IDENTIFICATION & MOLECULAR CHARACTERISATION OF A NOVEL *recA* FROM *MYCOBACTERIUM TUBERCULOSIS*

SHAMILA NAIR

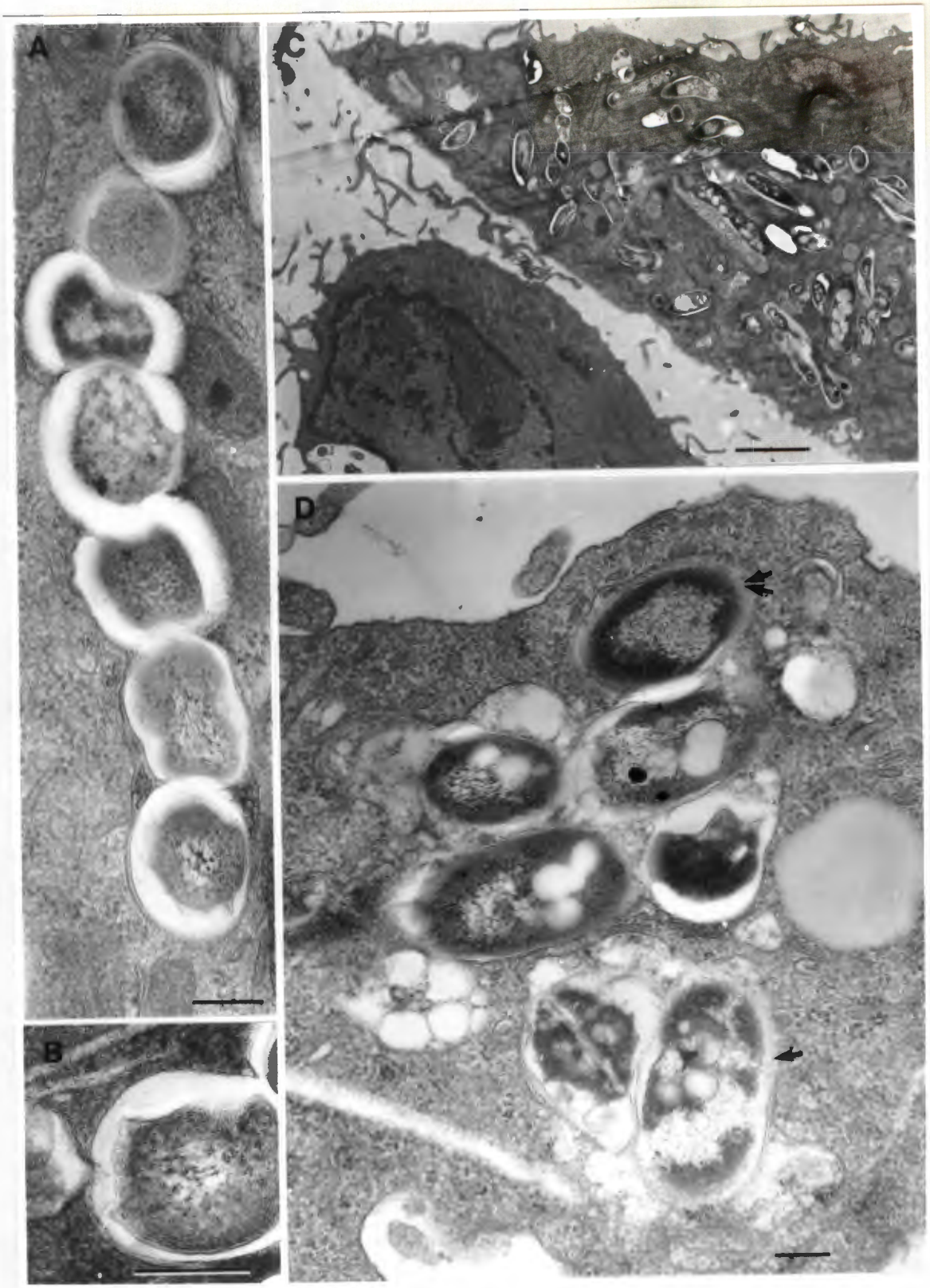
A dissertation submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy in the Faculty of Medicine, University of Cape Town, South Africa

Cape Town, January 1993

CONTENTS

		Pages
ACKNOWLEDGEMENTS		5
LIST OF TABLES		8
LIST OF FIGURES		9
ABBREVIATIONS		12
ABSTRACT		14
SECTION A		
CHAPTER 1:	General Introduction	16
CHAPTER 2:	Isolation of a *recA* homologue from *M. tuberculosis*	54
CHAPTER 3:	Location of the *recA* homologue in pSNS201	67
CHAPTER 4:	The nucleotide sequence and identification of a novel *recA*-like gene: *recS*	77
SECTION B		
CHAPTER5:	General Introduction	111
CHAPTER 6:	Isolation, characterisation and sequencing of a second DNA fragment able to complement *E. coli DK rec*$^-$ cells	121
CHAPTER 7:	General Conclusions	136
LITERATURE CITED		
APPENDIX		

Frontispiece: GY7066 carrying pSNS206 showing *lac*$^{+}$ papillae and internal recombination.



Frontispiece: Electron micrograph showing tubercle bacilli within macrophages (courtesy N. Rastogi, Dept. Mycobacteries et Tuberculosis)

ACKNOWLEDGEMENTS

My most sincere gratitude goes to my supervisor, Dr Lafras Steyn, for his encouragement and support. His advice and guidance during this study were greatly appreciated.

I am deeply indebted to Dr Brigitte Gicquel of the Unité de Génie Microbiologique, Département des Biotechnologies, Institut Pasteur, Paris, whose kindness and understanding helped me to complete the outstanding experimental part of this thesis in her laboratory. She provided me with invaluable assistance and advice.

My deepest gratitude goes to Dr Gerard Guglielmi of the Unité de Génie Microbiologique, for his expert technical assistance, proof-reading of the manuscript, artistic abilities with the preparation of diagrams and figures as well as computer analysis. His friendship, honesty and critical evaluations I value highly.

I am very grateful to Professor A. Forder, Head of the Department of Medical Microbiology, for his constant interest in my work and for finding financial support for me.

Special thanks are due to my colleagues and friends of the Department of Medical Microbiology for their patience and encouragement, Dr Gay Elisha for her help with the proof-reading of the manuscript, Patrick Tippoo for unlimited advice and especially Helen Everett for her friendship and support.

Special thanks are also due to my colleagues and friends of Unité de Génie Microbiologique, Institut Pasteur, for their interest and encouragement, especially Madame Helena Sinno for her help with the typing.

And lastly, but not least of all, I owe much to R.J. Gajjar, for his patience, companionship and enthusiasm during this study.

I acknowledge financial support from: The Medical Research Council (1 year); the Foundation for Research and Development of the Council for Scientific and Industrial Research (2 years); and a fellowship from the Deutscher Akademischer Austauschdienst (DAAD), for a 6 month period spent during this study in the Department of Bacteriology, Institut fur Experimenteele Biologie und Medizin, Borstel, D2061, Germany; and the Groote Schuur Hospital, Cape Town.

For my parents: for the sacrifices they have made to make this possible

LIST OF TABLES

Table 1.1 Observations which showed the requirements for RecA in various UV induced activities in *E. coli* and lambda phage (adapted from Witkin, 1991).

Table 1.2 Role of the RecA in SOS regulation (from Witkin, 1991).

Table 1.3 SOS inducible phenotypes requiring RecA* for activities other than LexA cleavage (from Witkin, 1991).

Table 1.4 Analysis of the *recA* gene (modified from Witkin, 1991).

LIST OF FIGURES

Fig. 1.1 Model of the SOS regulatory system (from Walker, 1985).

Fig. 1.2 SOS regulation in wild type *E. coli* (adapted from Witkin, 1991).

Fig. 1.3 A genetic map of the *E. coli* chromosome showing many of the genes needed for genetic recombination and recombinational repair of DNA damage (from West and Connolly, 1992).

Fig. 1.4 Structure of the RecA monomer (from Story *et al*, 1992).

Fig. 1.5 Schematic representation of DNA rearrangements during recombination. Shown is the recombination of chromosomes that differ at three loci (A or a, B or b, C or c) (from Weinstock, 1987).

Fig. 1.6 Formation of a D loop by RecA protein (from Weinstock, 1987).

Fig. 2.1 Southern blot of DNA isolated from *M. tuberculosis* H37Rv and *Streptomyces coelicolor*, probed with labelled DNA of the insert from pSNS201.

Fig. 2.2 Effect of the recombinant plasmid pSNS201 on the survival of *E. coli* DK cells which were plated on various EtMes concentrations.

Fig. 2.3 Detection of RecA protein in crude cell extracts by Western blot analysis.

Fig. 3.1a Restriction map of pSNS201.

Fig. 3.1b Restriction pattern of pSNS201 on gel electrophoresis.

Fig. 3.2 ExoIII shortening strategy of pSNS201.

Fig. 3.3a Western blot analysis of exoIII deletion constructs of pSNS201.

Fig. 3.3b Western blot analysis of exoIII deletion constructs of pSNS201.

Fig. 3.4a Untransformed GY7066 plated on MacConkey plates.

Fig. 3.4b GY7066 carrying pSNS206.

Fig. 4.1a A map of pSNS201, shaded region indicates the recombination/rearrangement region in pSNS201.

Fig. 4.1b	PCR amplification using primers and cosmid T256 as a template, of the 400 bp recombination/rearrangement region.
Fig. 4.2	Restriction map of pSBG300/pSNS201.
Fig. 4.3	Cosmids T256 and T276 digested with *Bgl*II. Gel electrophoresis and hybridisation with the *Bgl*II/*Pst*I probe.
Fig. 4.4a	Restriction pattern and subclones derived from pSBG100.
Fig. 4.4b	Restriction map of pSBG101, pSBG102, pSBG300/pSNS201.
Fig. 4.5	Restriction pattern of pSBG100 confirmed by gel electrophoresis.
Fig. 4.6	Sequencing strategy of pSBG300 and pSBG100.
Fig. 4.7	The complete nucleotide sequence of the contiguous *Pst*I fragments of pSBG300 and pSBG100 of cosmid T256.
Fig. 4.8	Analysis of the sequence of pSBG400 for ORFS in all phases using DNA Strider.
Fig.4.9	The complete nucleotide sequence and deduced proteins for ORF463.
Fig. 4.10	Comparisons between the *E. coli recA* promoter region and the regulatory region of ORF463.
Fig. 4.11	Hydrophobicity profile obtained with DNA Strider of ORF463.
Fig. 4.12	Codon usage of ORF463 compared to the *recA* from *M. tuberculosis.*
Fig. 4.13a	Codon preference programme of ORF463 compared to the 65 kDa protein of *M. tuberculosis.*
Fig. 4.13b	Analysis of the codon usage of *recS* by the TestCode programme.
Fig.4.14a	Amino acid sequence alignment between the *E. coli recA*, the entire ORF of the *M. tuberculosis recA* and ORF463.
Fig. 4.14b	Amino acid sequence alignment between the ORF463 and the *M. tuberculosis recA* without its "intron" region.
Fig. 4.15	Genomic DNA isolated from different species of mycobacteria, digested with *Sma*I.

Fig. 4.16 *M. tuberculosis* H37Rv cosmid DNA digested with *Sma*I and *Pst*I and hybridisation with the internal *Bgl*II/*Pst*I fragment.

Fig. 4.17 Hybridisation of the Southern blot of Fig. 4.16 with the insert of pSBG102.

Fig. 5.1 Mode of repair of H_2O_2 induced DNA damage.

Fig. 6.1 The effect of the recombinant plasmid pSNS100 on the survival of *E. coli* DK cells which were plated on various concentrations of EtMes.

Fig. 6.2 The effect of the recombinant plasmid pSNS100 on the survival of *E. coli* DK cells which were exposed to a UV irradiation of 254nm.

Fig. 6.3 Restriction map of the cloned *Pst*I fragment of pSNS100.

Fig. 6.4 Sequencing strategy used to sequence pSNS100.

Fig. 6.5 Nucleotide sequence of the *Pst*I insert of pSNS100.

Fig. 6.6 Analysis of the sequence of pSNS100 by DNA Strider for open reading frames.

Fig. 6.7 Nucleotide and deduced amino acid sequence of the *Pst*I fragment of pSNS100.

Fig. 6.8 Alignment of COOH terminal domains of *katG* genes from different mycobacteria and ORF242.

Fig. 6.9 Southern blot of DNA isolated from different species of mycobacteria, digested with *Kpn*I and probed with the *Alu/Pst*I fragment of pSNS100 containing a domain of the *katG* gene.

ABBREVIATIONS

A	adenine
$Ap^{(R)}$	ampicillin (resistance)
ATP	adenosine triphosphate
BET	Ethidium bromide
bp	base pair(s)
C	cytosine
DNA	deoxyribonucleic acid
ds	double stranded
e.o.p.	efficiency of plating
EtES	ethyl ethane sulphonate
EtMes	ethyl methane sulphonate
G	guanine
kb	kilobase pair(s)
kD	kiloDalton
LA	Luria Agar
LB	Luria Broth
M9 MM	M9 minimal medium
M_r	relative molecular mass
MES	methyl ethane sulphonate
MMS	methyl methane sulphonate
ORFs	open reading frame(s)
PCR	polymerase chain reaction
Pol I	polymerase I
Pol II	polymerase II
Pol III	polymerase III
RNA	ribonucleic acid
rRNA	ribosomal RNA

RNase	ribonuclease
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
$Sm^{(R)}$	streptomycin resistance
ss	single stranded
T	thymine
UV	ultraviolet

ABSTRACT

A *Pst*I genomic library of *Mycobacterium tuberculosis* H37Rv, constructed in pEcoR252, and used to transform the *E. coli recA* mutant *(E. coli* DK11), was screened in the presence of the chemical mutagen ethyl methane sulphonate (EtMes), for resistant clones. Two recombinant plasmids were identified, pSNS201 and pSNS100, which, on repeated transformation of *E. coli* DK1 cells, confirmed consistent resistance to EtMes. Further characterisation of these clones revealed the following:

A 3.4 kb *Pst*I insert was contained within pSNS201 which conferred resistance to EtMes and to long wave (302 nm) but not short wave (254 nm) UV radiation on *E. coli* DK1 cells. Homologous recombination functions were restored by pSNS201 to the *recA* mutant *E. coli* HB101. *E. coli* DK1 (pSNS201) produced a protein of 38-40 kDA which on Western analysis cross-reacted with *E. coli* polyclonal RecA antiserum.

Nucleotide sequence analysis of pSBG300 (which contains the equivalent *Pst*I insert of pSNS201), revealed 4 major ORFs. Screening of these ORFs against the Genebank database using the Fasta programme revealed no homology with any known or putative proteins. RecA activity was located to the first ORF. This ORF was truncated and lacked the 5' end. Recovery of the 5' end and the regulatory region of ORF463 was achieved using *M. tuberculosis* H37Rv cosmids. The complete ORF (ORF463) was examined and compared to *M. tuberculosis* and *E. coli recA*. Amino acid sequence analysis of ORF463 revealed 13% identity to the *E. coli* and cloned *M. tuberculosis* RecA. A conserved, ATP binding domain was observed in ORF463. ORF463 showed no significant DNA or amino acid similarities with other cloned *recA's*, and was thus designated *recS* and the putative protein product RecS. Examination of the regulatory region of *recS* showed no

motifs with identity to the LexA repressor binding site of *E. coli*, the SOS box. Hybridisation studies using an internal fragment of *recS* as a probe showed that *recS* is duplicated in the genome of both the virulent strain of *M. tuberculosis* H37Rv and the avirulent strain H37Ra. However, only one copy is present in BCG. *recS* was also observed in all other mycobacterial species: *M. smegmatis, M. fortuitum, M. aurum* and *M. kansasii.*

Functional characterisation of the second recombinant pSN100 showed that it conferred resistance to EtMes on *E. coli* DK1 cells and resistance to UV radiation (both long and short wave length). This resistance is much higher than that described for pSNS201. Sequence analysis revealed 3 ORFs. ORF242 was shown to be a truncated *katG* gene and ORF168 and ORF63 did not encode any putative proteins with identifiable functions.

SECTION A

CHAPTER 1

Pages

GENERAL INTRODUCTION

SECTION A1

THE *E. COLI MODEL*

1.1 DNA DAMAGE AND INDUCED DNA DAMAGE 20

1.1.1 U.V. DNA damage 20

1.1.1a Photoreactivation 21

1.1.1b Excision repair 21

1.1.1c Recombination repair 22

1.1.2 Alkylating agents, 23

1.1.2a Alkylating agents and DNA alkylation products 23

1.1.2b Mechanisms for transversions 24

1.2 THE REC A PROTEIN: SOS RESPONSE AND GENETIC RECOMBINATION 25

1.2.1 Major historical events relating to the SOS response 26

1.2.2 Molecular genetics of *recA* 31

1.2.3 Location of the Rec A protein 32

1.2.4 The RecA protein: folding and functions 34

1.2.5 Regulation of expression of the *rec A* gene 36

1.2.6 Mysteries of the RecA protein 38

1.3 THE ADAPTIVE RESPONSE 41

1.4 GENETIC RECOMBINATION 41

1.4.1 General recombination 42

1.4.2 RecA protein and strand transfer 44

1.4.3 RecBCD recombination pathway 45

1.4.4 RecE and RecF pathways 46

SECTION A2

THE MYCOBACTERIA

1.5.1 Nomenclature and characterisation 47

1.5.2 Pathogenic mycobacteria 48

1.5.3 Genetic recombination and DNA repair 49

CHAPTER 1

GENERAL INTRODUCTION

Tuberculosis is a destructive, debilitating disease, and is responsible for one in four of all avoidable adult deaths in developing countries. Approximately 1.7 billion people are infected with *Mycobacterium tuberculosis* and annually three million deaths result from tuberculosis (Kochi, 1991). The treatment of patients suffering from tuberculosis is prolonged and expensive, and requires the use of combinations of antibiotics to prevent the emergence of drug resistant mutants of the causative organism *M. tuberculosis*. At present the only feasible tuberculosis control method is mass BCG vaccination. However, the ability of BCG to protect against tuberculosis has varied considerably in a number of human vaccine trials (ten Dam, 1984). The emergence of multi-drug resistant strains (Bloom, 1992) indicates the need for continued efforts towards newer and better drugs (Nordeen and Godal, 1988).

Tuberculosis is not the only major mycobacterial disease, ten to fifteen million people suffer from leprosy caused by *Mycobacterium leprae* (Bloom & Godal, 1983), while the *Mycobacterium avium*, *Mycobacterium intracellulare* complex (MAC) organisms are the most common bacterial isolates and the most frequent cause of systemic bacterial infections in patients with acquired immunodeficiency syndrome (AIDS) in the United States (Young, 1988). Our knowledge of the genetics and molecular biology of the genus *Mycobacterium* lags behind that of *E. coli* due to a number of reasons. First, their slow growth rate does not make them attractive for study by the geneticist; second, systems for the exchange of genetic material between mycobacterial strains have not proved nearly

so easy to establish as in *E. coli*, and third, they are major pathogens. The use of molecular biological techniques, however, will overcome these problems and lead to the cloning of enzyme targets for drugs; the understanding of pathogenicity; and the development of recombinant vaccines. Mycobacterial DNA has been cloned and expressed in *E. coli*. (Hopwood *et al.*, 1988). However, it has been shown that mycobacterial promoters are recognised poorly or not at all by the *E. coli* transcriptional-translational machinery, thus there is a need to find a more suitable surrogate host capable of expressing many mycobacterial genes relevant to pathogenesis.

Genetic analysis of *E. coli* has identified pathways and genes whose products participate in the survival of the bacterial cell under adverse conditions. These include DNA repair enzymes which have also been implicated in the process of genetic recombination (Imlay and Linn, 1988; Walker, 1984). The major regulator of these processes is the multifunctional RecA protein. This protein is known to play an essential role in DNA repair and recombination, having both protease and recombinase activities. This unusual protein has recently been crystalised by Story *et al.* (1992), thus offering some suggestions as to how this protein uses its structural forms to modulate its functional diversity. Little is known about recombination and repair systems in Mycobacteria and it is possible that an efficient DNA repair system in the slow growing pathogenic mycobacteria may have evolved to protect the integrity of the genome during survival within the radical-rich environment of the macrophage cells in the host.

Genetic recombination, DNA repair, and antioxidant defense mechanisms have been extensively studied in *E. coli*, thus proving to be invaluable in investigations of, and comparisons with, other bacterial systems. In an attempt to understand how these pathways function in the genus *Mycobacterium*, the *E. coli* systems are used as a model and are described in this chapter.

1.1 DNA DAMAGE AND INDUCED DNA REPAIR

The biochemical components of DNA are susceptible to damage by a variety of agents that may break the phosphodiester backbone, cleave N-glycosidic bonds, alter purine or pyrimidine structures, as well as covalently cross-link DNA strands. The bacterial response to DNA damage is dramatic. It involves an active cellular system in which the damaged DNA is processed. This processing is referred to as "error-prone repair", "SOS repair", or misrepair (Radman, 1974; Walker, 1984-5-7).

In *E. coli*, inducible genes which code for DNA repair proteins are members of two major regulatory pathways:

The SOS error-prone repair system which is regulated by the RecA and the LexA proteins;

The adaptive response which is controlled by the Ada protein (Walker, 1985).

In order to understand the repair pathways mentioned above, examples of DNA damage involving specific repair mechanisms are given below.

1.1.1 UV DNA damage

Pyrimidine dimers which arise from the formation of a cyclobutane ring across the 5-6 double bond of adjacent pyrimidines are the best studied effects of UV (254 nm) induced DNA damage. Thymine-thymine dimers are formed preferentially, however, cytosine-thymine and cytosine-cytosine dimers also occur (Gordon and Heseltine, 1982). It seems likely that the exposure to UV light generates additional lesions including single and double strand breaks (Kushner, 1987).

A variety of repair mechanisms exist in *E. coli* by which UV induced photoproducts are recognised and repaired. These repair mechanisms are specific for certain photoproducts and they can be either error-free or error-prone as part of the SOS response (Kushner *et al.*, 1978). Examples of these repair systems are described below.

1.1.1a Photoreactivation.

This process involves the enzymatic cleavage of cyclobutane dimers in the presence of visible light (Rupert *et al.*, 1958). The process of photoreactivation is an error-free mechanism. Two photoreactivating enzymes photolyase R (Sutherland and Chamberlin, 1973) and a second similar enzyme (Snapka and Sutherland, 1980) which bind to cyclobutane dimers, have been purified from *E. coli*.

1.1.1b Excision repair

Excision repair is a process in which the UV induced pyrimidine dimers are excised via one of the following mechanisms:

a) the introduction of incisions at or near the site of the lesion;

b) the excision of the fragment of the DNA containing the lesion.

This is then followed by the resynthesis of the excised DNA using the complementary strand (Walker, 1985).

Three proteins have been identified to play an important role in excision repair: they are the products of the *uvrA* (Sancar *et al.*, 1981a), *uvrB* (Sancar *et al.*, 1981b) and *uvrC* genes (Sancar *et al.*, 1981c).

The reaction of the UvrABC protein complex involves three separate steps. UvrA interacts with ATP to form a protein dimer which binds to the

damaged DNA. UvrA has a high affinity for single-stranded and UV irradiated DNA. The UvrB then binds to the UvrA-DNA complex. The binding of UvrC to the stable DNA-UvrA-UvrB complex, results in phosphodiester bond cleavage (Yeung *et al.*, 1983). In the presence of helicase II, a product of the *uvrD* gene, the new complex UvrABC excinuclease is released (Ogawa *et al.*, 1968; Kuemmerle and Masker, 1980). Enzymes encoded by the *uvrA* (Kenyon and Walker, 1981), *uvrB* (Fogliano and Schendel, 1981) and *uvrD* are under the control of the LexA repressor protein. UvrABC excinuclease also recognises other photoproducts and is not specific for UV induced pyrimidine dimers (Husain *et al.*, 1985). Thus it would appear that the specificity of the UvrAB proteins is their ability to recognise alterations in the DNA helix (Kushner, 1987).

1.1.1c Recombinational repair (Postreplication repair)

Recombination also plays a role in the ability of bacterial cells to survive exposure to DNA damaging agents such as UV irradiation and chemicals. These agents induce single or double stranded breaks in DNA; such lesions are repaired by DNA recombination (Fig. 1.5e). This type of recombination is referred to as recombinational repair. The role of RecA protein (described in more detail in section 1.2) in recombinational repair is illustrated by experiments which show that RecA mutants are defective in recombinational repair (Weinstock, 1987).

Pyrimidine photoproducts can cause a temporary cessation of DNA synthesis presumably because the replication complex is unable to use the nucleotides of the photoproduct as a template. This results in DNA synthesis moving on to the next Okasaki fragment and the gap that is left has to be filled in by genetic recombination. This hypothesis, first suggested by Rupp and Flanders (1968), was supported by initial evidence from studies of UV survival of strains deficient in genetic recombination and excision repair. The *uvrA recA* double mutants were more sensitive to UV light than either of the single mutants,

suggesting a role for genetic recombination in the repair of UV-induced photoproducts (Howard Flanders and Boyce, 1966). The process of postreplication repair does not remove photoproducts from the DNA but simply dilutes them out and it is possible that these photoproducts are not recognised by other repair systems. It is also possible that this class of DNA lesions do not make distortions in the DNA helix large enough to be recognised by the UvrA and UvrB proteins (Kushner, 1987).

1.1.2 Alkylating agents

1.1.2a Alkylating agents and DNA alkylation products

Ethyl methane sulphonate (EtMes), ethyl ethane sulphonate (EtES) and methyl methane sulphonate (MMS) are alkylating agents known to transfer ethyl or methyl groups to a variety of positions on the purine and pyrimidine rings as well as to the phosphodiester backbone. The major reactions between alkylating agents and nucleotides occur at N7 of guanine, at N1 and N3 positions of adenine and at the N1 position of cytosine and thymine. The N7 alkylation product readily ionises, thus the loss of a proton could result in thymine:guanine pairing through 2 hydrogen bonds (Lawley and Brooks, 1961). However, alkylations at lower frequencies are also known to occur at any nitrogen, oxygen or phosphorous molecule (Drake, 1970).

EtES and EtMeS strongly induce G:C-A:T transitions as observed in bacteriophage T4 (Bautz and Freese, 1960). It is likely that the guanine reaction will produce a G:C-A:T transition (Lawley, 1966). However, transversions and probably deletions can be induced by EtMes (Malling and De Serres, 1968). A much wider range of base pair substitutions were detected in *E. coli* tryptophan mutants exposed to EtES (Yanofsky *et al.*, 1966). The transition G:C-A:T was not

observed at all whereas the transition A:T-G:C was strongly induced. The transversions A:T-T:A and G:C-C:G were also observed.

Tessman *et al.*, (1964) suggested that the alkylated pyrimidines often mispair and that alkylation of adenine and pyrimidine ring nitrogens (which are involved in base pairing) would interfere with pairing of any kind. Thus it is possible that alkylated bases could block hydrogen bonding to a certain extent, but still allow the insertion of a purine opposite an alkylated pyrimidine with the increased probability of an erroneous insertion (Drake, 1970).

1.1.2b Mechanisms for the production of transversions

The treatment of DNA with heat (at a low pH) or with an alkylating agent results in depurination (Bautz and Freese, 1960). Alkylation induced depurination of guanine is a relatively slow process (Lawley and Brookes, 1961), thus it is probable that an organism, once treated with an alkylating agent, accumulates post treatment incubation damage. Strauss (1962) observed reversion of *E. coli* auxotrophs following treatment with EtMeS but not with MMS.

Mechanisms for generating transversions can be divided into either a null-base scheme or a mispairing scheme.

Null-base schemes are due to modifications of the DNA resulting in the entire removal of a base (depurination) or the extensive degradation of a base.

Mispairing schemes depend upon mispairing between normal bases or mispairing promoted by the chemical modification of a normal base (Drake, 1970). The treatment of cells with chemical DNA damaging agents elicits an increased capacity for DNA repair and stable DNA replication. Mutagenesis by UV and chemical agents process damaged DNA in such a way that mutations result. This processing is often referred to as the "error-prone repair", "SOS repair", "misrepair" or "SOS processing" (Walker, 1987).

1.2 THE REC A PROTEIN, SOS RESPONSE AND GENETIC RECOMBINATION

The SOS response controls the expression of genes involved in recombination and DNA repair (daughter-strand gap repair, double-strand break repair), excision repair, mutagenesis (error-prone repair) and mismatch repair of DNA (Walker, 1985).

The following section deals with the role(s) of the RecA protein in the SOS response in *E. coli* and is traced from the discovery of this protein to our current knowledge of the regulation of this DNA damage-inducible system.

"Few proteins rival the biochemical versatility of the RecA enzyme found in E. coli.... It possesses an extraordinary portfolio of activities that range from homologous recombination to mutagenesis and control of gene expression during periods of cellular stress" (McEntee, 1992).

The product of the *recA* gene was first shown in 1965 to play an essential role in genetic recombination and in resistance to UV irradiation (Clark and Margulies, 1965). It has since been shown to play an important role in the induction of prophage (Clark and Margulies, 1965) and in the repair of various kinds of damage to DNA (Witkin, 1976 and 1991).

The SOS gene expression is controlled by a complex interaction between the *lexA* and the *recA* gene products (Fig. 1.1) (Walker, 1985). RecA protein binds to ssDNA and to a nucleotide triphosphate cofactor: dATP, ATP or ATP(-S). In this ternary complex the activated form of RecA (RecA*) mediates the inactivation of the LexA repressor, by cleavage at an ala-gly peptide bond near the middle of the protein, leading to induction of the SOS response.

Inactivation of LexA results in the derepression of the SOS genes, allowing the cell to repair the DNA lesions. As the cells recover, the inducing signals disappear and RecA molecules are no longer activated (Fig. 1.1) (Walker, 1985). Under non-inducing conditions, the LexA protein represses the expression of at least 20 genes in the SOS regulon and binds to a conserved sequence in the promoter region of each gene, the "SOS box", thereby repressing genes of the SOS regulon (Walker, 1985). A simplified model of the SOS regulatory system is given in Fig.1.2. (Witkin, 1991). In undamaged cells there is a low-level synthesis of SOS gene products including LexA and RecA..

1.2.1 Major historical events in the discovery of the SOS response

The history of the major events that have contributed to the discovery of the SOS response, traced back from 1965 are briefly related in the following section.

During the period 1965-1970 (Table1-1), UV induced mutagenesis in *E. coli* provided observations that a functional product of the *recA* gene was required. However, there was no obvious link to homologous genetic recombination (Witkin, 1991).

Table 1.1. Observations which showed requirement for RecA protein in various UV induced activities in *E. coli* and lambda phage (Adapted from Witkin, 1991).

Year	Observation	Reference
1967	Prophage induction by UV	Hertman and Luria
1968	Weigle reactivation and phage mutagenesis	Miura and Tomizawa
1968	Post replication repair	Rupp and Howard-Flanders
1969	Filamentous growth following UV radiation	Green *et al*
1969	Bacterial UV mutagenesis	Witkin
1970	Reinitiation of DNA replication after UV radiation	Worcel

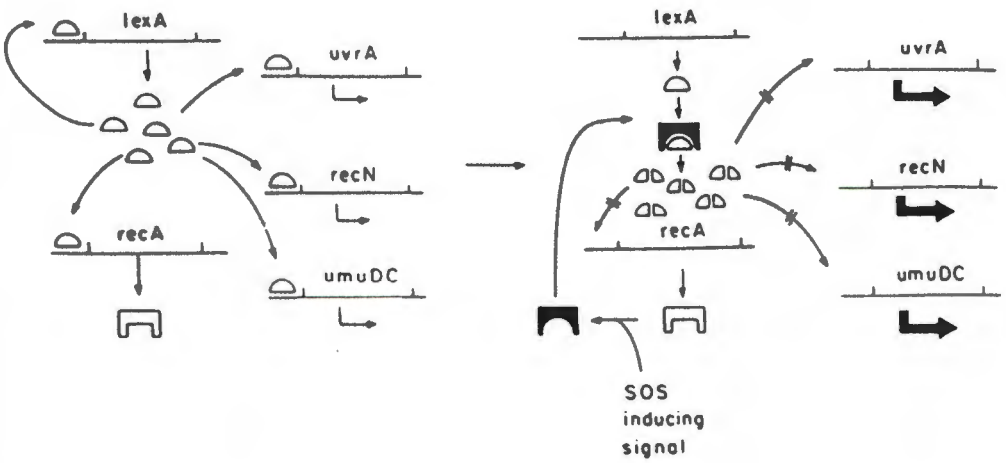


Fig.1.1 Model of the SOS regulatory system. The generation of an inducing signal following DNA damage leads to an activation of RecA. The interaction of activated RecA with LexA results in cleavage of LexA. As the LexA pool decreases, the SOS genes are expressed at higher levels (From, Walker, 1985)

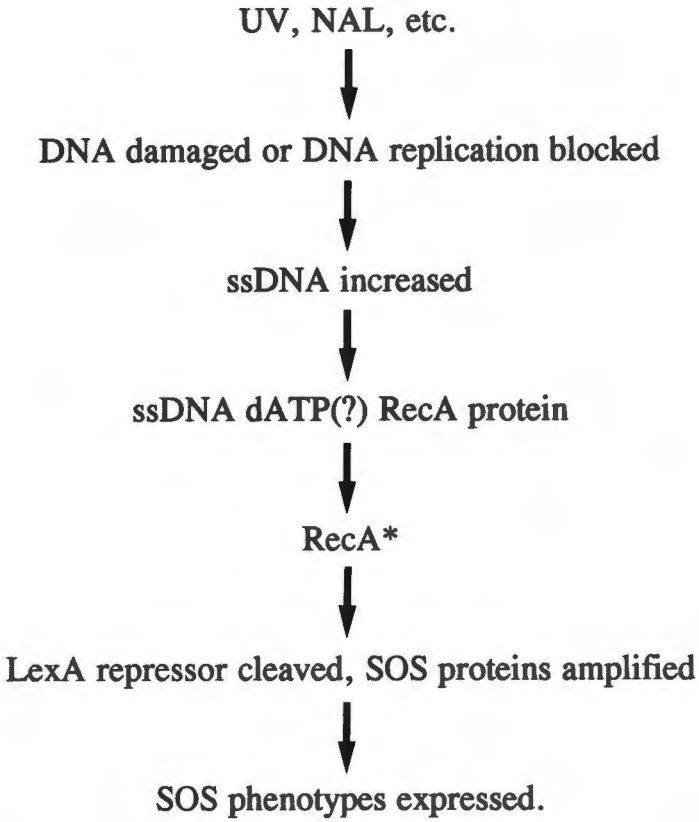


Fig. 1.2 SOS regulation in wild type *E. coli* (Adapted from Witkin, 1991).

In 1967, Witkin found that *lex* mutants were not only UV-sensitive but were also non-mutable by UV. These results suggested that the *lexA* gene probably encodes a modified DNA polymerase which is capable of inserting nucleotides opposite UV lesions. Further investigations proposed that translesion replication is the mechanism of UV mutagenesis.

In 1971, Defais and co-workers suggested that irradiation of bacterial cells induced a *recA lexA* -dependent, error-prone, DNA replication activity that may be responsible for bacterial UV mutagenesis. During that same year Radman coined the term 'SOS replication' to describe the inducible error-prone mode of DNA synthesis relating to UV bacterial mutagenesis. The SOS hypothesis was developed between 1971 and 1974. By this time it was clear that RecA was not

only necessary for recombination but that together with LexA, it was required for the induction of lambda prophage, filamentous growth and UV mutagenesis. Witkin, (1974) showed by using the *tif* mutant (which is an allele of the *recA* gene=*recA441*) that UV mutagenesis was a component of the inducible SOS response. Table 1.1 shows UV induced activities in *E. coli* and the requirement for RecA (Witkin, 1991). Roberts and Roberts, (1975) then demonstrated that lambda repressor was proteolytically cleaved. This suggested that RecA protein might be a protease or control a protease that cleaves repressors. The "First SOS conference on RecA" followed in 1975.

The first model of SOS regulation was based on the studies by Gudas and Pardee, in 1975, of "protein X" (today known as RecA). The model proposed that LexA is the common repressor of inducible SOS genes and that the anti-repressor possibly a protease was "protein X". The late 70's and the early 80's saw a rapid progress in the understanding of the role of RecA in genetic recombination and in the SOS regulation (Table 1.2).

Table 1.2 The role of the RecA protein in SOS regulation (From Witkin, 1991).

Year	Observation	Reference
1977	RecA is protein X	Gudas and Mount
1978	RecA cleaves repressor	Roberts *et al*
1979	RecA is purified and partially characterised	Ogawa *et al*
1980	RecA cleaves LexA protein	Little *et al*
1984	RecA facilitates autodigestion of LexA	Little

In 1988, a second role for RecA* was demonstrated, namely. the processing of UmuD, a protein required for most UV and chemical mutagenesis to generate an active fragment UmuD' in translesion replication (Burkhardt *et al.*, 1988).

RecA* promotes the proteolytic cleavage of the UmuD protein, λcI and LexA by a common proteolytic mechanism called autodigestion (Little, 1987). Little, (1984) has raised the possibility that the interaction of RecA* with LexA may change the conformation of LexA rendering it susceptible to hydrolysis, or alternatively, RecA* may cause conformational changes in LexA facilitating its autodigestion. This mechanism appears to be similar for all of these proteins and it is thought that the RecA protein interacts with the target protein in a positive allosteric manner. The allosteric effector surfaces of RecA protein are non-identical and can be independently altered by mutation (Ennis *et al.*, 1989). New *E. coli recA* mutants have been identified whose RecA proteins mediate cleavage of some but not all of the above protein substrates (Dutreix *et al.*, 1989). Biochemical and structural analyses have suggested that the RecA protein acts as a multimer (Moreau and Roberts, 1984; Peterson and Mount, 1987) and that RecA mutagenesis functions are genetically separable from LexA and cI proteolysis, the latter two activities are also separable from each other (Ennis *et al.*, 1989).

RecA mutants with altered phenotypes made it possible to study, independently, by mutation, the recombinase and protease activities of RecA. In addition to the cleavage of LexA repressor, it is thought that RecA* plays other roles in DNA replication following DNA damage (Table 1.3) (Witkin, 1991).

Table 1.3 SOS inducible phenotypes requiring RecA* for activities other than LexA cleavage (From Witkin, 1991).

SOS phenotype	Required RecA* activity	Reference
SOS mutagenesis	Processing UmuD-UmuD` + a third unknown activity	Shinagawa *et al*, 1988, Sweasy *et al*, 1990
Stable DNA replication	unknown	Kogoma *et al*, 1979
Induced replisome reactivation after UV (IRR)	unknown	Khidir *et al*, 1985
RecA association with cell membrane	unknown	Garvey *et al*, 1985

1.2.2 Molecular genetics of *recA*

The RecA protein is encoded by the *recA* gene which has been mapped on the *E. coli* chromosome. Fig.1.3 shows the genetic map of *E. coli* with most of the genes required for genetic recombination and recombinational repair of damaged DNA (West and Connolly, 1992). Also indicated are the enzymes involved in the maintenance of DNA eg. topoisomerases, ligases and polymerases.

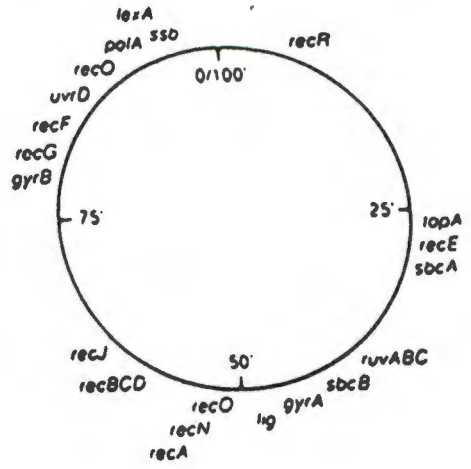


Fig.1.3 A genetic map of the *E. coli* chromosome showing the position of many of the genes needed for genetic recombination and recombinational repair of DNA damage. Approximate locations of the genes are indicated in minutes (From West and Connolly, 1992).

Sancar and co workers (1980) found the *E. coli recA* gene product to be a single polypeptide with a calculated molecular mass of 37.8 kDa. The importance of the *E. coli* RecA protein in recombination as well as DNA repair suggests that RecA analogs could be conserved among procaryotes. Genes encoding RecA-like activity have been cloned from many Gram-negative bacteria, including *Pseudomonas aeruginosa* (Kokjohn and Miller, 1985,), *Shigella flexneri* (Keener *et al.*, 1984), *Vibrio anguillarum* (Singer, 1989), *Proteus vulgaris* (Keener *et al.*, 1984), *Agrobacterium tumefaciens* (Farrand *et al.*, 1989), *Haemophilus influenzae* (Setlow *et al.*, 1988), Thiobacillis ferrooxidans (Ramesar *et al.*, 1988) and the Gram-positive organism, *Bacillus subtilis* (Marrero and Yasbin 1988); these different genes, highly conserved at the DNA level, were cloned by complementation of *E. coli rec*$^-$ mutants. Using hybridisation studies, a *recA* gene was isolated from *M. tuberculosis* (Davis *et al.*, 1991).

1.2.3 The location of the RecA protein

RecA constitutes 3-4% of the total protein content of the cell (Gudas and Pardee, 1976). The majority of the protein (90%) is located within the cytoplasm

presumably involved in DNA repair and recombination. However, in SOS induced cells it has been shown that RecA* is associated with the cell membrane.

The association of RecA* with the cell membrane in SOS induced cells was first observed by Gudas and Pardee, (1976) when RecA was still known as protein X. They found that 10% of the amplified protein X induced in DNA damaged cells remained associated with the membrane in cell fractions. Garvey *et al.*, (1985) have shown that only activated RecA* is associated with the membrane. It is possible that the RecA*-membrane association in SOS-induced cells could be due to its binding to ssDNA attached to the membranes. It is also possible that only the RecA bound to a DNA membrane complex is activated after SOS induction and that the rest of the protein in the cytoplasm probably participates in DNA repair. The DNA bound to the membrane of *E. coli* includes the origin of replication Hendrickson *et al.* (1982). Damaged DNA that cannot initiate replication stimulates a weak, indirect, SOS induction as compared to the induction signal generated by damaged DNA (plasmids etc) that are replicating or attempting to initiate replication (D'Ari and Huisman, 1982). Thus RecA-DNA-membrane complexes may be required for *in vivo* activation of RecA. Such a complex may play an important role not only in SOS induction but also in the expression of SOS phenotypes that require activated RecA for alternate functions to that of LexA cleavage. These include SOS mutagenesis and stable DNA replication. It is interesting to note that the damage-inducible protein UmuC is also membrane associated .

Whether RecA* is transported to the membrane is unknown but it may be due to a RecA* dependent "hitch-hiking" association to either a membrane protein or a cytoplasmic protein whose final destination is the membrane (Garvey *et al.*, 1985).

The effect of SOS induction on the synthesis of three major outer membrane proteins (Garvey *et al.*, 1985) raised new questions about the relationships between the SOS system and cell membrane metabolism. These outer membrane protein changes observed in SOS-induced cells deserve further investigation.

1.2.4 The RecA protein: functions and folding

In addition to its protease activity, the RecA protein catalyses the filling in of DNA gaps and by its capacity of polymerisation (5'-3' polymerisation on ss DNA or duplex DNA having a ss gap to form a helical filament) the pairing and the strand exchange of homologous DNA molecules (Walker, 1985). It is also involved in the hydrolysis of ATP in the presence of ssDNA (Ogawa *et al.*, 1979), and the ATP-dependent uptake of ssDNA by duplex DNA (McEntee *et al.*, 1979). RecA is activated *in vitro* when it forms a complex with single-stranded DNA and a nucleoside triphosphate. Roberts and Devoret, (1983) have proposed that single-stranded regions generated by SOS inducing treatments could be part of the *in vivo* signal for SOS induction. These same authors have also suggested that RecA is activated when it binds to "gaps caused by the replication fork encountering a lesion".

Story *et al.*, (1992) recently provided the crystal structure of the RecA protein at 2,3A° resolution (Fig.1.4). RecA protein consists of a major central domain, flanked by two smaller sub-domains at the amino and carboxy termini respectively, which protrude from the protein and stabilise the formation of the polymer and interpolymer bundles respectively. An important feature for an understanding of the mechanism of RecA protein function is the location of the site(s) of DNA binding. RecA protein is thought to have 2 DNA binding sites that

can bind 2 DNA molecules, one for primary binding to ss or gaped duplex DNA, and the other for binding to homologous duplex DNA (Flory *et al.*, 1984; Stasiak *et al.*, 1981). One would expect these regions to be on the surface of the RecA filament and to be highly conserved (Fig 1.4).

Evidence from the crytallographic studies has suggested that residues around the disordered loops L1 (157-164) and L2(195-209) are for DNA binding. Mutant *recA430* which has a mutation Gly204→Ser indicates that this region is involved in ss DNA binding. Larminat *et al.*, 1992 have recently shown that mutations in the vicinity of the L2 disordered loop play a role in the conformational modifications of the protein and thus affect RecA functions to various degrees. Residues around L1 may be involved in the binding of homologous duplex DNA necessary for recombination and mutations in and around L1 affect DNA binding.

Helix G is implicated in binding ssDNA as it is on the most conserved region of bacterial *recA* sequences. At the amino terminal (211-212), there are two conserved glycines which could play a structural role in mediating ATP induced conformational change. A possible binding site for LexA, UmuD and phage repressors may be in the "notch" region between the adjacent lobes of the RecA polymer. Mutant RecA phenotypes can be explained by the assumption that contacts between RecA monomers both interpolymer and intrapolymer are biologically relevant (Story *et al.*, 1992).

The co-operative polymerisation of the RecA protein in the presence of ATP on ssDNA forms a helical nucleoprotein (the pre-synaptic phase) (Flory *et al.*, 1984; Stasiak *et al.*, 1981). Synapsis involves alignment of the nucleoprotein filament with duplex DNA and finally strand exchange. The details of the ability of nucleoprotein filaments to search for homology and the pairing with the duplex DNA, which is bound by other proteins, is not clearly understood. This issue has been addressed by Muniyappa *et al.*, (1991), who have investigated the role of

nucleosome-like structures on homologous pairing, specifically the HU protein. Their results indicate that the binding of HU protein to linear duplex DNA differentially affects homologous pairing *in vitro*.

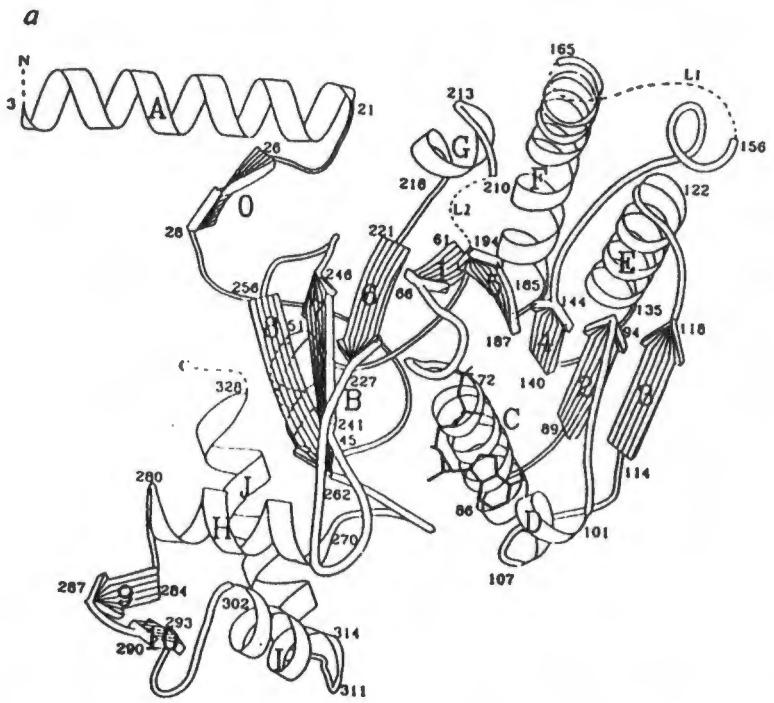


Fig.1.4 Structure of the RecA monomer. Schematic representation of the three-dimensional structure of the RecA protein (using the computer programme RIBBON). β strands are numbered 0-10 and α-helices as A-J. Dashed lines indicate the two disordered loops (L1 and L2), proposed to be involved in DNA binding (From Story *et al.*, 1992).

1.2.5 Regulation of expression of the *recA* gene

Regulation of the RecA protein is under the control of the LexA protein which binds to the SOS box in the promoter region of the gene. Repression of *recA* by the LexA protein has been demonstrated both *in vitro* as well as *in vivo* (Little *et al.*, 1981; Sassanfar & Roberts, 1990; Phizicky & Roberts, 1981), and these studies have shown that LexA recognises a site in the *recA* promoter which lies between the -35 and -10 promoter region. Table 1.4 lists some recent important

and historical highlights in the analysis of the *recA* gene. The relationship between the LexA and RecA proteins at the DNA level is discussed below.

Table 1.4 Analysis of the *recA* gene (Modified from Witkin, 1991).

Year	Observation	Reference
1980	RecA gene sequenced	Hori *et al*; Sancar *et al*
1981	20 bp SOS box identified LexA protein binds recA and lexA operators	Brent& Ptashne, Little *et al*
1984-	Mutant alleles of recA seq definition	*Kawashima et al*
1990	active domains of RecA protein	*Dutreix et al*
1991	novel recA gene isolated from *M.tuberculosis*	*Davis et al*
1992	L2 loop modulates recombination/coprotease activity	Larminat *et al*
1992	crystal structure of RecA	Story *et al*

Following the sequencing of the *recA* gene and the identification of its 20bp SOS box, it was shown that purified LexA protein binds both the *recA* and *lexA* operators. Sequencing of the *recA* gene uncovered a region with very strong homology to the consensus sequence for *E. coli* promoters (Sancar *et al.*,1980; Horii *et al.*,1980). It has been shown in *E. coli* that a promoter corresponding to a consensus is not as strong as a promoter differing in one position in the region between -35 and -10 (Grana *et al.*, 1988). It is thus possible that the single change observed from consensus in the *recA* promoter may actually strengthen this promoter.

The ability of the cell to produce high levels of RecA as part of the SOS response is very important. Also important is the ability to maintain a significant level of basal (uninduced) protein. How is this basal level of expression maintained? Is it due to a second LexA independent promoter, or to incomplete

repression by LexA or perhaps another protein involved in the regulation of *recA*? The isolation, characterisation and sequencing of mutant alleles of the *recA* promoter region have indicated DNA sequences important for *recA* transcription and have raised the possibility that a large part of basal expression was due to another promoter region (Weisemann and Weinstock, 1991). Single or double mutations constructed within the promoter region of the recA gene did not shut down the *recA* promoter. Two models have been proposed for the expression of *recA*. In the first model, a main promoter is regulated by the SOS response and a second weaker promoter, that is independent of the LexA repressor, is also functional. This second promoter hypothesis is supported by the observation that double mutations do not reduce expression to the extent expected. The second model predicts a single functional promoter controlled by LexA in the following way: LexA controls the formation of the open promoter complex, but not the binding of RNA polymerase (Weisemann and Weinstock, 1991).

1.2.6 RecA protein mysteries

The ability to recover a normal rate of DNA synthesis after replication blockage by UV photoproducts is an inducible SOS activity requiring an unidentified function of RecA. This recovery is called "induced replisome reactivation" (IRR) (Khidir *et al.*, 1985), and is also referred to as "replication restart" (Echols, 1982). This process requires the amplification of RecA and the synthesis of at least one other protein and is not dependent on *recB, umuC* or *uvrA*. Defects in the ability to recover post-UV replication activity were studied using recombination proficient *recA* mutants, *recA718* and *recA430* (Witkin *et al.*, 1987). The replication recovery after inhibition in RecA718 was totally dependent on the presence of UmuDC protein whereas other RecA proteins can perform replication recovery without the presence of functional Umu proteins. RecA430 strains are UV non-mutable and unconditionally deficient in their ability to restart DNA replication after UV irradiation and also have a reduced ability to bind DNA at lesion sites.

Thus it would appear that post UV replication requires an additional un-identified function of RecA.

An inducible SOS function called "stable DNA replication" has been described by Kogoma *et al.*, (1979). Induced stable DNA replication (iSDR) differs from constitutive stable DNA replication (cSDR) in that transcription is required to reinitiate replication in cSDR in the absence of RecA, whereas transcription is unnecessary in iSDR which occurs in the presence of RecA*. Magee and Kogoma, (1990) have shown that iSDR occurs constitutively at 42°C in a *recA441* strain but not in its *recB* derivative, thus they concluded that the RecBCD enzyme is required for iSDR.

iSDR- determines the ability of SOS-induced cells to utilise secondary origins of replication and this same ability may be reflected by the ability to restart UV-blocked replication. However, for the onset of iSDR a prerequisite could be the unblocking of a replication fork. A possibility is that a *recB* mutation prevents iSDR indirectly by decreasing the size of ssDNA regions. Khidir *et al.*, (1985), have shown that replication restart after UV does not require the RecBCD enzyme. It therefore would be interesting to see whether RecBCD is necessary for the operation of iSDR (Cooper, 1982).

Recovery of DNA synthesis is far more rapid in excision proficient *uvr*$^+$ strains than in *uvr*$^-$ strains. Thus it would be interesting to determine whether long-patch excision repair (which occurs primarily at replication forks) affects the replication restart process.

It is clear that DNA replication is profoundly altered in SOS-induced cells; we still need to determine how RecA* alters reinitiation to re-activate lesion blocked replication forks in addition to inserting nucleotides opposite non-coding template lesions (Witkin, 1991).

Nohmi *et al.* (1988) and Dutreix *et al.* (1989), examined the possibility of a third unknown RecA* function in SOS mutagenesis and they have concluded that a third RecA* role is necessary for UV mutagenesis. However, this third role has not yet been defined. A possible role for the inhibition of DNA polymerase proofreading *in vivo* may occur. An additional role for RecA in SOS mutagenesis is probably a direct role in translesion DNA replication (Sweasy *et al.*, 1990; Nohmi *et al.*, 1988; Dutreix *et al.*, 1989). Several additional roles have been proposed for RecA in SOS mutagenesis, namely the activation of another protein proteolytically, inactivation of a repressor other than LexA or the probable interaction with DNA polymerase III which has been implicated in SOS mutagenesis (Hagensee *et al.*, 1987; Bridges *et al.*, 1976).

It is possible that RecA* may be necessary to cleave an unknown repressor or to activate another protein as a prerequisite for SOS mutagenesis; or based on the ability of RecA to bind dsDNA containing lesions, RecA* may provide lesion recognition for the UmU proteins. RecA* could participate directly in translesion synthesis by interacting with components of the replisome or with the Umu proteins to form a multiprotein "mutasome"(Woodgate *et al.*, 1989), or by binding to the target and altering the configuration of the lesion to facilitate misincorporaton (Witkin, 1991).

Finally it is possible that proteins other than RecA and LexA may play a role in the induction of the SOS response either directly or indirectly. It was found that strains carrying mutations in the ssb (encoding single-strand DNA binding protein) have defects in a number of SOS responses (Liebermann and Witkin, 1981).

1.3 THE ADAPTIVE RESPONSE

Some bacterial cells use a system which does not involve the SOS response to repair DNA damage. Evidence for the existence of this response was first suggested by Samson and Cairns, (1977). They observed that *E. coli* cells exposed to low concentrations of the methylating agent, N- Methyl-N`-nitro-N-nitrosoguanidine became resistant to a subsequent challenge with a higher dose of this agent. This induced resistance that repairs DNA damage from methylating and ethylating agents is termed the adaptive response. The response is regulated by the *ada* gene product (a 37 kDa protein) and at least four other genes have been observed in the *ada* response (Walker, 1987).

The adaptive response is a RecA independent system and is specifically produced by ethylating or methylating agents since UV light and agents that induce the SOS response do not induce the adaptive response. Two types of DNA repair enzymes are induced during the adaptive response. A broad spectrum DNA glycosylase is produced which initiates excision repair of methylated bases (Karran, 1982) and the second type of repair activity involves methyltransferases which act directly on ethyl and methyl groups, removing them from the DNA molecules. This process involves self-methylation and results in the suicide inactivation of the methyltransferase and each molecule can thus only act once (Walker, 1987).

1.4 GENETIC RECOMBINATION IN *E. COLI*

Genetic recombination is at the heart of the adaptation and evolution of bacterial populations. It is responsible for the lateral transfer of genetic information and the repair of mutated genes which are essential for survival. Recombination processes include illegitimate or site specific recombination which depends on special DNA sequences and general recombination which depends upon homologous DNA sequences.

The following section concentrates on general recombination where the versatile RecA protein, with its recombinase and protease activities, plays a central role.

1.4.1 General recombination

General recombination is a process whereby genetic information is exchanged via a physical exchange of strands of DNA during recombination. This genetic exchange requires the presence of homologous DNA sequences, thus the process is sometimes referred to as homologous recombination. Basically, recombination occurs by the breaking and rejoining of strands of DNA at regions of homology, thus the recombinant structure formed depends upon how the DNA fragments are reconnected (Fig. 1.5). Strand transfer occurs at the heteroduplex joint which is an intermediate for all models of recombination. This joint was first proposed by Holliday (1964) (Fig1.5 c,e) and formation of the joint enables general recombination.

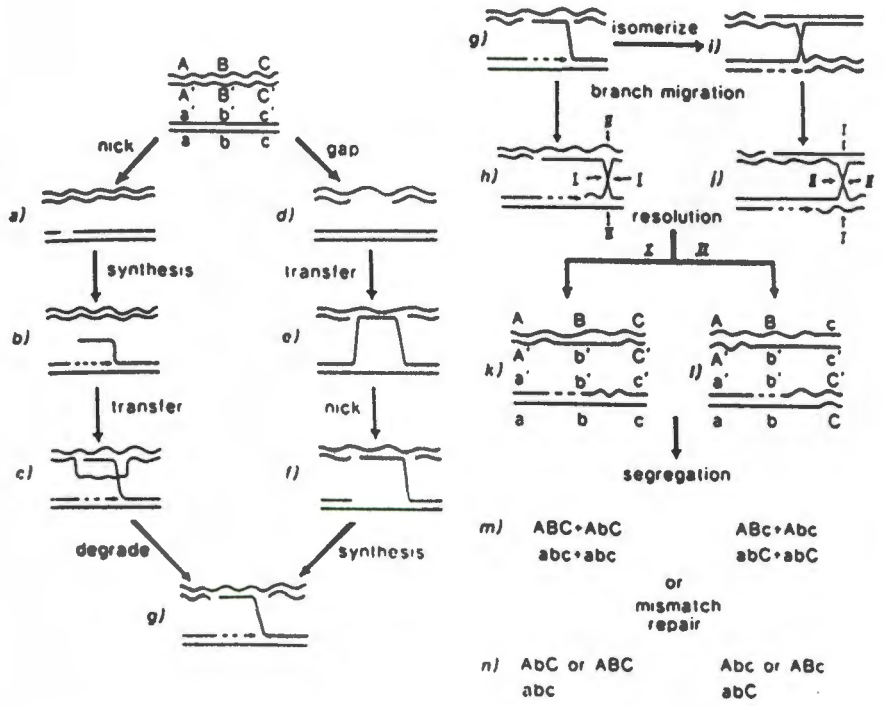


Fig. 1.5. Schematic representation of DNA rearrangements during recombination. Shown, is recombination that differs at three loci (A or a, B or b, C or c) (from Weinstock, 1987).

Following DNA synthesis and strand cleavage (Fig; 1.5 b,c,g and 1.5 e, f, g), a central recombination intermediate is formed (Fig. 1.5 g).

The process of branch migration occurs by breaking base pairs adjacent to the cross over point and reforming them with the opposite parent strand (Fig. 1.5 h), and by creating more heteroduplex DNA by branch migration (Fig. 1.5 j). Breaks occur at the position of the cross over joint, thus determining the nature of the recombinants (Fig. 1.5 k). Resolution can be processed in one of two ways. Firstly, molecules are not recombined for markers outside the heteroduplex region (Fig. 1.5 k,m, n). Secondly, molecules are recombined for outside markers (Fig. 1.5 k,mn). Segregation results in the recombinants Abc, ABc and abC (Fig. 1.5 n). The heterozygous regions in the heteroduplex DNA are susceptible to mismatch

repair in which nucleotides on one strand are excised and replaced with those that are complementary to the other strand (Fig. 1.5 n).

The above model illustrates some of the key steps that occur during recombination in *E. coli*, however, different forms of intermediate DNA structures, combined with different enzymatic reactions at one or more of the steps indicate that more than one pathway exists. Two major routes described for recombination after conjugation are the RecBC and the RecF pathways, of lesser importance is the RecE pathway. These pathways require the RecA function for recombination (Weinstock, 1987).

1.4.2 RecA protein and strand transfer

The RecA protein alone catalyses the central steps in recombination: the pairing and strand exchange of homologous DNA molecules. It is indispensable for homologous recombination because it is the enzyme that helps form the heteroduplex region in the cross over intermediate. Different recombining systems may vary in their modes of initiation and resolution, however, they all depend upon the formation of the heteroduplex joint (Weinstock, 1987).

The assimilation of ssDNA into a homologous double stranded molecule to form a D loop (3 strand structure) accompanied by ATP hydrolysis explains the strand transfer reaction. In *E. coli* single stranded DNA can be complexed to single stranded binding protein (ssB) (Fig. 1.6). The ssB-DNA complexes have been shown to be substrates for RecA promoted strand transfer. ssDNA plays a critical role in the initiation of recombination and it is to this ss region that RecA first binds. Fig. 1.6 shows the role of RecA in strand exchange. It is thought that RecA protein may also facilitate branch migration (Weinstock, 1987).

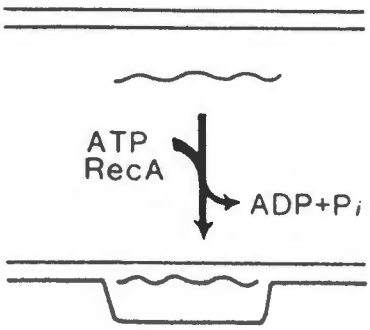


Fig. 1.6. Formation of a D loop by RecA protein (from Weinstock, 1987).

1.4.3 The RecBCD recombination pathway

Apart from the RecA protein, the RecBCD enzyme is the principal component of the major route of homologous recombination (Chaudhury and Smith, 1984). The *recB*, *recC* and *recD* genes encode subunits of the RecBCD enzyme, exonucleaseV (Amundsen *et al.*, 1986). Mutations in these genes results in recombination deficient phenotypes as well as causing a deficiency in DNA repair; however, the sensitivity of these strains to UV radiation is not as great as in *recA* mutant strains (Clark, 1973). The RecBCD enzyme has multiple enzymatic activities including ATP-dependent exonuclease for double and single stranded DNA, ATP dependent DNA unwinding, and ATP endonuclease for single stranded DNA (Telender - Muskavity and Linn, 1982).

Chi-sites consisting of the nucleotide sequence 5'-GCTGGTGG-3', are known to enhance RecBCD-dependent recombination (Weinstock, 1987). Incubation of the RecBCD enzyme with duplex DNA containing a *chi*-site results in nicking near the *chi*-site. However, some RecBCD mutants are deficient in *chi*-promoted nicking. RecBCD enzyme nicking at *chi*-sites and the unwinding of DNA are possible mechanisms for initiating recombination (Taylor *et al.*, 1985). Rinken

et al. (1992) suggested that the helicase activity of the RecBCD enzyme could play a central role in the RecBCD pathway of recombination by the production of "recombinogenic" single strands at the ends of duplex DNA. RecBC dependent unwinding produces single strands and ssDNA is an SOS inducing signal through the binding of the RecA protein. (Little, 1984).

1.4.4 The RecF and RecE recombination pathway

The RecF and RecE pathways are not independent and they represent variations on a common recombination pathway that is distinct from the RecBCD pathway (Weinstock, 1987).

The RecF pathway appears to be more active in cells exposed to DNA damaging agents (Lovett and Clark, 1983). The residual recombination which occurs in RecBCD mutants is due to the RecF pathway. The RecA protein in strand exchange reactions is necessary for the function of the RecBCD and the RecF pathways (Weinstock, 1987). Eight genes including *recA, recF, recJ, recN, recQ, recO, ruv* and *uvrD* have been identified as members of the RecF pathway (Peterson *et al.*, 1988). The *recN, recQ, uvrD* and *ruv* genes are part of the SOS regulon (discussed later). The RecF recombination pathway can be blocked by mutations in the *recA* or *lexA* genes that prevent induction of the SOS response. (Weinstock, 1987).

Daughter strand gap repair appears to be mediated through the RecF pathway of recombination. The RecA, RecF and the Ruv proteins function directly in daughter strand gap repair (Walker, 1985). The Ruv proteins are encoded by the *ruvA, ruvB* and *ruvC* genes. The products of these genes are required for genetic recombination and the recombinational repair of damaged DNA. Studies by West and Connolly (1992), suggest that these proteins function late in recombination and process Holliday junctions made by RecA protein mediated strand exchange.

SECTION A2

THE MYCOBACTERIA

1.5.1 Nomenclature and characterisation

The order Actinomycetales comprises the families: Actinomycetaceae (non-acid fast, mol% G+C=57-69) and Mycobacteriaceae (acid fast, mol% G+C=62-70 with the exception of *M. leprae* which has a mol% G+C=58). These organisms are Gram-positive and the cells are generally rod shaped and sometimes filamentous or branched (Bergey's Manual, 1986).

The genus *Mycobacterium* is the only one in the family Mycobacteraceae. Within this genus a natural division occurs between slowly and (relatively) rapidly growing species. For practical, determinative purposes, slow growers are those that require over 7 days incubation at optimal temperature to produce readily visible, isolated colonies on solid media. Rapid growers are visible in less than 7 days under optimal conditions. The exception is *M. leprae* which divides very slowly- about once in every two weeks- and attempts to cultivate it *in vitro* have been unsuccessful. The rapid and slow growers are too similar in terms of DNA homology, antigenic composition, lipid composition and bacteriophage susceptibility, to justify their separation into 2 sub-genera. Among the slow growers are species such as *M. tuberculosis, M. bovis* and *M. kansasii.* Included in the list of rapidly growing species are *M. smegmatis, M. fortuitum* and *M. aurum.* (Bergey's Manual of Systematic Bacteriology, 1986). The growth rate ranges from slow to very slow with generation times ranging (by species) from 2 to more than 20h (David, 1973).

The slow growth of the mycobacteria may be due in part to the hydrophobic nature of their high lipid content cell walls. It may also reflect the relatively low activity of their RNA polymerase and their low ratio of RNA to DNA (Harshey and Ramakrishnan, 1977). Also related to growth rate is the number of rRNA genes in relation to the size of the genome. Bercovier *et al.* (1986), hypothesised that a low number of copies of rRNA genes in mycobacteria may be one of the reasons that the bacilli grow more slowly than other prokaryotes. The number of rRNA genes is surprisingly low in mycobacteria; only one copy in the slow growers and two copies in the fast growers, compared to seven copies in *E. coli* (Bercovier *et al.*, 1986). Genetic studies of mycobacteria have been hampered by the slow growth of these organisms and the difficulty in obtaining auxotrophs.

1.5.2 Pathogenic mycobacteria

There are two major recognised pathogens, *M. tuberculosis* and *M. leprae*. Other species of mycobacteria are saprophytic and can occasionally cause disease. One of the best studied species is *M. tuberculosis*, the causative agent of tuberculosis in man, primates, dogs and animals which have contact with man. The cells of *M. tuberculosis* are rod shaped, ranging in size from 0,3-0,6 x 1-4 mm and may be straight or slightly curved. Growth tends to be in serpentine or in cordlike masses. The generation time in vitro is 14-15 h at 37°C (Bergey's Manual, 1986).

M. bovis, is generally more pathogenic than *M. tuberculosis* for animals, however, *M. bovis* does produce a tuberculosis - like disease in man and other primates. The bacillus of Calmette-Guerin (BCG) (1908) conforms to the properties described for *M bovis* but is attentuated in pathogenicity. *M. kansasii* has been isolated from human pulmonary lesions. It causes chronic human pulmonary

disease resembling tuberculosis, however, it is not considered contagious from man to man (Bergey's Manual, 1986).

M. fortuitum has been isolated from patients with pulmonary disease, postoperative sternal wound infections and endocarditis as well as from lymph glands of cattle. *M. smegmatis* is not pathogenic for man or mice, however, positive cultures may be obtained from the spleens of mice. *M. aurum* , like the majority of the described species, can be readily recovered from the soil and may occasionally be seen in the sputum of humans but is not associated with disease (Bergey's Manual, 1986).

M. avium and *M. intracellulare* are the prevalent causes of the "opportunist" mycobacterial diseases in man. Mycobacteria are a major cause of secondary infection in AIDS victims and in the United States of America, the great majority are due to *M. avium-intracellulare*.

M. paratuberculosis is the cause of chronic hypertrophic enteritis of cattle and, less often, of sheep and other ruminants.

1.5.3 Genetic recombination and DNA repair in the actinomycetes

DNA repair mechanisms have not been extensively studied in the Actinomycetes. Available information on Streptomyces, which are a member of the Actinomycetes, may serve as an appropriate model for the mycobacteria. Evidence for the existence of DNA repair mechanisms in the Actinomycetes, has been provided by the isolation of UV-sensitive mutants of *Streptomyces coelicolor* (Harold and Hopwood, 1970a), and several DNA repair pathways have been shown to exist in *Streptomyces fradiae* (Baltz, 1987). The interrelationship between the *rec* and *uvr* systems in Streptomycetes may be essential to an understanding of genetic instability in these organisms (Usdin *et al.*, 1985). The homologous recombination

system in *Streptomyces* appears to be distinct from the DNA repair system, since no UV-sensitive mutants of *Streptomyces* have been isolated which are recombination negative (Harold and Hopwood, 1970b,1972).

The presence of an inducible system involved in the process of recombination in the mycobacteria was first identified by Mizuguchi (1974). Recombination deficient (*rec*$^-$) *M. smegmatis* mutants were isolated by selecting for colonies that were very sensitive to UV irradiation. The correlation between recombination deficiency and UV sensitivity indicated that the same mechanism was involved in both recombination and DNA repair. Subsequently mutants of *M. smegmatis* that were more resistant than the wild type to DNA damaging agents including UV iradiation, mitomycin C and methyl methane sulphonate, have been isolated (Norgard and Imaeda, 1978).

The 2 main genes involved in recombination in *E. coli* are the *recA* and the *recBC* loci. Winder and Coughlan (1969) described an ATP dependent deoxyribonuclease similar to the enzyme coded by the *recBC* gene of *E. coli* in *M. smegmatis*. The exact nature of the ATP dependency of the enzyme is not fully understood. This enzyme has been characterised and *rec*$^-$ mutants of *M. smegmatis* were found to have high activities of this enzyme (Winder and Sastry, 1971; Johnson *et al.*, 1974).

A DNA polymerase having properties similar to DNA polymerase I of *E. coli*, but being more involved in DNA repair than in replication has been partially purified from *M. smegmatis* (Mc Nulty and Winder, 1971; Cambell *et al.*, 1979). The concentration of this enzyme was found to increase within the mycobacterial cell under conditions of iron limitation and following damage to DNA. The inducibility of the enzyme in the presence of DNA damage strongly suggested an error-prone inducible repair system analogous to the SOS system present in *E. coli*.

Using DNA hybridisation experiments, a *recA*-like gene has been cloned from *M. tuberculosis* (Davis *et al.*, 1991). Expression of this gene partially complemented *rec*$^-$ mutants of *E. coli* for recombination, DNA repair and mutagenesis. The *M. tuberculosis recA* locus specifying the RecA protein comprises a single open reading frame encoding an 85 kDa product. This putative product is twice the size of the RecA from *E. coli* (38-42 kDa). The carboxy and amino ends of the complete 85 kDa protein sequence revealed homologies to the sequence of the *E. coli* RecA protein. Davis *et al.* (1992) have recently shown that although no RNA processing was detected, the 85 kDa precursor protein was spliced thus releasing a 47 kDa spacer protein, followed by joining of the terminal fragments to form mature RecA protein.

PURPOSE OF THIS STUDY

The purpose of this study was to isolate, identify and characterise by complementation studies, gene(s) involved in DNA repair and recombination in *M. tuberculosis*. This dissertation describes the isolation and functional characterisation of an unusual *recA* like gene designated *recS* which differs from that already isolated from *M. tuberculosis* (Davies *et al.*, 1991). Analysis of the nucleotide sequence obtained revealed major differences in comparison with that of other cloned *recA*s.

Also described in this dissertation is the isolation and cloning of a 1.7kb DNA fragment, able to complement defects in an *E. coli rec* mutant. This fragment was shown to contain a truncated *katG* gene and two small ORFs with no sequence homology to known proteins.

Results included in this dissertation have been published in The Journal of General Microbiology (Nair and Steyn, 1991).

CHAPTER 2

ISOLATION OF A *recA* HOMOLOGUE FROM *M. TUBERCULOSIS*

2.1 SUMMARY 54

2.2 INTRODUCTION 54

2.3 MATERIALS AND METHODS

2.3.1 Bacterial strains and plasmids 55
2.3.2 Media, culture conditions and DNA isolation. 55
2.3.3 Resistance to UV irradiation. 56
2.3.4 DNA hybridization. 56
2.3.5 Recombination studies. 57
2.3.6 Efficiency of plating (e.o.p.) of phage P1 57
2.3.7 Protein isolation and western blot analysis. 57

2.4 RESULTS

2.4.1 Isolation of a *M. tuberculosis* chromosomal fragment that complements an *E. coli recA* deletion. 58
2.4.2 EtMes resistance conferred by the *M. tuberculosis* chromosomal fragment 59
2.4.3 UV sensitivity. 60
2.4.4 Complementation of homologous recombination in *E. coli recA* mutants. 61
2.4.5 Replication of phage P1. 61
2.4.6 Western blotting.

2.5 DISCUSSION 63

CHAPTER 2

ISOLATION OF A *recA* HOMOLOGUE FROM *M. TUBERCULOSIS*

2.1 SUMMARY

A *Pst*I DNA library of *M. tuberculosis* was transformed into an *E. coli* deficient host *(rec⁻)*. The transformants were screened for EtMes resistance. A recombinant containing a 3.8 kb *Pst*I fragment of *M. tuberculosis* DNA was cloned. The cloned fragment restored homologous recombination in Hfr crosses and confered resistance to long wave (302 nm) but not short wave (254 nm) UV light. *E. coli* containing the 3.8 kb *Pst*I fragment produced a 38-40 kDa protein which cross-reacted with *E. coli* RecA antiserum. The cloned DNA appears to encode a RecA homologue.

2.2. INTRODUCTION

In *E. coli* several enzymes are involved in DNA repair and replication (Little *et al.*, 1980; McEntee *et al.*, 1979). One of these enzymes, RecA has been shown to be associated with several different processes, including the mediation of recombination between homologous DNA fragments, recombination repair of damaged DNA and the initiation of the SOS response (Walker *et al.*, 1987). RecA mutants are recombination deficient and extremely sensitive to UV radiation and to DNA-damaging chemicals (Walker, 1984). Complementation studies of *E. coli*

recA mutants have been used in the isolation of analogous *recA* genes from Gram-negative bacteria (see section 1.2.2).

We have extended these complementation assays to isolate the *E. coli* *recA* counterpart from *M. tuberculosis*.

2.3 MATERIALS AND METHODS

2.3.1 Bacterial strains and plasmids

The bacterial strains used in this study are listed in Appendix 1; *E. coli* strain DK is a K-12 host which carries a deletion that spans the *recA* structural gene (Willis *et al.*, 1981) and was used to test for clones which complement the *rec* mutation. The parent strain of *E. coli* DK is the *recA*$^+$ strain MC1060 (Casadaban *et al.*, 1980). Cloning vectors and plasmids not originating from this study are listed in Appendix 2A while all plasmids constructed during the course of this study are listed in Appendix 2B.

2.3.2 Media, culture conditions and DNA isolation.

E. coli cultures were grown in Luria agar or M9 minimal media (Maniatis *et al.*, 1982) and when required ampicillin (50 mg/ml) or streptomycin (25 mg/ml) were added. Freeze-dried cultures of the *M. tuberculosis* strain H37RV were reconstituted and streaked onto Lowenstein-Jensen (Difco) slopes. After 6 weeks incubation at 37^{o}C the cells were harvested and suspended in SDS buffer (10mM Tris, 1mM EDTA, 0.15mM NaCl and 10% SDS) for DNA isolation as described by de Wit *et al.*, 1990. H37RV DNA was digested to completion with *Pst*I and ligated to *Pst*I-cut pEcoR252 (Zabeau and Stanley, 1982). *E. coli* DK cells

cells were transformed (Hanahan, 1983) and plated onto Luria agar containing ampicillin and ethyl methane sulphonate (EtMes)(0.1%, v/v).

2.3.3 Resistance to UV irradiation.

Bacterial cultures (10 ml) were grown with aeration at 37°C for 2hrs (A_{600}=0.2) and the cells were harvested by centrifugation and suspended in physiological saline (1ml). Aliquots (100 ml) of a series of 10 fold dilutions were exposed to UV light (1J/m^2 at 254nm or 302nm). All dilutions were performed in the dark and the plates were incubated overnight at 37°C to determine cell survival.

2.3.4 DNA hybridisation.

The digested DNA fragments were separated by agarose gel electrophoresis and transferred to a nylon membrane (Hybond N, Amersham) (Southern, 1975). Probe fragments were radiolabelled with [α-^{32}P]dATP (Amersham) by nick translation (Boehringer Mannheim) and DNA hybridisations and post hybridisation washes were performed as described by Johnson *et al.*, 1984. Hybridisations were performed in hybridisation buffer (6 x SSC, 5 x Denhardt's solution, 0.5% SDS, 1 mg/ml salmon sperm DNA) containing 50% formamide at 42°C for 16h. The membranes were then washed three times in 2 x SSC, 0.1% SDS at room temperature for 20 min, once in 1 x SSC, 0.1% SDS for 30 min at 56°C and once in 0.1 x SSC, 0.1% SDS for 30 min at 56°C. In hybridisations involving *Streptomyces coelicolor* DNA, the temperature of the post hybridisation washes was reduced from 56°C to 42°C. SSC (1x) is 150mM NaCl and 15mM NaCitrate.

2.3.5 Recombination studies

Homologous recombination experiments using *E. coli* HB101 containing pSNS201 or pSNS202 (see Appendix 2B) as recipients and the Hfr strain, CSH62 (Appendix 1A), as donor, were performed as described by Ramesar *et al.*, 1988. *E. coli* HB101 is a proline auxotroph and is streptomycin resistant, while CSH62 is streptomycin sensitive. The number of colony forming units per 50 ml of CSH62 donor cells was determined on M9 agar supplemented with lysine. *E. coli* HB101 recombinants (proline prototrophs) were selected on M9 agar containing streptomycin (25 mg/ml), leucine (40 mg/ml) and thiamine (3 mg/ml). The recombination frequency is expressed as the number of recombinants obtained per 1000 donor cells.

2.3.6 Efficiency of plating (e.o.p.) of phage P1

The e.o.p. of phage P1 was tested on *E. coli* strains RR1 (*recA*$^+$), HB101 (*recA*$^-$) and HB101(pSNS201). Cells were grown in Luria broth to an OD $_{600}$=0.2, harvested by centrifugation and resuspended in SM buffer (Maniatis *et al.*, 1982). Cell aliquots (1ml) were mixed with diluted phage and adsorption was allowed to proceed at room temperature for 20min. Top agar, (3ml, 0,8%, w/v, 45°C) was added to the cell/phage mix and poured onto Luria agar plates. The number of plaques was determined after 18-24h incubation at 37°C.

2.3.7 Protein isolation and western blot analysis

Cells were harvested from cultures (1.5ml, A_{600}= 0.5) of *E. coli* MC1060, *E. coli* DK and *E. coli* DK (pSNS201 and pSNS202) and were suspended in 300 ml of SDS/mercaptoethanol (10% w/v) and placed in a boiling water bath for 5min. Denaturing polyacrylamide gels (10%) (Laemmli, 1970) were

used to separate the cellular proteins at a constant current of 35mA for 2.5h and the proteins were electroblotted onto nitrocellulose membranes (Schleicher and Schuell) at 200mA for 3-4h in transfer buffer (25mM Tris, 0.19M glycine, pH 8.3, 25% methanol), (Towbin *et al.*, 1979). The membrane was blocked for 2h at 37°C with 10mM Tris-Cl/150mM NaCl containing 1% low-fat milk powder and 0.05% Tween-20, pH 7.4 (Johnson *et al.*, 1984). Purified antiserum raised in rabbits against *E. coli* RecA protein (Goodman *et al.*, 1987) was diluted 1 in 20 and used for the detection of cross-reacting proteins with alkaline phosphatase conjugated goat anti-rabbit immunoglobulin as the reporter molecule (Rybicki and Wechmar, 1982).

2.4 RESULTS

2.4.1 Isolation of a *M. tuberculosis* chromosomal fragment that complements an *E. coli recA* deletion.

H37Rv DNA was digested to completion with *Pst*I and ligated to *Pst*I-cut pEcoR252 (Zabeau and Stanley, 1982). *E. coli* DK cells were transformed (Hanahan, 1983), and plated onto Luria agar containing ampicillin and ethyl methane sulphonate (EtMes, 0.1% v/v). One recombinant plasmid, pH15, was isolated which on retransformation allowed consistent growth of *E. coli* DK on EtMes. pH15 contained a *Pst*I fragment of 3.8 kb. Southern transfer and DNA hybridisation were used to confirm that the insert originated from the *M. tuberculosis* chromosome (Fig. 2.1). The insert does not hybridise to *Streptomyces coelicolor* (strain A3(2)M130, John Innes Institute, Norwich) genomic DNA digested with *Pst*I even under moderate stringency conditions (Fig. 2.1).

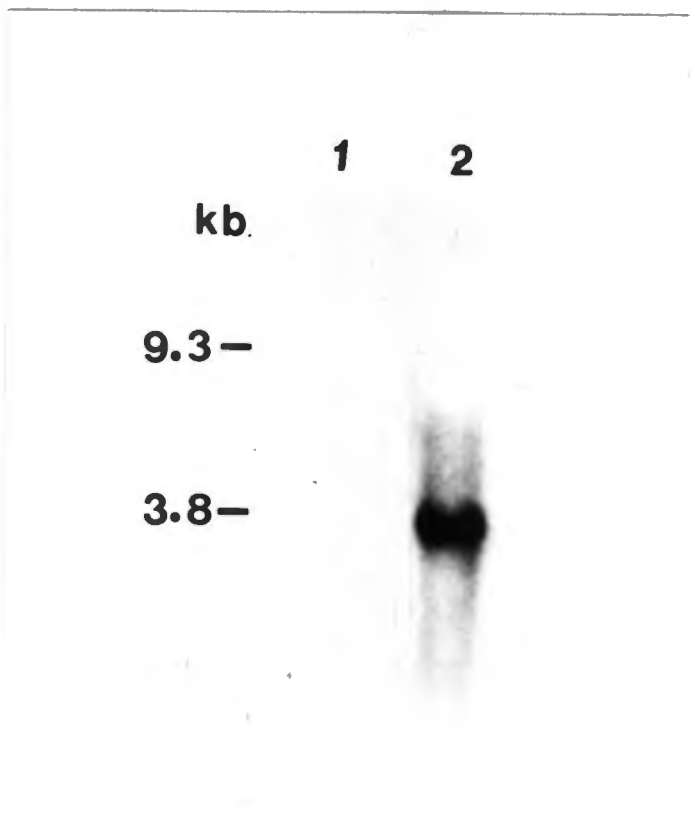


Fig.2.1 Southern blot of DNA isolated from *M. tuberculosis* H37Rv and *Streptomyces coelicolor*, probed with labelled DNA of the insert from pSNS201. Lanes:1, *S. coelicolor* DNA (5mg) digested with *Pst*I; 2, *M. tuberculosis* DNA (5 mg) digested with *Pst*I (Nair and Steyn, 1991).

2.4.2 EtMes resistance conferred by pSNS201

The plasmids pSNS201 and pSNS202 are pUC19 derivatives containing the 3.8 kb fragment of pH15 in opposite directions. Both plasmids conferred sufficient EtMes resistance on DK1 cells to allow growth on 0.1% EtMes while DK1 without plasmid did not grow on this percentage of EtMes. Fig. 2.2 shows that *E.coli* DK1(pSNS201) was less resistant to EtMes than *E. coli* MC1060 the

recA$^+$ parent of *E. coli* DK1; 6% of the cells containing pSNS201 but only 2% of those containing pSNS202 survived on 0.1% EtMes (data not shown). The percentage of EtMes resistant cells containing these recombinant plasmids was obtained from 5 independent experiments. Selection was always performed with both ampicillin and EtMes to ensure that no loss of plasmid DNA had occurred.

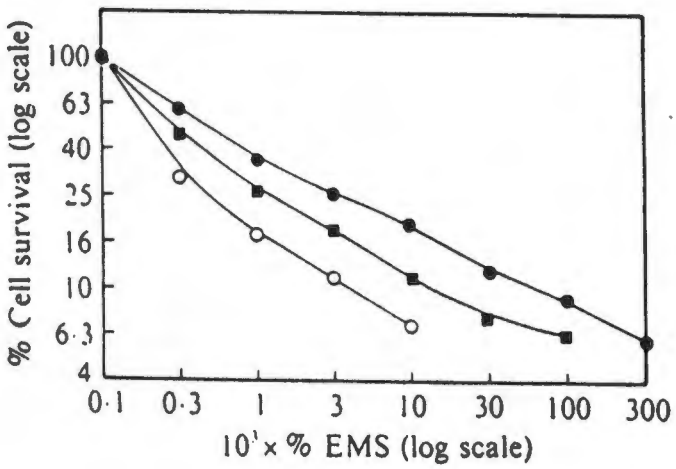


Fig. 2.2 Effect of the recombinant plasmid pSNS201 on the survival of *E. coli* DK1 cells. which were plated on various EtMes concentrations. Symbols: •, MC1060; O, DK1; ♦,DK1(pSNS201) (Nair and Steyn, 1991).

2.4.3 UV sensitivity

No significant difference in UV resistance at 254nm (1J/m^2) between *recA*$^-$ strains and strains carrying the *M. tuberculosis* recombinant plasmid pSNS201 was found. Clear differences in UV sensitivity were detected at a longer wavelength. Exposure of *E. coli* DK1(pSNS201) to UV light of (302nm) (1J/m2) failed to kill these cells whereas no survivors were detected from the plasmid free *E. coli* DK1 under the same conditions (data not shown).

2.4.4 Complementation of homologous recombination in *E.coli recA* mutants.

The recombinant plasmids pSNS201 and pSNS202 were tested for their ability to complement the homologous recombination functions of various *E. coli recA* mutants. *E. coli* HB101 containing the recombinant plasmid pSNS201 restored recombination to levels approaching those of *E. coli* RR1 (77-81 as opposed to 95-102 recombinants/1000 donor cells), while the recombinant plasmid with the insert in the opposite orientation (pSNS202), produced fewer transconjugants (18-21/1000 donor cells). The ranges indicated were obtained from 4 individual experiments. No revertants were obtained when similar conjugation experiments were performed with HB101 cells containing pUC19.

2.4.5 Replication of phage P1

Efficient phage P1 lytic growth is controlled by the bacterial general recombination system and requires a functional *recA* gene, thus *E. coli* strains containing *recA* mutations are unable to support the growth of phage P1 (Cohen, 1983). Phage P1 was unable to form plaques on *E. coli* HB101 cells. The plasmid pSNS201 was not capable of restoring the e.o.p. of phage P1 by *E. coli* HB101.

2.4.6 Western blotting

Rabbit antiserum prepared against *E. coli* RecA protein (Goodman *et al.*, 1987) was used to detect RecA protein in crude extracts of *E. coli* MC1060, *E. coli* DK and *E. coli* DK(pSNS201) (Fig.2.3). As the antiserum was not preabsorbed to *E. coli* DK cell extracts, it reacted with several polypeptides in the control lysates as well as in the test samples. The antiserum reacted with new polypeptide bands

with an apparent M_{rs} of 38 000 - 40 000 in extracts prepared from *E. coli* DK (pSNS201) as well as in extracts of *E. coli* MC1060 (*recA*$^+$) cells. The *E. coli* *recA* gene product is a polypeptide with a M_r of 37,800. RecA proteins were not detected in DK (pSNS202), which has the 3.8 kb insert in the opposite orientation to *E. coli* DK (pSNS201) (results not shown).

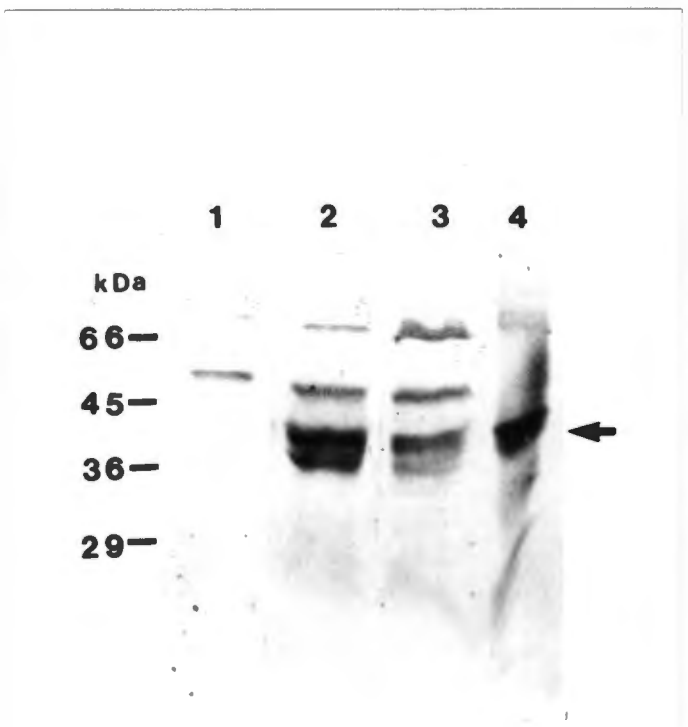


Fig.2.3 Detection of RecA protein in crude cell extracts by Western blot analysis. Lanes: 1, *E. coli* DK; 2, *E. coli* DK(pSNS201); 3, *E. coli* DK MC1060; 4, *E. coli* RecA protein (Sigma) (arrowed) (Nair and Steyn, 1991).

2.5 DISCUSSION

Although the cloned *E. coli recA* has been shown to complement defects in DNA repair and mutagenesis in the Actinomycetes *Streptomyces fradiae* JS6 (*mcr-6*) (Matsushima and Baltz, 1987), it is not known if DNA repair systems in an organism such as *M. tuberculosis* can complement those of *E. coli*. To address this possibility the functional complementation of *E. coli recA* mutants with a recombinant plasmid encoding a *M. tuberculosis* RecA homologue was examined.

Several lines of evidence indicate that the recombinant plasmid, pSNS201, contains a gene from *M. tuberculosis* that encodes activity similar to that of the *E. coli recA* gene. The cloned DNA fragment restored near wild-type levels of resistance to the DNA damaging agent EtMes to *E. coli recA* mutants, it promoted homologous recombination between a Hfr donor and a *recA* mutant, it encoded a protein which cross-reacts with antiserum against the *E. coli* RecA protein and it confirmed low level resistance to UV light (302nm) on *recA E. coli*. Southern hybridisation confirmed that the insert originated from the *M. tuberculosis* genome.

E. coli DK (pSNS201) cells had increased resistance to the DNA damaging agent, EtMes (0.1%), whereas cells without the recombinant plasmids were killed at EtMes concentrations above 0.01%. Based on studies on DNA repair in *Streptomyces cattleya*, Hromic and Kirby, 1987, suggested that two repair systems exist in *Streptomyces*. The first was evident when the cells were exposed to low levels of EtMes. This may represent a constitutive, error-free repair system which may be sufficient when the integrity of the genome is not disrupted. At high levels of EtMes when extensive DNA damage occurs a second DNA repair system can be demonstrated which is error prone and is similar to that found in *E. coli* (Baltz and Stonesifer, 1985). Thus in *Streptomyces*, EtMes mutagenesis occurs by

induction of an error prone repair pathway, while in contrast *E. coli* EtMes mutagenesis occurs by inducing base pair mismatching. At present the type of system that may be induced by EtMes in *M. tuberculosis* is not known.

pSNS201-mediated homologous recombination in a *recA13* recipient following conjugation with a Hfr donor strain. There were quantitative differences in the number of recombinant prototrophs obtained when the insert was in opposite orientations in the cloning vector (pSNS201 vs pSNS202). *E. coli* DK (pSNS202) cells also had fewer survivors on EtMes. This polarity may reflect the effects of plasmid promoters on the expression of the *M. tuberculosis recA* gene. Most mycobacterial promoters are recognised poorly, if not at all, by the *E. coli* transcriptional machinery (Thole *et al.*, 1985). This could explain the lack of UV induction of pSNS201 expression in *E. coli*.

Based on western analysis, the 3.8 kb *M. tuberculosis* chromosomal fragment (pSNS201) encodes polypeptides of M_r 38 000 to 40 000, which cross-react with antisera raised against *E. coli* RecA protein. With *E. coli* DK (pSNS202), however, the antiserum failed to detect polypeptides in this molecular weight range. This may reflect the relative insensitivity of the western blot technique compared with the functional complementation of RecA deficiency *in vivo*. Previous studies on RecA proteins from different microorganisms have shown evidence for polypeptides with a M_r of 37 000 to 38 000 in *Anabaena variabilis* (Owttrim and Coleman, 1989); 37 000, 39 000 and 40 000 in *Bacteroides fragilis* (Goodman *et al.*, 1987) and 35 500 and 38 000 in *Legionella pneumophila* (Zhao and Dreyfuss, 1990). According to Zhao and Dreyfuss, 1990, the higher molecular weight form could correspond to the *E. coli recA* gene product and the lower molecular weight polypeptide to a proteolytic degradation product of the RecA protein.

Exposure of DNA to a UV-light source of 302nm results predominantly in the formation of pyrimidine dimers, while exposure to a source of 254nm results in damage due to the formation of ionising radicals (Parrish *et al.*, 1981). The effects of these two wavelengths on cells containing the cloned insert in *E. coli* DK1 were studied. Unlike cells exposed to the longer wavelength UV light, *E. coli* mutants carrying the *M. tuberculosis* recombinant plasmid were sensitive when exposed to short wavelength UV radiation. Similar results were observed for the cloned *recA* homologs from the Gram-negative bacterium *Vibrio anguillarum* (Singer, 1989) and the obligate anaerobe *Bacteroides fragilis* (Goodman *et al.*, 1987). The results obtained with *Vibrio anguillarum* indicate that complementation becomes less efficient when doses of UV radiation are higher. Thus the cloned *RecA* homologue from *M. tuberculosis* did not fully complement defects in *E. coli recA* strains.

In *E. coli*, repair to UV induced damaged DNA is due to the cleavage of the protein UmuD. Dutreix *et al.*, 1989, have described *recA* mutants which do not cleave UmuD. They suggest that the RecA protein has different domains which interact to mediate protein cleavage or DNA recombination. It has been shown that *recA* mutagenesis functions are genetically separable from *lexA* and cI proteolysis, the latter activities are also separable from each other (Ennis *et al.*, 1989). It is possible that the *M. tuberculosis* RecA protein is defective as regards UmuD cleavage or that a separate protein fulfills this role in *M. tuberculosis*. Tessman and Peterson (1985) have also provided evidence for the separation of recombination and mutagenesis functions of the *E. coli* RecA protein. Evidence for functional conservation of RecA has been provided by the complementation of *E. coli recA* mutants by cloned genes from a variety of bacteria. This study indicates that this functional conservation extends to members of the Actinomycetes as exemplified by *M. tuberculosis*. Cloned *recA* genes that have been sequenced show relatededness in their functional domains (Ramesar *et al.*, 1989). Thus comparison of the

M. tuberculosis recA-like gene sequence with those of other sequences would provide a further indication of the degree of conservation of this gene.

This study has demonstrated the existence of a *recA* homologue in *M. tuberculosis*. Investigations by Davis *et al.*, 1991, have shown that a *recA* like gene does exist in *M. tuberculosis*, and that the open reading frame encodes a protein of molecular weight of 85 kDa. A comparison between pSNS201 and the *recA* described by Davis *et al.*, 1991, is given in Chapter 7. At present it is not clear if all the *E. coli* RecA functions are contained within one protein in *M. tuberculosis* or whether these are performed by more than one enzyme in this organism.

CHAPTER 3

LOCATION OF THE *recA* ANALOG IN pSNS201

3.1 SUMMARY 68

3.2 INTRODUCTION 68

3.3 MATERIALS AND METHODS

3.3.1 Bacterial strains and plasmids 69
3.3.2 ExoIII nuclease deletion 69
3.3.3 Western blot analysis 69
3.3.4 Homologous recombination 69

3.4 RESULTS

3.4.1 Restriction map of pSNS201 70
3.4.2 ExoIII deletions 70
3.4.3 Western blot analysis 72
3.4.4 Homologous recombination 74

3.5 DISCUSSION 76

CHAPTER 3

LOCATION OF THE *recA* ANALOG IN pSNS201

3.1 SUMMARY

A restriction map of pSNS201 was prepared for different restriction enzymes. In an attempt to localise the RecA activity in pSNS201, the 3.8 kb *Pst*I fragment was shortened by exonuclease III digestion from one end and the fragments obtained were analysed for RecA activity by the following: presence of RecA protein on Western blot analysis which cross-reacted with the *E. coli* RecA polyclonal antiserum; and the ability to mediate homologous recombination in an *E. coli rec*⁻ mutant. pSNS206 containing an insert of approximately 1.1 kb encoded RecA activity, whereas pSNS207, which contained a further shortened insert of 750 bp, encoded no detectable Rec activity.

3.2 INTRODUCTION

The cloned *Pst*I fragment (pSNS201) described in Chapter 2 was shown to complement *E. coli recA* mutants for recombination, DNA repair and mutagenesis.

This chapter deals with the localisation of the *recA* analog in pSNS201 by exoIII deletion constructs. These constructs were subsequently investigated for Rec activity on Western analysis and their ability to mediate internal recombination in a *rec*⁻ strain.

Clones containing recombinants with deletions of greater than 3 kb did not encode any RecA activity. pSNS206 (1.1 kb) showed RecA activity as compared to pSNS207 (750 bp) which was lacking in RecA activity

3.3 MATERIALS AND METHODS

3.3.1 Bacterial strains and plasmids

E. coli strains used in this chapter are listed in Appendix 1A. Plasmids and constructs made during this study are listed in Appendix 2A and 2B.

3.3.2 ExoIII nuclease deletion

No restriction sites for *Kpn*I were present in pSNS201. Thus the *Kpn*I site of the vector (pUC19) was used to protect the vector DNA, and the *Bam*H1 site was used to initiate activity of the insert DNA by exoIII digestion (Henikoff, 1984).

3.3.3 Western blot analysis

ExoIII deletion derivatives of pSNS201 were used to transform *E. coli* DKI cells and Western blot analysis was performed as described in Chapter 2.

3.3.4 Homologous recombination

Recombination was measured by a papillation assay (Konrad, 1977). Strain GY7066 (a gift from R. Devoret and A. Bailone, Laboratoire d'

Enzymologie, CNRS, F-91198, Gif-sur-Yvette, France), is a *rec* mutant which also has a *lac* deletion.

GY7066 was transformed with pSNS201, pSNS202, pSNS203, pSNS204 and pUC19. Dilutions of an overnight culture ranging from 10^{-2} - 10^{-8} were plated onto MacConkey lac^+ plates containing 100μg/ml ampicillin. Untransformed GY7066 was also diluted and plated onto MacConkey lac^+ plates with and without ampicillin and streptomycin (25μg/ml). Plates were incubated for 4 days at 37C.

3.4 RESULTS

3.4.1 Restriction enzyme map of pSNS201

A unique *Bgl*II site is shown approximately 750 bp from the left *Pst*I site and principal *Sma*I and *Hinc*II sites are shown in Fig. 3.1a. Hybridisation with the internal *Bgl*II/*Pst*I fragment confirmed the restriction map (Fig. 3.1b).

3.4.2 ExoIII deletion

ExoIII deletion constructs are given in Fig. 3.2. The plasmid derivatives obtained by the exoIII shortening are listed below:

pSNS209 - 530 bp; pSNS208 - 550 bp; pSNS207 - 750 bp; pSNS206 - 1.1 kb; pSNS205 - 1.9 kb; pSNS204 - 2.5 kb and pSNS203 - 3 kb

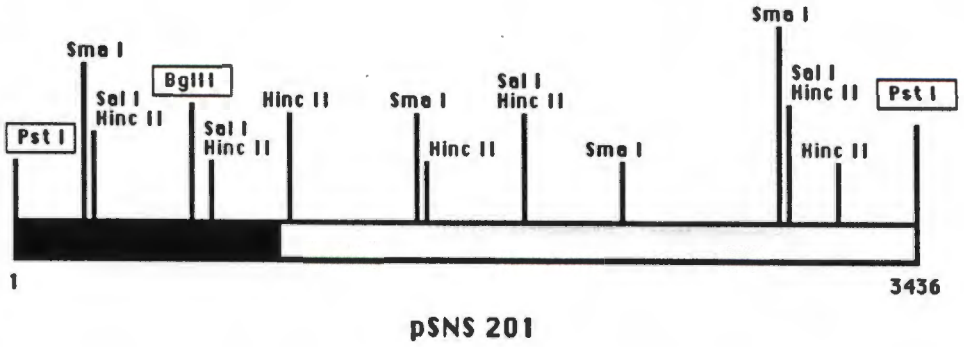


Fig.3.1a Restriction map of pSNS201. A unique *Bgl*II site is shown and principal sites for *Sma*I, *Sal*I and *Hinc*II are indicated. The shaded region denotes the truncated major ORF (see Chapter 4).

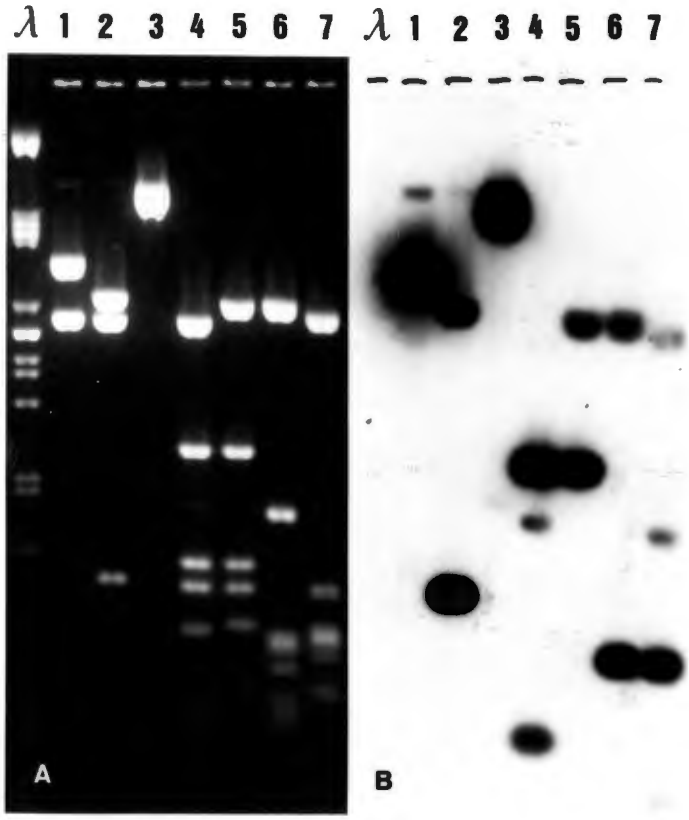


Fig.3.1b Restriction pattern of pSNS201; A, gel hybridisation with the *Bgl*II/*Pst*I fragment. Lane 1,*Pst*I; lane 2,*Pst*I/*Bgl*II; lane 3, *Bgl*II; lane 4,*Pst*I/*Sma*I; lane 5, *Sma*I; lane 6, HincII; lane 7, *Ava*I/*Acc*I. Molecular weight marker:λ/*Pst*I.

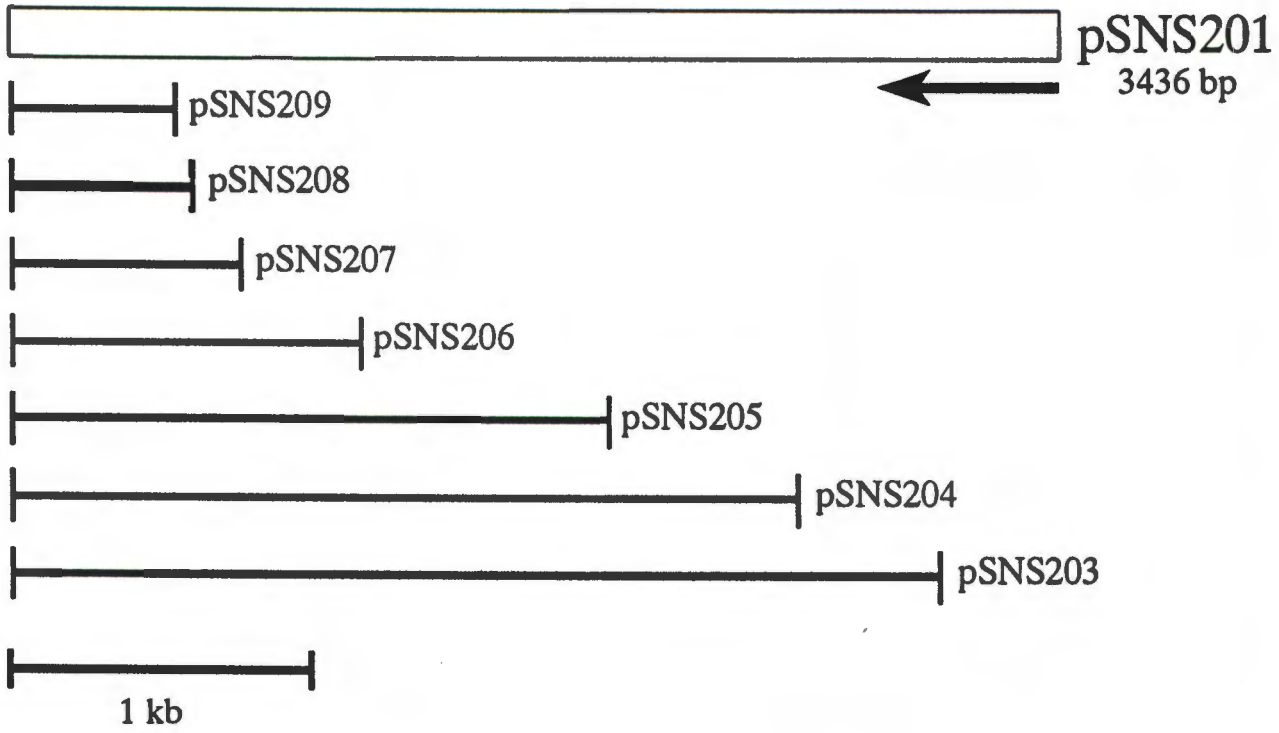


Fig. 3.2 ExoIII shortening strategy of pSNS201. Arrow indicates the direction of exonuclease shortening.

3.4.3 Western blot analysis

Western blot analysis of the exoII deletion constructs in *E. coli* DKI showed that clones containing the *Bgl*II/*Pst*I fragment (750 bp, pSNS207) and smaller fragments (pSNS208, pSNS209) did not produce a cross reacting protein on Western analysis (Fig. 3.3a). Recombinant *E. coli* DK cells containing DNA fragments larger than the *Bgl*II/*Pst*I fragment produced a protein of 38 kDa which cross-reacted with the *E. coli* RecA polyclonal antiserum (Fig. 3.3b).

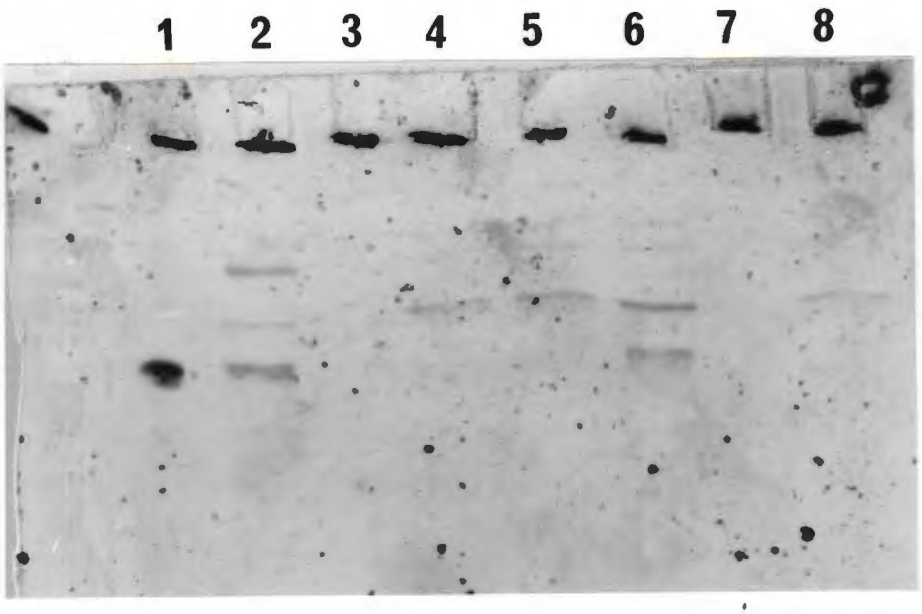


Fig. 3.3a Western blot analysis of exoIII deletion constructs of pSNS201. Lane1, E. coli RecA protein (Sigma); Lane 2, *E. coli* MC1060; Lane 4, *E. coli* DK1 pSNS208 (550 bp); Lane 5, *E. coli* DK1 pSNS207; Lane 6, *E. coli* DK1 pSNS206 (1.1 kb insert); Lane 8, *E. coli* DK1.

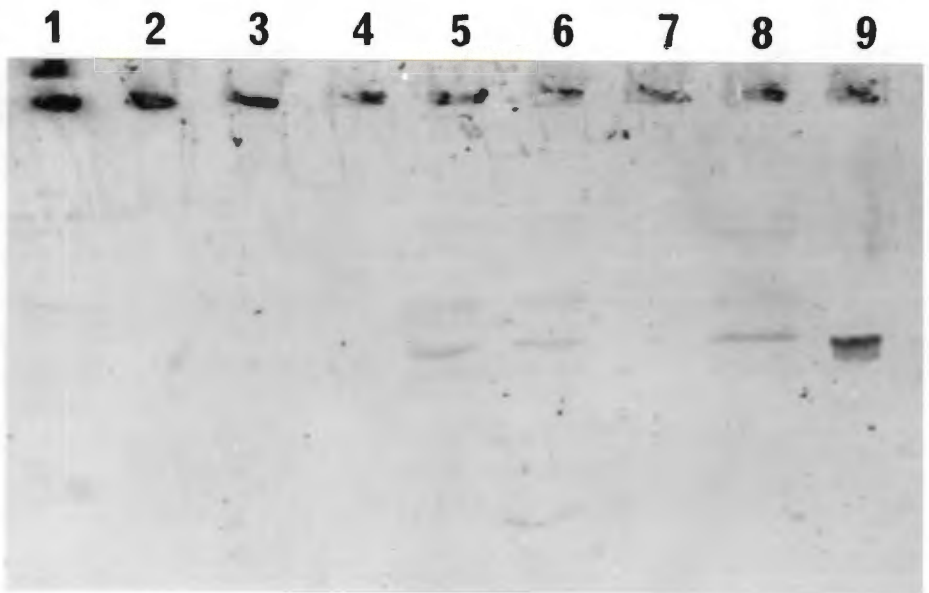


Fig. 3.3b Western blot analysis of different ExoIII deletion constructs used to transform *E. coli* DK11 cells. Lane 1, *E. coli* DK1; Lane 5, pSNS205 (1.9 kb); Lane 6, pSNS204 (2.5 kb); Lane 7, pSNS209 (530 bp); Lane 8, pSNS203 (3 kb); Lane 9, RecA protein, Sigma.

3.4.4 Homologous recombination

No papillae were observed in GY7066 untransformed and transformed with pSNS202, pSNS207 and pUC19 (Fig 3.4a). *Lac*$^+$ papillae formed by GY7066 recombinant bacteria GY7066(pSN201) and GY7066(pSNS203) were observed after 4 days. Colonies were grown on plates containing 4% (wt/vol) MacConkey agar base (Difco Laboratories) supplemented with 1% (wt/vol) lactose, containing ampicillin (Fig. 3.4b). These results show that pSNS206 is able to mediate recombination in the *rec*$^-$ host.

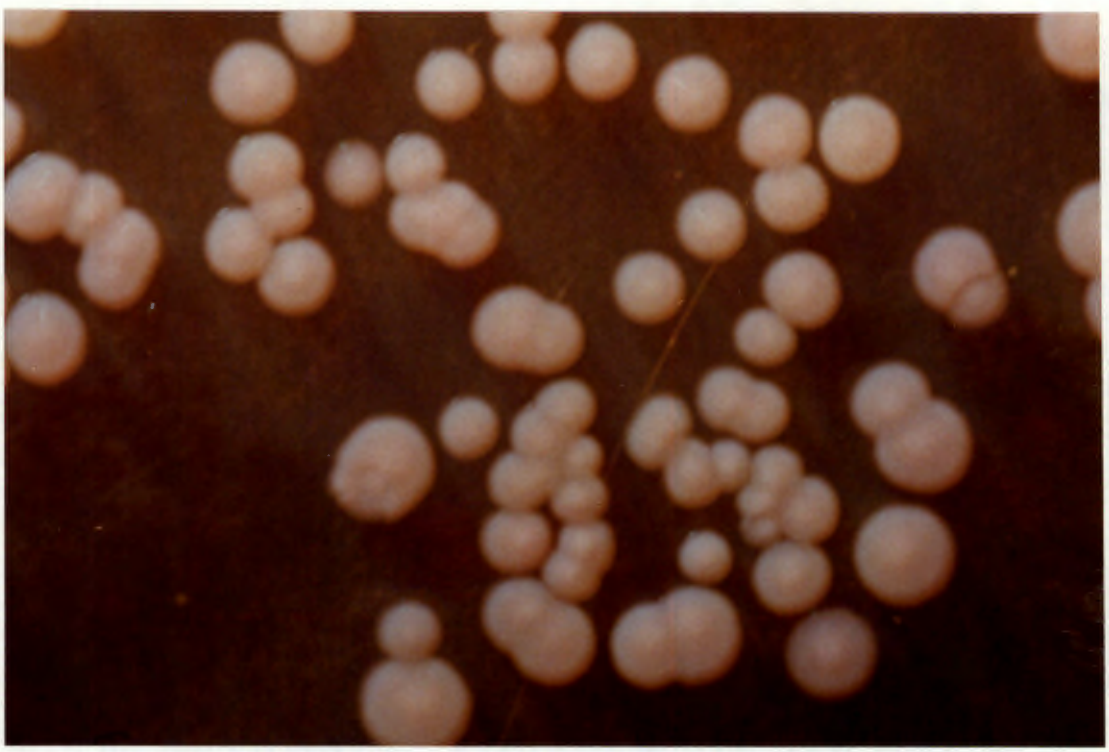


Fig. 3.4a Untransformed GY7066 plated onto MacConkey plates. No differences were observed in colony morphology between GY7066 untransformed and transformed with pUC19 and pSNS207.

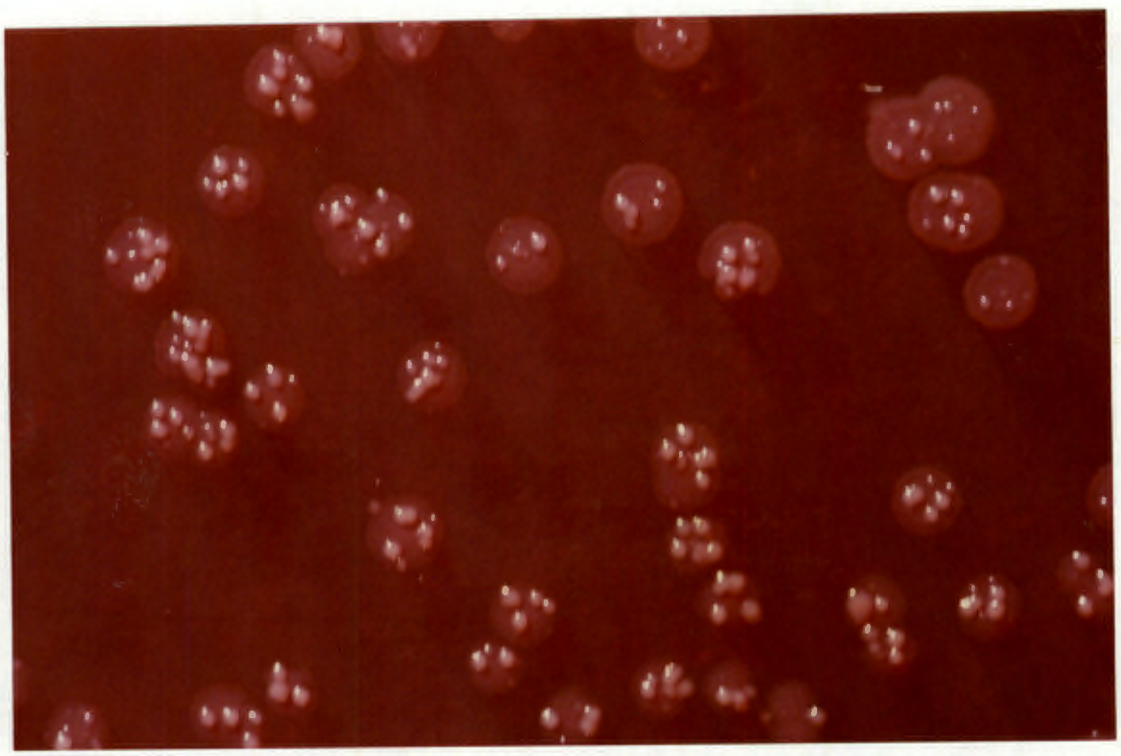


Fig. 3.4b GY7066 carrying pSNS206. *Lac*$^{+}$ papillae was also observed in GY7066 pSNS201.

3.5 DISCUSSION

Western blot analysis indicates that pSNS206 (1.1 kb) is the shortest pSNS201 derivative to encode a RecA cross-reacting protein. It is possible that certain regions responsible for folding of the protein are absent in pSNS207 (750 bp). The antigenic epitopes responsible for cross-reaction with the RecA antibody are either not present in this deletion or not correctly formed.

Intra-chromosomal *lac* recombination was observed in GY7066 bacteria carrying plasmid pSNS201 and pSNS206 but not pSNS207 and pUC19. This result confirms the presence of RecA activity encoded by pSNS206. These data indicate that the gene encoding the RecA-like activity is located at the 5' end of the construct.

CHAPTER 4

THE NUCLEOTIDE SEQUENCE AND IDENTIFICATION OF A NOVEL *recA*-LIKE GENE: *recS*

4.1 SUMMARY 78

4.2 INTRODUCTION 79

4.3 MATERIALS AND METHODS

4.3.1 Bacterial strains and plasmids 80

4.3.2 DNA sequencing 80

4.3.3 DNA hybridisation 81

4.3.4 PCR reactions 81

4.4 RESULTS

4.4.1 Nucleotide sequence of the 3.8 kb *Pst*I fragment (pSBG300) encoding the *recA*-like activity 81

4.4.2 Sequencing strategy used to sequence pSBG300 and pSNS201 83

4.4.3 Cloning of the 5' end of the putative *recA*

4.4.3a Screening of a H37Rv cosmid library 85

4.4.3b Restriction map and subcloning of pSBG100 86

4.4.3c Sequencing of pSBG101, pSBG102 and pSBG103 88

4.4.4 Analysis of the sequence of pSBG400 92

4.4.5 Computer analysis of the putative product of ORF463

4.4.5a Hydrophobicity 95

4.4.5b Codon usage of *recS* 96

4.4.6 Multiple sequence alignment 101

4.4.7 Occurrence of *recS* in myobacterial strains 104

4.5 DISCUSSION 108

CHAPTER 4

THE NUCLEOTIDE SEQUENCE AND IDENTIFICATION OF A NOVEL *recA*-LIKE GENE: *recS*

4.1. SUMMARY

The nucleotide sequence of the 3.8 kb *Pst*I fragment (pSNS201) was obtained partially from a series of overlapping clones produced by restriction cleavage with *Sal*I and *Sma*I. Sequencing was attempted on M13 sub-clones and double stranded DNA using universal primers and custom synthesised oligonucleotides. However, this strategy was not sufficient to determine a sequence of approximately 400 base pairs where recombinations and/or rearrangements occurred during the course of cloning. This sequence was resolved by using PCR reactions on H37Rv cosmid DNA (T252), with synthesised oligonucleotides. Sequencing was performed directly on the PCR products. To ensure that the sequence is correct, the equivalent 3.8 kb *Pst*I fragment was re-isolated from cosmid T256 to generate pSBG300.

The entire sequence of pSBG300 was obtained from double stranded DNA sequencing as well as of sections of the cosmid T256. Analysis of this sequence showed that the major ORF (ORF463), which was present in pSNS206, was truncated at the 5' end. An overlapping cosmid clone (T256) was identified by hybridisation with a probe corresponding to a portion of ORF463. A *Bgl*II fragment (4.8 kb) containing the 5' portion of ORF463 was subcloned and sequenced (pSBG100).

Analysis of the complete ORF showed an open reading frame of 1389 bp that could encode a protein of 463 amino acids. There are 5 potential initiation codons present and it is not clear which one is involved in initiation of the RecA-like protein encoded by ORF463. No significant DNA or amino acid similarities with other cloned *recA*'s were demonstrated for ORF463 or its putative protein product, thus it was designated *recS* and the product RecS.

An internal *Bgl*II/*Pst*I fragment of *recS* was used to probe chromosomal DNA isolated from different mycobacterial strains. Hybridisation demonstrated the presence of the gene in all strains studied. Unexpectedly there were 2 copies of *recS* in H37Rv and H37Ra and only one copy in BCG.

4.2 INTRODUCTION

The recombinant plasmid (pSNS201) described in Chapter 2 & 3 encoded a RecA activity as shown by its ability to partially complement *E. coli recA* mutants for recombination, DNA repair and mutagenesis. The entire insert of pSNS201 was sequenced. Due to the presence of DNA recombinations and rearrangements apparent in this clone, the *Pst*I insert was re-isolated, (pSBG300) from cosmid T256.

This chapter deals with the analysis of the nucleotide sequence of pSBG300, the cloning of the 5' end of the major ORF (ORF463) and the identification of conserved motifs within the amino acid sequence of the putative product (RecS) compared to the *M. tuberculosis recA* (Davies *et al.*, 1991) and the E. coli *recA* (Horii *et al.*, 1980). The results described here suggest the existence of a novel *recA*-like gene, *recS,* within the genome of *M. tuberculosis*. The *recS* gene is present in all the mycobacterial strains tested.

4.3. MATERIALS AND METHODS

4.3.1 Bacterial strains and plasmids

E. coli and mycobacterial strains used in this chapter are listed in Appendix 1A and 1B. The genomic DNA isolation was as described by de Wit *et al.*, 1990. Plasmid DNA for sequencing was prepared using the Quiagen method (Qiagen, Inc, USA).

The cosmid vector pYUB (12kb), was used in the construction of the *M. tuberculosis* H37Rv cosmid library (S. Cole *et al.*, Institut Pasteur) from partial *Sau*3A DNA fragments cloned into the *Bam*H1 site of pYUB.

4.3.2 DNA sequencing

The DNA sequencing strategy is outlined in Fig. 4.6. The sequencing of the templates was done with [α-^{35}S]dATP (Amersham) by the dideoxy chain-termination method (Sanger *et al.*, 1977) using Sequenase 2.0 (United States Biochemical Corporation, USB) for M13 DNA, and the Taq Track system (Promega) for double stranded DNA template. In each case the sequencing was started initially with universal primers and continued with custom synthesised oligonucleotide primers. Electrophoresis was as described by Maniatis *et al.* (1989) with 6% acrylamide gels, or 8% acrylamide, containing 50% formamide (Martin, 1987) or with acrylamide gradient gels (Biggin, *et al.*, 1983). Computer analyses were performed using Clustal V (Higgin *et al.*, 1991), GCG package (Devereux *et al.*, 1984), DNA Strider (Kate and Doolittle programme, Christain Marck, Departement de Biologie Cellulaire et Moleculaire, Direction des Sciences de al

Vie, C.E.A., France), and Fasta (Pearson and Lipman, 1988) programmes. The Genebank database and its daily updates were used for sequence comparisons.

4.3.3 DNA Hybridisation

DNA probe labelling reactions were performed as described in the Amersham Megaprime DNA Labelling Systems RPN protocol 1604/5/6/7. The rapid hybrisation protocol of Amersham, UK (Feinber, 1983, 1984; Southern, 1975) was used for the DNA hybridisations. Hybridisations were performed at 65°C for 2hrs. High stringency post hybridisation washes were carried out at 65°C and 70°C in 2x SSC, 0.1%SDS, to ensure between 63 - 68% homology with the hybridisation probes.

4.3.4 PCR reactions

PCR reactions were performed using the Perkin Elmer Cetus DNA Thermal Cycler with Cetus Taq polymerase (New England Biolabs).

4.4. RESULTS

4.4.1 Nucleotide sequence of the 3.8 kb *Pst*I insert (pSBG300) corresponding to the insert of pSNS201

Difficulties were experienced with sequencing the insert of pSNS201, due to genetic instability and rearrangements. Initial sequencing of exoIII deletion clones, on ssDNA or dsDNA, and on M13 subclones gave different sequences for a region 1 kb from the *Pst*I site (Fig. 4.1a). In one of the subclones, a sequence with 80% homology to the ColE1 origin of replication of *E. coli,* was identified. The occurrence of this sequence was further investigated by PCR reactions on a *M.*

tuberculosis H37Rv cosmid, T256. The use of oligonucleotide primers for the PCR reactions was as follows: two oligonucleotides were designed for annealing to one strand based on sequences in pSNS201 upstream of the recombination/rearrangement region and within the ColE1 sequence (the ColEI oligonucleotide). A third oligonucleotide was synthesised for the complementary strand 400 bp down stream from the first primer (Fig. 4.7). These primers were used in two separate PCR reactions and the combination of the first primer upstream of the recombination/rearrangement and complementary strand primer region gave a specific reaction. The reaction with the "ColE1 oligonucleotide" primer produced almost six amplified fragments (Fig. 4.1b).

The specific PCR product of lanes 3-4 was sequenced, as was the cosmid DNA, using the same primers. No ColEI sequences could be found.

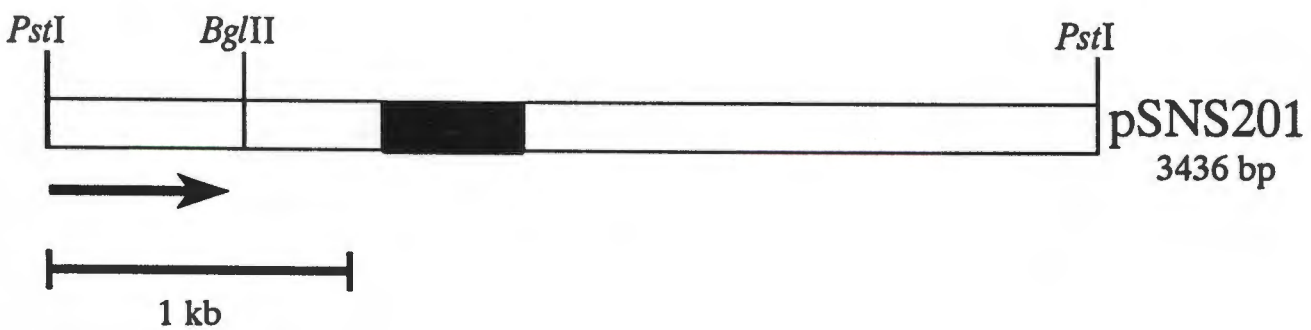


Fig. 4.1a A map of pSNS201. The shaded region indicates the region of recombination/rearrangement.

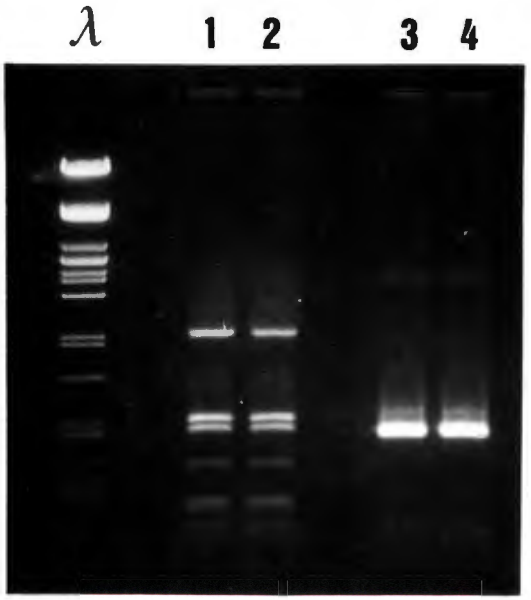


Fig.4.1b PCR amplification, using cosmid T256 as the template, of the 400 bp recombination/rearrangment region One primer was chosen downstream and two upstream of this region. Lanes 1 and 2, reactions with primers within the region of the recombination (ColEI oligonucleotide) and upstream of this region; lanes 3-4, reactions with primers downstream of the recombination region and the upstream primer. The product demonstrated in lanes 3-4 was used for sequencing.

4.4.2 Sequencing strategy used to sequence pSBG300 and pSNS201

*Sal*I, *Sma*I and *Hinc*II fragments generated from the *Pst*I inserts of both pSNS201 and pSBG300 (a subclone obtained from the cosmid T256 see restriction map Fig. 4.2) were subcloned into pUC18 and 19 as well as M13 and used as templates for DNA sequencing (see sequencing strategy in Fig. 4.4b). The complete nucleotide sequence of the 3.8 kb *Pst*I insert (pSBG300) was determined and confirmed by direct sequencing on the cosmid T256.

The analysis of the sequence of pSBG300 with DNA Strider demonstrated several ORFs in the 6 reading frames and showed one major ORF of 957 bp. The ORF is incomplete and appears to be truncated at the 5' end.

The sequence of pSNS206 showed no differences to the corresponding sequence of pSBG300 (data not shown). The subclone pSNS206 contains the truncated ORF referred to above and 120 bp distal to the termination codon (see Chapter 3).

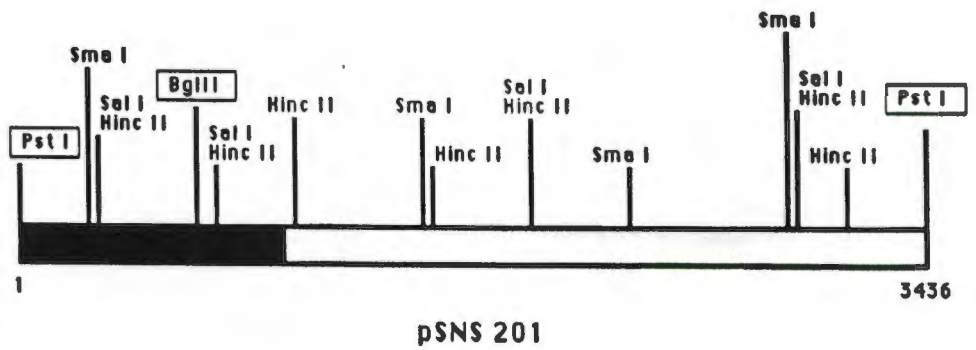


Fig 4.2 Restriction map of pSBG300/pSNS201.

4.4.3 Cloning of the 5' end of the putative *recA*

4.4.3a Screening of a H37Rv cosmid library

The internal *Bgl*II/*Pst*I fragment of the truncated ORF was used to probe a *Sau*3A cosmid library of H37Rv (courtesy of S. Cole, Institut Pasteur) in order to isolate an overlapping fragment containing the 5' end of the gene. Hybridisations revealed 7 positive clones; T733, T565, T481, T180, T252, T256 and T276.

Cosmid T256 and T276 DNA restricted with *Bgl*II (Fig. 4.3) and hybridised with the *Bgl*II/*Pst*I probe showed that the probe hybridised to a band of 4.8 kb in T256 and a smaller band of 2.8 kb in T276. Further investigation by hybridisation with the same probe and with the recovered 5' end of the ORF463 (see later) showed that the ORF was also truncated and lacking the 5' end in T276. The 4.8 kb *Bgl*II fragment of T256 was gel purified and sub-cloned in the *Bam*H1 site of pUC18 to generate pSBG100.

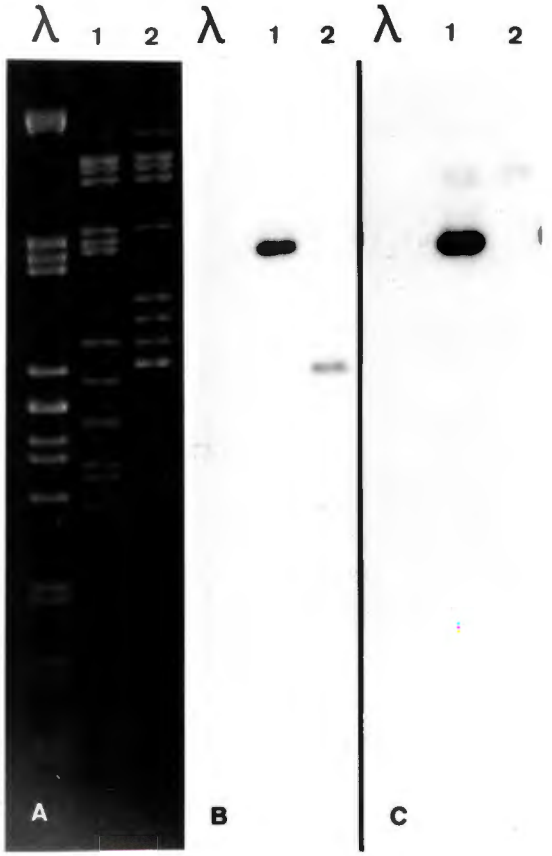


Fig.4.3 Cosmids T256 (lane 1) and T276 (lane 2) digested with *Bgl*II. A, agarose gel electrophoresis of the restricted cosmid DNA; B, autoradiograph of the same gel blotted and hybridised with the *Bgl*II/*Pst*I probe; C, same blot hybridised with the insert of pSBG102 (containing the 5' end of the gene) showing that in the cosmid T276 the gene is truncated and lacking its 5' end.

4.4.3b Restriction map and subcloning DNA fragments of pSBG100

A restriction map of pSBG100 is given in Fig. 4.4a and Fig. 4.4b. Restriction with *Bgl*II/*Pst*I (700 bp), PstI (900bp) and the *Hinc*II (400 and 500bp) fragments, provided DNA fragments that corresponded to similar fragments in pSNS201. The *Pst*I fragment (0.9 kb) and the *Bgl*II/*Pst*I fragment (0.7 kb), were

gel purified and subcloned into pUC18 to generate pSBG102 and pSBG101, respectively.

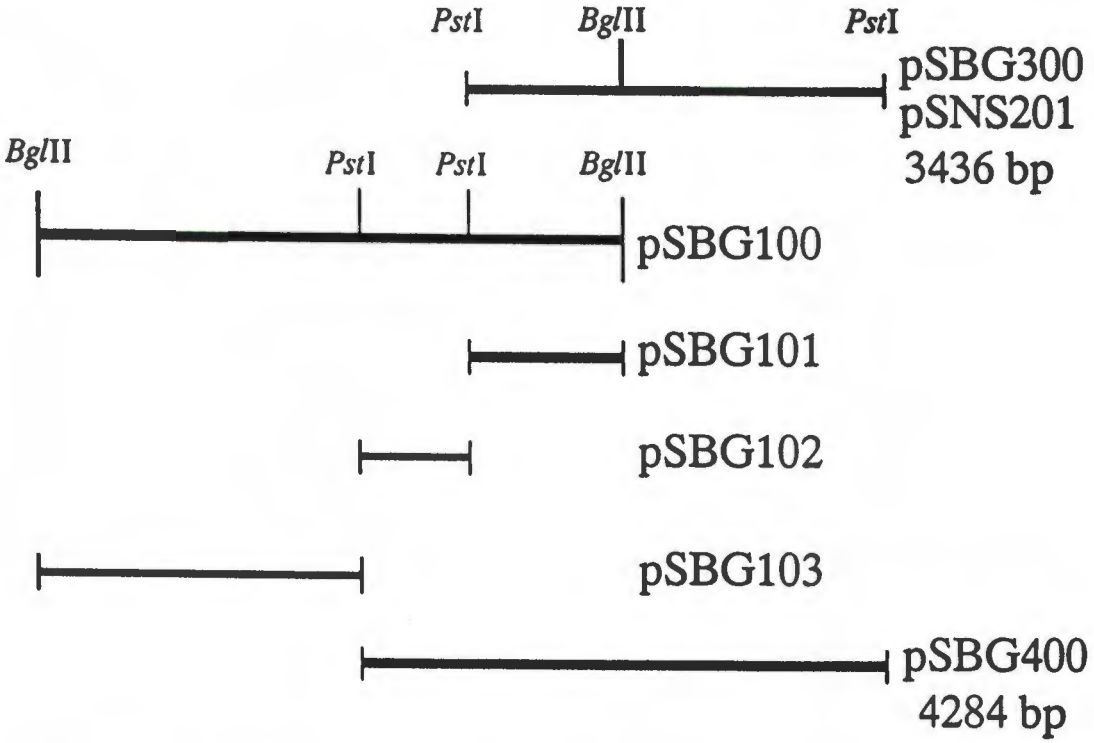


Fig.4.4a Restriction pattern and subclones derived from the recovered fragment pSBG100.

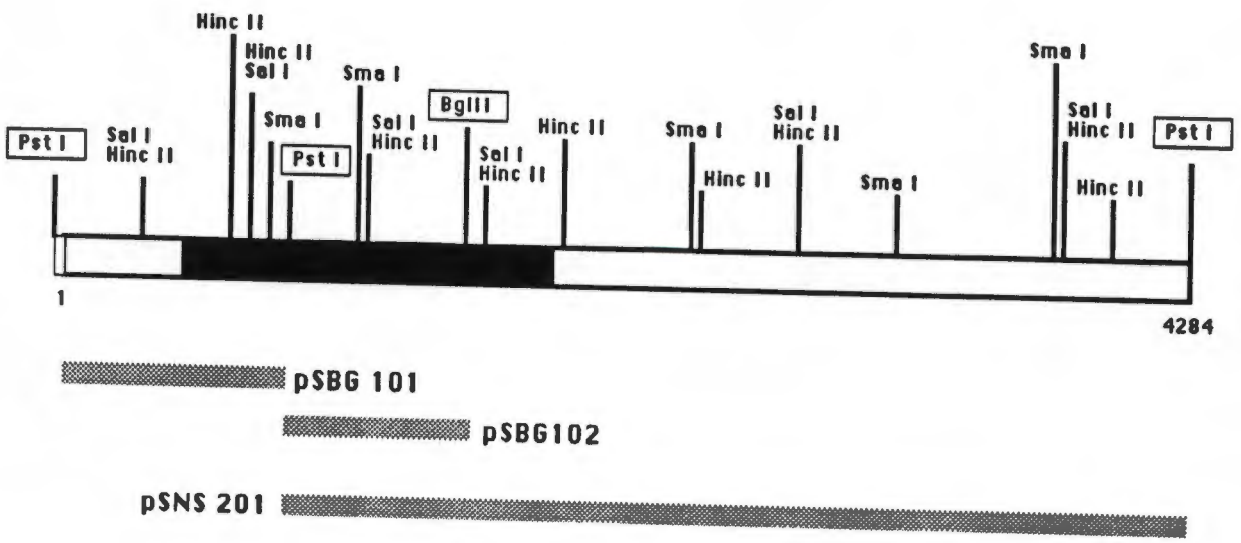


Fig. 4.4b Restriction map of pSBG101, pSBG102, pSBG300/pSNS201

4.4.3c Sequencing of pSBG101, pSBG102 and pSBG103

Double stranded DNA sequencing was performed partially on pSBG100, pSBG101 and pSBG102 in both directions. The sequences showed that the *Bgl*II/*Pst*I fragment of pSBG100 was identical to the *Bgl*II/*Pst*I fragment of pSNS201 and pSBG300 (Fig. 4.2), and permitted the isolation of an overlapping fragment of the 5' end of the truncated ORF from pSBG102. Hybridisation experiments using the *Bgl*II/*Pst*I probe of pSNS201 confirmed that pSBG100 contained the identical *Bgl*II/*Pst*I fragment (Fig. 4.5).

The DNA sequencing strategy using pSBG300, pSBG100 and the constructs pSBG101, pSBG102, is given in Fig. 4.6. The entire DNA sequence obtained from the contiguous PstI fragments (pSBG300 and pSBG102) was 4286 bp (Fig. 4.7) and is designated pSBG400. Identification of the open reading frames with DNA Strider programme is shown in Fig.4.8. The DNA sequence for the completed major ORF463 (1389 bp) and the deduced amino acid sequence is given in Fig. 4.9.

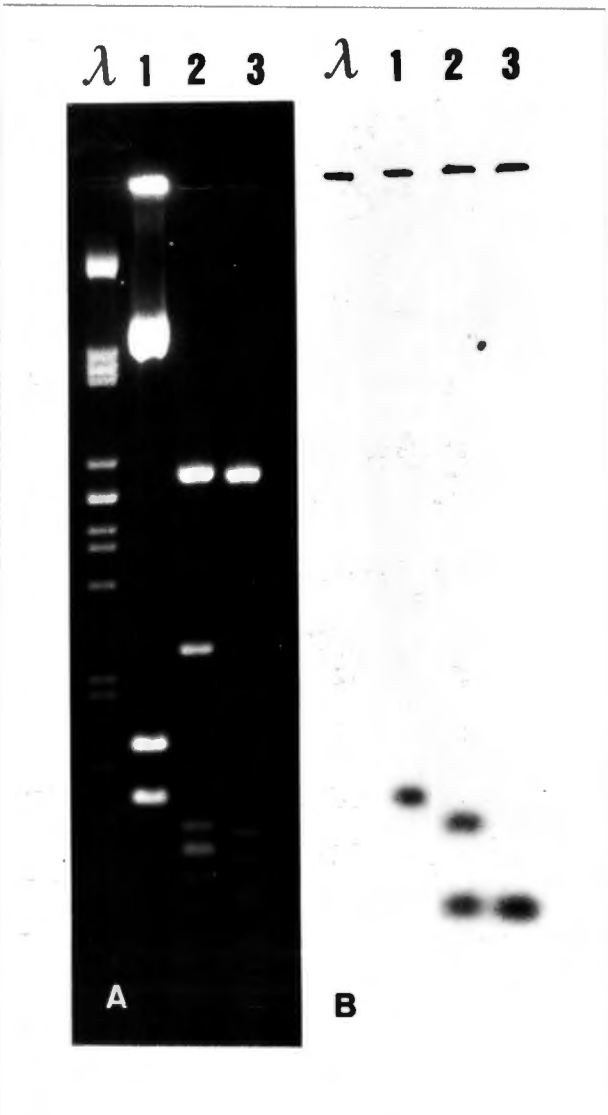


Fig. 4.5 Restriction pattern of pSBG100 containing the cloned *Bgl*II fragment; A, Electrophoresis of restricted DNA B, autoradiograph after hybridisation with the *Bgl*II/*Pst*I fragment. Lane 1;PstI; lane2 *Hinc*II; lane 3, *Ava*I/*Acc*I.

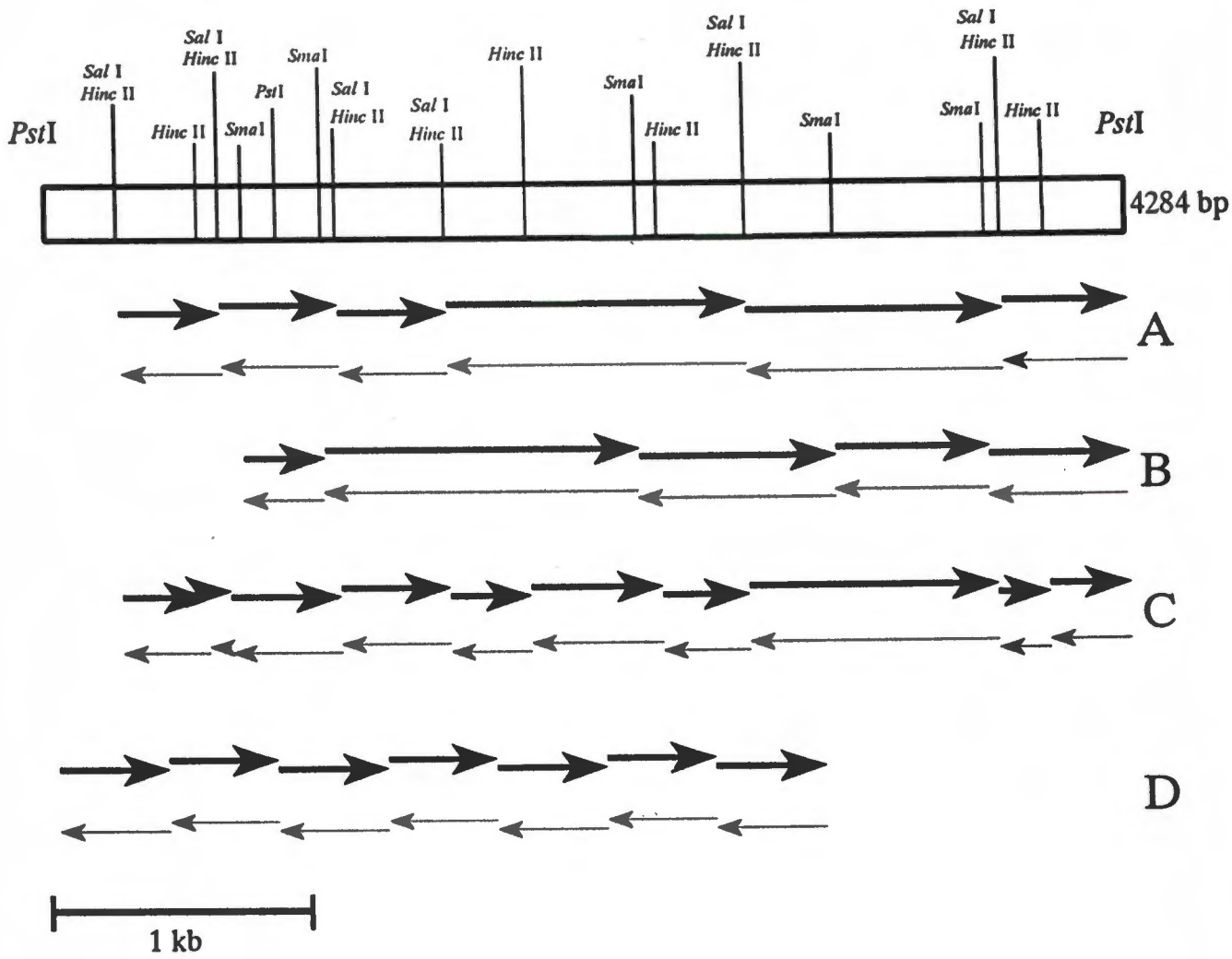


Fig. 4.6 DNA sequencing strategy using pSBG300 and pSBG100. Fragments used for sequencing A=*SalI*, B=*SmaI*, C=*HincII* and D=synthesised primers used for sequencing.

```
        |  10        |  20        |  30        |  40        |  50        |  60
   1 TTTCTTGACC GCAGCCAATG CCTCGGCCGA GCGCACGCCG TCTTGCAGTA CAGCACGGCG 60
  61 GTGCGGTCCT GGGGGAGCTT GGCCAGACCC TCACCCGAGT TGATCAACGA TTTCGGAATC 120
 121 AGTTGGGCTC CGTCGATATG CACGATGTCC CACTCACGGG ATCGCGAACG TCGATCAGTG 180
 181 CCAGCTTACG GCCGGAGTCC AGCCAGTCGC GCACGTCGCG CGGCGTGATG GTGGAACCTT 240
 241 TGGCCGCCTG GGCCGGCATC GTCAGCAACC ACGCCCGAGA ACTGTTCGTA GTCGACCAGC 300
 301 TCGGTGATCT TCGGTGTCGA TGGGTCCTTG CGGATGGTGA TCGTGCGATA GCTCATCTCC 360
 361 AGCGCGTCGT ACACCAGCAA CCGGCCAAGC AGTGTTTCAC CTATCCCGGT GATCAGCCTG 420
 421 ATTCGCCTCA GTGCCCATCA CCGATGCGAC CCAGGCACAG ATAATGCCCA GCACCCCGCC 480
 481 CTCAGCACAG GACGGCACCA TGCCCGGCGG CGGCGCGCTC GGGATACAGG TCGCGGTAGT 540
 541 TGATCTACCC AACCCGTCGG GGCGTCCTCC CAAAACACCG ATGCCTGGCC CTCGAAGCGG 600
 601 TAAATCGACC CCCACACGTA CGGCTTGCCA GCCAGCACCG CGGCGTTCGT TGACCAGATA 660
 661 CCGGGTGGCG AAGTTGTCGG TGCCATCCAA GATCAGGTCG TACTGCTTGA ACAGGTCGAC 720
 721 GGCGTTGCTC GGCGCAAGCC GCAGCTCGTG TAGTCGCACC CGGGATCAGC GGGTTGATCG 780
 781 CGACAATCGA ATCGCGCGCC GACTGAGCCT TGGAGCGCCC GACGTCAGCT ACCCCATGGA 840
 841 TGACCTGGCG CTGCAGGGGT GGGATCGTCG CGTGGGCCAA GCAGATGCAG CCCGACATGG 900
 901 TGATGTACTA GGACCGATAC AGGCCCAATA CACCCCGATG GGCTCTACTA GGCTGACCGG 960
 961 TGTGAATGTC GAGCCGCCGC CCGAGCACGT GTTGGTGGCG TTCGGTTTGG CAGGTGCGCA 1020
1021 ACCCATCCTG TTGGGTGCCG GTTGGGAAGG TGGCTGGCGA TGCGGCGAGG TGGTGTTGTC 1080
1081 GATGGTGGCC GACAACGCCC GCGCGGCCTG GTCGGCCCGG GTGCGCGAGA CGTTGTTCGT 1140
1141 CGACGGCGTA CGCTGGCTCG GCCCGTCCGA TCGACCGACG GCCGGTACGT GGTGTGTCTG 1200
1201 GTTGGCGGCA GACACGTCCG CCGGCGCGCC GGAGCCTAGG CATGATGAGG TCGTCTCGGC 1260
1261 GGCGGTGCGG CTGCATGAGG CCACCGGAAA ACTGGAACGC CCCCGATTCT TGACCCAGGG 1320
1321 ACCCGCGGCG CCCTGGGCCG AGATCGATGT GTTCGTCGCC GCAGACCGAG CTGGGTGGGA 1380
1381 GGAGCGGCCA TTACAGTCGG TCCCACCGGG CGTGCCGACC GCCCCCCCGG CAGCGGACCC 1440
1441 CCAGCGATCG ATCGATCTGA TCAATCAGCT TGCCGGGTTG CGTAAGCCGA CCAAGAGCCC 1500
1501 GAACCAGCTG GTGCACGGAG ATCTTTACGG TACAGTGCTT TTGGCGGGCA CCGCCCTCCA 1560
1561 GGGATCACCG ACATCACGCC TACTGGCGGC CCGCATCCTG GGCGGCCGGG GTGGCCGTCG 1620
1621 TCGACGCGCT GTCCTGGGGT GCGGCCGACG AACGGCTCAT CGAGCGGTGG AACGCGCTGC 1680
1681 CGGAGTGGCC CCCAGAATGT TGTTGCGCGC GTTGATGTTC CGCCTAGCAG TGTACGCGCT 1740
1741 GCACCCACGA TCCACCGCCG AAGCGTTTCC CGGCCTGGCC CACACCGCGG CCCTAGTGCG 1800
1801 GCTAGTGCTC TAAACCTACT CAGCCCGGCG GAAACTCATA GCGGATCTCG CTCAGCGCCA 1860
1861 CCCGGCCGTC AACACTGAGA ACGCCCTCTG CGCGCAACAA CTCCAACTGC CGGGTGGCCA 1920
1921 GGTGCTGTGC TGGGCGCCCG GAGGCTCTGA TCACCCGGTG CCAGGGCAGA TCCGAGGAAT 1980
1981 CGGTCCGCAT AATCCAGCCG ACAATACGCG GACTGGAAAG CCCTGTGAAG AGCTGCGAAT 2040
2041 GTCTACAGGG CGCGACCTCA GTGGCTTGGC TTGGCACATT GGCCGGCAGT CGAAGTCTAG 2100
2101 TGCTAAAATG CTCCTAGCCG TTGTCCAGGG CGCGAGGCAG CTGGTCGCTG AGTACGCCGG 2160
2161 GTGAGCACGG ACCGCCCGAA CCAGGTGGTT GCCGTTCCCC ACCGGCGGCA GCACGATGTC 2220
2221 GCGGGCCAGA TTCGCTCCAG TAGCACACCA TCGCCGGCCA GCTATACGCC AACCGTACCG 2280
2281 ACCGTTGGGC AATGCACGAG GTGCTCGTCG AGTTCGGCGT CGAGCACTCC CCCCTCCACG 2340
2341 TCCGCCCAGG CACGGTCGCG GCGCGGCCCC GGGACCTGCC GGGTCGAGTC AACACCAGCG 2400
2401 CCGCGACCTG GTCTGGGCGG GCCGCGGCCA GGTGCATAGC GACAGCGCCG CCGAAGGAGT 2460
2461 GTCCGACCAC CACTACCGGA CCGTCGCCCT GATTGTCGAG GAGTGCTGCC AGGGCGGACA 2520
2521 CGTTGGCGTC GATGGTCCAC GCGCGGCCCA TGGTGACCTA CCGTGGCCCA GCAGATCGGG 2580
2581 TGCGGCGATG GGGATTTCGG GCAAGTGATG GGCTAACCGG TGCCAGATGC GCCCGTGCTC 2640
2641 GGTACCTCCG TGGATGGTCA GCACCCGCGC GCCGTGACGG GCCGTAGCGC TGTACGTGAA 2700
2701 GGTCGATGAT CACGCGTCGA TGATGCCAGC CCGGCGACGC CTGCTACCGG TCGATTTCTT 2760
2761 CGTCGTCGGC GACCGGAACG ACGAAGGCTT GGTCGACGAC GTCGCGTTCG CTGGCGTCTC 2820
2821 GGTCGCCGGC GCCGCCGCTG AGGTAGGCGG TGTCCAACCC GGCTTCGTCG TCGAAATCGA 2880
2881 CAGCACGCCC TTGCTCCACC GCATCGGCTT CGGGTATTTC GTCCTGTGGA CCGCCGCCTA 2940
2941 CGACCGCCGC CATGGTAGCC ACCCCTCACA TCGCGTTTTC GGTTTTCTTG CCAGCTCCAG 3000
3001 TATCCGAGCT TGTCAGACCC TCGTGATGTC ATGCGGCTAT GTCACATATC TGGGGTGTCG 3060
3061 AGGCGGGTGC GGCCCTCGCG CCGGGTTTGC GCGGTCCGGT GCTTGTGCTG GGCGGCCCGG 3120
3121 GCACCGTAAG AGCACACTGT TGGTTGAGGC CGCGGTCGCT CACATCGGCG CCGGCACCGA 3180
3181 CCCGGAGTCG GTTCTGCTGC TGACCGGTTC CGGCCGAATG GGCATGCGGG CCCGCAGTGC 3240
3241 GCTGACGACG GCGCTGCTGC GGTCGCGCAC CAACGGCCCT TGCCGGGCGG CGATCCGCGA 3300
3301 ACCGGTGGTA CGCACCGTGC ACACGTACGC CTATGCGGTC TTGCGCAAGG CGCACAGCGC 3360
3361 GCCGGTGACG CCTTGCCGCG GCTGCTTACC AGCGCCGAGC AGGACGCCAT CATTCGGGAA 3420
3421 CTGCTGGCCG GGGACGCCGA AGACGGACCG GCCGCCACCA CCACCTGGCC TGCGCATCTG 3480
3481 CGGCCCGCGC TGACTACCGC CGCTTCGCCA CCGAGCTGCG AAACCTGTTG GCACGTTGCG 3540
3541 CCGAACGCGG CCTGGACCCG CTGGAGTTGC AGCAACTGGC CGTCGCCGCG GCCGTCCGGA 3600
3601 ATGGTACGCC GCCGGTCAAT TCGCCCAGCG GTACGAGCAG GTGATGTTGT TGCGGGGTGC 3660
3661 GGTGGGGCTG CGGCGCAGGC CACGGCGCCG GCGCTGAGTG CCGCCGAACT GGTGGGGGCA 3720
3721 GCTTTGGAGG CCTTCGCGGT CGATCCTGAG TTACTGGCCG CCGAACGCGC CCGGGTCCGG 3780
3781 ACCCTTTTGG TCGACGACGC CCAACAACTT GATCCGCAGG CGGCACGCCT GGTCCGGATG 3840
3841 CTGGCGGCGG GCACCGAGCT GGCCCTGATC GCCGGTGATC CGAACCAAGC GGTGTTCGGG 3900
3901 TTCCGCGGCG GCGAGCCCAC CGGCCTGCTG GCCGACGATC CGCCGCCAGC AGGTGGTGCC 3960
3961 CCCATTCCGT CGGTGACGTT GACGGTCTCT CATCGGTGTG CACCCGCCGT GGCGCGGGCG 4020
4021 GTCACCGGCA TCGCACGACG GCTACCGGGT CGAAGCGTCG GTCGGCGAAT CGAGGGCACC 4080
4081 GGGACCGAGG TCGGATCGGT GACGGTGCGC CTGGCCGGTT CGGCGCACGC TGAGGCAGCG 4140
4141 ATGATTGCCG ACGCGCTGCG ACGCGCGCAC CTAGATCGAT GGGGTGCCTT GGTCGCAGAT 4200
4201 GGCGGTGATC GTCAGGTCGG TGCCGCGCGC TGTGCGGTTG CCGCGCGCTC TGGCCGCCGC 4260
4261 CGGGGTGCCG GTGGCCCCAC CTGCAG                                       4286
```

Fig.4.7 The complete nucleotide sequence obtained in this study. In essence it consists of the two contiguous *Pst*I fragments from T256. A total of 4286 bp was obtained (pSBG400). ORF463 lies between nucleotides 422-1811. Oligonucleotides used for the PCR reaction are under lined.

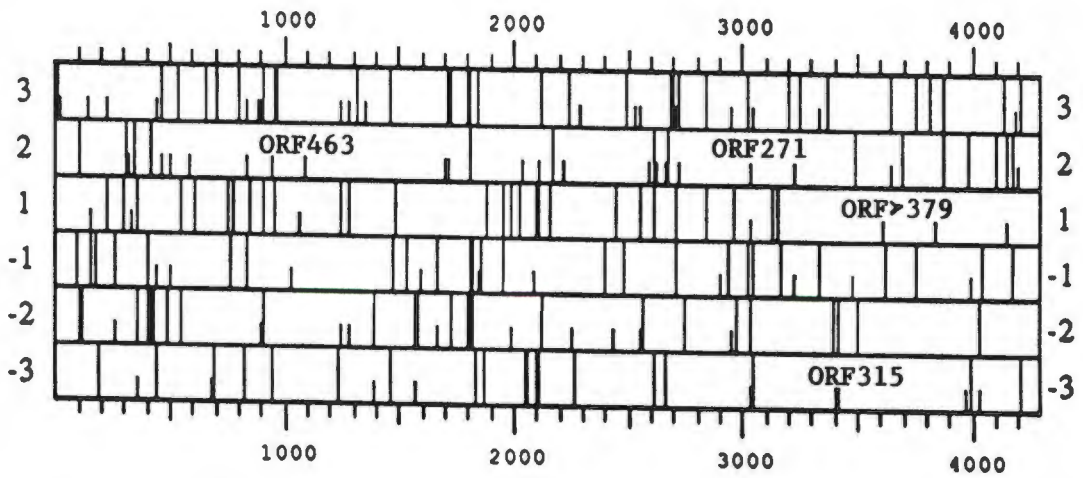


Fig. 4.8 Analysis of the open reading frames in all the phases with DNA strider program for the entire sequence of pSBG400. Major ORFs are labelled (ORF463, ORF271, ORF>379, which is truncated and ORF315 on the complementary strand).

4.4.4 Analysis of the sequence of pSBG400

pSBG400 consists essentially of two contiguous *Pst*I fragments of the cosmid T256. This sequence contains the complete sequence of the truncated ORF463, present in pSNS206 and pSBG300, which encodes the functional RecA-like protein (Chapter 3). The total sequence presented here is 4286 bp in length (Fig. 4.7). An analysis for open reading frames is shown in Fig. 4.8. There are 4 open reading frames of greater than 800 bp; three are present in the same orientation and the fourth, in the opposite orientation, is read off the complementary strand (Fig. 4.8.) The ORFs designated ORF463 (nucleotides 422-

1811), ORF271 (nucleotides 2678-3491), ORF>379 (truncated, nucleotides 3148-4286), ORF315 (complementary strand, nucleotides 3988-3043) respectively. The numbers correspond to the total number of amino acids potentially encoded by each ORF.

It was shown in Chapter 3 that pSNS206, which contains a truncated portion of ORF463 alone, produced a protein on Western blot analysis which cross-reacted with polyclonal *E. coli* RecA antiserum. pSNS206 was also shown to mediate internal homologous recombination in an E. coli *rec lac* host (Konrad, 1977). Thus ORF463 is presumed to encode the RecA homologue. No homology to any published *recA* genes with ORF463 is present. Furthermore, sequence alignment of the putative product at the amino acid level with RecA proteins shows 11-13% identity (Fig. 4.14a and b).

ORF463, contains 5 possible methionine sites before base 852 (*Pst*I site, Fig. 4.9). The possible proteins initiated at the different start sites would have the following predicted molecular weights: 49,6 kDa (463 aa, nucleotide 431), 47, 811 kDa (449 aa, nucleotide 464), 46,641 kDa (437 aa, nucleotide 500), 44,035 kDa (410 aa, nucleotide 581) and 34,626 kDa (325 aa, nucleotide 836).

Analysis of ORF463 for putative ribosome binding sites by comparison with ribosome binding sites in GC% rich organisms (Strohl, 1992), seem to indicate that only two of the putative initiation codons may serve for protein initiation. These are at position 500 (encoding a 46,64 kDa protein) and position 836 (encoding a 34,62 kDa protein). these sequences, AGGA and GGAGC, are 6 and 18 nucleotides upstream from the methionine codons (Fig. 4.9).

```
CGG TGA TCT TCG GTG TCG ATG GGT CCT TGC GGA TGG TGA TCG TGC GAT AGC TCA TCT CCA
GCG CGT CGT ACA CCA GCA ACC GGC CAA GCA GTG TTT CAC CTA TCC CGG TGA TCA GCC TGA
                                                                          *

TTC GCC TCA GTG CCC ATC ACC GAT GCG ACC CAG GCA CAG ATA ATG CCC AGC ACC CCG CCC
 F   A   S   V   P   I   T   D   A   T   Q   A   Q   I   M   P   S   T   P   P   6

TCA GCA CAG GAC GGC ACC ATG CCC GGC GGC GGC GCG CTC GGG ATA CAG GTC GCG GTA GTT
 S   A   Q   D   G   T   M   P   G   G   G   A   L   G   I   Q   V   A   V   V  26

GAT CTA CCC AAC CCG TCG GGG CGT CCT CCC AAA ACA CCG ATG CCT GGC CCT CGA AGC GGT
 D   L   P   N   P   S   G   R   P   P   K   T   P   M   P   G   P   R   S   G  46

AAA TCG ACC CCC ACA CGT ACG GCT TGC CAG CCA GCA CCG CGG CGT TCG TTG ACC AGA TAC
 K   S   T   P   T   R   T   A   C   Q   P   A   P   R   R   S   L   T   R   Y  66

CGG GTG GCG AAG TTG TCG GTG CCA TCC AAG ATC AGG TCG TAC TGC TTG AAC AGG TCG ACG
 R   V   A   K   L   S   V   P   S   K   I   R   S   Y   C   L   N   R   S   T  86

GCG TTG CTC GGC GCA AGC CGC AGC TCG TGT AGT CGC ACC CGG GAT CAG CGG GTT GAT CGC
 A   L   L   G   A   S   R   S   S   C   S   R   T   R   D   Q   R   V   D   R 106

GAC AAT CGA ATC GCG CGC CGA CTG AGC CTT GGA GCG CCC GAC GTC AGC TAC CCC ATG GAT
 D   N   R   I   A   R   R   L   S   L   G   A   P   D   V   S   Y   P   M   D 126

GAC CTG GCG CTG CAG GGG TGG GAT CGT CGC GTG GGC CAA GCA GAT GCA GCC CGA CAT GGT
 D   L   A   L   Q   G   W   D   R   R   V   G   Q   A   D   A   A   R   H   G 146

GAT GTA CTA GGA CCG ATA CAG GCC CAA TAC ACC CCG ATG GGC TCT ACT AGG CTG ACC GGT
 D   V   L   G   P   I   Q   A   Q   Y   T   P   M   G   S   T   R   L   T   G 166

GTG AAT GTC GAG CCG CCG CCC GAG CAC GTG TTG GTG GCG TTC GGT TTG GCA GGT GCG CAA
 V   N   V   E   P   P   P   E   H   V   L   V   A   F   G   L   A   G   A   Q 186

CCC ATC CTG TTG GGT GCC GGT TGG GAA GGT GGC TGG CGA TGC GGC GAG GTG GTG TTG TCG
 P   I   L   L   G   A   G   W   E   G   G   W   R   C   G   E   V   V   L   S 206

ATG GTG GCC GAC AAC GCC CGC GCG GCC TGG TCG GCC CGG GTG CGC GAG ACG TTG TTC GTC
 M   V   A   D   N   A   R   A   A   W   S   A   R   V   R   E   T   L   F   V 226

GAC GGC GTA CGC TGG CTC GGC CCG TCC GAT CGA CCG ACG GCC GGT ACG TGG TGT GTC TGG
 D   G   V   R   W   L   G   P   S   D   R   P   T   A   G   T   W   C   V   W 246

TTG GCG GCA GAC ACG TCC GCC GGC GCG CCG GAG CCT AGG CAT GAT GAG GTC GTC TCG GCG
 L   A   A   D   T   S   A   G   A   P   E   P   R   H   D   E   V   V   S   A 266

GCG GTG CGG CTG CAT GAG GCC ACC GGA AAA CTG GAA CGC CCC CGA TTC TTG ACC CAG GGA
 A   V   R   L   H   E   A   T   G   K   L   E   R   P   R   F   L   T   Q   G 286

CCC GCG GCG CCC TGG GCC GAG ATC GAT GTG TTC GTC GCC GCA GAC CGA GCT GGG TGG GAG
 P   A   A   P   W   A   E   I   D   V   F   V   A   A   D   R   A   G   W   E 306

GAG CGG CCA TTA CAG TCG GTC CCA CCG GGC GTG CCG ACC GCC CCC CCG GCA GCG GAC CCC
 E   R   P   L   Q   S   V   P   P   G   V   P   T   A   P   P   A   A   D   P 326

CAG CGA TCG ATC GAT CTG ATC AAT CAG CTT GCC GGG TTG CGT AAG CCG ACC AAG AGC CCG
 Q   R   S   I   D   L   I   N   Q   L   A   G   L   R   K   P   T   K   S   P 346

AAC CAG CTG GTG CAC GGA GAT CTT TAC GGT ACA GTG CTT TTG GCG GGC ACC GCC CTC CAG
 N   Q   L   V   H   G   D   L   Y   G   T   V   L   L   A   G   T   A   L   Q 366

GGA TCA CCG ACA TCA CGC CTA CTG GCG GCC CGC ATC CTG GGC GGC CGG GGT GGC CGT CGT
 G   S   P   T   S   R   L   L   A   A   R   I   L   G   G   R   G   G   R   R 386

CGA CGC GCT GTC CTG GGG TGC GGC CGA CGA ACG GCT CAT CGA GCG GTG GAA CGC GCT GCC
 R   R   A   V   L   G   C   G   R   R   T   A   H   R   A   V   E   R   A   A 406

GGA GTG GCC CCC AGA ATG TTG TTG CGC GCG TTG ATG TTC CGC CTA GCA GTG TAC GCG CTG
 G   V   A   P   R   M   L   L   R   A   L   M   F   R   L   A   V   Y   A   L 426

CAC CCA CGA TCC ACC GCC GAA GCG TTT CCC GGC CTG GCC CAC ACC GCG GCC CTA GTG CGG
 H   P   R   S   T   A   E   A   F   P   G   L   A   H   T   A   A   L   V   R 446

CTA GTG CTC TAA ACC TAC TCA GCC CGG CGG AAA CTC ATA GCG GAT CTC GCT CAG CGC CAC
 L   V   L   *                                                              449

CCG GCC GTC AAC ACT GAG AAC GCC CTC TGC GCG CAA CAA CTC CAA CTG CCG GGT GGC CAG

GTG CTG TGC TGG GCG CCC GGA GGC TCT GAT CAC CCG GTG CCA GGG CAG ATC CGA GGA ATC
```

Fig. 4.9 Complete nucleotide sequence and deduced proteins of ORF463. The sequence includes 162bp upstream from the ORF. Potential initiation codons are in bold, and inverted repeats are arrow overlined. Putative RBS are overlined.

Comparison of the 5' region of ORF463 to the corresponding region of the *E. coli recA* gene did not identify any typical promoter sequences. In particular, there is an absence of an *E. coli* SOS-like box (Fig. 4.10). A potential promoter sequence was identified 60 bp upstream from the initiation codon at position 836. Initiation at this codon would lead to the production of the protein of 325aa and 34,6 kDa.

```
 -35                        -10

ACTTGATACTGTATGAGCATACAGTATAATTCGTTCAA--GGAGTAAAAATGGCT

GGTTGATCGCGACAATCGAATCGCGCGCCGACTGAGCCTTGGAGCn19ATG

                                          Shine-  Start

                                          Dalgarno
```

Fig. 4.10 The top sequence shows the *E. coli recA* promoter region with the LexA binding site in bold; the -35 and -10 regions are underlined (Weisemann and Weinstock, 1991). The bottom sequence shows a possible promoter region in ORF463 detected 60 bp upstream from initiation codon 836. No Lex A-like binding sites were found in relation to ORF463.

4.4.5 Computer analysis of the putative product of ORF463

4.4.5a Hydrophobicity

Computer analysis of the putative product of ORF463, for hydrophobicity (Kate and Doolittle programme with DNA Strider) (Fig.4.11) indicates two predominantly hydrophobic regions separated by a hydrophilic

indicates two predominantly hydrophobic regions separated by a hydrophilic region. The 3' end of the hydrophilic region contains the fifth initiation codon and would initiate translation of a protein with a molecular weight of 34,626 kDa (325 aa). This plot is shown along with the *recA* from *M. tuberculosis* (Fig. 4.11). The region with the strongest hydrophilicity in the RecA is in the "intron" which is excluded from the mature protein (Fig. 4.11).

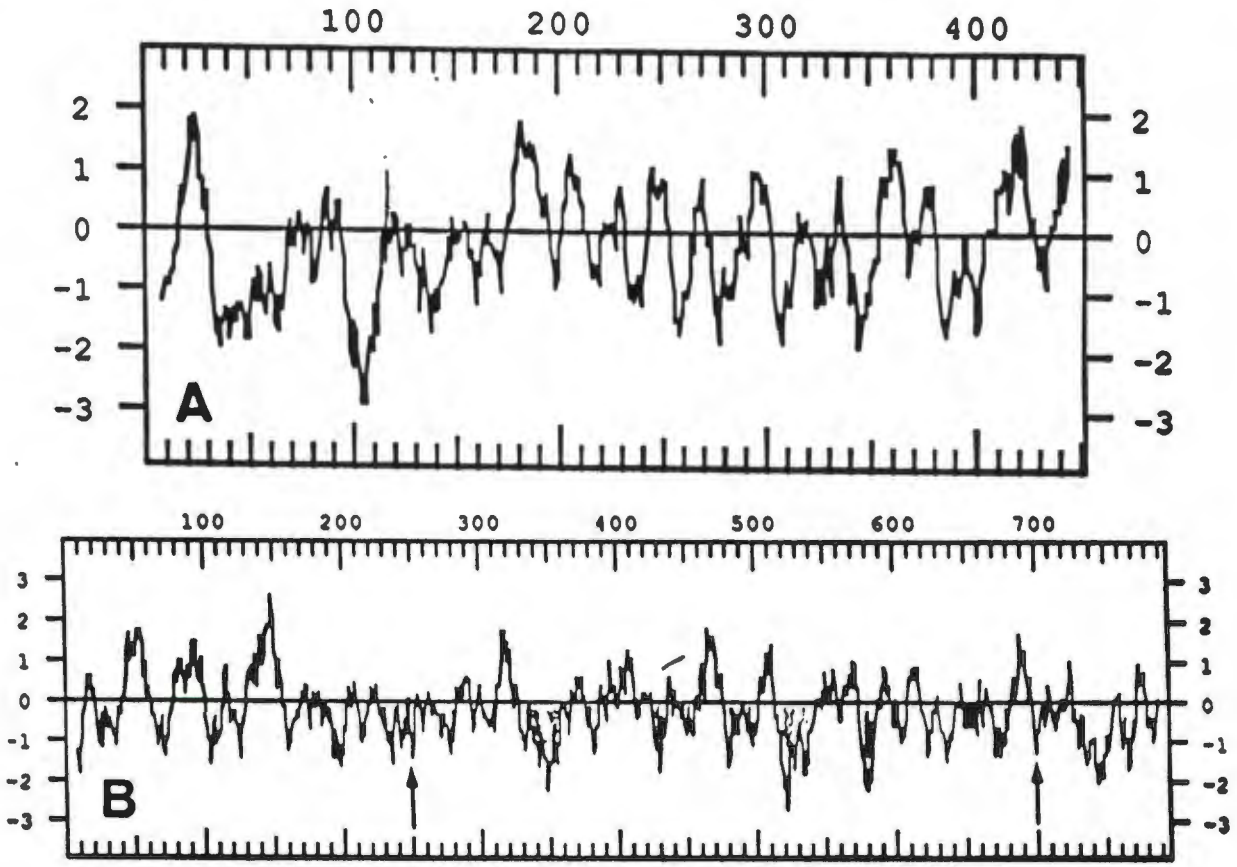


Fig. 4.11 Hydrophobicity profile obtained with DNA strider program. A, The putative product of ORF463; B, RecA from *M. tuberculosis*. Arrows define the intron region in the *M. tuberculosis recA*.

4.4.5b Codon usage of *recS*

ORF463 was shown to have a 66 % GC content. A comparison of the codon usage in ORF463 and the *M. tuberculosis recA* was done using the GCG Package - pattern recognition programme (Fig. 4.12). There is a bias for the TTG

A difference in codon bias is also seen for CGA encoding Arg. It is much higher in *recS* (25%) compared to 5% for the *M. tuberculosis recA*.

The Codon Preference programme was used to plot the similarity of the codon usage in *recS* to the pattern of codon usage for the gene encoding the 65 kDa protein of *M. tuberculosis* (Fig. 4.13a). The 65 kDa antigen of *M. tuberculosis* is a heat shock protein and a major T-cell immunogen (Shinnick, 1987). Codon Preference is a frame-specific gene finder; and recognises protein coding sequences by virtue of the similarity of their codon usage to a codon table. Codon Preference is also useful for locating sequencing errors. A sequencing error that causes a frame shift can be detected. Frameshifts were not detected in *recS*, thus *recS* has a similar codon usage as the gene encoding the 65 kDa protein and no sequencing errors were revealed. The codon usage of *recS* was confirmed by the TestCode programme (Fig. 4.13b).

	A	B		A	B		A	B		A	B
TTT phe F	1	-	TCT ser S	1	-	TAT tyr Y	-	1	TGT cys C	2	-
TTC phe F	5	9	TCC ser S	4	1	TAC tyr Y	6	5	TGC cys C	4	1
TTA leu L	1	-	TCA ser S	3	1	TAA OCH Z	-	-	TGA OPA Z	-	-
TTG leu L	16	2	TCG ser S	12	12	TAG AMB Z	-	-	TGG trp W	9	1
CTT leu L	4	2	CCT pro P	4	-	CAT his H	4	-	CGT arg R	7	2
CTC leu L	5	11	CCC pro P	17	6	CAC his H	4	3	CGC arg R	15	8
CTA leu L	6	1	CCA pro P	5	-	CAA gln Q	3	-	CGA arg R	14	1
CTG leu L	14	15	CCG pro P	17	7	CAG gln Q	12	13	CGG arg R	9	6
ATT ile I	-	4	ACT thr T	1	-	AAT asn N	3	4	AGT ser S	1	1
ATC ile I	7	16	ACC thr T	14	12	AAC asn N	4	7	AGC ser S	7	4
ATA ile I	2	2	ACA thr T	4	1	AAA lys K	3	2	AGA arg R	2	-
ATG met M	8	9	ACG thr T	7	5	AAG lys K	4	20	AGG arg R	4	1
GTT val V	2	1	GCT ala A	5	2	GAT asp D	12	6	GGT gly G	11	11
GTC val V	10	11	GCC ala A	21	16	GAC asp D	9	19	GGC gly G	20	20
GTA val V	3	-	GCA ala A	10	3	GAA glu E	4	4	GGA gly G	7	4
GTG val V	20	18	GCG ala A	24	17	GAG glu E	10	19	GGG gly G	6	3

Fig. 4.12 Codon usage of ORF463 compared to the *recA* from *M. tuberculosis*. Column A= ORF463 and column B= *M.tuberculosis recA*.

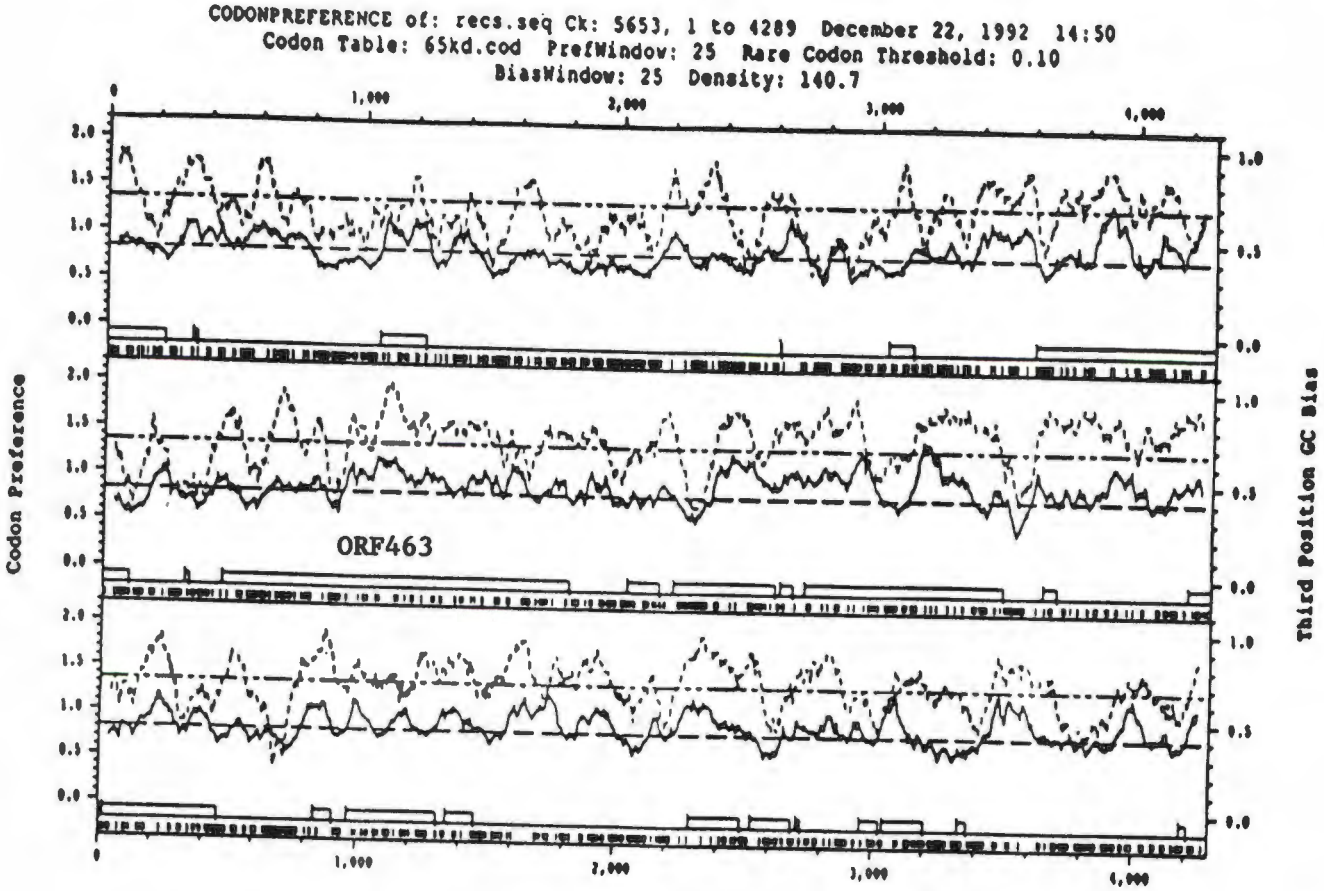


Fig. 4.13a Codon preference of ORF463 compared to the gene encoding the 65 kDa protein of *M. tuberculosis*.

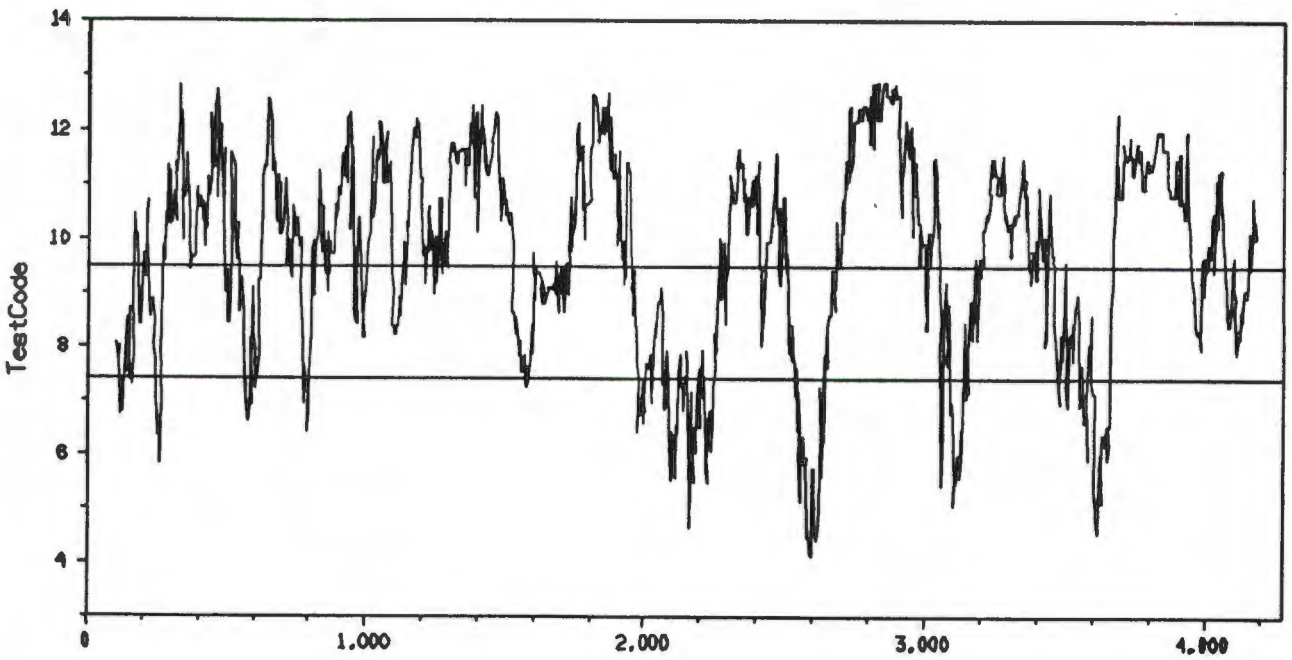


Fig. 4.13b Analysis of the codon usage of *recS* by the TestCode programme.

4.4.6 Multiple sequence alignment

An amino acid sequence alignment between the putative product of ORF463 and the proteins encoded by two *recA* genes *(E. coli* and *M. tuberculosis)* was done using the Clustal V programme (Higgen *et al.*, 1991). This comparison was done with the entire open reading frame of the *M. tuberculosis* RecA including the "intron" region (Fig. 4.14a). A comparison between the putative product of ORF463 and the *Mycobacterium tuberculosis* RecA excluding the "intron" region is given in Fig. 4.14b. In neither alignment does the similarity between ORF463 and the other two proteins exceed 13%.

Examination of the 5' regions of all three RecAs, (*E. coli, M. tuberculosis* and the product of ORF463) demonstrates the presence of three glutamines. This arrangement of amino acids appears to be important for polymerisation (Story *et al.*, 1992) and indicates that the protein encoded by ORF463 may be capable of polymerisation. A conserved motif was observed in the binding site for ATP (Q--n--LYG) (Story *et al.*, 1992). These results show that ORF463 encodes a novel RecA-like protein and consequently this protein is designated RecS and its corresponding gene *recS.*

```
CLUSTAL V multiple sequence alignment

E. coli          MAIDENKQK----------ALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLS---
RecA:            MTQTPDREK----------ALELAVAQIEKSYGKGSVMRLGDEARQPISVIPTGSIA---
Sequence         MPSTPPSAQDGTMPGGGALGIQVAVVDLPNPSGRPPKTPMPGPRSGKSTPTRTACQPAPR
                 *       .           ..  *. .. .  *. .   .                *.  .

E. coli          ----------LDIALGAGGLPMGRIVEIYGPESSG------------------------
RecA:            ----------LDVALGIGGLPRGRVIEIYGPESSG------------------------
Sequence         RSLTRYRVAKLSVPSKIRSYCLNRSTALLGASRSSCSRTRDQRVDRDNRIARRLSLGAPD
                           * ..    .    *   . *.  *.

E. coli          ----KTTLTLQV----IAAAQREGKTCAFIDAEHALDPIYARKL-GVDID----------
RecA:            ----KTTVALHA----VANAQAAGGVAAFIDAEHALDPDYAKKL-GVDTD----------
Sequence         VSYPMDDLALQGWDRRVGQADAARHGDVLGPIQAQYTPMGSTRLTGVNVEPPPEHVLVAF
                       ..*.     .. *.       .   .    *  . .* **. .

E. coli          NLLCSQP-----------DTGEQALEICDALARSG----AVDVIVVDSVAALTPKAEIEG
RecA:            SLLVSQP-----------DTGEQALEIADMLIRSG----ALDIVVIDSVAALVPRAELEG
Sequence         GLAGAQPILLGAGWEGGWRCGEVVLSMVADNARAAWSARVRETLFVDGVRWLGPSDRPTA
                  *  .**              **  * .    *..      . . .*.* .* *     .

E. coli          EI------GDSHMGLAAR----MMSQAMR--KLAGNLKQSNTLLIFINQIRMKIGVMFGN
RecA:            EM------GDSHVGLQAR----LMSQALR--KMTGALNNSGTTAIFINQLRDKIGVMFGS
Sequence         GTWCVWLAADTSAGAPEPRHDEVVSAAVRLHEATGKLERPRFLTQGPAAPWAEIDVFVAA
                        .*.  *        ..* *.*    .* *  .             .  *.*  .

E. coli          PETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYG
RecA:            PETTTGGKALKFYASVRMDVRRVETLKDGTNAVGNRTRVKVVKNKCLAPFKQAEFDILYG
Sequence         DRAGWEERPLQ---SVPPGVPTAPPAADPQRSIDLINQLAGLRKPTKSP-NQLVHGDLYG
                   .     .*.   **  ..       .    ..   ..  ...   .* .*     ***

E. coli          E----GINFYGELVDLGVKEKLIEKAGAWYSYKGEKIGQGKANATAWLKDNPETA-KEIE
RecA:            K----GISREGSLIDMGVDQGLIRKSGAWFTYEGEQLGQGKENARNFLVENADVA-DEIE
Sequence         TVLLAGTALQGSPTSRLLAARILGGRGGRR--RRAVLGCGRRTAHRAVERAAGVAPRMLL
                       *    *      .  ..  *..       .* *.  *   .   .  *   .

E. coli          KKVRELLLSNPN----STPDFSVDDSEGVAETNEDF
RecA:            KKIKEKLGIGAV----VTDD---PSNDGVLPAPVDF
Sequence         RALMFRLAVYALHPRSTAEAFPGLAHTAAL-VRLVL
                 . .   *   .      .       .      .
```

Fig. 4.14a Amino acid sequence alignment between the *E. coli* RecA[1], the entire sequence of the *M. tuberculosis* RecA[2] and RecS[3] using the Clustal V programme. Key: *, indicates base identity and •, conserved substitution (PAM250).

```
RecS:          MPSTPPSAQDGTMPGGGALGIQVAVVDLPNPSGRPPKTPMPGPRSGKSTPTRTACQPAPR
RecA (Mtb)     MTQTPDREK----------ALELAVAQIEKSYGKGSVMRLGD---------------EAR
               *  **   .          ....** .. .. *. .   . .                .*

RecS:          RSLTRYRVAKLSVPSKIRSYCLNRSTALLGASRSSCSRTRDQRVDRDNRIARRLSLGAPD
RecA (Mtb)     QPISVIPTGSIA---------LDVALGIGGLPRG-----------------RVIEIYGPE
               ....    . ..          *. . .. * .*.                  * . . .*.

RecS:          VSYPMDDLALQGWDRRVGQADAARHGDVLGPIQAQYTPMGSTRLTGVNVEPPPEHVLVAF
RecA (Mtb)     SS-GKTTVALHAVANAQAAGGVAAFIDAEHALDPDYA----KKL-GVDTDS---------
                *     .**..     . .. *   *   .....*.     .* **. ..

RecS:          GLAGAQPILLGAGWEGGWRCGEVVLSMVADNARAAWSARVRETLFVDGVRWLGPSDRPTA
RecA (Mtb)     -LLVSQP-----------DTGEQALEIADMLIRSG----ALDIVVIDSVAALVPRAE---
                *  .**              **  * .     *..      . . .*.*  * *

RecS:          GTWCVWLAADTSAGAPEPRHDEVVSAAVRLHEATGKLERPRFLTQGPAAPWAEIDVFVAA
RecA (Mtb)     ------LEGEMG-DSHVGLQARLMSQALR--KMTGALNNSGTTAIFINQLRDKIGVMFGS
                     * .. . ..    .  ..* *.*    ** *. .   .       .  *.*  ..

RecS:          DRAGWEERPLQ---SVPPGVPTAPPAADPQRSIDLINQLAGLRKPTKSP-NQLVHGDLYG
RecA (Mtb)     PETTTGGKALKFYASVRMDVRRVETLKDGTNAVGNRTRVKVVKNKCLAPFKQAEFDILYG
                 .    ..*.   **  .*       *   ...   ..  ...   .* .*   . ***

RecS:          TVLLAGTALQGSPTSRLLAARILGGRGGRRRRAVLGCGRRTAHRAVERAAGVAPRMLLRA
RecA (Mtb)     KGISREGSLIDMGVDQGLI-RKSGAWFTYEGEQ-LGQGKENARNFLVENADVADEIE-KK
                 .    .* .    . *  *  *..        ** *.  *.  .   *.**  .  .

RecS:          LMFRLAVYALHPRSTAEAFPGLAHTAALVRLVL
RecA (Mtb)     IKEKLGIGAVVTDDPSND--GVLPAP----VDF
               .  .*.. *.     ..   *.  ..    . .
```

Fig. 4.14b Amino acid sequence alignment between RecS and the *M. tuberculosis* RecA without its "intron" region.

4.4.7 Occurrence of the *recS* in mycobacterial strains

Genomic DNA isolated from different mycobacteria was digested with *Sma*I, gel electrophoresed (Fig.4.15), and transferred to Hybond N membranes. The restricted chromosomal DNAs were probed with the *Bgl*II/*Pst*I fragment of pSNS201. The probe hybridised to 2 bands in H37Rv and H37Ra of , 1.5 and 1.6 kb and to a fragment of 1.5 kb in BCG. The single bands observed in the other mycobacterial strains have a molecular weight ranging from 3.5 kb (*M. smegmatis)* to 2 kb (*M. kansasii)*.

Examination of the cosmids for the duplicated *recS* in *M. tuberculosis* revealed the following: the *Bgl*II/*Pst*I probe (of ORF463) hybridised to 2 *Sma*I bands of 1.5 and 1.6 kb and one *Pst*I band of 3.4 kb in 6 of the 7 cosmids (Fig. 4.16). On the same blots only one band (1.6 kb *Sma*I and 0.9 kb *Pst*I fragment) was detected in each of the cosmids after hybridisation with the total insert of pSBG102 which contains the 5' end of *recS* (Fig. 4.17). These results implied that 2 copies of the *rec S* gene are present on the 30 kb cosmid, but that they may have different 5'-ends. Furthermore the restricted *Pst*I (3.4 kb) and *Bgl*II (4.8 kb) DNA fragments are also present in two copies. This observation suggests a duplication of approximately 9 kb within the genome, each duplication containing only one gene with conserved *Hinc*II, *Sma*I and *Pst*I intra-genic restriction sites. The observed differences in size concerning the *Sma*I fragment is due to the location of one of the sites down stream the gene which probably takes place outside the region of genetic instability described previously (see previous sections).

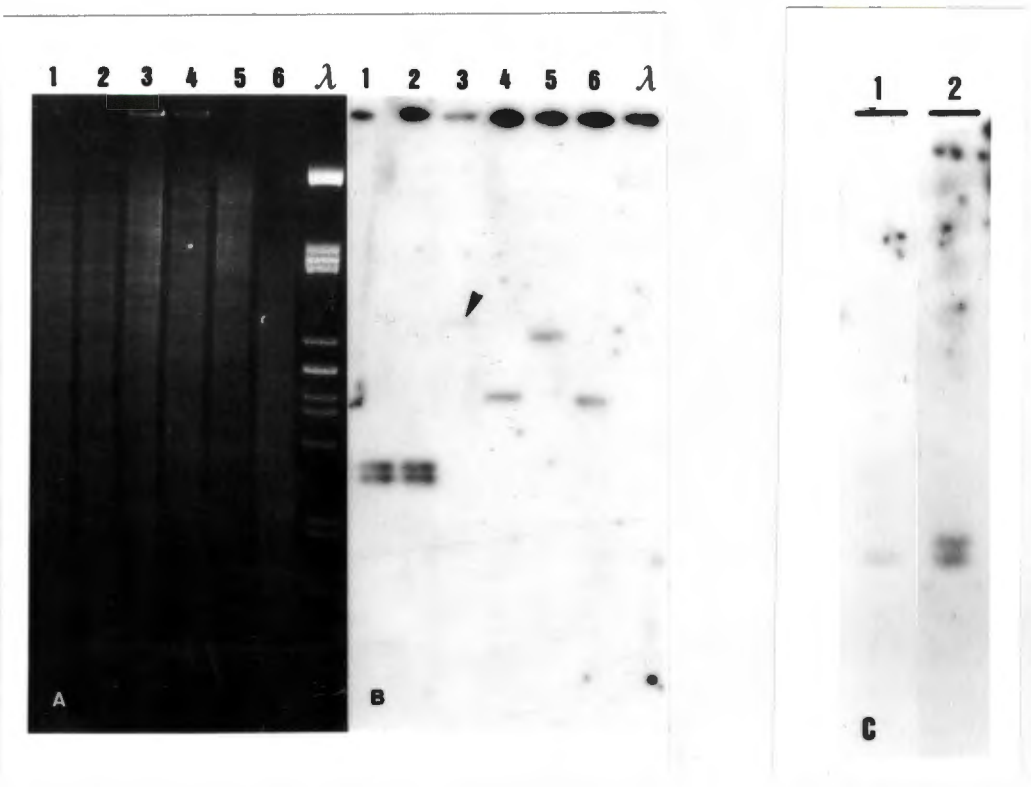


Fig. 4.15 A, Chromosomal DNA from different species of Mycobacteria, restricted with *Sma*I; B and C hybridisation with the *Bgl*II/*Pst*I fragment. A and B: lane 1, *M. tuberculosis* H37Ra; lane 2, *M. tuberculosis* H37Rv; lane 3, *M. smegmatis;* lane 4, *M. fortuitum*; lane 5, *M.aurum*; lane 6, *M. kansassi*. C. lane 1, *M. bovis* BCG; lane 2, *M. tuberculosis* H37Rv. Molecular weight marker: λ*Pst*I. The arrow in lane 3 indicates hybridisation to a band of 3.5 kb in *M. smegmatis*.

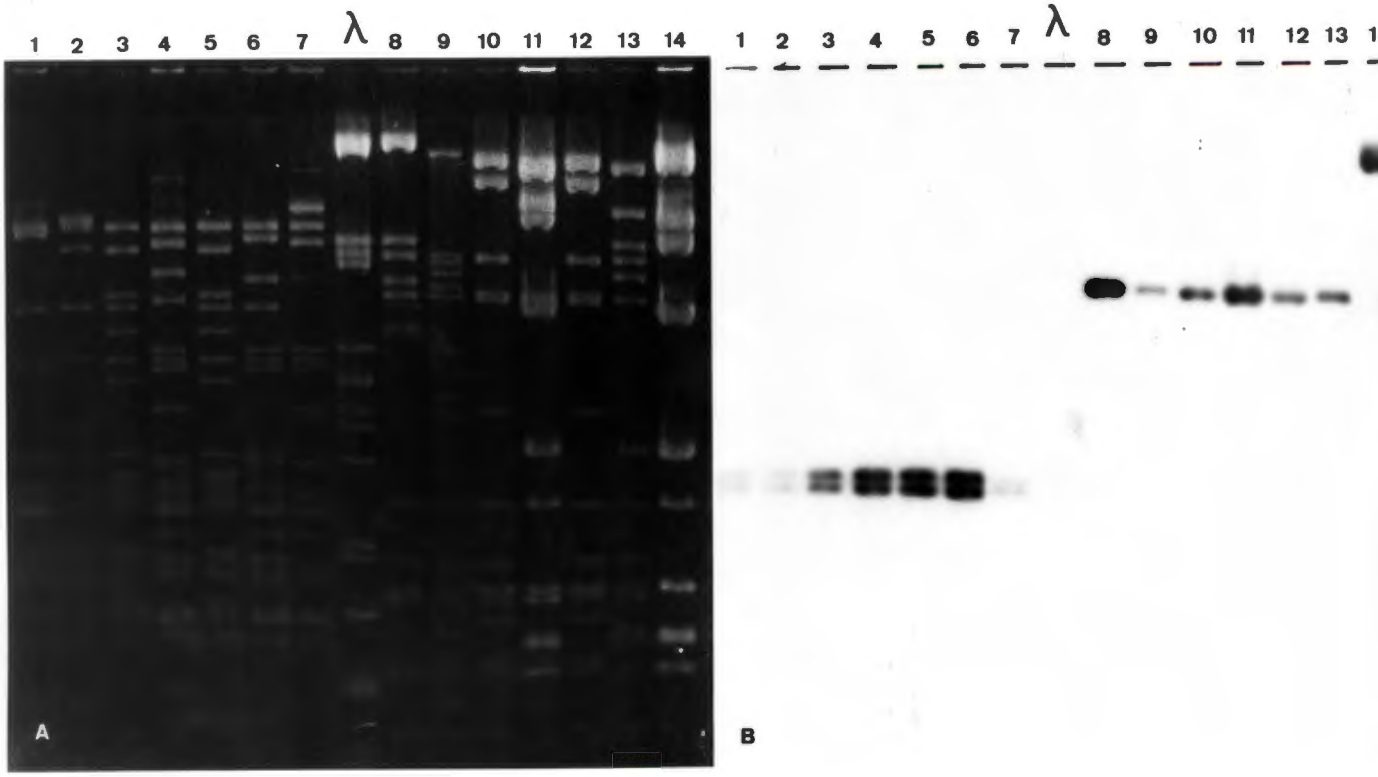


Fig. 4.16 A, gel electrophoresis of different cosmid DNA of *M. tuberculosis* restricted with *Sma*I (lanes 1 to 7) and *Pst*1 (lanes 8 to 14). B, hybridisation with the *Bgl*II/*Pst*I internal fragment of *recS;* lane 1, T733; lane 2, T565; lane 3, T481; lane 4, T180; lane 5, T252; lane 6, T256; lane 7, T276. From lanes 8 to 14, the cosmid DNA's were loaded in the same order. Molecular weight marker λ*Pst*I.

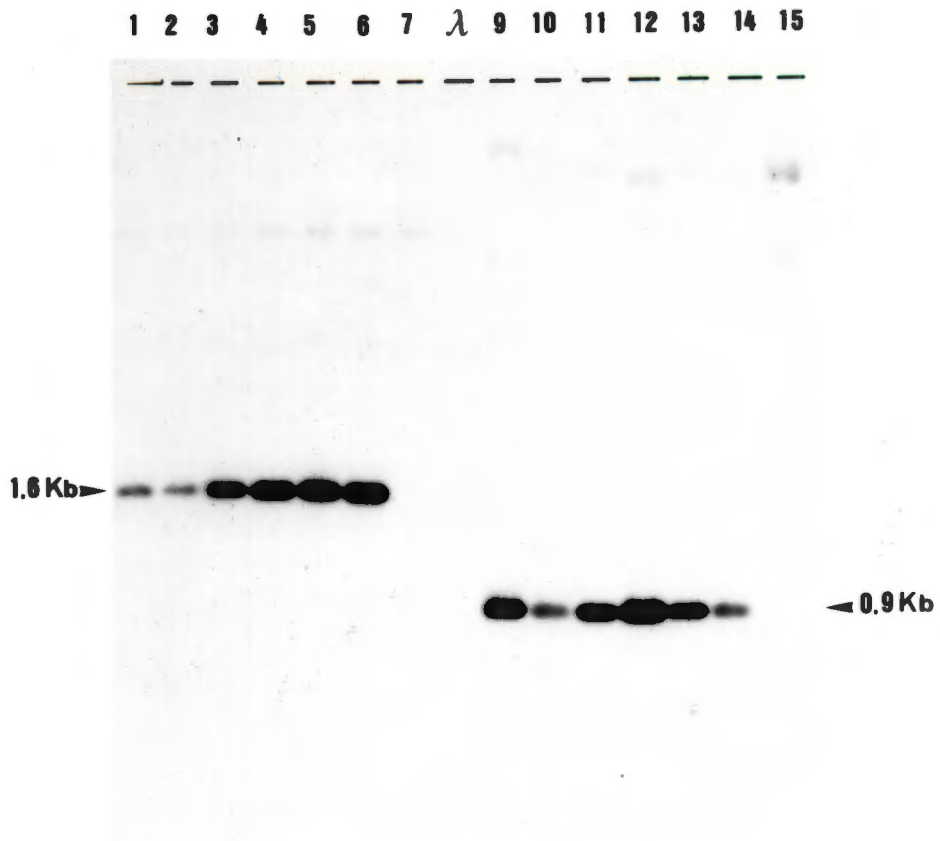


Fig. 4.17 Southern blot of Fig. 4.16, probed with the insert of pSBG102 containing the 5' end of the *recS*. Lanes 1-7, cosmids digested with *Sma*I; Lanes 8-14, cosmids digested with *Pst*I.

4.5 DISCUSSION

The 3.4 kb *Pst*I fragment (in pSNS201) was able, in part, to complement *E. coli rec*$^-$ mutants when being expressed in these cells. However, it was not known whether expression of the RecA homologue was off the *lacZ* promoter or off its own promoter. Sequence analysis of pSBG300 revealed an ORF that was truncated and lacking in its 5' end. The expression of the *recA* homologue in pSNS201 (observed in relation to EtMes, UV resistance and recombination) was in frame with and under the control of the *lacZ* promoter. This also explains why pSNS202, which has the *Pst*I fragment in the opposite orientation to pSNS201 was not expressed.

The partial degradation observed on expression of the protein in the Western blot analysis seems to show that the protein was not stable; probably due to the lack of the amino terminal region of the gene or alternatively being a foreign protein would account for the instability. It is surprising that with such a low % of identity at the amino acid level, there is a high cross reactivity with the *E. coli* RecA polyclonal antibody. A possible explanation could be that the antibody recognises conserved motifs.

Examination of the regulatory region of the *M. tuberculosis recA* gene (Davis *et al.*, 1991), does not show the presence of an SOS box for the binding of the LexA repressor protein. However, the absence of an *E. coli*-like SOS box in the 5' upstream region is not unusual, as the *recA* genes isolated from *Anabaenia variabilis* and *Thiobacillus ferrrooxidans* do not have a SOS box. (Owttrim and Coleman, 1989; Ramesar *et al.*, 1989). The *Anabaena variabilis recA* is inducible in contrast to the *Thiobacillus ferrooxidans* gene. Further investigations are required to elucidate the regulation of *recS*.

Two LexA binding sites have been identified for the *Legionella pneumophila recA* gene which is induced by UV (Xun Zhao and Dreyfuss, 1990). Nucleotide sequence analysis of the *L. pneumophila recA* gene has revealed 2 potential translation methionine start sites encoding proteins of 35 and 38 kDa, respectively. Each of the start sites has a potential -10 and -35 region and LexA binding sites. Of the 5 potential initiation codons in *recS* it is unclear which serves as the translation start site. The best candidate based on the sequence analysis is at position 836 which would encode a protein of 325aa and has a potential -35 and ribosome binding site associated with it. Complementation studies and the deduced amino acid sequence indicate that RecS is not a classical RecA protein when compared to published RecAs.

To show conclusively that the protein encoded by ORF1 is involved in DNA repair/recombination and is responsible for the phenotype observed in the *rec*$^-$ mutants, with respect to UV, EtMes and homologous recombination, it will be necessary to do induced mutagenesis studies. Similar studies have already been done for the *E. coli* systems and this knowledge could the applied to the *recS*.

Studies using mutations within the *E. coli recA* promoter region have proposed at least 2 models for the expression of the gene (see Chapter 1, Weissemann and Weinstock, 1991). Analysis of the putative promoter regions of *recS* and the cloned *M. tuberculosis recA* (Davis *et al.*, 1991) will provide valuable information regarding gene expression in *M. tuberculosis*.

The reason for the presence of two *recS* genes in *M. tuberculosis* is not known but one could speculate that; the regulation of these two genes may be different and expression may occur under different circumstances and may also require different inducing signals. The low rate of mutagenesis in *M. tuberculosis*

indicates that an efficient recombination/repair system exists, and it is possible that *recS* is involved in the SOS response or alternatively in another pathway evolved in these mycobacteria to protect the integrity of the genome. In this respect it is important to note that *M. tuberculosis* contains a significant number of repeated DNA sequences and this could lead to instability of the genome, it is possible that a second type of RecA protein is required to ensure that the repeated sequences do not lead to genome instability. It would be interesting to find a more precise role for this protein in the SOS response and recombination and possibly other functions.

In this respect the presence of two genes which differ in their amino terminal regions may lead to recombination of the conserved 3' end with a variety of 5' ends. This type of recombination of "expression cassettes" in *N. gonnorhoea* leads to phase variation (Meyer and Putten, 1989).

CHAPTER 5

GENERAL INTRODUCTION

5.1 SCAVENGING ENZYMES : OXYGEN RADICAL TOXICITY AND DNA REPAIR IN *E. COLI* 112

5.1.1 H_2O_2 induced DNA damage
5.1.1a Mode 1 killing 113
5.1.1b Mode 2 killing 114
5.1.2 DNA repair of H_2O_2 induced DNA damage 114
5.1.3 Catalase / Peroxidase 115

5.2 SCAVENGING ENZYMES IN THE MYCOBACTERIA 117

5.2.1 Isoniazid resistance 119
5.2.2 Role of Catalase/peroxidase in INH action 120

CHAPTER 5

GENERAL INTRODUCTION

5.1 SCAVENGING ENZYMES: OXYGEN RADICAL TOXICITY AND DNA REPAIR IN *E. COLI*

E. coli and *S. typhimurium* have been used to study the mechanisms of oxidative stress and cellular resistance. Oxidative stress is mediated by active oxygen species such as the superoxide radical (O_2^-), hydroxyl radical (OH^-) and hydrogen peroxide (H_2O_2). These oxygen species arise as by-products of aerobic metabolism, or from the exposure to compounds that can divert electrons from electron transport components to O_2 (redox-cycling compounds) (Greenberg and Demple, 1988). To prevent oxidative damage to proteins, membranes and DNA, cells utilise several protective enzymes, including the enzymatic scavenging activities of catalase, superoxide dismutase (SOD) and peroxidase (Walkup and Kogoma, 1989). Oxygen radicals can attack DNA at either the sugar or the ribose bases purine or pyrimidine resulting in a number of DNA damaging products (Imlay and Linn, 1988). Exposure of bacteria to H_2O_2, or O_2^- results in a base loss or strand breakage with a fragmented sugar residue. These single stranded DNA breaks accumulate and the *E. coli* mutants deficient in recombinational repair, are unable to repair the DNA once exposed to toxic oxygen molecules such as H_2O_2 (Halliwel and Gutteridge, 1984).

5.1.1 H_2O_2 induced DNA damage

5.1.1a Mode 1 killing

Exponentially growing *E. coli* cells exposed to H_2O_2 are known to undergo 2 kinetically distinguishable modes of killing (Imlay and Linn, 1986). In mode one, killing occurs only at low concentrations of H_2O_2 and mutagenèsis and induction of the SOS responses are largely involved during challenges at these doses. Thus *E. coli* mutants, defective in recombinational repair are particularly sensitive to mode 1 killing.

An H_2O_2 adaptive response (different to that described in 1.3) is seen in bacteria exposed to sub-lethal levels of H_2O_2 enabling them to survive subsequent lethal doses of oxidising agents (Demple and Halbrook, 1983). During this response there is an increased synthesis of approximately 30 proteins. The bacterial *oxyR* gene has been identified as a regulator of acquired resistance to oxidative stress. The *oxyR* gene is required for the expression of many stress proteins that are induced in response to sublethal levels of hydrogen peroxide (Christman *et al.*, 1985). Eight of these proteins including the catalase enzymes (*kat*) and alkyl hydroperoxide reductase (*Ahp*) are under the positive control of the *oxyR* gene (Christman *et al.*, 1985; Morgan *et al.*, 1986). *oxyR* deletion mutants are sensitive to H_2O_2 and alkyl peroxides (Christman *et al.*, 1985), to redox cycling agents. (Greenberg and Demple, 1988) and show a high spontaneous mutation rate (Storz *et al.*, 1987).

Also part of the Mode 1 killing is the response to $O_2^{-}\cdot$ The induced response to O_2^{-} generating conditions differs from that of the response to H_2O_2. H_2O_2 induces the SOS response (Imlay and Linn, 1987) whereas O_2^{-} generating

conditions do not (Farr *et al.*, 1985). Walkup and Kogoma, (1989), have proposed that *E. coli* possesses an inducible oxidative stress response involving O_2^- that is distinct from the response involving H_2O_2 following the exposure to redox cycling agents. This response is termed the superoxide inducible (Sox) response. Greenberg and Demple, (1989) have provided evidence to show that there is an overlap between the response elicited by H_2O_2 and the oxidative stress activated by superoxide generating agents. However, both agents induce many of the same *oxyR*-independent proteins and in addition, redox cycling drugs induce a global response resulting in changes in the production of many proteins which overlap with that induced by peroxide stress.

5.1.1b Mode 2 killing

Mode 2 killing, on the other hand, requires a relatively higher dose of H_2O_2 (at least 2 or 3 x's higher) as compared to that of Mode 1 killing and the exact sites of toxicity are not clear. The DNA damage is generated at a rate that is linearly proportional to the concentration of H_2O_2 (Imlay and Linn, 1988).

5.1.2 H_2O_2 induced DNA repair

It has been shown in *E. coli* that hydrogen peroxide causes strand cleavage of DNA directly by the collapse of the deoxyribose ring following the removal of a hydrogen atom, and produces a repair response involving DNA polymerase and DNA ligase. The relative importance of catalase and the RecA system in the protection of *E. coli* from H_2O_2 is not clear and observations imply that there is no direct role for the RecA protein in the repair of hydrogen peroxide-induced DNA damage (Hagensee *et al.*, 1987).

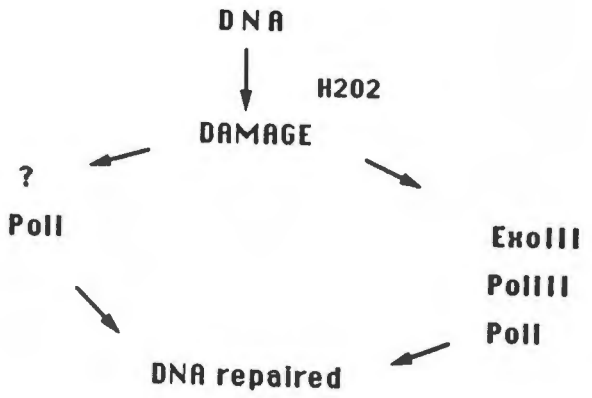


Fig. 5.1 Mode of repair of hydrogen peroxide induced. DNA damage (Hagensee and Moses, 1989).

It appears that multiple pathways exist for the repair of hydrogen peroxide induced DNA damage (Fig. 5.1). The one pathway utilises DNA PolI either alone or with other components. It is also possible that RecA may interact with these enzymes or other cellular enzymes and that different types of hydrogen peroxide-induced DNA damage involve different enzymes (Hagensee and Moses, 1989).

5.1.3 Catalase/peroxidase

The catalase and peroxidase enzymes are haem proteins responsible for the decomposition of hydrogen peroxide in cells by the following reactions:

Catalase

$$2\ H_2O_2 \Rightarrow 2\ H_2O + O_2$$

Peroxidase

$$2\ RH + H_2O_2 \Rightarrow 2\ H_2O + 2\ R$$

RH = reducing substrate

Most purified catalases have been shown to consist of 4 protein subunits, each of which contains a haem (Fe(III)-protoporphyrin) group bound to its active site.

Two catalases have been characterised in *E. coli*, HPI (Clairborne and Fridovich, 1979) and HPII (Clairborne *et al.*, 1979; Triggs-Raine and Loewen, 1987). HPI is encoded by the *katG* gene (which has been sequenced) (Triggs-Raine *et al.*, 1988), and occurs as a tetramer of 84 kDa monomers with 2 associated proto-heme groups, and is hydrogen peroxide inducible. HPI has dual catalase and ortho-dianisidine peroxidase activities (Clairborne and Fridovich, 1979). HPII which is encoded by non contiguous loci, the *katE* and *katF* genes (Loewen, 1984; Loewen and Triggs-Raine, 1984); is induced by aeration, and has no peroxidase activity (Clairborne and Fridovich, 1985). The relationship between the *katE* and *katF* genes is still unclear. It is possible that one locus has a regulatory function while the other is a structural gene. *katF* plays a role in bacterial adaptation to starvation and its role as a regulator of *katE* seems to be accepted. It is possible that the 2 genes code for different subunits of the KatE tetrameric enzyme (Schellhorn and Hassan, 1988).

There has been no report of DNA repair functions assigned to these genes or or to the different domains of the catalase enzyme.

5.2. SCAVENGING ENZYMES IN THE MYCOBACTERIA

Tubercle and leprosy bacilli infect their hosts and divide inside host macrophages. Phagosome-lysosome fusion which includes the activation of hydrolytic enzymes, the lowering of intraphagosomal pH, as well as the production of peroxide by the macrophages, is important in the killing of infecting micro-organisms and probably *M. tuberculosis* and *M. leprae* (Sharp and Banerjee, 1985). Microorganisms which infect host cells are known to elaborate the enzymes catalase and superoxide desmutase to destroy harmful peroxide and other toxic oxygen metabolites produced by the host. The functional interaction between mycobacterial catalases with the immune system have not been investigated. It is possible that toxic O_2 metabolites produced by the host are not sufficient to kill mycobacteria (Wayne and Diaz, 1982; Lowrie and Andrew, 1988).

Extensive evidence supports the view that hydrogen peroxide and the associated relative forms of reduced oxygen produced by the macrophages/phagocytes are responsible for its bactericidal powers. However, there is no evidence that "peroxidation of microbial substance" is the cause of death of the mycobacteria (Lowrie and Andrew, 1988). It has been suggested that the peroxidase susceptibility of micro-organisms is independent of catalase content and may be caused by a deficiency in the mechanisms which repair damage to DNA (Carlsson and Carpenter, 1980). Peroxide toxicity for tubercle bacilli can be enhanced by peroxidase (present in monocytes but not in macrophages) or by catalase but the involvement of these mechanisms in the destruction of the bacilli by the hosts defense mechanisms is not known (Lowrie, 1983).

Two types of catalases may be found in mycobacterial lysates and can be distinguished from one another by their sensitivities to thermal inactivation. The heat-labile T-class catalases have been detected in *M. tuberculosis, M. bovis, M.*

avium and M. intracellulare (Wayne and Diaz, 1982). The T catalases have substrate specificities similar to the *E. coli* HPI bifunctional peroxidase-catalase (Morris *et al.*, 1992). However, the heat-stable M-class catalases have not been detected in the above mycobacterial species. The mycobacterial M catalase is a monofunctional HPII-like catalase (Morris *et al.*, 1992). Both the M and the T catalases have been detected in extracts of *M. kansasii and M. scrofulaceum* (Wayne and Diaz, 1982)). Purified catalase enzyme from *M. tuberculosis* has been shown to be a tetramer of a molecular weight of 160,000 (Diaz and Wayne, 1974) and to possess both catalase as well as peroxidase activities (Gayathri Devi *et al.*, 1975).

The gene (*M185*) encoding the *M. intracellulare* T catalase-peroxidase has recently been shown to encode a tetramer of 335 kDa. As mentioned earlier (section 5.1.1), the expression of proteins in response to oxidative stress is mediated by the *oxyR* regulon. Analysis of sequences upstream from the putative *M185* initiation codon showed similarity to the putative binding domain of the *oxyR* trans-activating protein in the *katG* gene of *E. coli* (Morris *et al.*,1992). The *katG* gene from *M. tuberculosis* has been isolated by Zhang *et al.*, (1992); analysis of the catalase activity of the protein indicated that it resembled *katG* of *E. coli* in having an associated peroxidase activity.

5.2.1 Isoniazid resistance

Isonicotinic acid hydrazide (isoniazid, INH) is a key agent for the treatment of tuberculosis and is the most active of a number of carboxylic acid hydrazides. INH has a remarkably high activity against *M. tuberculosis* and *M. bovis*. The bactericidal action of INH is accompanied by loss of acid fastness and is dependent on the active growth of the mycobacteria. The initial effect is a slight stimulation of respiration which then changes to inhibition with the bactericidal effect. There appears to be no evidence that INH inhibits nucleic acid or protein synthesis (Winder, 1982).

The pathway primarily affected by INH was believed to be the mycolic acid synthesis (Winder and Collins, 1970; Quemard *et al.*, 1991) However, recent evidence by Heym and Cole, 1992), indicates that INH resistance may not be due to an alteration in mycolic acid production. The effects of INH on the cell wall of mycobacteria have been known to include the accumulation of certain classes of carbohydrates, damage to the bacterial envelope (Winder, 1964b) and a decrease in the synthesis of the acyl components of the "bound lipid fraction" (Winder and Rooney, 1968).

Mutants with a high resistance to INH occur with a frequency of 10^{-8} in *M. tuberculosis* and *M. bovis* BCG and these mutants lack peroxidase and catalase activity. Evidence suggests that strains of *M. tuberculosis* and *M. bovis* resistant to INH lack catalase and peroxidase activities (Middlebrook, 1954) and is probably due to a single mutation (Hedgecock and Fancher, 1957; Dunbar *et al.*, 1959).

5.2.2 Role of catalase-peroxidase in INH action

Following the original findings by Middlebrook *et al.* (1954) there were many reports that resistance to high concentrations of INH in *M. tuberculosis* and *M. bovis* involved loss of catalase and peroxidase activity (Cohn *et al.*, 1954; Hedgecock *et al.*, 1957; Dunbar *et al.*, 1959; Heym *et al.*, 1992). Evidence by Heym *et al.* (1992) suggested that the loss of peroxidase and catalase activities may be due to the absence of the *katG* product. Wayne *et al.* (1968) showed that *M. gastri* and some strains of *M. kansasii* resembled *M. tuberculosis* and *M. bovis* in that the catalase activity was lost on the development of resistance to INH.

Two possible roles have been proposed for the catalase-peroxidase in INH activity (Winder, 1964a). Firstly, catalase-peroxidase may have a carrier role in the transport of INH into the mycobacteria (Youatt, 1969), however, there is no clear evidence for this. Approximately one-fifth of the catalase in *M. bovis* is located on the surface of the bacterial cell wall. Secondly, it is possible that catalase-peroxidase is involved in the chemical modification of INH by converting it to an active form. Investigations by Heym *et al.* (1992) and Gayathri *et al.* (1975), have shown that not only does the catalase enzyme from *M. tuberculosis* possess both catalase and peroxidase activities but that it also catalyses peroxidation of INH. This reaction is known as the "Youatt" or "Y" reaction and it is possible that the radicals generated by this reaction contribute to the antimycobacterial activity of INH (Youatt, 1969)

CHAPTER 6

ISOLATION, CHARACTERISATION AND SEQUENCING OF A SECOND *M. TUBERCULOSIS* RECOMBINANT PLASMID ABLE TO COMPLEMENT *E. COLI* DK *rec*⁻ CELLS

6.1 SUMMARY 122

6.2 INTRODUCTION 122

6.3 MATERIAL AND METHODS

6.3.1 Bacterial strains 123
6.3.2 Isolation of EtMes resistant clones 123
6.3.3 Resistance to UV radiation 124
6.3 4 Restriction map of pSNS100 and nucleotide sequencing of pSNS100 124
6.3.5 DNA Hybridisation 124

6.4 RESULTS

6.4.1 EtMes resistance 125
6.4.2 UV radiation study 126
6.4.3 Restriction map and sequence analysis of pSNS100 127
6.4.4 Presence of *katG* in different strains of mycobacteria 132

6.5 DISCUSSION 137

CHAPTER 6

IDENTIFICATION, CHARACTERISATION AND SEQUENCING OFA SECOND *M. TUBERCULOSIS* RECOMBINANT PLASMID ABLE TO COMPLEMENT *E. COLI* DKI *rec*⁻ CELLS

6.1 SUMMARY

The *Pst*I genomic library of *M. tuberculosis* H37Rv described in Chapter 2 was screened further for recombinant plasmids that were able to confer EtMes resistance on their host cells. A second clone pSNS100, containing a 1.6 kb insert, was isolated. The DNA insert contained an ORF (ORF242, 726 bp) which encoded a conserved domain of a *katG* gene and 2 smaller ORFs (ORF168 - 504 bp and ORF63 - 187 bp) which did not encode any known proteins. pSNS100 conferred resistance to EtMes and short wave length, near UV radiation on *E.coli* DK cells. The truncated *katG* gene (ORF242) hybridises to genomic DNA from INH resistant *M. tuberculosis* isolates and the INH sensitive strain of *M. tuberculosis*, H37Rv.

6.2 INTRODUCTION

Treatment of DNA with EtMes results in the depurination of DNA. Depurination has often been suggested to be the source of mutations (Bautz, and Freese, 1960). However, it is not clear how alkylated bases can engage in mutagenic mispairing, and it is known that depurinations which cannot be repaired, become lethal.

The alkylating agents EtMes and EtES and MMS are classical mutagens used for the isolation of clones that complement defects in *recA* mutants. Thus by selecting for EtMes, EtES or MMS resistant clones, genes involved in general recombination and repair can be isolated. These agents have been used to isolate *recA*-like genes from different bacteria (see Chapter 1). UV induced DNA damage is also known to activate repair mechanisms.

This chapter deals with the functional characterisation of the recombinant plasmid pSNS100. Nucleotide sequencing of pSNS100 revealed 3 potential ORFs. Of these ORF242 revealed an in frame, truncated *katG* gene.

6.3 MATERIAL AND METHODS

6.3.1 Bacterial strains

Bacterial strains and plasmids used in this study are listed in Appendix 1A,B and 2A,B.

The *Pst*I genomic library of H37Rv described in Section A, Chapter 2 was used in this study. *M. avium, M. intracellulare, M kansasii, M. Bovis* BCG (WHO) as well as two hospital strains of *M. tuberculosis* (INH resistant and INH sensitive) were included. These latter two strains were isolated by the Dept of Medical Microbiology at the Groote Schuur Hospital, Cape Town.

6.3.2 Isolation of *E. coli* DK1 EtMes resistant clones

In addition to pSNS201, which is described in Section A of this thesis, a second recombinant plasmid, was isolated from the genomic library, which on transformation of *E. coli* DK1 cells allowed consistent growth of the transformant on 0.1% EtMes. This plasmid contained a *Pst*1 insert of 1.6kb. The plasmid pSNS100 is a pUC19 derivative containing the 1.6 kb insert.

6.3.3 Resistance to UV radiation

E. coli DK pSNS100 cells were examined for their ability to recover from exposure to UV irradiation (254 nm). Bacterial cell cultures were treated as described in Chapter 2.

6.3.4 Restriction map nucleotide sequencing of pSNS100

The restriction enzyme sites in pSNS100 were mapped by standard methods.

DNA templates for sequencing were obtained by exoIII (BRL) digestion of pSNS100. Universal primers and custom synthesised oligonucleotides were used. DNA sequencing was done with Sequenase 2.0 (USB) and Taq Track (Promega).

6.3.5 DNA Hybridisation studies

Total genomic DNA isolated from the different mycobacteria was digested with *Kpn*1. The fragments were separated by gel electrophoresis and transferred to a nylon membrane (Hybond N, Amersham) (Southern, 1975). The *Alu-Pst*1 fragment containing the truncated *katG* gene was radiolabelled with ^{32}Pα-dATP (Amersham) by nick-translation (Boehringer Mannheim) and DNA hybridisations and post-hybridisation washes were done as described in Chapter 2.

6.4 RESULTS

6.4.1 EtMes resistance

The percentage of *E. coli* DK (pSNS100) survivors on EtMes were calculated as a fraction of the total number of cells plated as controls, on plates lacking EtMes. The recombinant plasmid pSNS100 conferred resistance on DK cells with a growth rate of 30% on 0,1% EtMes. The percentage of EtMes resistant cells was determined 2 days post-inoculation and were obtained from 4 independent experiments. Fig. 6.1 shows that *E.coli* MC1060, the *recA*$^+$ parent of *E.coli* DK, was less resistant to 0,1% EtMes (only 6% survivors) as compared to *E. coli* DK (pSNS100), 25 %.

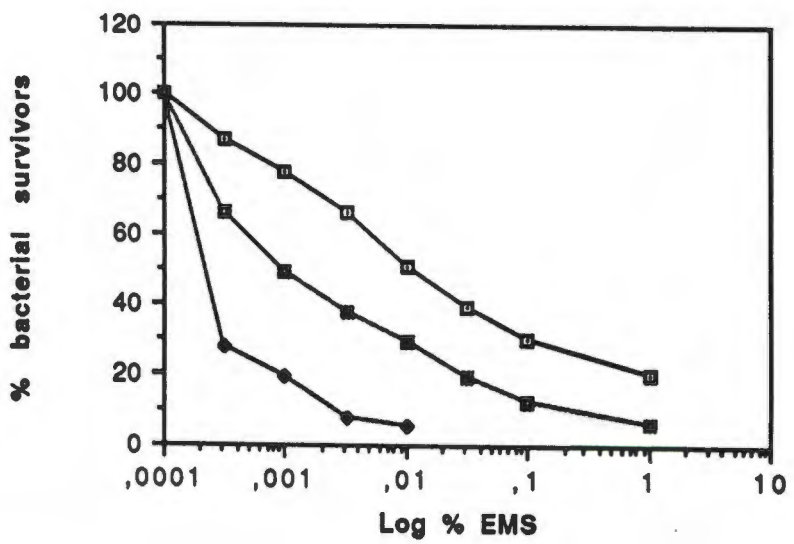


Fig.6.1 The effect of the recombinant plasmid pSNS100 on the survival of *E.coli* DK cells which were plated on various concentrations of EtMes. Symbols; —♦— *E. coli* DK cells, —□— *E. coli* DK (pSNS100) and —■— *E. coli* MC1060.

6.4.2 UV radiation study

The UV survival curve in Fig. 6.2 is expressed as a fraction of the non-irradiated control. A survival rate of >80% was observed for *E. coli* DK (pSNS100).

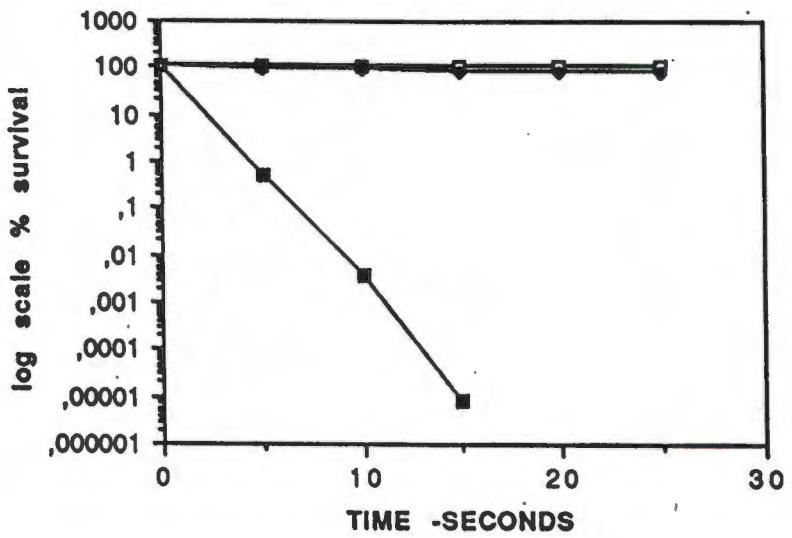


Fig.6.2 The effect of the recombinant plasmid pSNS100 on the survival of *E.coli* DK cells which were exposed to a UV radiation of 254nm. Symbols; —♦— *E. coli* DK (pSNS100), —□—*E. coli* MC1060 and —■— *E. coli* DK cells.

6.4.3 Restriction map and Sequence analysis

Restriction endonuclease map of pSNS100 is given in Fig. 6.3. A unique *Alu*I site is located at position 783 from the left *Pst*I site. The sequencing strategy is also indicated in Fig. 6.4.

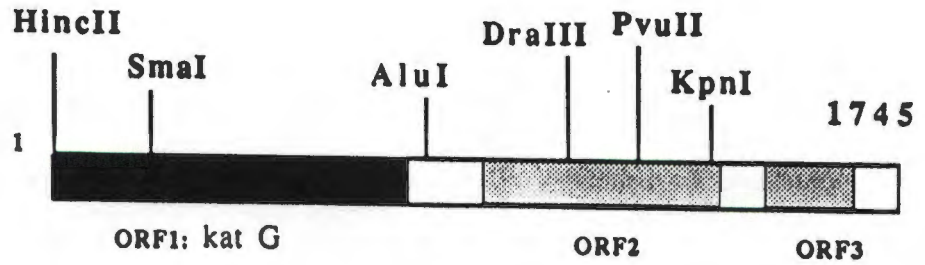


Fig. 6.3 Restriction map of the cloned *Pst*1 fragment with unique sites indicated.

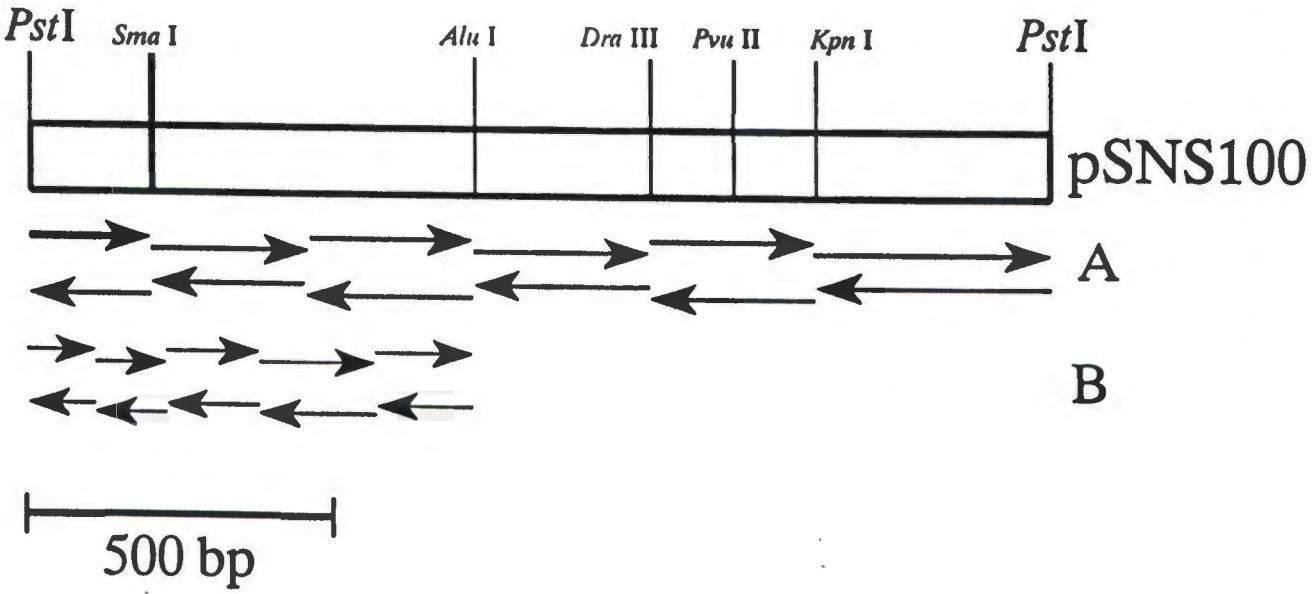


Fig. 6.4 Sequencing strategy use to sequence pSNS100. The top arrows indicate sequencing done using overlapping clones and synthesised oligonucleotides. The bottom arrows indicate sequencing using only synthesised oligonucleotides to confirm the *katG* sequence.

The entire sequence of pSNS100 is given in Fig. 6.5. Analysis of the sequence of pSNS100, with DNA Strider, for ORFs in all phases is given in Fig.6.6. Three ORFS were identified: ORF242 (726 bp), ORF168 (504 bp) and ORF63 (187 bp). A comparison of these ORFs with the Genebank Database was made using Fasta.

```
        |   10      |   20      |   30      |   40      |   50      |   60
   1 CTGCAGCCAC AAGTCGGGTG GGAGGTCAAC GACCCGACGG GATCTGCGCA AGGTCATTCG 60
  61 CACCCTGGAA GAGATCCAGG AGTCATTCAA CTCCGCGGCG CCGGGGAACA TCAAAGTGTC 120
 121 CTTCGCCGAC CTCGTCGTGC TCGGTGGCTG TGCCGCCATA GAGAAAGCAG CAAAGGCGGC 180
 181 TGGCCACAAC ATCACGGTGC CCTTCACCCC GGGCCGCACG GATGCGTCGC AGGAACAAAC 240
 241 CGACGTGGAA TCCTTTGCCG TGCTGGAGCC CAAGGCAGAT GGCTTCcGAA ACTACCTCGG 300
 301 AAAGGGCAAC CCGTTGCCGG CCGAGTACAT GCTGCTCGAC AAGGCGAACC TGCTTACGCT 360
 361 CAGTGCCCCT GAGATGACGG TGCTGGTAGG TGGCCTGCGC GTCCTCGGCG CAAACTACAA 420
 421 GCGCTTACCG CTGGGCGTGT TCACCGAGGC CTCCGAGTCA CTGACCAACG ACTTCTTCGT 480
 481 GAACCTGCTC GACATGGGTA TCACCTGGGA GCCCTCGCCA GCAGATGACG GGACCTACCA 540
 541 GGGCAAGGAT GGCAGTGGCA AGGTGAAGTG GACCGGCAGC CGCGTGGACC TGGTCTTCGG 600
 601 GTCCAACTCG GAGTTGCGGG CGCTTGTCGA GGTCTATGGC GCCGATGACG CGCAGCCGAA 660
 661 GTTCGTGCAG GACTTCGTCG CTGCCTGGGA CAAGGTGATG AACCTCGACA GGTTCGACGT 720
 721 GCGCTGATTC GGGTTGATCG GCCCTGCCCG CCGATCAACA CAACCCGCCG CAGCACCCCG 780
 781 CGAGCTGACC GGCTCCGGGG CTGTGTTTGC CCGGCGCGAT TTGTCAGACC CCGCGTGCAT 840
 841 GGTGGTCGCA GGCACGACGA GACGGGGATG ACGAGACGGG GATGAGGAGA AAGGGCGCCG 900
 901 AAATGTGCTG GATGTGCGAT CACCCGGAAG CCACCGCCGA GGAGTACCTC GACGAGGTGT 960
 961 ACGGGATAAT GCTCATGCAT GGCTGGGCGG TACAGCACGT GGAGTGCGAG CGACGGCCAT 1020
1021 TTGCCTACAC GGTTGGTCTA ACCCGGCGCG GCTTGCCCGA ACTGGTGGTG ACTGGCCTCT 1080
1081 CGCCACGACG TGGGCAGCGG TTGTTGAACA TCGCCGCTCG CAGGGCTCTG GTCGGTGACT 1140
1141 TGCTGAACTC CCGGTATGCA GACCACCCTC CCAGCCGGCC CTCTTGTCGA AACGGTCCAG 1200
1201 GTTACACATC CGGACGCGCA TTTGTATTGT GCGATCGCCA TCTTTGGCGA CAAGGTGACG 1260
1261 GCCTTGCAGT TGGTGTGGGC CGACCGGTGG TCGCTGGCCG TGGGGCGGCG GACTTCGACG 1320
1321 AAGGTCGGTA CCCAGCCGGT GCTCGGGATG CGAGCCACCA GGAGGTCAGC CTGACGCGAG 1380
1381 CGCTCTAAAC GTGTTGCCGG AGAACGATTT GTGACAGCAA TGCAGTGACA GCGGTGGTTG 1440
1441 CTCGACACGG AGGTGGCCAA CGCTTCCGGA CATCCTCCGA TACATGCGGT CTGGGCGGTG 1500
1501 TGCTGGCGGA CAAGGCTGTG ACTCGGTGGA AGCGTTGGCG CCGTCCGCGA TTTACCGCAT 1560
1561 CATTGCCGGA CACTGCGTCG GTGGCGCGCC GTCCGTGCGC TCGCAGGCGC ATATGGGGAT 1620
1621 CGAACAAGCC CATCGGCGGC GGGAGAACGA GTGATTGCTT CCAGGCCGAG TGAGTTCCGG 1680
1681 ATACCGGTAG GCCCCAGTAG GCTATTGTGT GATGCGCTTG AAGCCAGCCC CATCTCCTGC 1740
1741 TGCAG                                                            1745
        |   10      |   20      |   30      |   40      |   50      |   60
```

Fig. 6.5 The entire sequence of pSNS100, (ORF242, nucleotide 1-726).

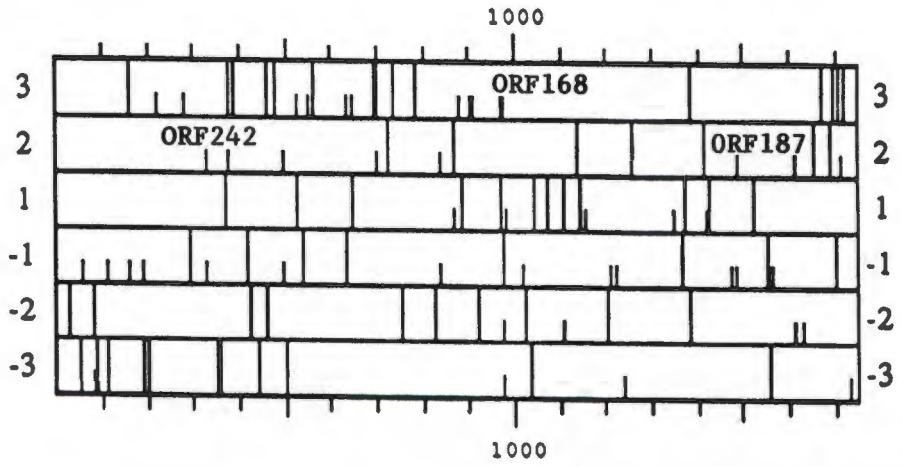


Fig. 6.6 Analysis of the sequence of pSNS100 for open reading frames by the DNA Strider programme.

Nucleotide sequence and deduced amino acid sequence of the *Pst*I insert of pSNS100 is shown in Fig. 6.7. ORF242 encodes a protein with 60% identity to the terminal 241 amino acids of the *E. coli katG* gene. The stop codon (nucleotide 726 bp) is followed by a stem-loop structure. Similarities between the COOH-terminal of cloned bacterial *katG* proteins is given in Fig. 6.8. The *M. tuberculosis katG* in this alignment shows an amino acid identity of 55% with that of *B. stearothermophilus* (Loprasert *et al.*, 1980) and 50% with that of *E. coli* (Triggs-Raine *et al.*, 1988) showing a high degree of conservation in the compared sequences. ORF168 and ORF63 do not encode for any proteins identifiable by sequence homology.

The *katG* gene of *E. coli* encodes a protein which has both catalase and peroxidase activities. It is not known whether separate functional domains exist which are responsible for the catalase and peroxidase functions, respectively, or whether more than one domain is responsible for these functions.

```
TGC AGC CAC AAG TCG GGT GGG AGG TCA ACG ACC CGA CGG GAT CTG CGC AAG GTC ATT CGC
 C   S   H   K   S   G   G   R   S   T   T   R   R   D   L   R   K   V   I   R

ACC CTG GAA GAG ATC CAG GAG TCA TTC AAC TCC GCG GCG CCG GGG AAC ATC AAA GTG TCC
 T   L   E   E   I   Q   E   S   F   N   S   A   A   P   G   N   I   K   V   S

TTC GCC GAC CTC GTC GTG CTC GGT GGC TGT GCC GCC ATA GAG AAA GCA GCA AAG GCG GCT
 F   A   D   L   V   V   L   G   G   C   A   A   I   E   K   A   A   K   A   A

GGC CAC AAC ATC ACG GTG CCC TTC ACC CCG GGC CGC ACG GAT GCG TCG CAG GAA CAA ACC
 G   H   N   I   T   V   P   F   T   P   G   R   T   D   A   S   Q   E   Q   T

GAC GTG GAA TCC TTT GCC GTG CTG GAG CCC AAG GCA GAT GGC TTC CGA AAC TAC CTC GGA
 D   V   E   S   F   A   V   L   E   P   K   A   D   G   F   R   N   Y   L   G

AAG GGC AAC CCG TTG CCG GCC GAG TAC ATG CTG CTC GAC AAG GCG AAC CTG CTT ACG CTC
 K   G   N   P   L   P   A   E   Y   M   L   L   D   K   A   N   L   L   T   L

AGT GCC CCT GAG ATG ACG GTG CTG GTA GGT GGC CTG CGC GTC CTC GGC GCA AAC TAC AAG
 S   A   P   E   M   T   V   L   V   G   G   L   R   V   L   G   A   N   Y   K

CGC TTA CCG CTG GGC GTG TTC ACC GAG GCC TCC GAG TCA CTG ACC AAC GAC TTC TTC GTG
 R   L   P   L   G   V   F   T   E   A   S   E   S   L   T   N   D   F   F   V

AAC CTG CTC GAC ATG GGT ATC ACC TGG GAG CCC TCG CCA GCA GAT GAC GGG ACC TAC CAG
 N   L   L   D   M   G   I   T   W   E   P   S   P   A   D   D   G   T   Y   Q

GGC AAG GAT GGC AGT GGC AAG GTG AAG TGG ACC GGC AGC CGC GTG GAC CTG GTC TTC GGG
 G   K   D   G   S   G   K   V   K   W   T   G   S   R   V   D   L   V   F   G

TCC AAC TCG GAG TTG CGG GCG CTT GTC GAG GTC TAT GGC GCC GAT GAC GCG CAG CCA AAG
 S   N   S   E   L   R   A   L   V   E   V   Y   G   A   D   D   A   Q   P   K

TTC GTG CAG GAC TTC GTC GCT GCC TGG GAC AAG GTG ATG AAC CTC GAC AGG TTC GAC GTG
 F   V   Q   D   F   V   A   A   W   D   K   V   M   N   L   D   R   F   D   V

CGC TGA TTCGGGGTTGATCGGCGCTGCCCGGCCGATCAACACAACCCGCCGCAGCACCCCGCGAGCTGACCGGCTCCG
 R  Stop

GGGCTGTGTTTGCCCGGCGCGATTTGTCAGACCCCGCGTGCATGGTGGTCGGCAGGCACGACGAGACGGGGATGACGAGA

CGGGG   ATG AGG AGA AAG GGC GCC GAA ATG TGC TGG ATG TGC GAT CAC CCG GAA GCC ACC
          M   R   R   K   G   A   E   M   C   W   M   C   D   H   P   E   A   T

GCC GAG GAG TAC CTC GAC GAG GTG TAC GGG ATA ATG CTC ATG CAT GGC TGG GCG GTA CAG
 A   E   E   Y   L   D   E   V   Y   G   I   M   L   M   H   G   W   A   V   Q

CAC GTG GAG TGC GAG CGA CGG CCA TTT GCC TAC ACG GTT GGT CTA ACC CGG CGC GGC TTG
 H   V   E   C   E   R   R   P   F   A   Y   T   V   G   L   T   R   R   G   L

CCC GAA CTG GTG GTG ACT GGC CTC TCG CCA CGA CGT GGG CAG CGG TTG TTG AAC ATC GCC
 P   E   L   V   V   T   G   L   S   P   R   R   G   Q   R   L   L   N   I   A

GCT CGC AGG GCT CTG GTC GGT GAC TTG CTG AAC TCC CGG TAT GCA GAC CAC CCT CCC AGC
 A   R   R   A   L   V   G   D   L   L   N   S   R   Y   A   D   H   P   P   S

CGG CCC TCT TGT CGA AAC GGT CCA GGT TAC ACA TCC GGA CGC GCA TTT GTA TTG TGC GAT
 R   P   S   C   R   N   G   P   G   Y   T   S   G   R   A   F   V   L   C   D

CGC CAT CTT TGG CGA CAA GGT GAC GGC CTT GCA GTT GGT GTG GGC CGA CCG GTG GTC GCT
 R   H   L   W   R   Q   G   D   G   L   A   V   G   V   G   R   P   V   V   A

GGC CGT GGG GCG GCG GAC TTC GAC GAA GGT CGG TAC CCA GCC GGT GCT CGG GAT GCG AGC
 G   R   G   A   A   D   F   D   E   G   R   Y   P   A   G   A   R   D   A   S

CAC CAG GAG GTC AGC CTG ACG CGA GCG CTC TAAACGTGTTGCCGGAGAACGATTTGTGACAGCAATGCA
 H   Q   E   V   S   L   T   R   A   L  Stop

GTGACAGCGGTGGTTGCTCGACACGGAGGTGGCCAACGCTTCCGGACATCCTCCGATAC ATG CGG TCT GGG CGG
                                                              M   R   S   G   R

TGT GCT GGC GGA CAA GGC TGT GAC TCG GTG GAA GCG TTG GCG CCG TCC GCG ATT TAC CGC
 C   A   G   G   Q   G   C   D   S   V   E   A   L   A   P   S   A   I   Y   R

ATC ATT GCC GGA CAC TGC GTC GGT GGC GCG CCG TCC GTG CGC TCG CAG GCG CAT ATG GGG
 I   I   A   G   H   C   V   G   G   A   P   S   V   R   S   Q   A   H   M   G

ATC GAA CAA GCC CAT CGG CGG CGG GAG AAC GAG TGA TTGCTTCCAGGCCGAGTGAGTTCCGGATACC
 I   E   Q   A   H   R   R   R   E   N   E  Stop

GGTAGGCCCCAGTAGGCTATTGTGTGATGCGCTTGAAGCCAGCCCCATCTCCTGCTGCAG
```

ORF242

ORF168

ORF187

Fig.6.7 Nucleotide sequence and deduced amino acid sequences of the *Pst*1 insert of pSNS100. Three open reading frames are indicated. The first ORF corresponds to the COOH domain of a KatG like protein. Arrows indicate a potential stem loop.

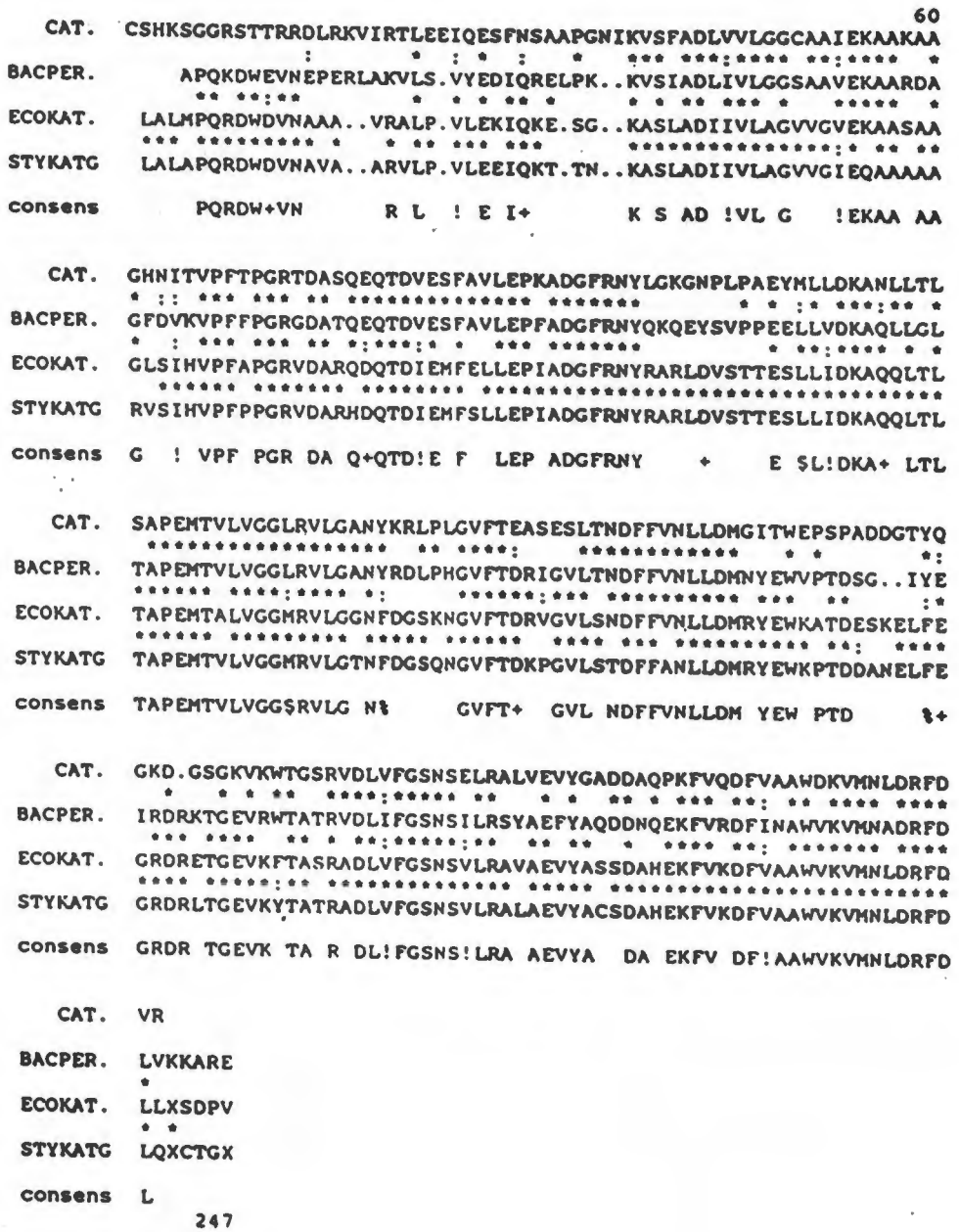


Fig. 6.8 Alignment of cloned COOH terminal domains of *katG* genes from *B. stearothermophilus*, *E.coli*, and *Salmonella typhimurium*. The upper lane indicates the *katG* of *M. tuberculosis*.. The alignment was done using Multalin programme: stars show conservation of similar amino acids.

4.4 Presence of *katG* in different mycobacterial strains

The *Alu-Pst*1 fragment of pSNS100, encoding a truncated *katG* gene, hybridised to two bands of 4,4 kb and 3 kb in *M. tuberculosis* H37Rv, an INH resistant clinical isolate and an INH sensitive isolate (Fig.6.9). The presence of the gene in the INH resistant (catalase negative) and INH sensitive (catalase positive) strains, is in keeping with the idea that loss of catalase activity is due to a mutation within the gene (Wayne *et al.*, 1968; Zhang *et al.*, 1992). The probe also hybridised to *M. Bovis* BCG but not to *M. kansasii, M. avium* and *M. intracellulare* genomic DNA. Morris *et al.* (1992), have recently isolated a catalase-peroxidase gene from M. *intracellulare*, which hybridised to *M. kansassi, M. avium* and *M. tuberculosis* H37Rv and Ra. The gene is 2452 bp. In contrast to this the probe used to detect catalase-peroxidase genes in this study was only 783 bp and contained only the carboxy-terminal end of the *KatG*. The hybridisations were washed under high stringency conditions (same as for the hybridisation in Chapter 3). This indicates that the similarity between *M. intracellulare, M. kansasii, M. avium* and the *M. tuberculosis* gene is less than 80%.

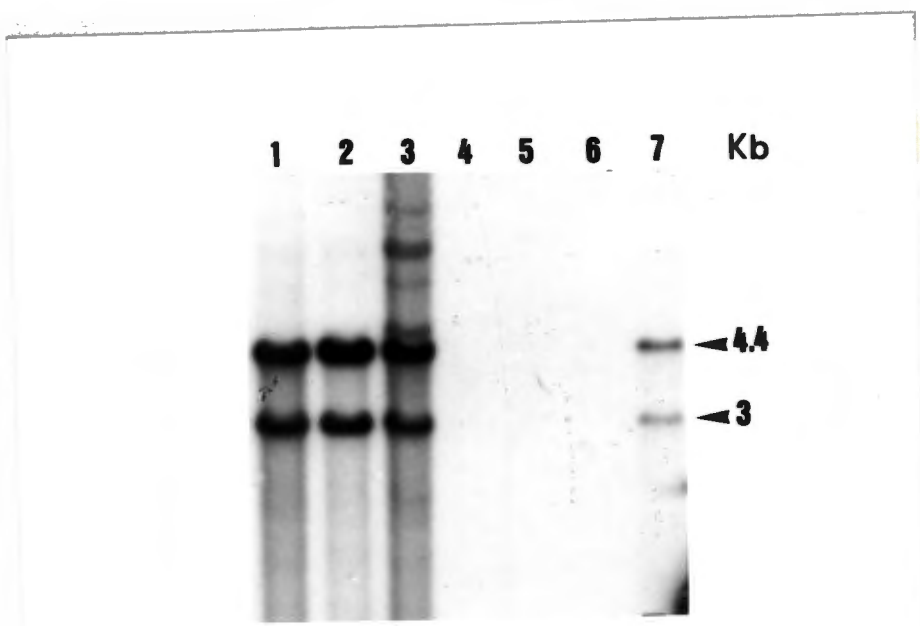


Fig.6.9 Southern blot of DNA isolated from *M. tuberculosis* H37RV (lane1), *M. tuberculosis* INH resistant strain (lane2) and an INH sensitive strain (lane3), *M. kansassi* (lane4), *M. avium* (lane5), *M. intracellulare* (lane6) and *M. bovis* BCG (lane7). Genomic DNA was digested with *Kpn*1 and probed with the *Alu-Pst*1 fragment containing a domain of the *katG* gene.

6.5 DISCUSSION

Little is known about DNA repair mechanisms, induced mutagenesis and the genes involved in these processes in the mycobacteria. It is possible that the repair mechanism(s) in the mycobacteria are more complex than that of the *E.coli* system. The finding of a second clone pSNS100 that complemented some of the defects in the *E.coli* mutant was serrendipidous. pSNS100 complements defects in the *E.coli* mutant (EtMes and UV) to a surprisingly higher level than did the clone of a *recA*-like homologue (*recS*).

It is possible that EtMes is degraded by the catalase-peroxidase activity or EtMes damages DNA by a mechanism that involves the generation of toxic oxygen species and that these are indicated by the *katG* activity.

Near UV radiation is known to exert its effects through the intermediary of oxygen species (Imlay and Linn, 1988). The ability of an exogenous catalase-peroxidase to protect against near UV damage suggests that radiation-excited chromophores could be part of the Fenton reaction (for the decomposition of H_2O_2), by either producing H_2O_2 or reducing equivalents (Imlay and linn, 1988). However, UV induced mutagenesis also generates DNA damage by producing pyrimidine dimers and this damage would need to be repaired by processes such as excision repair or post- recombinational repair.

A possible role in DNA repair may be provided by the 1.6 kb insert. However, it is not possible to say whether such DNA repair activity is due to the COOH domain of the *katG* gene (ORF242) or due to the presence of ORF168 and / or ORF63. In order to assign functions to the different domains of the *katG* gene it

would be necessary to study the entire gene with mutations or cassettes inserted in different domains, thus the unmodified domains could be assigned a function. The sub-cloning of ORF168 and ORF63 could provide insight as to whether the proteins they encode for play a role in the EtMes and UV resistance observed in the transformed *E.coli* host cells.

CHAPTER 7

GENERAL CONCLUSIONS

DNA repair and recombination have been extensively studied in the Gram-negative bacteria *E. coli.* A number of genes and their products have been isolated and shown to play important roles in these pathways. The RecA protein has been shown to have a pivotal role in both DNA repair and recombination. In contrast, the Gram-positive bacteria and in particular the pathogenic mycobacteria have only recently been investigated for these important pathways involved in cell survival and evolution. The aim of this investigation was to isolate and characterise gene/s involved in DNA repair and recombination in *M. tuberculosis.*

A 3.8 kb *Pst*I fragment pSNS201 (later shown to be 3425 bp- on sequencing), partially complemented defects in *E. coli recA* mutants. pSNS201 conferred resistance to EtMes on *E. coli* DK cells and resistance to long wave (302 nm) but not short wave (254 nm) UV light. Recombinase activity of pSNS201 was demonstrated by homologous recombination mediated in a *recA13* recipient following conjugation with a Hfr donor strain. *E. coli.* DK pSNS201 cells produced a protein (38-40 kDa), which on Western analysis cross-reacted with *E. coli* RecA antibody.

Sequencing of pSNS201 showed possible recombination/rearrangements that were taking place within the insert. The equivalent *Pst*I fragment (pSBG300) was re-isolated from an *M. tuberculosis* cosmid library.

Sequencing of pSBG300 showed only one significant ORF (ORF463) which was truncated at its 5' end; thus the complementation activity, was under control of the *lacZ* promoter. pSNS206 contains the truncated ORF436 alone and encoded homologous recombination activity and produced a RecA like protein on Western blot analysis. Recovery of the 5' end of the gene was possible using *M. tuberculosis* H37Rv cosmids. A comparison of the amino acid sequence of the entire ORF463 with that of other cloned RecA proteins showed little similarity. A 13% identity in conserved motifs was observed in the comparison with the *M. tuberculosis* and *E. coli* RecA proteins. Of particular note was the observation of a few conserved motifs in the duplex DNA binding region and a conserved ATP binding domain, Q - LYG (348aa -356aa), which appears to be conserved for RecA proteins isolated from different bacteria.

There are 5 potential start codons at the 5' end of the sequence. Of the 5 methionines observed, only one (at position 836) was associated with ribosome binding (GGAGC). The predicted molecular weight of a hypothetical protein indicated here would be 34 ,626 kDa.

The examination of the regulator/promoter region, upstream of this ribosome binding site revealed no motifs similar to the *E. coli* SOS box. The SOS box is also absent in the *M. tuberculosis recA* gene. Analysis of ORF463 shows no evidence for an "intron" as found in the cloned *recA* gene of Davis *et al.* (1990), who found an ORF twice the size of the conserved *recA* gene, even though the protein produced was approximately the same size as that of *E. coli* RecA.

Previous results obtained from functional complementation studies of pSNS201 (in an *E. coli rec* mutant), in particular the production of a protein which cross reacts with the *E. coli* polyclonal RecA antibody, led us to believe that pSNS201 encoded a RecA like protein (Nair and Steyn , 1990). However, the additional studies presented here, show that the gene is distinct from other *recA* genes and therefore ORF463 is designated *recS.*

The low percentage of homology observed between the DNA sequence of *recS* and that of the *recA* of *M. tuberculosis* probably explains the differences observed in the complementation assays between the two genes. pSNS201 was unable to restore growth of phage P1 on *E. coli* HB101, thus showing partial functional complementation of recombinase activity. In contrast, the cloned *recA* gene of Davis *et al*, from *M. tuberculosis,* demonstrated homologous recombination by a high plating efficiency of lambda red gam phages and lower levels of recombination of Hfr DNA in bacterial mating. The differences in homologous recombination complementation assays observed between the two *rec* genes from *M. tuberculosis* may indicate different abilities for recombination and DNA repair on different types of substrates. The *M. tuberculosis*, *recA* (Davis *et al.*, 1991) also demonstrated recovery after UV irradiation. However, the wavelength used to expose the transformed cells is not stated. Recovery after UV irradiation involves the direct action of RecA in recombinational repair; it has an indirect action in derepressing other DNA repair genes and the cleavage of UmuD to UmuD'. The UV resistance observed with pSNS201 in *recA* hosts does indicate the presence of RecA activity but could be due to any of the above activities. pSNS201 contains the truncated ORF463, thus it is possible that the deleted sequences are necessary for activities such as restoring growth of phage P1 and for recovery after exposure to UV radiation.

Southern hybridisation experiments revealed two copies of *recS* in H37Rv and H37Ra strains. This is in contrast to other mycobacteria and the reasons for the duplication are unknown.

Sequence analysis of the second recombinant, pSNS100, revealed 3 ORFS. Fasta comparisons of ORF168 and ORF63 with the Genebank programme did not show homology with any known or putative proteins. Analysis of ORF242 revealed that it encoded for the truncated *katG*. Hybridisation with ORF242 has demonstrated the presence of the *katG* gene in both INH resistant and INH sensitive *M. tuberculosis* clinical isolates.

This study has identified at least two DNA clones involved either directly or indirectly in repair and recombination in *M. tuberculosis*; one containing a *recA*-like gene: *(recS)* and another fragment containing a truncated *katG* gene. Furthermore, data suggests that *M. tuberculosis,* contains another gene involved in DNA repair and recombination (Davis *et al.*, 1991). It is not known whether the *recS* and *recA* genes are involved in the same or different pathways of repair and recombination and whether there in an interplay between these two genes. It is possible that a complex repair/recombination system distinct from that of *E. coli* has evolved in these organisms to protect the integrity of the mycobacterial genome.

Techniques for the genetic manipulation of mycobacteria and the cloning of genes that are essential for bacterial survival will allow an understanding of mycobacterial pathways and gene expression providing insight into the mechanisms of pathogenicity.

LITERATURE CITED

AMUNDSEN, S.K., TAYLOR, A.F., CHAUDHURY, A.M. & SMITH, G.R. (1986). *recD*: the gene for an essential third subunit of Exonuclease V. *Proceedings of the National. Academy of Sciences of the United States of America* **83**, 5558-5562.

BALTZ, R.H. (1987). Mechanism of mutation and DNA repair in *Streptomyces fradiae*. In *Proceedings of the Fifth International symposium for the Genetics of Industrial Microorganisms*. Part A, pp. 85-94. Edited by M. Alacevic, D. Hranueli & Z. Toman. Pliva, Zagreb.

BALTZ, R.H. & STONESIFER, J. (1985). Adaptive response and enhancement of N-methyl-N'-nitro-N-nitrosoguanidine mutagenesis by chloramphenicol in *Streptomyces fradiae*. *Journal of Bacteriology* **164**, 944-946.

BAUTZ,E. & FREESE, E. (1960). On the mutagenic effect of alkylating agents. *Proceeding of the National Academy of Sciences of the United States of Amercia* **46**, 1585-1594.

BERCOVIER, H., KAFRI, O. & SEAL, S. (1986). Mycobacteria possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. *Biochem Biophys Res Comm* **136**, 1136-1141.

BERGEY'S MANUAL OF SYSTEMATIC BACTERIOLOGY. Vol 2 p. 1437-1457. Eds: P.H.A. Sneath, N.S. Mair, M.E. Sharpe and J.G. Holt. (1986). Williams and Wilkins, Waverly Press.

BIGGIN, M.D., GIBSON, T.G. & HONG, G.F. (1983). Gradient sequencing gels. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3963.

BLOOM, B.R. (1992) Tuberculosis back to a frightening future. *Nature* **358**, 538-539.

BLOOM, B.R. & GODAL,T. (1983). Selective primary strategies for control of disease in the developing world. *Reviews of Infectious Diseases* **5**, 765.

BOLIVAR, F., RODRIGUEZ, R.L., GREENE, P.J., BETLACH, H.L., HEYNECKER, H.L., BOYER, H.W., CROSA, J.H. & FALKOW, S.

(1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**, 95-113.

BOYER, H.W. & ROULLAND-DUSSOIX, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *Journal of Molecular Biology* **41**, 459-472.

BRENT, R. & PTASHNE, M. (1981). The *lexA* gene product represses its own promoter. *Proceedings of the National Academy of Science of the United States of America* **78**,4204-4208.

BRIDGES, B.A., MOTTERSHEAD, R.P. & SEDGWICK, S.G. (1976). Mutagenic DNA repair in *E.coli* III. Requirement for a function of DNA polymerase III in ultraviolet-light mutagenesis. *Molecular & General Genetics* **144**, 53-58.

BURKHARDT, S.E., WOODGATE, R., SCHEUERMANN, R.H. & ECHOLS, H. (1988). The UmuD protein of *E.coli*: overproduction purification and cleavage of RecA. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 1811-1815.

CAMBELL, G.R., CARTY, P. & WINDER, F.G. (1979). Preparation and some properties of deoxyribonucleic acid polymerase of high specific activity from *Mycobacterium smegmatis*. *Biochemical Society Transactions* **7**, 23-28.

CARLSSON, J, &. CARPENTER, V.S. (1980). The *recA*$^+$ gene product is more important than catalase and superoxide dismutase in protecting E. coli against hydrogen peroxide toxicity. *Journal of Bacteriology* **142**, 319-321.

CASADABAN, M.J. &COHEN, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular BIology* **138**, 179-207.

CHAUDHURY, A.M. & SMITH, G.R. (1984). A new class of *E.coli recBC* mutants:. Implications for the role of RecBC enzyme in homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 7850-7854.

CHRISTMAN, M.F., MORGAN, R.W., JACOBSON, F.S. & AMES, B.N. (1985). Positive control of a regulon for defenses against oxidative stress and some heat shock proteins in *S. typhimurium. Cell* **41**, 753-762.

CLAIRBONE, A. & FRIDOVICH, I. (1979). Purification of the O-dianisidine peroxidase from *E.coli B. Journal of Biological Chemistry* **254**, 4245-4252.

CLAIRBONE, A., MALINOWSKI, D.P., & FRIDOVICH, I. (1979). Purification & characterisation of hydroperoxidase II of *E.coli B. Journal of Biological Chemistry* **254**, 11664-11668.

CLARK, A.J & MARGULIES, A.D. (1965). Isolation and characterisation of recombination-deficient mutants of *Escherichia coli. Proceedings of the National Academy of Sciences of the Unites States of America* **53**, 451-459.

CLARK, A.J. (1973). Recombination deficient mutants of *E. coli* and other bacteria. *Annual Review of Genetics* **6**, 67-86.

COHEN, G. (1983). Electron microscopy study of early lytic replication forms of bacteriopgahe P1 DNA. *Virology* **131**, 159-170.

COHN, M.L., ODA, U., KOVITZ, C. & MIDDLEBROOK, G. (1954a). Studies on isoniazid tubercle bacilli. I the isolation of isoniazid resistant mutants *in vitro. American Review of Tuberculosis* **70**, 465-475.

COOPER, P.K. (1982). Characterisation of long-patch excision repair of DNA in ultraviolet irradiated *Escherichia coli*: an inducible function under Rec-Lex control. *Molecular & General Genetics* **185**, 189-197.

D'ARI, R. & HUISMAN, O. (1982). DNA replication and indirect induction of the SOS response in *E.coli. Biochemie* **64**,824-835.

DAVID, H.L. (1973). Response of mycobacteria to ultraviolet radiation. *American. Review of Respiratory Diseases* **108**, 1175-1185.

DAVIS, E.O., SEDGWICK, S.G. & COLSTON, M.J. (1991). Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product. *Journal of Bacteriology* **173**, 5653-5662.

DAVIS, E.O., JENNER, P.J., BROOKS, P.C., COLSTON, M.J. & SEDGWICK, S.G. (1992). Protein splicing in the maturation of *M.tuberculosis* RecA protein: a mechanism for tolerating a novel class of intervening sequence. *CELL* **71**, 210-210.

DEFAIS, M., FAUQUET, P., RADMAN, M. & ERRERA, M. (1971). Ultraviolet reactivation and ultraviolet mutagenesis of lambda in different genetic systems. *Virology* **43**,495-503.

DEMPLE, B. & HALLBROOK, J. (1983). Inducible repair of oxidative DNA damage in *E.coli, Nature* (London), **304**, 446-448.

DEVEREUX, J., HAEBERLI,P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387-395.

DIAZ, G.A., & WAYNE, L.G., (1974). Isolation and characterisation of catalase produced by *M. tuberculosis. American Review of Respiratory Diseases* **110**, 312-319.

DRAKE, J.W (1970). The Molecular Basis of Mutagenesis. University of Illinois. Holden Day.

DUNBAR, F.P., MCALISTER, E. & GEFFERIES, M.B. (1959). Catalase and peroxidase activity of isoniazid susceptible and resistant strains of *M. tuberculosis.American Review of Tuberculosis* **79**, 664-667.

DUTREIX, M., MOREAU, P.L., BAILODE, A., GALIBERT, F., BATTISTA,J.R., WALKER, G.C. & DEVORET, R., (1989), New *recA* mutations that dissociate the various RecA protein activities in *E. coli* provide evidence for an additional role for RecA protein UV mutagenesis. *Journal of Bacteriology* **171**, 2415-2423.

ECHOLS, H. (1982). Mutation rate: some biological and biochemical considerations. *Biochimie*. **64**, 57575.

ENNIS, D.G., OSSANNA, N., & MOUNT, D.W. (1989). Genetic separation of *Escherichia coli recA* functions for SOS mutagenesis and repressor cleavage. *Journal of Bacteriology* **171**, 2533-2541.

FARR, S.B., NATVIG, D.O., KOGOMA, T. (1985). Toxicity and mutagenicity of plumbagin and the induction of a possible new DNA repair pathway in *E.coli*. *Journal of Bacteriology* **164**, 1309.

FARRAND, S.K., O'MORCHOE, S.P., & MCCUTCHAN, J. (1989). Construction of an *Agrobacterium tumefaciens* C58 *rec*A mutant. *Journal of Bacteriology* **171**, 5314-5321.

FLORY, J., TSANG, SS., & MUNIYAPPA, K.K. (1984). Isolation and visualisation of active presynaptic filaments: recA protein and single stranded DNA. *Proceedings of the National Academy of Sciences of the Unites States of America* **81**, 7026-7030.

FOGLIANO, M & SCHENDEL, P.F. (1981). Evidence for the inducibility of the uvrB operon. *Nature* **289**,196-198.

GARVEY, N. ST JOHN, A.C. & WITKIN, E.M.. (1985). Evidence for recA protein association with the cell membrane and for changes in the levels of major outer membrane proteins in SOS induced *Escherichia coli* cells. *Journal of Bacteriology* **163**, 870-876.

GAYATHRI-DEVI, B., SHAILA M.S., RAMAKRISHNAN T. & GOPINATHAN K.K. (1975), Purification and properties of peroxidase in *Mycobacterium tuberculosis*. H37Rv and its possible role in the mechanism of action of isonicotinic acid hydrazide. *Biochemical Journal*, **149**, 187-197.

GOODMAN, H.J.K., PARKER, J.R., SOUTHERN, J.A., & WOODS, D.R. (1987). Cloning and expression in *Escherichia coli* of a *rec*A-like gene from *Bacteroides fragilis*. *Gene* **58**, 265-271.

GORDON, L.K., & HESELTINE, W.A. (1982). Quantitation of cyclobutane pyrimidine dimer formation in double and single stranded DNA fragments of defined sequence. *Radiation Research* **89**,99-112.

GRANA, D., GARDELLA, T.& SUSSKIND, M.M.(1988). The effects of mutations in the *ant* promoter of phage P22 depend on context. *Genetics* **120**, 319-327.

GREEN, M.H., GREENBERG, J. & DONCH,J. (1969). Effect of a recA gene on cell division and capsular polysaccharide production in a strain of *E.coli*. *Genetic Research* **14**, 159-162.

GREENBERG, J.T. & DEMPLE, B. (1988). Overproduction of peroxidic-scavenging enzymes in *E.coli* suppresses spontaneous mutagenesis and sensitivity to redox-cycling agents in oxyR$^-$ mutants *EMBO Journal* **7**, 2611-2617.

GREENBERG, J.T. & DEMPLE, B. (1989). A global response induced in *E.coli* by redoxcycling agents overlaps with that induced by peroxide stress. *Journal of Bacteriology* **171**,3933-3939

GUDAS, L.J. & PARDEE, A.B.(1975). Model for the regulation of *E.coli* DNA repair functions. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 2330-2334.

GUDAS, L.J. & PARDEE, A.B.(1976). DNA synthesis inhibition and the induction of protein X in *E.coli*. *Journal of Molecular Biology* **101**, 459-477.

GUDAS, L.J. & MOUNT, D.W. (1977). Identification of the recA(tif) gene product of *E.coli*. *Proceedings of the National Academy of Science of the United Stataes of America* **74**, 5280-5284.

HAGENSEE, M.E., TIMME, T.L. BRYAN, S.K. & MOSES, R.E. (1987). DNA polymerase III of *E.coli* is required for UV and ethylmethanesulfonate mutagenesis. *Proceedings of the National Academy of Sciences of the United Stated of America,* **84**, 4195-4199.

HAGENSEE M.E., & MOSES, R.E. (1989). Multiple pathways for repair of hydroxide induced DNA damage in *E.coli*. *Journal of Bacteriology* **171**, 991-995.

HALLIWEL, B. & GUTTERIDGE, J.M.C. (1984). Lipid peroxidation, oxygen radicals, transition metals and disease.*Biochemical Journal* **219**, 1.

HANAHAN, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557-580.

HAROLD, R.J. & HOPWOOD, D. A. (1970a). Ultraviolet- sensitive mutants of *Streptomyces coelicolor* I. Phenotypic characterisation. *Mutational Research* **10,** 427-438.

HAROLD, R.J., & HOPWOOD, D.A. (1970b). Ultraviolet-sensitive mutants of *Streptomyces coelicolor* II. Genetics. *Mutational Research* **10**, 439-448.

HAROLD, R.J. & HOPWOOD, D.A. (1972). A rapid method for complementation testing of ultraviolet sensitive (uvs) mutants of *Streptomyces coelicolor. Mutation Research***16,** 27-34.

HARSHEY, R.M. & RAMAKRISHNAN, T. (1977). Rate of ribonucleic acid chain growth in *Mycobacterium tuberculosis H37Rv. Journal of Bacteriology* **129**, 616-622.

HEDGECOCK, L.W. & FANCHER I.O. (1957). Relation of pyrogallol-peroxidative activity to isoniazid resistance in *M.tuberculosis. American Review of Tuberculosis & Pulmonary Diseases* **75**, 670-764.

HENDRICKSON, W., KUSANO, G.,YAMAKI, H., BALAKRISHNAN, R., KING, M., MURCHIE, J. & SCHAECHTER, M. (1982). Binding of the origin of replication of *E.coli* to the outer membrane. *Cell* **30**, 915-923.

HENIKOFF, S. (1984). Unidirectional shortening with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28,** 351-359.

HERTMAN, I. & LURIA,S.E. (1967). Transduction studies on the role of a *rec* gene in ultraviolet induction of prophage lambda. *Journal of Molecular Biology* **23**, 117-133.

HEYM, B. & COLE, S.T. (1992).Isolation and characterization of isoniazid resistant mutants of *Mycobacterium smegmatis* and *Mycobacterium aurum. Research in Microbiology* **143**, 721-730.

HIGGIN, D.G., BLEASBY, A.H. & FUCHS, R. (1991). Improved software for multiple sequence alignment. *Comp. Appl. Biol. Sci.* **5,** 151-153.

HOLLIDAY, R., (1964). A mechanism for gene conversion in fungi. *Genetical Research* **5**, 282-304.

HOPWOOD, D.A, KIESER, T., COLSTON, M.J. & LAMB, F.I. (1988). Molecular Biology of mycobacteria. *British Medical Bulletin* **44**,528-546.

HORII, T., OGAWA,T. & OGAWA, H.(1980). Organisation of the *recA* gene of *E.coli*. *Proceeding of the National Academy of Sciences of the United States of America* **77**, 313-317.

HOWARD-FLANDERS, P. & BOYCE, R.P. (1966). DNA repair and genetic recombination:studies on mutants of *Escherichia coli* defective in these processes. *Radiation Research (supplement)* **6**, 156-184.

HROMIC, A. & KIRBY, R. (1987). Isolation study of three mutants of *Streptomyces cattleya* affecting DNA repair and genetic instability. *FEMS Microbiology Letters* 60, 150-155.

HUSAIN, I., VON HOUTEN, B. , THOMAS, D.C., ABDEL-MONEM, M. & SANCAR, A. (1985). Effect of DNA polymerase I and helicase II on the turnover rate of UvrABC excision nuclease. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 6774-6778.

IMLAY, J A & LINN, S. (1988). DNA damage and oxygen radical toxicity. *Science* **240**, 1302-1309.

IMLAY, J.A., & LINN, S. (1986). Mutagenesis & stress responses induced in *E.coli* by hydrogen peroxide.*Journal of Bacteriology* **169**, 2967-2976.

.JOHNSON, S.H., CREEDON, T. & WINDER, F.G. (1974). Complexes between deoxyribonucleic acid and adenosine triphosphate dependant deoxyribonuclease from *Mycobacterium smegmatis*. *Biochemical Society Transactions* **2**, 1334-1336.

JOHNSON, D.A., J.W. GAUTSCH, J.R. SPORTSMAN, & J.H. ELDER. (1984). Improved technique using non fat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose.*Gene Analysis Techniques* **1**, 3-8.

KARRAN, P., HJELMGREN, T. & LINDAHL, T. (1982). Induction of a DNA glycosylate for N-methylated purines is part of the adaptive response to alkylating agents. *Nature* **296**, 770-773.

KAWASHIMA, H., HORII, T., OGAWA, T. & OGAWA, H. (1984). Functional domains of *E.coli recA* protein deduced from the mutational sites in the gene. *Molecular and General Genetics* **193**, 288-292.

KEENER, S.L., McNAMEE, K.P. & McENTEE, K. (1984). Cloning and characterisation of *recA* genes from *Proteus vulgaris, Erwinia carotovora, Shigella flexneri,* and *Escherichia coli B/r. Journal of Bacteriology* **163**, 153-160.

KENYON, C., WALKER, G.C. (1981). Expression of the *E.coli* uvrA gene is inducible. *Nature* **289**, 808-810.

KHIDIR, M.A., CASAREGOLA, S. & HOLLAND, I.B. (1985). Mechanism of transient inhibition of DNA synthesis in ultraviolet-irradiated *E.coli*: inhibition is independent of *recA* whilst recovery requires RecA protein itself and an additional inducible SOS function. *Molecular & General Genetics* **199**, 133-140.

KOCHI, A., (1991). The global tuberculosis situation and the new control strategy of the World Health Organization. *Tubercle* **72**, 1-6.

KOGOMA, T., TORREY, T.A. & CONNAUGHTON, M.J. (1979). Induction of UV resistant DNA replication in *E.coli:* induced stable DNA replication as an SOS function. *Molecular & General Genetics* **176**, 1-9.

KOKJOHN, T.A. & MILLER, R.V. (1985). Molecular cloning and characterisation of the *recA* gene of *Pseudomonas aeruginosa* PAO. *Journal of Bacteriology* **163**, 568-572.

KONRAD, E.B. (1977). Method for the isolation of *E. coli* mutants with enhanced recombination between chromosomal duplications. *Journal of Bacteriology* 130, 167-172.

KUEMMERLE, N.B. & MASKER, W.E. (1980). Effect of the *uvrD* mutations on excision repair. *Journal of Bacteriology* **142**, 535-546.

KUSHNER, S.R., SHEPERD, J., EDWARDS, G., & MAPLES,V.F., (1978). *uvrD*, *uvrE* and *rec* represent a single gene, p251-254. In: DNA repair mechanisms. ed: P.C. Hanawalt, E.C. Friedberg & C.F. Cox. *Academic Press. Inc., New. York*

KUSHNER S.R. (1987). DNA repair in *E.coli*. In *E.coli* and *S typhimurium* Vol 2, p.1044, 1052-1053. Ed. Frederick C. Neidhardt. American Society for Microbiology.

LAEMMLI, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

LARMINAT, F., CAZAUX, C., GERMANIER, M. & DEFAIS, M. (1992). New mutations in and around the L2 disordered loop of the RecA protein modulate recombination and/or coprotease activity. *Journal of Bacteriology* **174**, 6264-6269.

LAWLEY, P.D. & BROOKES, P.(1961). Acidic dissociation of 7:9 dialkylguanines and its possible relation to mutagenic properties of alkylating agents. *Nature* **192**, 1081-1082.

LAWLEY, P.D. (1966). Effects of some chemical mutagens and carcinogens on nucleic acids. *Progress Nucleic Acids Research in Molecular Biology* **5**, 89-131

LIEBERMANN, H.B. & WITKIN, E.M. (1981). Variable expression of the ssb-I allele in different strains of *E.coli* K-12 and B: differential suppression of its effects on DNA replication, DNA repair and UV mutagenesis. *Molecular & General Genetics* **183**, 348-355.

LIPMAN, D.J. & PEARSON, W.R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435-1441.

LITTLE, J.W., EDMISTON, S.H., PACELLI, L.Z. & MOUNT, D.W. (1980). Cleavage of the *Escherichia coli* LexA protein by the *RecA* protease. *Proceeding of the National Academy of Sciences of the United States of America* **77**, 3225-3229.

LITTLE, J.W., MOUNT, D.W. & YANISCH-PERRON, Cr. (1981). Purified LexA protein is a repressor of the *recA* and *lexA* genes. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4199-4203.

LITTLE, J.W. (1984). Autodigestion of LexA and phage repressors. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 1375-1379.

LOEWEN, P.C., (1984). Isolation of catalase deficient *E.coli* mutants and genetic mapping of *KatE*, a locus that affects catalase activity. *Journal of Bacteriology* **157**, 622-626.

LOEWEN, P.C. & TRIGGS-RAINE, B.L. (1984). Genetic mapping of *KatF*, a locus that with *KatE* affects the synthesis of a second catalase species in *E.coli*. *Journal of Bacteriology* **160**, 668-675.

LOPRASERT, S., NEGRO, S. & OKADO, H. (1989). Cloning, nucleotide sequence and expression in *E. coli* of *Bacillus stearothermophilus* peroxidase gene (*per*). *Journal of Bacteriology* **171,** 4871-4875.

LOVETT, ST. & CLARK, AJ. (1983). Genetic analysis of regulation of the RecF pathway of recombination in *E. coli* K-12. *Journal of Bacteriology* **153**, 1471-1478.

LOWRIE, D.B. (1983). How macrophages kill tubercle bacilli. *Journal of Medical Microbiology* **16**, 1-2.

LOWRIE, D.B. & ANDREW, P.W. (1988). Macrophage antimycobacterial mechanisms in: Tuberculosis and Leprosy Ed: R.J.W. Rees. *British Medical Bulletin* **44**(3), 624-634.

MAGEE, T.R. & KOGOMA T. (1990). Requirement of RecBC enzyme and an elevated level of activated RecA for induced stable DNA replication in *Escherichia coli*. *Journal of Bacteriology* **172**, 1834-1839.

MALLING, H.V. & DE SERRES, F.J. (1968). Identification of genetic alterations induced by ethylmethane sulphonate in *Neurospora crassa. Mutation Research* **6**, 181-193.

MANIATIS, T., FRITSCH, E.F. & SAMBROOK, J. (1982). *Molecular cloning: a Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MARRERO, R. & YASBIN, R.E. (1988). Cloning of the *Bacillus subtillis* *recA*$^+$ gene and functional expression of *recA*$^+$ in *B.subtilis*. Journal of Bacteriology **170**, 335-344.

MARTIN, R. (1987). Overcoming DNA sequencing artefacts: stops and compressions; in *FOCUS* **9**(1), 8-10.

MATSUSHIMA, P. & BALTZ, T.H. (1987). RecA gene of E. coli complements defects in DNA repair and mutagenesis in *Streptomyces fradiae* JSE *(mcr-6). Journal of Bacteriology* **169,** 4834-4836.

McENTEE, K., WEINSTOCK, G.M. & LEHMAN, I.R., (1979). Initiation of general recombination catalysed *in vitro* by the *RecA* protein of *Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America* **76**, 2615-2619.

McENTEE, K. (1992). *recA*: from locus to lattice. *Nature*. **355**, 302-303.

McNULTY, M.S. & WINDER, F.G. (1971). Partial purification and properties of a DNA polymerase from *Mycobacterium smegmatis. Biochemica et Biophysica Acta*. **254**, 213-225.

MEYER, T.F. & PUTTEN, J.P.M. (1989). Genetic mechanisms and biological implications of phase variation in pathogenic Neisseriae. *Clinical Microbiology Reviews* **2,** s139-145.

MIDDLEBROOK G. (1954). Isoniazid resistance and catalase activity of tubercle bacilli. *American Review of Tuberculosis* **69**, 471-472.

MILLER, J.H. (1972). *Experiments in Molecular Genetics*. p21. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

MIURA, A. & TOMIZAWA, J. (1968). Studies on radiation sensitive mutants of *E.coli* III. Participation of the Rec system in induction of mutation by ultraviolet radiation. *Molecular and General Genetica* **103**, 1-10.

MIZUGUCHI, Y. (1974). Effect of ultraviolet sensitive mutants on gene inheritance in mycobacterial matings. *Journal of Bacteriology* **117**, 914-916.

MOREAU, P.L. & ROBERTS J.W. (1984). RecA protein- promoted lambda repressor cleavage: complementation between RecA441 and RecA430 proteins *in vitro*. *Molecular & General Genetics* **198**, 25-34.

MORGAN, R.W., CHRISTMAN, M.W., JACOBSON, F.S., STORY,G. & AMES, B.S. (1986). Hydrogen-peroxide inducible proteins in *S.typhimurium* overlap with heat shock and other stress proteins. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 8059-8063.

MORRIS, L.S., NAIR, J. & ROUSE, D.A. (1992). The catalase-peroxidase of *Mycobacterium intracellulare*: nucleotide sequence analysis and expression in *E.coli*. *Journal of General Microbiology* **138**, 2363-2370.

MUNIYAPPA, KK., RAMDAS, E., MYTHILI, E., & GALANDE, S. (1991). Homologous pairing between nucleosome cores on a linear duplex DNA and nucleoprotein filaments of RecA-protein-single stranded DNA. *Biochemie* **73**, 187-190.

NAIR, S. & STEYN, L.M. (1991). Cloning and expression in *Escherichia coli* of a *recA* homologue from *Myobacterium tuberculosis*. *Journal of General Microbiology* **137**, 2409-2414.

NOHMI, T., BATTISTA, J.R., DODSON, L.A. & WALKER, G.C. (1988). RecA mediated cleavage activities, UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 1816-1820.

NORDEEN, S.K. & GODAL, T. (1988). *British Medical Bulletin* **44**, 523-527.

NORGARD, M. & IMAEDA, T. (1978). Physiological factors involved in the transformation of *Mycobacterium smegmatis*. *Journal of Bacteriology* **133**, 1254-1262.

NORRANDER, J., KEMPE, T. & MESSING, J. (1983). Improved M13 vectors using oligonucleotide directed mutagenesis. *Gene* **26**, 101-106.

OGAWA, H., SHIMADA, K. & TOMIZAWA,J. (1968). Studies on radiation sensitive mutants of *E.coli* defective in repair synthesis. *Molecular & General Genetics* **101**, 227-244.

OGAWA, T., WABIKO, H., TSURIMOTO, T., HORII, T., MASUKATA, H. & OGAWA, H. (1979). Characteristics of purified RecA protein and the regulation of its synthesis *in vivo*. *Cold Spring Harbor Symposium Quant.Biology* **43**, 909-915.

OWTTRIM, G.W., & COLEMAN, J.R. (1989). Regulation of expression and nucleotide sequence of the *Anabaena variabilis rec*A gene. *Journal of Bacteriology* **171**, 5713-5719.

PARRISH, J.A., ANDERSON, R.R., URBACH, F. & PITTS, D. (1981). Effects of ultraviolet radiation on microorganisms and animal cells. *Biological Effects of Ultraviolet Radiation with Emphasis on Human Responses to Longwave Ultraviolet,* Chptr 5. New York and London: Plenum Press.

PETERSON, K.R. & MOUNT, D.W. (1987). Differential repression of SOS genes by unstable *lex*A41(*tsl-l*) protein causes a split-phenotype in *E.coli* K-12. *Journal of Molecular Biology* **193**, 27-40.

PETERSON, K.R., OSSANNA, N., THLIVERIS, A.T., ENNIS, D.G. & MOUNT, D.W. (1988). Derepression of specific genes promotes DNA repair and mutagenesis in *E.coli*. *Journal of Bacteriology* **170**, 1-4. Minireview.

PHIZICKY, E.M. & ROBERTS, J.W. (1981). Induction of SOS functions: regulation of proteolytic activity of *E. coli* RecA protein by interaction with DNA and nucleoside triphosphate. *Cell* **25**, 259-267.

QUEMARD, A., LACAVE, C. & LANCELLE, G. (1991). Isoniazid inhibition of mycolic acid synthesis by cell extracts of sensitive and resistant strains of *Mycobacterium aurum*. *Antimicrobial Agents & Chemotherapy* **35**, 1035-1039.

RADMAN, M. (1974). Phenomenology of an inducible mutagenic DNA repair pathway in *E.coli*: SOS repair hypothesis. *In: Molecular and Environmental Aspects of Mutagenesis* (Prakash L, Sherman F, Lawrence T,Tabo HW, eds) Charles C. Thomas, Springfield, IL, 128-142.

RAMESAR, R.S., WOODS, D.R., & RAWLINGS, D.E. (1988). Cloning and expression in *Escherichia coli* of a *rec*A-like gene from the acidophilic

autotroph *Thiobacillus ferrooxidans. Journal of General Microbiology* **134**, 1141-1146.

RAMESAR, R.S., ABRATT, V., WOODS, D.R. & RAWLINGS, D.E. (1989). Nucleotide sequence and expression of a cloned *Thiobacillus ferrooxidans recA* gene in *E. coli. Gene* **78**, 1-8.

RINKEN, R., THOMS, B. & WACKERNAGEL, W. (1992). Evidence that *recBC*-dependent degradation of duplex DNA in *E. coli recD* mutants involve DNA unwinding. *Journal Bacteriology* **174**, 5424-5429.

ROBERTS J.W. & ROBERTS C.W. (1975). Proteolytic cleavage of bacteriophage lambda repressor in induction. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 147-151.

ROBERTS, J.W., ROBERTS, C.W. & CRAIG, N.L. (1978). *E.coli recA* gene product inactivates lambda repressor. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 4714-4718.

ROBERTS, J. & DEVORET, R. (1983). *Lysogenic induction,* **p.123-144**. *In Lambda II.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. ed: R.W. Hendrix, J.W. Roberts, F.W. Stahl & R. A. Weisberg.

RUPERT, C.S., GOODGAL, S.H. & HERRIOT, R.M. (1958). Photoreaction *in vitro* of ultraviolet inactivated Hemophilus influenzae transforming factor. *Journal of General Phisology* **41**, 451-471.

RUPP, W.D. & HOWARD-FLANDERS, P. (1968). Discontinuities in the DNA synthesised in an excision defective strain of *E.coli* following UV irradiation. *Journal of Molecular Biology* **31**, 291-304.

RYBICKI, E.P., & VON WECHMAR, M.B. (1982). Enzyme assisted immune detection of plant virus proteins blotted onto nitrocellulose paper. *Journal of Virological Methods* **5**, 267-278.

SAMSON, L., & CAIRNS, J. (1977). A new pathway for DNA repair in *E.coli*. *Nature* **267**, 281-282.

SANCAR, A., WHARTON, R.P., SELTZER, S., KACINSKI, B.M. CLARK, N.D. & RUPP, W.D. (1981a). Identification of the *uvrA* gene product. *Journal of Molecular Biology* **148**, 45-62.

SANCAR, A., CLARK, N.D., GRISWOLD, J., KENNEDY, W.J. & RUPP,W.D. (1981b). Identification of the uvrB gene product. *Journal of Molecular Biology* **148**, 63-76.

SANCAR, A., KACINSKI, B.M., MOTT, D.L. & RUPP, W.D. (1981c) Identification of the *uvrC* gene product. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 5450-5454.

SANCAR, A., STACHELEK, C., KONMIGSBERG, W. & RUPP, W.D. (1980). Sequences of the *recA* gene and protein. *Proceedings of National Academy of Sciences of the United States of America* **78**, 4274-4278.

SANGER, F., NICKLEN, S. & COULSON, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74,** *5463-5467.*

SASSANFAR, M. & ROBERTS, J.W. (1990). Nature of the SOS inducing signal in E. coli: the involvement of DNA replication. *Journal of Molecular Biology* **212,** 79-76.

SCHELLHORN, H.E. & HASSAN, H.M. (1988). Transcriptional regulation of *katE* in *E.coli K-12. Journal of Bacteriology* **170**, 4286-4292.

SETLOW, J.K., SPIKES, D. & GRIFFIN K. (1988). Characterization of the *rec-I* gene of *Haemophilus influenzae* and behavior of the gene in *Escherichia coli. Journal Bacteriology* **170**, 3876-3881.

SHARP, A.K &. BANERJEE, D.K. (1985). Hydrogen peroxide and superoxide production by periphereal blood monocytes in leprosy. *Clinical & Experimental Immunology* **60**, 203-206.

SHINAGAWA, H., IWASAKI, H., KATO, T. & NAKATA, A. (1988). RecA protein dependent cleavage of UmuD protein and SOS mutagenesis. *Proceeding of the National Academy of Sciences of the Unites States of America* **84**, 1806-1810.

SHINNICK, T.M. (1987). The 65-Kilodalton Antigen of *Mycobacterium tuberculosis. Journal of Bacteriology* **169**, 1080-1088.

SINGER, J.T. (1989). Molecular cloning of the *recA* analog from the marine fish pathogen *Vibrio anguillarum* 775. *Journal of Bacteriology* **171**, 6367-6371.

SNAPKA, R.M. & SUTHERLAND, B.M. (1980). *E.coli* photoreactivating enzyme: purification and properties. *Biochemistry*. **19**, 4201-4208.

SOUTHERN, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503-517.

STASIAK, A., DI CAPUA, E., & KOLLER, T. (1981). Elongation of duplex DNA by RecA protein. *Journal of Molecular Biology* **151**, 557-564.

STORY R.M., WEBER, T.I. & STEITZ, T.A. (1992).The structure of the *E.coli recA* protein monomer and polymer.*Nature* **355**, 318-325.

STORZ, G., CHRISTMAN, M.F., SIES, H. & AMES, B.N. (1987). Spontaneous mutagenesis and oxidative damage to DNA in *Salmonella typhimurium*. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 8917-8921.

STRAUSS, B.S. (1962). Response of *E.coli* auxotrophs to heat after treatment with mutagenic alkyl methanesulfonates. *Journal of Bacteriology* **83**, 241-249.

STROHL, W.R. (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Research* **20**(5), 961-974.

SUTHERLAND, B.M. & CHAMBERLIN, M.J. (1973). Deoxyribonucleic acid photoreactivating enzyme from *E. coli*. *Journal of Biological Chemistry* **248**, 4200-4205.

SWEASY, J.B., WITKIN, E.M., SINHA, N. & ROEGNER-MANISCALO, V. (1990). RecA protein of *E.coli* has a third essential role in SOS mutator activity. *Journal of Bacteriology* **172**, 3030-3036.

TAYLOR, A.F., SCHULTZ, D.W., PONTICELLI, A.F. & SMITH, G.R. (1985). RecBC enzyme nicking at *Chi* sites during DNA unwinding: location and orientation-dependence of the cutting. *Cell* **41**, 153-163.

TELENDER-MUSKAVITY, K.M. & LINN,S. (1982). A unified mechanism for the nuclease and unwinding activities of the recBC enzyme of *E.coli*. *Journal of Biological Chemistry* **257**, 2641-2648.

TEN DAM, H.G. (1984). Research on BCG vaccination. Advances in *Tuberculosis Research* **21**, 79-106.

TESSMAN, I., PODDAR, R.K., & KUMAR, S. (1964). Identification of the altered bases in mutated single-stranded DNA, I. *In vitro* mutagenesis by hydroxylamine, ethyl methanesulfonate and nitrous acid. *Journal of Molecular Biology* **9**, 352-363.

TESSMAN, E.S. & PETERSON, P.K. (1985). Isolation of protease-proficient, recombinase-deficient *recA* mutants of *E. coli* K-12. *Journal of Bacteriology* **163**, 688-685.

THOLE, J.E.R., DAUWERSE, H.G., DAS, P.K., GROOTHUIS, D.G., SCHOULS, L.M. & VON EMBDEN, J.D.A. (1985). Cloning of *Mycobacterium bovis* BCG DNA and expressiion of antigens in *Escherichia coli. Infection & Immunity* **50**, 800-806.

TOWBIN, H., STAEHELIN, T. & GORDON, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of National Academy of Sciences of the United States of America* **76**, 4350-4354.

TRIGGS-RAINE, B.L. & LOEWEN, P.C. (1987). Physical characterisation of *KatG* encoding catalase HPI of *E.coli*. *Gene* **52**, 121-128.

TRIGGS-RAINE B.L., DOBLE, B.W., MULVEY, M.R., SORBY, P.A. & LOEWEN, P.C. (1988). Nucleotide sequence of *KatG*, encoding catalase HPI of *E.coli*, **170**, 4415-4419.

USDIN, K., GERTSCH, K. & KIRBY, R. (1985). The loss of a large DNA fragment is associated with an aerial mycelium negative (Am-) phenotype of *Streptomyces cattleya*. *Journal of General Mirobiology* **131**, 979-981.

VIEIRA, J. & MESSING, J. (1982). The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259-268.

WALKER, G.C. (1984). Mutagenesis and inducible responses to deoxyribonucleic acid damage in *E.coli*. *Microbiological Reviews* **48**, 60-93.

WALKER, G.C. (1985). Inducible DNA repair systems. *Annual Revieview of Biochemistry* **54**, 425-57.

WALKER, G.C., SHEVELL, D.E. & BATTTISTA, J.R. (1987). Mutagenesis and cellular response to DNA damage. In *Proceedings of the Fifth International Symposium on the Genetics of Industrial Microorganisms*, Part A, pp. 65-71. Edited by M. Alacevic, D. Hranueli and Z. Toman. Pliva: Zagreb.

WALKER, G.C. (1987). The SOS response of *E.coli*. In: *E. coli and Salmonella.* Vol 2, 1346-1353. Ed. Frederic, C. Neidhardt. American Society for Microbiology.

WALKUP, L.K.B. & KOGOMA, T. (1989). *E.coli* proteins inducible by oxidative stress mediated by the superoxide radical. *Journal of Bacteriology* **171**, 1476-1484.

WAYNE, L.G., DIAZ, G.A. & DOUBEK, J.R. (1968). Acquired isoniazid resistance and catalase activity of mycobacteria. *American Review of Respiratory Diseases* **97**, 903-913.

WAYNE, L.G., & DIAZ, G.A. (1982). Serological, taxonomic and kinetic studies of the T and M classes of Mycobacterial catalase. *International Journal of Systematic Bacteriology* **32**, 296-304.

WEINSTOCK, G.M. (1987). General Recombination in *E.coli*. In: *E.coli and Salmonella* Vol.**2.**, 1034-1043.Ed. by Frederic C. Neidhardt. American Society for Microbiology.

WEISEMANN, J.M. &. WEINSTOCK, G.M. (1991). The promoter of the *recA* gene of *E.coli. Biochemie* **73**, 457-470.

WEST, S.C. & CONNOLLY, B. (1992). Biological roles of the *E.coli* RuvA, RuvB and RuvC proteins revealed. *Molecular Microbiology* **6**(19), 2755-2759.38.

WILLIS, D. K., UHLIN, B. E., AMINI, K.S. & CLARK, A.J. (1981). Physical mapping of the *Srl* and *recA* region of *E.coli*: analysis of the Tn10 generated insertions and deletions. *Molecular & General Genetics* **183**, 497-504.

WINDER, F.G. (1964a). The antibacterial action of Streptomycin, isoniazid and PAS. In: *Chemotherapy of tuberculosis* (Ed: V.C. Bary), Butterworths, London, p 111-149.

WINDER, F.G. (1964b). Early changes induced by isoniazid in the composition of *M.tuberculosis*. *Biochemical & Biophysical Acta.* **82**, 210-212.

WINDER, F.G. & ROONEY, S.A. (1968). Effects of isoniazid on the triglycerides of BCG. *American Review of Respiratory Diseases* **97**, 938-940.

WINDER, F.G. & COUGHLIN, M.P. (1969). A nucleoside triphosphate-dependant deoxyribonucleic acid breakdown system in *Mycobacterium smegmatis* and the effect of iron limitation on the activity of this system. *Biochemical Journal* **111**, 189-196.

WINDER, F.G. & COLLINS, P.B. (1970). Inhibition by isoniazid of synthesis of mycolic acids in *Mycobacterium tuberculosis*. *Journal of General Microbiology* **63**, 41-48.

WINDER, F.G. & SASTRY, P.A. (1971). The formation of a long-lived complex between an ATP-dependent deoxyribonuclease and DNA. *FEBS Letters* **17**, 27-30.

WINDER, F.G. (1982). Mode of action of antimycobacterial agents and the associated aspects of the molecular biology of the mycobacteria. In: The biology of the mycobacteria. Vol.1. *Physiology identification and classification* ed. by Colin Ratledge & John Stanford. Academic Press London, 354-425.

de WIT, D., STEYN, L.M., SHOEMAKER, S. & SOGIN, M. (1990). Direct detection of *Mycobacterium tuberculosis* in clinical specimens by DNA amplification. *Journal of Clinical Microbiology* **28,** 2437-2441.

WITKIN, E.M. (1967). Mutation-proof and mutation-prone modes of survival in derivatives of *E.coli B* differing in sensitivity to UV light. Bookhaven Symposum on Biology **20**, 17-55.

WITKIN, E.M. (1969). The mutability toward ultraviolet light of recombination deficient strains of *E.coli*. *Mutational Research.* **8**, 9-14.

WITKIN, E.M. (1974). Thermal enhancement of ultraviolet mutability in a *tif-1 uvrA* derivative of *E.coli* B/r: evidence that ultraviolet mutagenesis depends upon an inducible function. *Proceedings of the National Academy of Sciences of the United States of America* **7**,1930-1934.

WITKIN, E.M. (1976). Ultraviolet mutagenesis and inducible DNA repair in *E.coli*. *Bacterial Review* **40**, 869-907.

WITKIN, E.M., V. ROEGNER-MANISCALCO, J.B. SWEASY & J.O. MCCALL. (1987). Recovery from ultraviolet light-inhibited inhibition of DNA synthesis requires *umuDC* gene products in *recA718* mutant strains but not in recA+ strains of *E. coli*. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 6805-6809.

WITKIN, E.M. (1991). RecA protein in the SOS response: milestones and mysteries. *Biochemie* **73**, 133-141.

WOODGATE, R., RAJAGOPALAN, M., LU, C. &. ECHOLS, H. (1989). UmuC mutagenesis protein of *E.coli*: purification and interaction with UmuD and UmuD'. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 7301-7305.

WORCEL, A. (1970). Induction of chromosome reinitiation in a thermosensitive DNA mutant of *E.coli*. *Journal of Molecular Biology* **52**, 371-386.

YANOFSKY, C., ITO, J. & HORN, V. (1966). Amino acid replacements & the genetic code. Cold Spring Harbor *Symposum of Quant. Biology* **31**, 151-162.

YEUNG, A.T., MATTES, W.B, OH, E.Y., YOAKUM G.H. &. GROSSMAN, L.(1983). Enzymatic properties of purified *E. coli* UvrABC protein. *Procedings of the National Academy of Sciences of the United States of America* **80**, 6157-6161.

YOUATT. J. (1969). A review of the action of isoniazid. *American Review of Respiratory Diseases* **99**, 729-749.

YOUNG, L.S. (1988). *Mycobacterium avium* complex. *Journal of Infectious Diseases* **157**, 863-867.

ZABEAU, M. & STANLEY, K.K. (1982). Enhanced expression of *cro*-B-galactosidase fusion proteins under the control of the Pr promoter of bacteriophage lambda. *EMBO Journal* **1**, 1217-1224.

ZHANG, Y., HEYM, B., ALLEN, B., YOUNG, D. & COLE, S.T. (1992). The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*.*Nature* **358**, 591-593.

ZHAO, X. & DREYFUSS, L.A. (1990). Expression and nucleotide analysis of the *Legionella pneumophila recA* gene. *FEMS Microbiology Letters* **70**, 227-232.

APPENDIX 1

Bacterial strains used during the course of this work

1A. *E. coli* - laboratory strains

Strain	Relevant markers	Reference
E. coli LK111	*E. coli* K514 derivative (*lacZ*ΔM15)	Zabeau & Stanley, 1982
E. coli HB101	*leuB6, proA2, recA13, rpsL20*	Boyer & Roulland-Dussoix, 1969
E. coli DK	Δ(*ara,leu*)7697 Δ(*srt-recA*)306	ATCC 35691 Willis *et al.*, 1981
E. coli MC1060	*recA*$^+$ parent strain of DK	Casadaban & Cohen, 1980
E. coli RR1	*recA*$^+$ parent strain of HB101	Bolivar *et al.*, 1977
E. coli CSH62 CA8000	HfrH (protroph, str^S) *thr-*, *Leu-*, *Lac-*, *Lys-*	Miller, 1972
E. coli GY7066	Δ*recA306* *lacBK1lacMS286* *sfiB114*	Dutreix *et al.*, 1989

1B. Mycobacterial strains

M. tuberculosis H37Rv, WHO strain
M. tuberculosis H37Ra, WHO strain
M. bovis BCG, WHO strain

The following mycobacterial strains were isolated at the Mycobacteriology Laboratory, Groote Schuur Hospital, Cape Town

M. smegmatis
M. fortuitum
M. kansassi
M. aurum
M. avium
M. intracellulare

APPENDIX 2A

pEcoR252	Ap^R *EcoR1 ("suicide vector")*	Zabeau & Stanley, 1982
pUC19 and 18	Ap^R	Vieira & Messing, 1982
M13mp18/19	Ap^R	Norander *et al.*, 1983

APPENDIX 2B

PSNS201	*M. tuberculosis recS* in pUC19	This study
pH15	*M. tuberculosis recS* in pEco R252	This study
pSBG100	4.8 kb	This study
pSBG101	750 bp	This study
pSBG102	900 bp	This study
pSBG103	3150 bp	This study
pSBG300	3436 bp	This study
pSNS203	3 kb	This study
pSNS204	2.5 kb	This study
pSNS205	1.9 kb	This study
pSNS206	1.1 kb	This study
pSNS207	750 bp	This study
pSNS208	550 bp	This study
pSNS209	530 bp	This study

APPENDIX 3

GENERAL METHODS

3.1 MEDIA

3.1.1 Luria Broth/agar

Luria	Broth	Agar
Tryptone	10g	10g
Yeast extract	5g	5g
NaCL	5g	5g
Agar	-	15g

Make up to 1 L with distilled water

3.1.2 Minimal medium (M9)

10 x M9 minimal salts solution

NA2HP04	60g
KH2PO4	30g
NH4C1	10g
NaCl	5g

Distilled water to lL

The following reagents were autoclaved separately before mixing

M9 salts	100ml
1M MgS04	1ml
0.1 M CaC12	1ml
1M Thiamine-HCL	1ml
20% glucose	10ml

distilled water to 1 L

4.2 BACTERIAL CULTURE

4.2.1 *E. coli*

Unless otherwise stated all strains of *E. coli* were cultured in Luria Broth (L/B) or on Luria Agar (LA) plates at 37°C. Antibiotic stocks were prepared as described in Maniatis *et al*, 1989, and routine concentrations used are shown in Table X.

Working concentrations of antibiotics used in this study

Antibiotic	Concentration μg/ml
Ampicillin	100
Kanamicin	35
Streptomycin	25