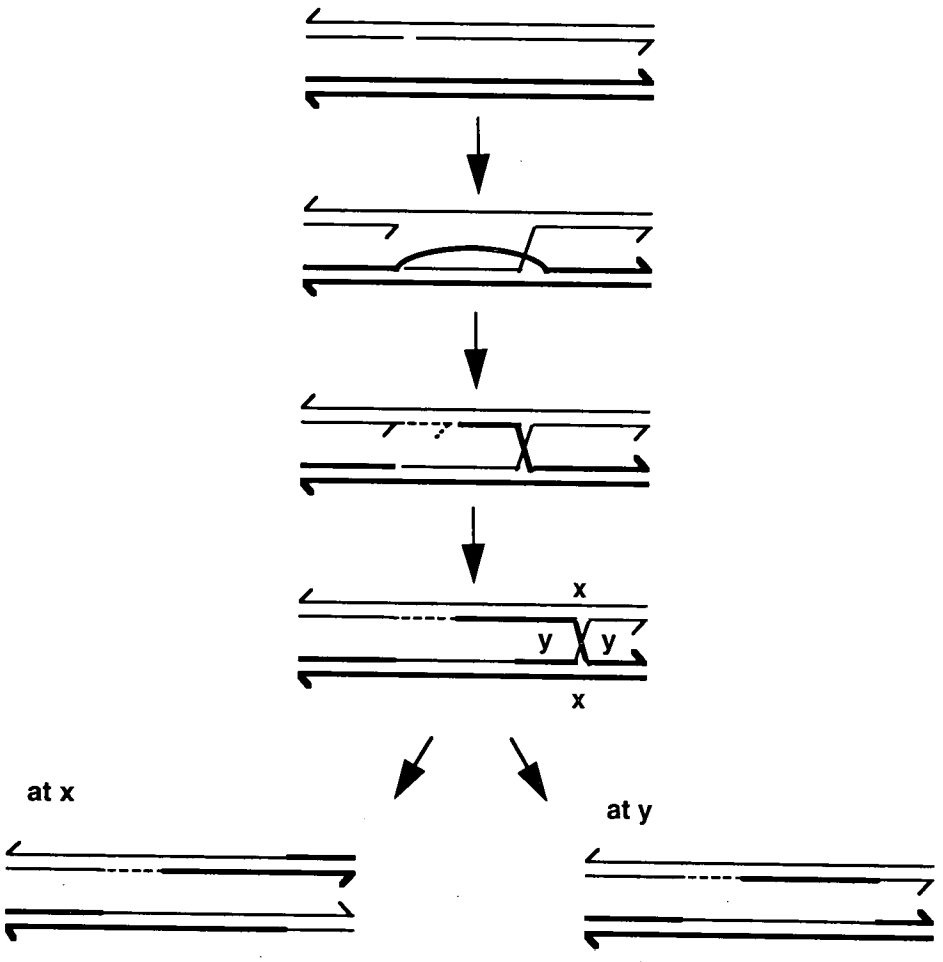


Figure 2a The Meselson-Radding Model of homologous recombination (Meselson and Radding 1975). A single-strand nick is repaired by asymmetric invasion. A single-strand tail invades an unpaired region in the homologous strand to form a D-loop. The resulting gap is repaired by DNA synthesis. Meanwhile strand crossover may occur to generate a Holliday Junction. Then branch migration may extend the region of heteroduplex DNA. Resolution at "x" or "y" generates "splice" and "patch" recombinants respectively. This model was worked out by fungal tetrad analysis. Mismatch repair affects the spore ratio. (The invading strand was originally depicted as a 5′ end, but is now thought to be a 3′ end, see text. This means that DNA synthesis would be primed in the other direction).

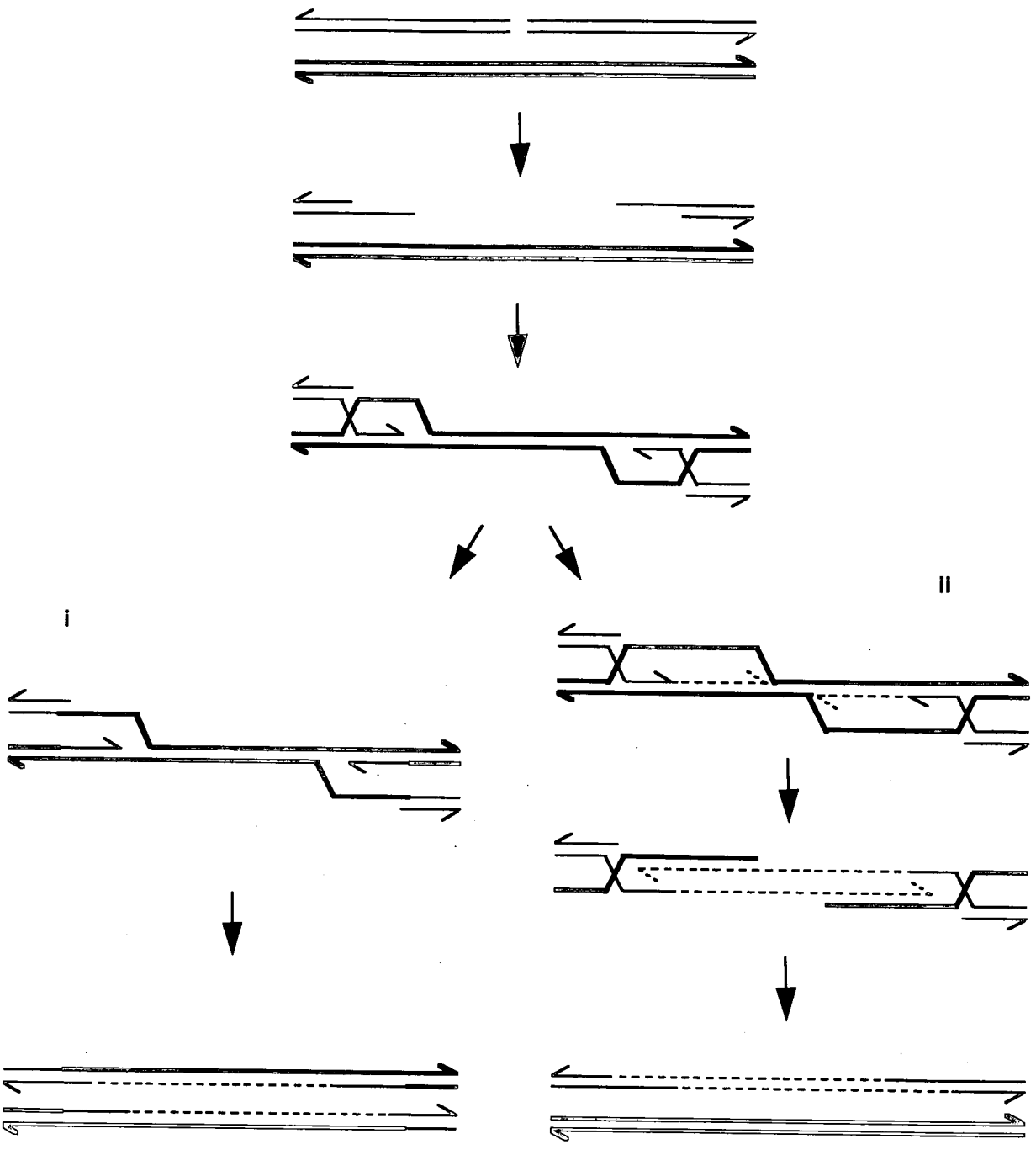


Figure 2b DSB repair by RecABCD-mediated homologous recombination. A DSB is resected and the 3′ single-strand tails invade homologous duplexes to generate Holliday Junctions. DNA synthesis is primed at the same 3′ ends. Recombinants can be formed either by (i) Holliday Junction resolution and DNA synthesis (Szostak *et al* 1983) or by (ii) further branch migration and DNA synthesis but no cleavage (Thaler *et al* 1987). The latter mechanism occurs where Holliday Junction resolvases are absent.

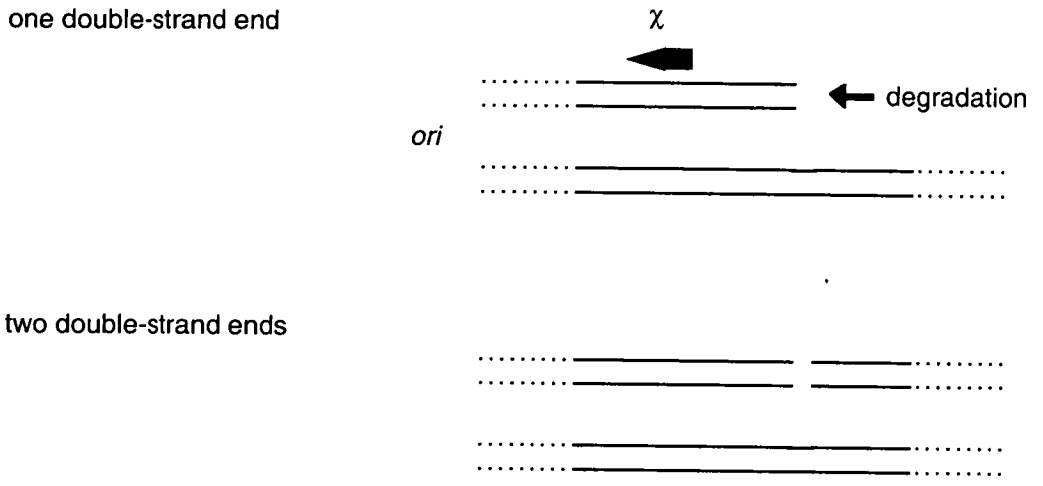


Figure 2c Types of DSB. The RecBCD exonuclease degrades DNA as it processes towards a χ site, which is oriented towards the replication origin. Therefore most DSBs repaired by RecABCD have only one double-strand end.

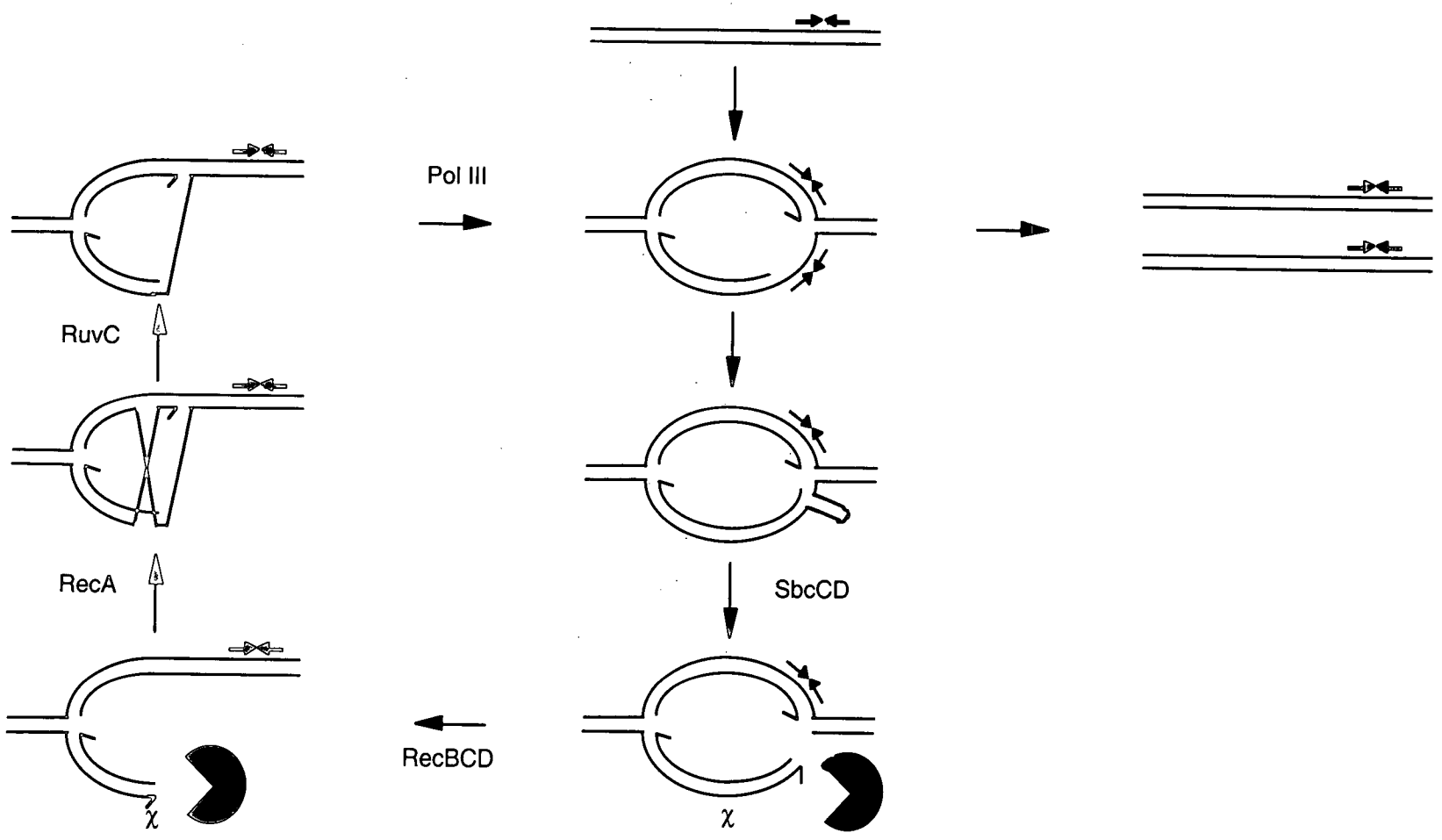


Figure 2d Model for the regulation of palindrome maintenance by SbcCD and homologous recombination (Leach 1994)