

# **A Study of the Expression of a Protein Proteinase Inhibitor from Sweet Corn**

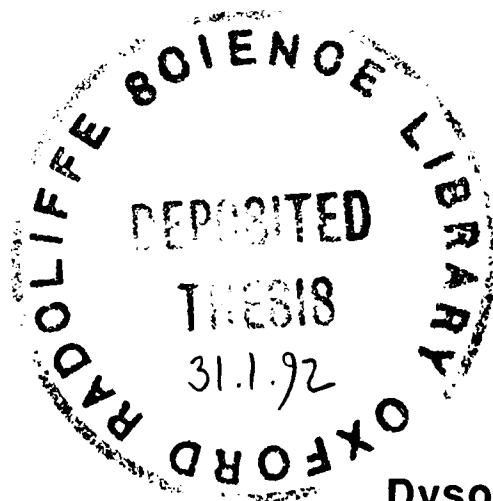
**A Thesis submitted to the board of the Faculty of Physical Sciences  
in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy  
in the  
University of Oxford**

**by**

**H.A. Rohan de Silva**

**Lincoln College**

**Oxford**



**Trinity Term, 1991**

**Dyson Perrins Laboratory**

# ABSTRACT

## A STUDY OF THE EXPRESSION OF A PROTEIN PROTEINASE INHIBITOR FROM SWEET CORN

H.A. Rohan de Silva  
Lincoln College  
Oxford

Submitted for the degree of  
Doctor of Philosophy  
Trinity Term 1991

=====

Sweet Corn Inhibitor (SCI), a small (11811Da.) protein from the seeds of opaque-2 corn is a potent and specific inhibitor of trypsin and the activated Hageman Factor (Factor  $\beta$ XIIa) of the human blood plasma coagulation system. With the eventual aim of obtaining insight into the structure-function relationships of the selective SCI- $\beta$ XIIa interaction, a synthetic gene for SCI was cloned into *Saccharomyces cerevisiae* (yeast) and *Escherichia coli* (*E.coli*) expression systems in an attempt to obtain overexpression of the recombinant gene product. The establishment of functional expression, together with an isolation and purification procedure for SCI would provide a system for obtaining selected reactive-site mutants of SCI by cassette- and oligonucleotide-directed mutagenesis.

A yeast secretion vector for a truncated form of SCI (tSCI) was constructed by cloning the gene for  $\alpha$ -factor prepro-tSCI fusion, downstream to the  $\alpha$ -mating factor (MF $\alpha$ 1) promoter of yeast. Yeast transformants containing the expression vector failed to express and secrete the desired product.

The synthetic gene encoding the complete SCI sequence was cloned into *E.coli* expression vectors that directed both cytoplasmic and periplasmic expression. In cytoplasmic expression, the SCI gene was cloned directly downstream to the powerful, inducible  $\lambda$ -phage P<sub>L</sub>- and *trc*-promoters. No expression was obtained with the latter. With the former, expression levels of up to 3% of the total bacterial protein were obtained. These levels were improved 3- to 4-fold on incorporation of the *E.coli dnaY* gene product. Solubilisation and refolding of the purified SCI inclusion bodies failed to yield the active, correctly folded product. Failure to obtain an N-terminal sequence indicated an incompletely processed N-terminal methionine. For periplasmic expression, SCI, fused in-frame to the signal sequence of OmpA, a major *E.coli* outer membrane protein, was cloned into the same  $\lambda$ -phage P<sub>L</sub> promoter vector. High levels ( $\approx$ 10%) of expression of insoluble SCI were obtained. The nearly homogeneous product was obtained by a two-step procedure, involving ion-exchange chromatography, followed by hydrophobic interaction chromatography. Characterisation by N-terminal sequencing, SDS-PAGE and electrospray mass spectrometry, confirmed the presence of correctly processed SCI in the form of covalently associated dimers. Refolding studies are at present in progress.

*We've covered ground since that awkward day  
When, thoughtlessly, a human mind  
Decided to leave the apes behind,  
Come pretty far, but who dare say  
If far be forward or astray,  
Or what we still might do in the way  
Of patient building, impatient crime,  
Given the sunlight, salt and time.*

W.H. Auden, 1964

# Acknowledgements

I would like to extend my thanks to the following:

**The Marquis de Amodio and the Berrow Foundation** without whose kind generosity all this would never have happened.

**The Roche Research Foundation** in Basel for helping me through my final year.

**Professor Gordon Lowe** for accepting me into his group, providing me with an interesting and challenging project and for his guidance and encouragement throughout.

**Drs. Yasumasa 'Gene' Kodo & Steve Field** who patiently led me through the ABCs of most I had to know about everything I have done here.

**Dr. John Sutherland & cohorts** for their help, advice, plasmids and the use of their state-of-the-art facilities.

**Richard Heath** for his lessons in the basics of black-and-white photography

**Drs. Wendy Loughlin & Sheena Radford** for setting me on the right track with their no-nonsense proofreading.

**Drs. Peter Esnouf and Anne Maisey** for their expert advice and help in matters protein isolation.

**Mr. Tony Willis** for the HPLC purification and N-terminal sequences.

**Dr. Robin Aplin**, for the electrospray MS analysis of my protein soups.

**The staff of the DP** for all their help at different times.

**The Robinson Family**; and members of the DeePee, past and present, for their friendly company and making my spell an interesting and memorable one. To **Wendy, Jackie, Anne, Nandita and Alison** for all their help and support in these last few days in a daze..

**Val**, a special pal.

**The Celliers: Aline, Marc, Sebastien and Dimitri** (and more yet to come) for letting me be one of them.

**Anna Frankum**, for "being there" and for the regular nutritional supplements.

And all those special creations including **Karim Abadir, Charlotte Chaliha, Dave "Mite"**

**Chalmers**, the intrepid Marlboro' adventure team: **MarilenaCarabatea, Leslie Glickman and Eleni Raftopolous**, my man at the pentagon: **Randy Hyer, Phil Kaplan, Eurydice Kefalidou**, the drude from Ojai: **Richard Kortum, Fiona McPhee, Jane Noble, Dimitris Plantzos, Jennifer Logan (née Roberts), Varun Sahni and Eli & Fania Salzberger-Oz**.

And further afield, the London Branch: **Ranjan & Enoka David, P. Thiyagarajah and Rohan Wijesinha** amongst others

**Karin Munasinghe**, my bonnie over the ocean, a special thanks for the invaluable use of her Mac.

And last, but most, my Mother, **Elsbeth de Silva**, if not for her sacrifice, I would still be selling "Durol".....

# Abbreviations

|               |                                        |
|---------------|----------------------------------------|
| A             | Adenine                                |
| α1-AT         | α1-antitrypsin                         |
| Ala           | Alanine                                |
| APS           | Ammonium persulphate                   |
| Arg           | Arginine                               |
| Asn           | Arparagine                             |
| Asp           | Aspartic Acid                          |
| AT-P          | α1-antitrypsin Pittsburgh              |
| ATP           | Adenosine triphosphate                 |
| b.p.          | Base pairs                             |
| BPTI          | Bovine pancreatic trypsin inhibitor    |
| BTI           | Barley trypsin inhibitor               |
| C             | Cytosine                               |
| C-terminus    | Carboxyl-terminus                      |
| C1-INH        | C1-Inhibitor                           |
| cDNA          | Complementary deoxyribose nucleic acid |
| Cys           | Cysteine                               |
| DAB           | Diaminobenzidine                       |
| DNA           | Deoxyribose nucleic acid               |
| DNaseI        | Deoxyribonuclease I                    |
| DTT           | Dithiothreitol                         |
| <i>E.coli</i> | <i>Escherichia coli</i>                |
| EDTA          | Ethylenediamine tetraacetic acid       |
| ER            | Endoplasmic reticulum                  |
| ES-MS         | Electrospray mass spectrometry         |
| FIX           | Factor IX (Christmas factor)           |
| fMET          | N-formyl methionine                    |
| FPLC          | Fast protein liquid chromatography     |
| FVIII         | Factor VIII (Anti-haemophilic factor)  |
| FXII          | Factor XII (Hageman factor)            |
| G             | Guanine                                |
| Gln           | Glutamine                              |
| Glu           | Glutamic acid                          |
| Gly           | Glycine                                |
| GSH           | Reduced glutathione                    |
| GSSG          | Oxidised glutathione                   |
| GuCl          | Guanidine hydrochloride                |
| HCC           | Hexamminecobalt (III) chloride         |
| HIC           | Hydrophobic interaction chromatography |

|                     |                                                                |
|---------------------|----------------------------------------------------------------|
| His                 | Histidine                                                      |
| HPLC                | High pressure liquid chromatography                            |
| IB                  | Inclusion body                                                 |
| IEC                 | Ion-exchange chromatography                                    |
| Ile                 | Isoleucine                                                     |
| IPNS                | Isopenicillin N synthase                                       |
| IPTG                | Isopropylthiogalactoside                                       |
| k.b.                | kilobase pairs                                                 |
| LB                  | Luria-Bertani                                                  |
| Leu                 | Leucine                                                        |
| Lys                 | Lysine                                                         |
| Met                 | Methionine                                                     |
| mRNA                | Messenger RNA                                                  |
| N-terminus          | Amino-terminus                                                 |
| OD <sub>n</sub>     | Optical density (subscript indicates wavelength in nanometres) |
| Orn                 | Ornithine                                                      |
| PAGE                | Polyacrylamide gel electrophoresis                             |
| PBS                 | Phosphate buffered saline                                      |
| Phe                 | Phenylalanine                                                  |
| PIPES               | Piperazine-N,N'-bis-[2-ethanesulphonic acid]                   |
| PMSF                | Phenylmethylsulphonyl fluoride                                 |
| ppαF                | Prepro-alpha-factor                                            |
| Pro                 | Proline                                                        |
| PS                  | Plasminostreptin                                               |
| PSTI                | Pancreatic secretory trypsin inhibitor                         |
| PT                  | Prothrombin                                                    |
| PTI                 | Pancreatic trypsin inhibitor                                   |
| recSCI              | Recombinant sweet corn inhibitor                               |
| rER                 | Rough endoplasmic reticulum                                    |
| RNA                 | Ribose nucleic acid                                            |
| RNaseA              | Ribonuclease A                                                 |
| rRNA                | Ribosomal RNA                                                  |
| <i>S.cerevisiae</i> | <i>Saccharomyces cerevisiae</i> (yeast)                        |
| SCI                 | Sweet corn inhibitor                                           |
| SD                  | Shine-Dalgarno                                                 |
| SDS                 | Sodium dodecyl sulphate                                        |
| Ser                 | Serine                                                         |
| SRP                 | Signal recognition particle                                    |
| SSI                 | <i>Streptomyces subtilisin</i> inhibitor                       |
| STI                 | Soybean trypsin inhibitor                                      |
| T                   | Thrombin, Thymine                                              |
| TEMED               | N,N,N',N'-tetramethylethylenediamine                           |
| TF                  | Tissue factor                                                  |
| TFA                 | Trifluoroacetic acid                                           |

|      |                                   |
|------|-----------------------------------|
| Thr  | Threonine                         |
| tPA  | Tissue-type plasminogen activator |
| tRNA | Transfer RNA                      |
| Trp  | Tryptophan                        |
| tSCI | Truncated sweet corn inhibitor    |
| Tyr  | Tyrosine                          |
| U    | Uracil                            |
| Val  | Valine, Valerie                   |

# Table of Contents

|                                                                                                                       |           |
|-----------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Chapter 1. Introduction.....</b>                                                                                   | <b>1</b>  |
| 1.1. Proteinases and Proteinase Inhibitors.....                                                                       | 3         |
| 1.1.1. Nomenclature & Classification .....                                                                            | 3         |
| 1.1.2. Serine Proteinases.....                                                                                        | 5         |
| 1.1.3. Mechanism of Serine Proteinases.....                                                                           | 6         |
| 1.1.4. Inhibitors of Serine Proteinases.....                                                                          | 8         |
| 1.1.5. The Standard Mechanism of Inhibition.....                                                                      | 10        |
| 1.2. Blood Coagulation.....                                                                                           | 12        |
| 1.2.1. Introduction.....                                                                                              | 12        |
| 1.2.2. Factor XII and the Contact Activation System.....                                                              | 16        |
| 1.3. Activated Hageman Factor Inhibitors.....                                                                         | 17        |
| 1.3.1. Plasma Inhibitors of the Activated Hageman<br>Factor.....                                                      | 18        |
| 1.3.2. Plant Inhibitors of the Activated Hageman Factor.....                                                          | 18        |
| 1.3.3. Sweet Corn Inhibitor.....                                                                                      | 19        |
| 1.3.4. The Putative Reactive-Site Loop of SCI<br>Contribution of Primary Sequence to FXIIa<br>Specificity?.....       | 20        |
| <b>Chapter 2. The Expression of a Truncated SCI Gene<br/>                in <i>Saccharomyces cerevisiae</i> .....</b> | <b>26</b> |
| 2.1. Introduction.....                                                                                                | 26        |
| 2.1.1. Heterologous Expression in <i>Saccharomyces<br/>                cerevisiae</i> .....                           | 26        |
| 2.1.2. The 2-Micron Circle.....                                                                                       | 27        |
| 2.1.3. The Expression Cassette.....                                                                                   | 27        |
| 2.1.4. Regulation of Heterologous Expression .....                                                                    | 28        |
| 2.1.5. Protein Secretion in Yeast.....                                                                                | 29        |
| 2.1.6. The $\alpha$ -Factor .....                                                                                     | 30        |
| 2.2. Expression of SCI in Yeast.....                                                                                  | 32        |
| 2.2.1. Construction of pUCSCR- $\alpha$ - an Expression Cassette<br>for a Truncated SCI Gene .....                    | 33        |
| 2.2.2. Construction of pRS-SC56 - an Expression Vector for<br>a Truncated SCI Gene.....                               | 35        |
| 2.2.3. Transformation of Yeast Spheroplasts.....                                                                      | 37        |
| 2.2.4. Expression Studies.....                                                                                        | 37        |

|                                                                                                  |           |
|--------------------------------------------------------------------------------------------------|-----------|
| 2.3. Discussion.....                                                                             | 41        |
| 2.4. Addendum.....                                                                               | 42        |
| <b>Chapter 3. Expression of SCI in <i>Escherichia coli</i> .....</b>                             | <b>44</b> |
| 3.1. Introduction.....                                                                           | 44        |
| 3.1.1. Expression of Heterologous Proteins in <i>Escherichia coli</i> - Expression Vectors.....  | 46        |
| 3.1.2. <i>E. coli</i> Promoters .....                                                            | 47        |
| 3.1.3. Regulation of Transcription.....                                                          | 48        |
| 3.1.4. The Bacteriophage Lambda P <sub>L</sub> Promoter.....                                     | 48        |
| 3.1.5. Translation Initiation .....                                                              | 50        |
| 3.2. Cytoplasmic Expression of SCI in <i>E.coli</i> .....                                        | 53        |
| 3.2.1. Inducible Expression of SCI.....                                                          | 53        |
| 3.2.2. Expression Studies Using The SCI Expression Plasmid pRS101.....                           | 55        |
| 3.2.3. The <i>dnaY</i> Gene Product.....                                                         | 59        |
| 3.2.4. Purification and Characterisation of recSCI Inclusion Bodies .....                        | 62        |
| 3.2.5. Attempted Refolding of Solubilised recSCI to Yield the Biologically Active Inhibitor..... | 66        |
| 3.2.6. Low-Temperature Expression.....                                                           | 70        |
| 3.3. Periplasmic Expression of SCI in <i>E.coli</i> .....                                        | 75        |
| 3.3.1. Introduction.....                                                                         | 75        |
| 3.3.2. Construction of pRS-om101 .....                                                           | 76        |
| 3.3.3. Expression Studies Using the Periplasmic Expression Vector pRS-om101.....                 | 79        |
| 3.3.4. Purification of Secreted recSCI.....                                                      | 83        |
| 3.3.4.1. Extraction and Fractionation.....                                                       | 83        |
| 3.3.4.2. Ion-Exchange Chromatography (IEC).....                                                  | 84        |
| 3.3.4.3. Analytical Hydrophobic Interaction Chromatography (HIC).....                            | 86        |
| 3.3.5. Characterisation of recSCI.....                                                           | 87        |
| 3.3.5.1. N-Terminal Sequence.....                                                                | 87        |
| 3.3.5.2. Non-Reducing SDS-PAGE.....                                                              | 88        |
| 3.3.5.3. Molecular Weight Analysis by Electrospray Mass Spectrometry.....                        | 88        |
| 3.3.6. Refolding .....                                                                           | 90        |
| 3.4. Discussion.....                                                                             | 92        |

|                                                                                                                                            |            |
|--------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Chapter 4. Future Directions .....</b>                                                                                                  | <b>97</b>  |
| <b>Chapter 5. Materials and Methods .....</b>                                                                                              | <b>101</b> |
| 5.1. Instrumentation.....                                                                                                                  | 101        |
| 5.2. Materials .....                                                                                                                       | 102        |
| 5.2.1. Water.....                                                                                                                          | 102        |
| 5.2.2. Chemicals, Biochemicals, Enzymes and Proteins.....                                                                                  | 102        |
| 5.2.3. Phenol .....                                                                                                                        | 104        |
| 5.3. DNA Preparation and Analysis.....                                                                                                     | 104        |
| 5.3.1. Buffers .....                                                                                                                       | 104        |
| 5.3.2. Enzyme Reactions.....                                                                                                               | 105        |
| 5.3.3. Large Scale Preparation of Plasmid DNA from<br><i>E.coli</i> - Alkaline-Lysis Method.....                                           | 106        |
| 5.3.4. Modified Alkaline-Lysis Method for the<br>Purification of <i>E.coli</i> Plasmid DNA Suitable For<br>Double-Stranded Sequencing..... | 107        |
| 5.3.5. Small-Scale Preparation of <i>E.coli</i> Plasmid DNA<br>("Minipreps") - Boiling Method.....                                         | 108        |
| 5.3.6. PEG Precipitation .....                                                                                                             | 109        |
| 5.3.7. Ethanol Precipitation .....                                                                                                         | 109        |
| 5.3.8. Analytical Gel Electrophoresis of DNA on<br>Agarose Gels.....                                                                       | 110        |
| 5.3.9. Preparative Gel Electrophoresis of DNA in<br>Low-Melting-Point Agarose Gels .....                                                   | 110        |
| 5.3.10. Purification of Synthetic Oligonucleotides by<br>Denaturing Urea-Polyacrylamide Gel<br>Electrophoresis.....                        | 111        |
| 5.4. DNA Sequence Analysis .....                                                                                                           | 114        |
| 5.4.1. Denaturing of Double-Stranded Plasmids.....                                                                                         | 114        |
| 5.4.2. Sequencing Reactions .....                                                                                                          | 114        |
| 5.4.3. Annealing Templates and Primers.....                                                                                                | 115        |
| 5.4.4. Labelling Reaction .....                                                                                                            | 115        |
| 5.4.5. Termination.....                                                                                                                    | 116        |
| 5.4.6. Denaturing Gel Electrophoresis.....                                                                                                 | 116        |
| 5.5.1. Construction of pUCSCR- $\alpha$ .....                                                                                              | 118        |
| 5.5.2. Construction of pRS-SC56.....                                                                                                       | 119        |
| 5.5.3. Construction of pRS101.....                                                                                                         | 120        |
| 5.5.4. Construction of pRS-om101 .....                                                                                                     | 122        |
| 5.6. Protein Analysis and Preparation .....                                                                                                | 124        |
| 5.6.1. Preparation of Proteins from Yeast.....                                                                                             | 124        |

|                                                                                                                                   |     |
|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.6.2. SDS-Polyacrylamide Gel Electrophoresis<br>(SDS-PAGE) of Proteins.....                                                      | 124 |
| 5.6.3. Western Blotting - Electrophoretic Transfer of<br>Proteins From SDS-Polyacrylamide Gels To<br>Nitrocellulose Membrane..... | 126 |
| 5.6.4. Determination of Protein Concentrations.....                                                                               | 128 |
| 5.6.5. Colourimetric Determination of Proteolytic<br>Activity of Trypsin and $\beta$ XIIa.....                                    | 128 |
| 5.7. Microbiological Techniques.....                                                                                              | 130 |
| 5.7.1. Strains.....                                                                                                               | 130 |
| 5.7.2. Growth of Strains.....                                                                                                     | 131 |
| 5.7.3. Media for Growth of <i>E.coli</i> . ....                                                                                   | 131 |
| 5.7.3.1. Rich Media .....                                                                                                         | 131 |
| 5.7.3.2. Minimal Media .....                                                                                                      | 131 |
| 5.7.3.3. Antibiotics .....                                                                                                        | 132 |
| 5.7.4. Media for Growth of <i>Saccharomyces cerevisiae</i> .....                                                                  | 132 |
| 5.7.5. Transformation of <i>E.coli</i> -Calcium Chloride<br>Method.....                                                           | 134 |
| 5.7.5.1. Preparation.....                                                                                                         | 134 |
| 5.7.5.2. Transformation.....                                                                                                      | 135 |
| 5.7.6. Transformation of <i>E.coli</i> - Electroporation .....                                                                    | 135 |
| 5.7.7. Transformation of Yeast Spheroplasts.....                                                                                  | 136 |

## **Appendices** **References**

## **Chapter 1**

# **Introduction**

# 1. Introduction

**Sweet corn inhibitor (SCI)**, a small protein proteinase inhibitor ( $M_r$  12000) isolated from the seeds of opaque-2 corn [Hojima *et al.*, 1980; Ratnoff & Moneme, 1981] has been shown to be a potent and specific inhibitor of trypsin as well as the activated Hageman Factor ( $\beta$ XIIa), a serine proteinase which plays a major role in the contact activation of the intrinsic pathway of the human blood plasma coagulation process.

The inhibitory properties and the specificity of the inhibitor are determined by the primary sequence of the **reactive-site loop**, together with the conformational restraints provided by the remainder of the protein. The reactive-site loop, a structural feature common to many protein proteinase inhibitors (excluding  $\alpha_2$ -macroglobulins) runs along the surface of the protein so as to be recognised by the proteinase as a potential peptide substrate. It binds very tightly along the active-site cleft of the proteinase and is cleaved very slowly.  $K_a$  values are typically around  $10^7$  to  $10^{13}$   $M^{-1}$  [Read & James, 1986].

To date, all attempts to obtain structural information on the inhibitor and its complexes by way of X-ray crystallography and N.M.R. methods have failed due to the poor solubility properties ( $<1$  mg/ml) of the native SCI and the molecular heterogeneity of the isolated protein (M.P. Esnouf; personal communication). Instead, by establishing a suitable microbial expression system for the heterologous over-production of the recombinant SCI and subsequent mutagenic modification of the reactive-site loop sequence, the contribution of individual amino acid residues to the specificity of SCI could be mapped.

By way of introduction, pertinent aspects of proteinases and proteinase inhibition will be dealt with (Section 1.1). This is followed by a review of the various processes of the blood coagulation system and, in particular, the seemingly ubiquitous role of the Hageman Factor (Section 1.2). This section also contains a detailed treatment of both physiological and plant inhibitors of the proteolytically active factor (FXIIa) (Section 1.3). The reactive-site loop regions of some of the many FXIIa inhibitors are compared so as to extract any common features.

## 1.1. Proteinases and Proteinase Inhibitors

### 1.1.1. Nomenclature & Classification

Enzymes that degrade proteins by peptide bond hydrolysis have been termed **proteases**. Grassmann & Dyckerhoff (1928) initially suggested that proteases that acted on intact proteins were called **proteinases** and those that acted on oligopeptides were called **peptidases**. Bergmann & Fruton (1937) defined peptide bond hydrolases as peptidases and divided them into the **exo-** and **endopeptidases**, which hydrolyse terminal peptide bonds and central peptide bonds, respectively. The unified modern scheme as suggested by Barrett & McDonald (1986) (EC 3.4) refers to peptidases (or proteases) as all those enzymes that degrade proteins (or peptides) by peptide bond hydrolysis. These enzymes are subsequently divided into the endopeptidases (or proteinases) (EC 3.4.21-99) and the exopeptidases (EC 3.4.11-19) as previously defined by Bergmann & Fruton (1937).

Hartley (1960) observed that proteinases can be divided into four distinct classes according to their catalytic mechanism of peptide bond hydrolysis. These four classes are now termed the '**serine**', '**cysteine**', '**aspartic**' and '**metallo-** proteinases (EC 3.4.21-24, respectively). Precise crystallographic investigation of all four mechanistic classes of proteinases show that their catalytic sites lie in the **active-site cleft** that runs along the surface of the enzyme. The substrate oligopeptide chain lies along this cleft, bound on either side of the active site to **specificity subsites**. These subsites are adapted to recognise and bind to side-chains and the polypeptide backbone of the substrate. Figure 1.1 illustrates the accepted terminology as suggested by Berger & Schechter (1970). Subsites starting from the catalytic site of the proteinase towards the N-terminus of the substrate are labelled  $S_1$ ,  $S_2$ ,  $S_3$  etc. and the corresponding domains of the substrate,  $P_1$ ,  $P_2$ ,  $P_3$  etc. Similarly, the

proteinase subsites starting from the catalytic site towards the C-terminus of the substrate and the corresponding domains of the substrate are labelled  $S'_1$ ,  $S'_2$ ,  $S'_3$  etc and  $P'_1$ ,  $P'_2$ ,  $P'_3$ , etc., respectively.

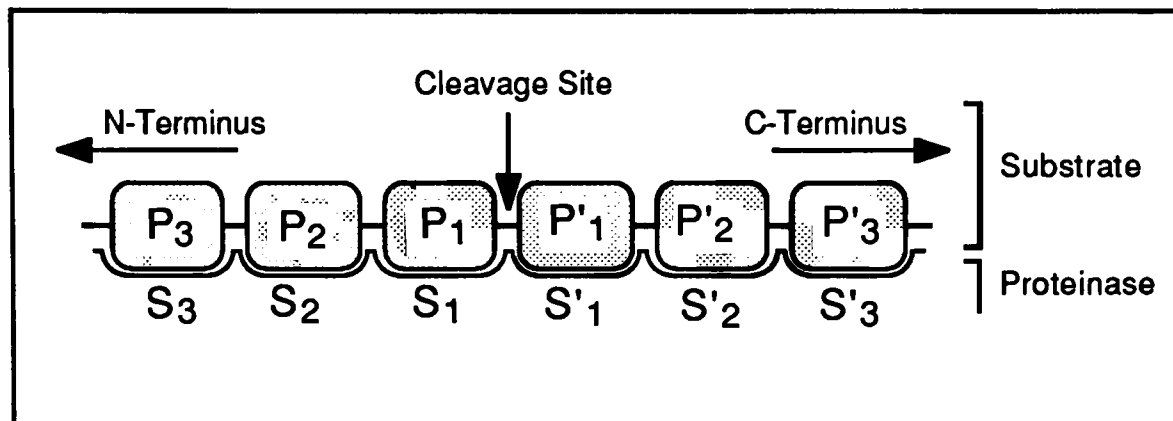


Figure 1.1: Active Site of Proteinases

As has been shown by several investigations, the  $S_1$  subsite is usually of prime importance in determining the specificity and function of the respective proteinase. In proteinases with a broader specificity such as those involved in protein degradation and digestion (e.g. trypsin, chymotrypsin and elastase), the sole requirement for recognition by the  $S_1$  subsite and subsequent cleavage of an oligopeptide substrate would be the  $P_1$  residue. For example, the digestive enzyme, trypsin cleaves all peptide bonds on the C-terminal side of the basic Arg and Lys residues [Read & James, 1986]. On the other hand, there are several proteinases involved in finer physiological regulatory processes that require the controlled activation of inactive zymogens by the recognition and cleavage of just one or a very small number of very specific sites along the entire sequence of a precursor protein. Examples of such regulatory functions are discussed in subsequent pages. With such a narrowing of specificity, the following subsites ( $S_2$ ,  $S_3$  etc.) become more discerning in that only those substrates with specific amino acid sequence motifs to the N-terminal side of the scissile peptide bond are recognised. Little is yet known of the importance of the  $S'_n$  subsites in the specificity of the proteinases though in certain cases the requirement of specific amino acid side-chains has been shown.

### 1.1.2. Serine Proteinases

The mechanistic class of serine proteinases is by far the most numerous and widespread of the proteinases and therefore the best studied with numerous examples to illustrate the very varied specificities and functional categories of this enzyme class.

There are two super-families of the serine proteinases that have formed as a result of convergent evolution [Neurath, 1984]. These are the **chymotrypsin** and the **subtilisin** super-families. Functionally, they include those enzymes involved in degradation and digestion of proteins and those involved with varied degrees of specialisation in physiological regulation. The subtilisin super-family is only found in bacteria whereas the chymotrypsin super-family is found in both prokaryotic and eukaryotic microorganisms, plants and invertebrate and vertebrate animals [Barrett, 1986]. The latter class which includes the digestive enzymes trypsin, chymotrypsin and elastase is characterised by less stringent requirements for productive interaction of the enzyme and substrate. For example, the sole requirement in chymotrypsin is a large hydrophobic P<sub>1</sub> side-chain (Phe, Trp or Tyr), whereas in trypsin, the P<sub>1</sub> residue should be basic (Arg and Lys). More stringent targeting of proteolytic activity is seen in the proteolytic activation of inactive precursor zymogens in processes such as blood coagulation and complement activation [Barrett, 1986], in hormone activation, secretory signal peptide cleavage, macromolecular assembly and the release of physiologically active peptides. [Neurath, 1984].

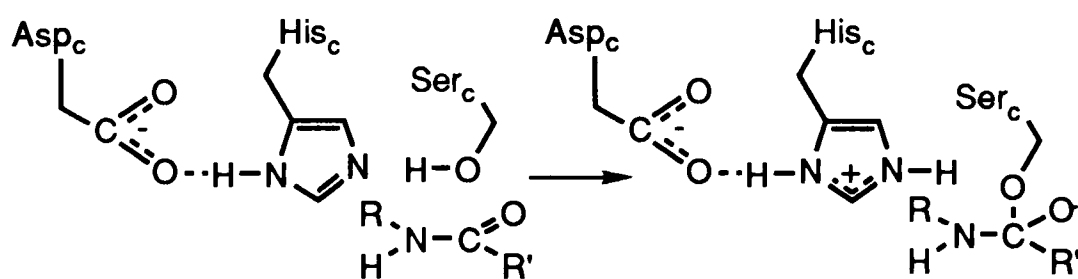
Most serine proteinases in higher animals are expressed as their enzymatically inactive zymogens. Activation is brought about by cleavage at a site 200-260 residues back from the C-terminus, resulting in the enzymatically active proteinase and the release of an activation peptide. These activation peptides are small in trypsinogen, chymotrypsinogen and the other digestive

enzymes but are large, relatively complex and confer new properties to the zymogens and enzymes of several of the proteins involved in blood coagulation and fibrinolysis [Barrett, 1986]. An example of the complexity of one such activation peptide of the blood coagulation cascade will be described in a subsequent section.

### 1.1.3. Mechanism of Serine Proteinases

In view of their widespread availability, physiological significance and relative simplicity, serine proteinases have been subject to extensive investigation as models for the determination of complex interactions involved in enzyme catalysis. The main premises of the transition state theory, suggested by Pauling (1946), are that enzymes, like other catalysts, function by reducing the activation barriers, that is, the activation free energies of the corresponding reactions. Warshel & co-workers (1989) using a combination of theoretical calculations and experimental information concluded that serine proteinases and other enzymes function by contributing electrostatic complementarity to the changes in charge distribution that occur during the reactions which they catalyse. Central to the catalytic machinery of serine proteinases is the "catalytic triad" composed of the three invariant residues Asp<sub>c</sub>-His<sub>c</sub>-Ser<sub>c</sub> (Asp<sub>102</sub>, His<sub>67</sub> and Ser<sub>195</sub> in chymotrypsin). These three residues are found adjacent to each other in the active-site of the proteinase, the Asp<sub>c</sub> being buried within the protein. This particular arrangement confers unusual nucleophilicity to the Ser<sub>c</sub> β-hydroxyl oxygen atom.

Catalysis of peptide bond hydrolysis by the catalytic triad involves a mechanism with a rate-limiting step defined as follows:



The transfer of the Ser<sub>c</sub> proton to the His<sub>c</sub> results in the formation of the imidazolium ion. This is markedly facilitated by the unsolvated carboxylate ion of Asp<sub>c</sub> which is H-bonded to the His<sub>c</sub> residue and so stabilises the positive charge of the imidazolium ion (see Figure 1.2). It is now accepted that the Asp<sub>c</sub> remains a carboxylate ion without abstracting a proton from the imidazolium ion to form an uncharged carboxylic acid group [Warshel *et al.*, 1989].

A second major contributory factor towards the catalytic power of serine proteinases is the stabilisation of the tetrahedral intermediate (the oxyanion) by the so-called "oxyanion hole". Conformational distortion that occurs upon the formation of the tetrahedral intermediate causes the carbonyl oxygen of the scissile peptide bond to move into the oxyanion hole resulting in hydrogen bonding to the -NH of Gly<sub>193</sub> and Ser<sub>195</sub> (chymotrypsin numbering). The serine proteinase therefore preferentially binds the tetrahedral intermediate and presumably therefore the transition state leading to it [Warshel *et al.*, 1989].

The driving force of serine proteinase catalysis can thus be summarised as the stabilisation of the {Asp<sub>c</sub><sup>-</sup>...His<sub>c</sub><sup>+</sup>...t<sup>-</sup>} transition-state by a "preorganised cryptate-like network of hydrogen-bonds" [Warshel *et al.*, 1989] (t<sup>-</sup> indicates the O-C-O<sup>-</sup> fragment of the tetrahedral intermediate).

With computer-aided modelling, Hahn & co-workers (1990) succeeded in the design of a 73-residue peptide "chymohelizyme-1" consisting of a bundle of four short parallel amphipathic helical peptides bearing the serine proteinase catalytic site triad at the amino end of the bundle in the same

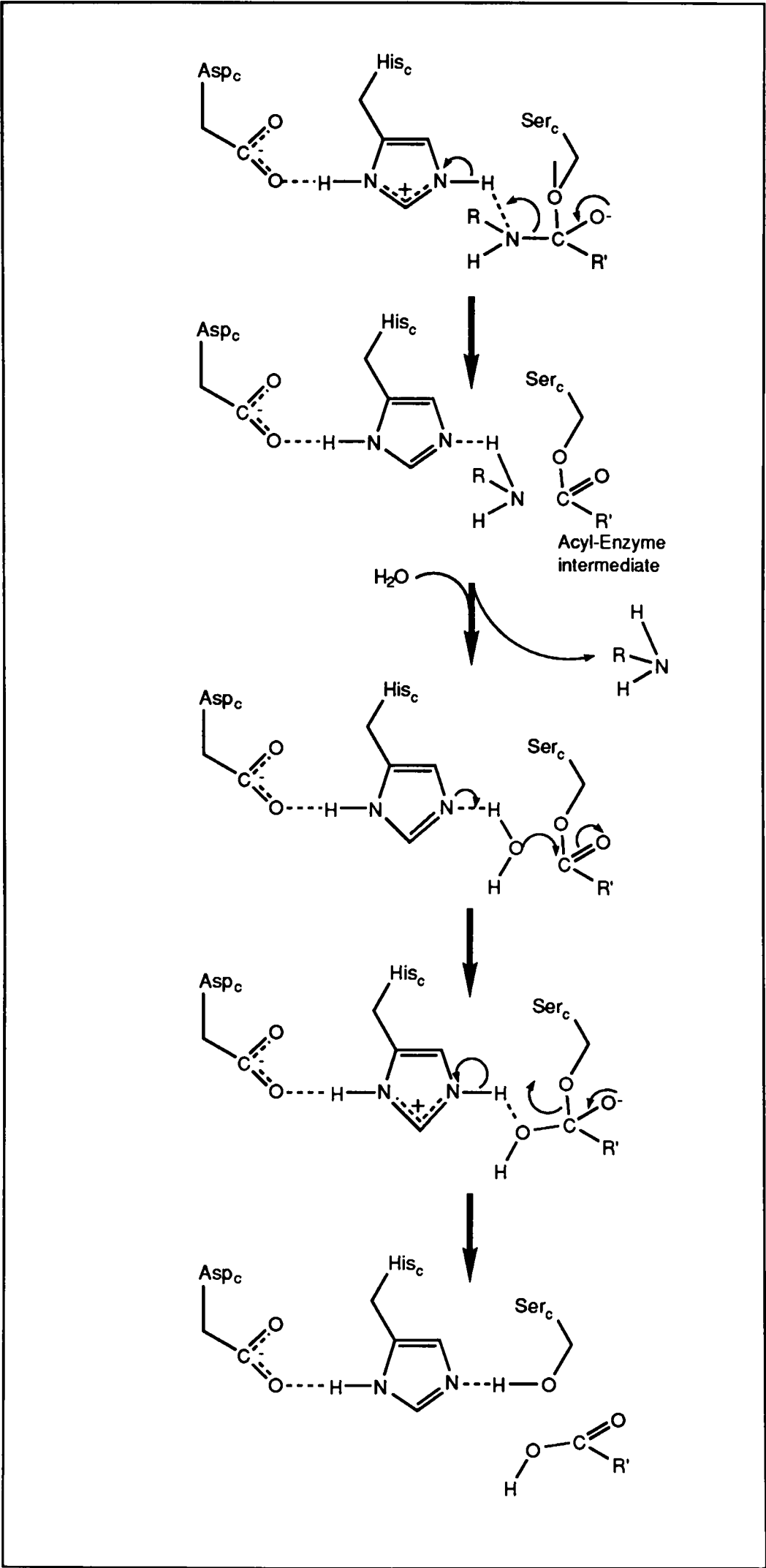


Figure 1.2: Mechanism of Serine Proteinases

spatial arrangement as in chymotrypsin. The catalytic triad of chymotrypsin was retained in its known spatial configuration having discarded the remaining protein structure of chymotrypsin. The supporting framework of three amphipathic helices was constructed so as to hold the three residues in their correct configuration. A fourth helix-forming chain with an N-terminal Glu residue was added so as to stabilise the entire structure and provide backbone hydrogen bonds to form the oxyanion hole. All four helices were covalently linked together at their C-termini through the side-chain amine groups of Orn and Lys residues. The correctly assembled chymohelizyme-1 was shown to measurably hydrolyse selected chymotrypsin substrates.

The substrate acetyltyrosine ester was hydrolysed by chymohelizyme-1 at the rate of catalysis  $k_{\text{cat}}$  of  $0.042 \text{ s}^{-1}$ , a value  $\approx 10^5$  times greater than spontaneous hydrolysis and about 0.03% that of chymotrypsin. It was inactivated by treatment with PMSF, an irreversible inhibitor (inactivator) of chymotrypsin and serine proteinases. This is possibly the first instance in which present knowledge and methods in the prediction and design of peptide structures have been used to successfully model and construct a synthetic polypeptide with catalytic properties.

#### **1.1.4. Inhibitors of Serine Proteinases**

Although proteinases are a physiological necessity, they are also potentially a great hazard to the organism. Uncontrolled proteolysis could lead to the destruction of essential protein components or the premature activation of zymogens. Examples of disease states involving serine proteinases include inflammation, rheumatoid arthritis, pulmonary emphysema and cerebral and coronary infarction resulting from thromboses. The importance of this hazard is reflected by the fact that nearly 10% of plasma proteins are proteinase inhibitors [Powers & Harper, 1986] constituting by weight, the third largest group of proteins in human plasma after albumin

and immunoglobulins [Travis & Salvesen, 1983]. Major plasma inhibitors include the serpins, the family of plasma serine proteinase inhibitors such as  $\alpha_1$ -proteinase inhibitor, antithrombin III, C1-inhibitor and  $\alpha_1$ -antichymotrypsin, all of which are strictly competitive inhibitors. The high levels of these plasma proteinase inhibitors are important in controlling blood clotting and other physiological processes that involve proteolytic cascade reactions such as hormone production and elimination and complement fixation [Laskowski Jr. & Kato, 1980]. Proteinase inhibitors in plants could serve to protect against parasite and insect infestation by inhibiting their degradatory proteinases.

Being proteins, the protein inhibitors of proteinases should in fact be substrates, not inhibitors of proteinases. Extensive studies to elucidate this paradox have focused mainly on the **reactive-site** of the protein inhibitor, a peptide bond that exists on the surface of the inhibitor molecule on the **reactive-site loop**, the primary contact region that projects out of the main framework of the protein [Powers & Harper, 1986]. This reactive-site loop which is a common feature in most protein proteinase inhibitors of known structure constitutes the relatively small portion of the inhibitor molecule that interacts with the enzyme in substrate-like manner [Laskowski Jr. & Kato, 1980]. The folding of the rest of the protein which provides the framework is dramatically different in the different inhibitor families.

Comparison of the geometry of the reactive-site loops of representatives of the various known families of serine proteinase inhibitors show that despite having significant variation in primary sequence they possess a canonical main-chain conformation, in particular, the region spanning  $P_3$  to  $P'_3$  [Laskowski Jr. & Kato, 1980; Read & James, 1986; Hubbard *et al.*, 1991]. The limited flexibility of the reactive-site loop as a function of the loop primary sequence and the conformational support provided by the

remainder of the protein is primarily responsible for the tight binding of the inhibitor to the enzyme ( $K_a \approx 10^7$ - $10^{13} \text{ M}^{-1}$ ). The relatively rigid conformation of the loop resembles essentially that of an "idealised substrate" with shape complementarity to the active site of the enzyme. This results in what could broadly be classed as a "lock and key" in that binding of the inhibitor by the enzyme is facilitated by salt-bridges, hydrogen-bond interactions and the burial of hydrophobic residues. The main-chain conformation is preserved although there are side-chain reorientations [Hubbard *et al.*, 1991]. In comparison, substrates for proteinases such as free peptides in solution or floppy external loops of globular proteins are considerably more "mobile" and require in most cases a gross change in conformation so as to bring the scissile peptide into an accessible position. Substrate binding is therefore thermodynamically less favourable as it is accompanied by a decrease in entropy by the loss of internal degrees of freedom. No such decrease in entropy is encountered in inhibitors with their comparatively rigid loops that are already complementary to the enzyme binding-site. Inhibitors therefore have a considerably higher binding energy than substrates with  $K_a$  values of the order of  $10^{10}$  to  $10^{13} \text{ M}^{-1}$  [Read & James, 1986].

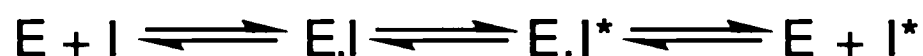
#### 1.1.5. The Standard Mechanism of Inhibition

In the standard mechanism of inhibition of serine proteinases by small protein inhibitors as proposed by Laskowski Jr. & Kato (1980), the inhibitors are hydrolysed very slowly and their cleavage does not proceed to completion.

The value for  $k_{\text{cat}}/K_m$  for the hydrolysis of the reactive-site peptide at neutral pH is very high, varying from  $10^4$ - $10^6 \text{ M}^{-1} \text{ s}^{-1}$  compared to a typical value of  $10^3 \text{ M}^{-1} \text{ s}^{-1}$  for a normal substrate [Laskowski Jr. & Kato, 1980].  $k_{\text{cat}}/K_m$  is used as a measure of the catalytic efficiency of an enzyme. Those enzymes having values of  $10^8$  to  $10^9 \text{ M}^{-1} \text{ s}^{-1}$  are said to have achieved a state of virtual catalytic perfection [Voet & Voet, 1990]. However, the values for

both the  $k_{\text{cat}}$  and  $K_{\text{m}}$  for inhibitors are many orders of magnitude lower than those of a substrate, resulting from very tight binding (low  $K_{\text{m}}$ ) and slow hydrolysis (low  $k_{\text{cat}}$ ) [Laskowski Jr. & Kato, 1980].

On interaction of an inhibitor with the enzyme, when equilibrium has been reached, most inhibitors remain functional. At neutral pH, the equilibrium can be described in simplified form as follows:



where, I and I\* are the virgin and modified inhibitor, with intact and cleaved reactive-sites, respectively. The equilibrium constant  $K_{\text{hyd}}$  between I and I\* is approximately unity in that the same stable complex is formed between the enzyme and the modified or virgin inhibitors [Laskowski Jr. & Kato, 1980]. However, the rate of formation of the former is much slower than that of the latter.

Conformational constraints on the regions around the reactive-site give them much less freedom to relax after hydrolysis as compared to the flexible regions of substrate cleavage sites which undergo almost total hydrolysis [Read & James, 1986]. The extensive involvement in  $\beta$ -sheet structure of the P' side of the reactive-site loop of inhibitors of known structure and, in many cases, a disulphide bridge close to the reactive-site on the P side are thought to contribute to the rigidity needed for minimising hydrolysis of the reactive-site [Read & James, 1986].

Another aspect of the inhibitor reactive-site structure considered important for its functionality is that, on binding, the reactive-site carbonyl adopted a conformation that was intermediate between that of a planar peptide and of the tetrahedral intermediate. The exact extent of distortion towards the  $\text{Ser}_c \text{O}^\beta$  has not been determined - but estimates show it to be too long for a normal covalent bond but shorter than van der Waals' contact.

Both reversible and irreversible low molecular weight synthetic inhibitors of serine proteinases have been reviewed in detail by Powers & Harper (1986). The most potent of the reversible inhibitors are the transition state analogues such as the peptide aldehydes and the peptide boronic acids. The minimal knowledge in the design of the inhibitors is the primary specificity of the enzyme, that is, the binding specificity of the  $S_1$  subsite. Knowledge of the extended specificity (subsites  $S_2$ - $S_n$ ) of the proteinase could enable one to design highly specific peptide inhibitors. Both peptide aldehydes and boronic acids bind to form tetrahedral adducts similar to those formed during peptide bond hydrolysis.

## **1.2. Blood Coagulation**

### **1.2.1. Introduction**

The complex series of sequential processes involved in the control of blood loss as a result of vascular injury is known as **haemostasis**. Haemostasis, triggered by blood coming into contact with damaged tissue, consists of four main processes, the completion of which depends on the extent of injury. Firstly, platelet adhesion results in the formation of a platelet plug where platelets become sticky and bind to endothelial tissue structure including the basement membrane and collagen fibres. Usually in minor injuries platelet plug formation is adequate to stem blood flow. Secondly, when platelets aggregate, they are induced to release several vasoactive amines (platelet release reaction) including serotonin (5-hydroxytryptamine) and thromboxane  $A_2$ . These cause vasoconstriction, thereby reducing blood flow. The third major process which will be discussed in this section is the

**blood coagulation cascade** which results in the formation of an insoluble fibrin clot. The fourth important effect is the activation of the fibrinolytic system which degrades the fibrin clot when healing and regeneration of the injured vessel occurs [Davie & Fujikawa, 1975].

Blood coagulation is the body's main defence against blood loss. An intricate and complex interaction of over twenty different liver-synthesised plasma glycoproteins (Figure 1.3) results in the final thrombin catalysed formation of an insoluble fibrin clot (thrombus). Essentially, the blood coagulation process consists of a bifurcated cascade of sequential proteolytic activations of inactive precursor zymogens [Davie & Ratnoff, 1964; MacFarlane, 1964]. These inactive zymogens, seven of which are precursors of serine proteinases, circulate in the blood plasma under normal physiological conditions where little or no intravascular coagulation occurs.

On coming into contact with damaged tissue or regions of impeded blood flow, essentially two possible pathways leading to clot formation can be initiated; the **intrinsic pathway**, where all the protein factors involved are present in the blood and the **extrinsic pathway**, where tissue factor, a cell membrane protein plays a critical role [Furie & Furie, 1988] (Figure 1.3). *In vitro*, both pathways can be functionally separated although *in vivo*, these processes are closely interrelated and function simultaneously [Davie & Fujikawa, 1975]. Both pathways merge at the proteolytic activation of Factor X (FX). The activated Factor X (FXa) goes on to initiate a common pathway which finally results in the thrombin catalysed cleavage of fibrinogen.

The initial cascade hypothesis Davie & Ratnoff (1964) and MacFarlane (1964) implied the step-wise amplification of a triggering signal leading to the formation of relatively massive quantities of fibrin which ensures the efficient arrest of blood flow. For example, moving down the extrinsic pathway (Figure 1.3), Factor VII (FVII), FX, prothrombin and fibrinogen are

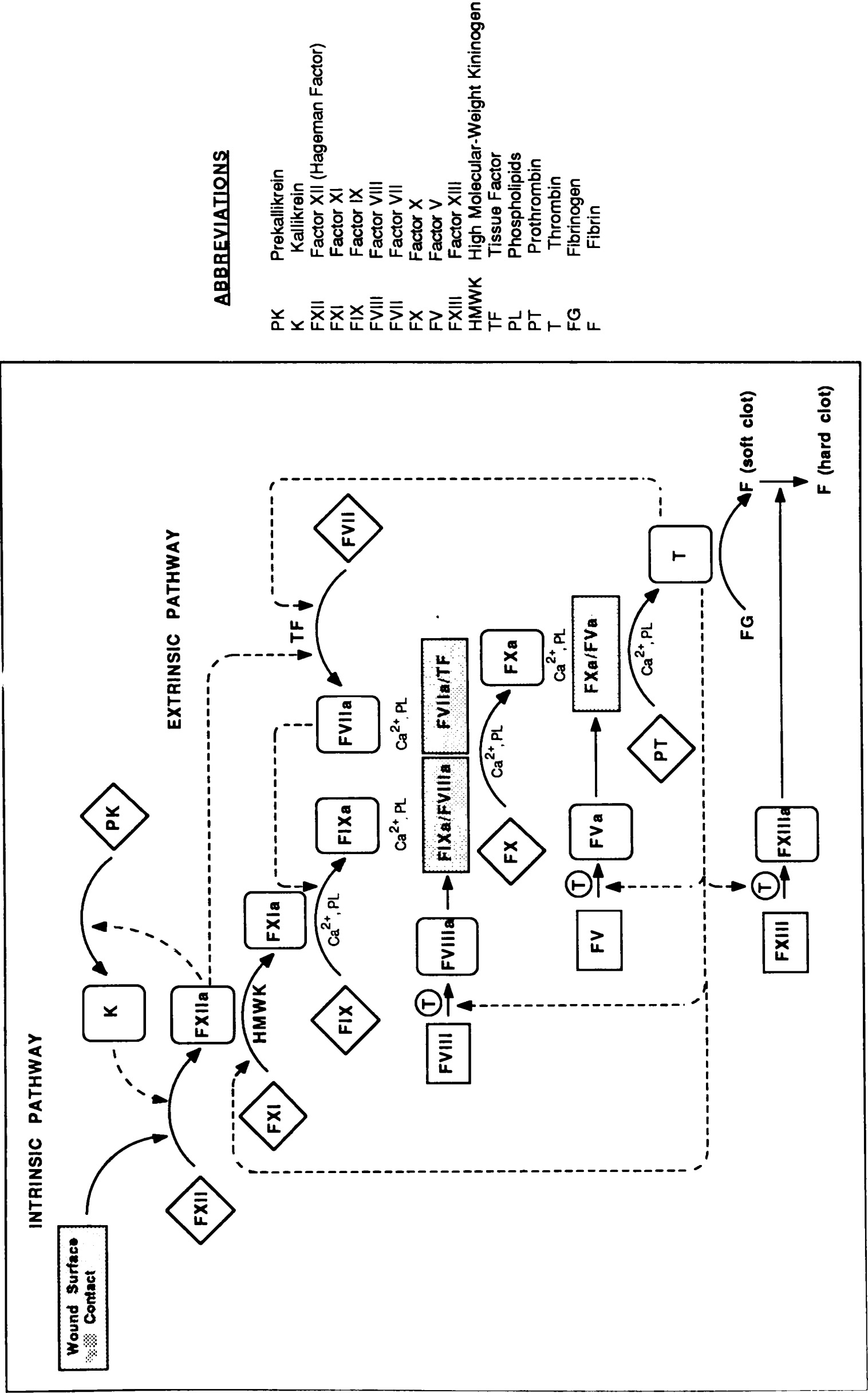


Figure 1.3: The Blood Coagulation Processes

present at plasma concentrations of <1, 8, 150 and up to 4000 µg per mL, respectively [Voet & Voet, 1990]. However, in the intervening time, studies have shown the involvement of an extremely complex network of positive feedback mechanisms (see Figure 1.3), self-regulation and a variety of factors that control clot growth. The stringent regulation of clotting activity is essential, as premature or abnormal clot formation can often have fatal or debilitating consequences. Strokes and heart attacks as a result of thromboses are two major causes of human death in industrialised nations.

The extrinsic pathway of blood coagulation is initiated as a result of vessel injury, involving the expression of tissue factor (TF) activity. TF, an integral membrane glycoprotein acts as an essential co-factor in a complex with activated Factor VII (FVIIa) which is necessary for the subsequent activation of FX. Unlike the other proteinase zymogens involved in blood coagulation, FVII shows substantial proteolytic activity but will not convert its substrate FX in the absence of TF or calcium ions. No specific information is yet available on the serine proteinase responsible for the activation cleavage of FVII in the extrinsic pathway but incubation with both thrombin and the activated Hageman Factor (FXIIa) of the intrinsic pathway caused FVII activation [Esnouf, 1984].

Of relevance to this work is the intrinsic pathway and, in particular, the activated Hageman Factor (FXIIa), a factor which plays a central role in the initiation of this pathway. Pioneering work by Lister in 1863 showed that blood, when exposed to glass, coagulated much faster than when enclosed in a blood vessel [Esnouf, 1984]. Ratnoff & Calopy (1955) reported a clotting defect in a Mr. Hageman in that his blood failed to clot on coming into contact with glass. This led to the discovery of the Hageman Factor (FXII) and the elucidation of the **contact activation system** of the intrinsic pathway (for detailed reviews, see Cochrane & Griffin (1982) and Bouma & Griffin (1986)).

The contact activation system is most effectively induced when blood plasma comes into contact with negatively charged surfaces including kaolin, glass, ellagic acid, dextran sulphates and cerebroside sulphatides. Of these, the cerebroside sulphatides were found to be the most effective. They are constituents of mammalian tissues and cell membranes and could therefore be important for the *in vivo* activation of the intrinsic pathway [Esnouf, 1984].

Together with the accessory high molecular-weight kininogen (HMWK; Fitzgerald Factor), FXIIa catalyses the *in vitro* proteolytic activation of Factor XI (FXI; Thromboplastin Antecedent) and FXIa in turn activates Factor IX (FIX; Christmas Factor). In the final step of the intrinsic pathway, FIX in association with a phospholipid membrane and the activated accessory Factor VIII (FVIIIa; anti-haemophilic factor) proteolytically cleaves FX leading to the formation of FXa. As yet, the precise physiological importance of the intrinsic pathway and FXII in particular is subject to debate as patients with congenital deficiencies in any of the component factors of contact activation, FXII, HMWK, or prekallikrein showed no major bleeding disorders [Revak *et al.*, 1977]. This suggests that a FXIIa-independent mechanism for the activation of FXI exists *in vivo* [Gailani & Broze Jr., 1991; Naito & Fujikawa, 1991]. In a revised model proposed by Gailani & Broze Jr. (1991), the thrombin produced through the initial action of FVIIa/TF in the extrinsic pathway (Figure 1.3) could be responsible for the activation of FXI. FXIa, possibly added to by autoactivation of additional FXI would sustain coagulation by activating more FIX [Gailani & Broze Jr., 1991].

In the common pathway that follows, prothrombin (PT) is converted to thrombin (T) in the presence of a lipoprotein complex containing FXa, FVa, calcium ions and phospholipids. The exact role played by FVa is not yet known although thrombin is not formed in its absence. The activated thrombin is then responsible for the conversion of the soluble fibrinogen to

the insoluble fibrin gel. For detailed reviews of the biochemistry involved in these basic mechanisms, see Davie & Fujikawa (1975), Esnouf (1982) and Furie & Furie (1988).

### 1.2.2. Factor XII and the Contact Activation System

FXII is a single-chain glycoprotein (16% carbohydrate) with a molecular weight of 74,000-80,000. The nucleotide sequence of human FXII cDNA showed that it encoded a protein of 596 amino acids [Cool *et al.*, 1985]. Sequence comparison revealed that the N-terminal residues 1-276, that is, the heavy-chain of FXII, have extensive sequence homology with both fibronectin and tissue plasminogen activator (tPA) which led to putative structure motifs for the protein being suggested [Furie & Furie, 1988]. Figure 1.4 gives a schematic representation of these domains. The type II homology region of fibronectin could by analogy to fibronectin be possibly responsible for FXII's ability to bind to collagen and anionic surfaces. Although their exact function is as yet unknown, both the the type I fibronectin domain and the kringle domain (prevalent in prothrombin, plasminogen and urokinase) could also be responsible for binding to fibrinogen and fibrin. This could be of importance to the function of FXII *in vivo* in that following activation of coagulation, FXII and  $\alpha$ FXIIa remain bound to the fibrin matrix and be available for the subsequent activation of the fibrinolytic pathway [Cool *et al.*, 1985]. Pixley & co-workers (1987) localised the region of the FXII heavy chain responsible for binding to negatively charged surfaces to a 20 residue portion of the region with fibronectin type I homology. Following an intervening Pro-rich connective region bearing no sequence homologies, the serine proteinase catalytic region forms the C-terminal light chain.

The FXII zymogen can be activated by selective proteolytic cleavage at two possible sites (see Figure 1.4); one within a disulphide loop (site A) generating a two-chain enzyme  $\alpha$ FXIIa consisting of disulphide linked

fragments of  $M_r$  52,000 and 28,000, the heavy and the light chains, respectively.  $\alpha$ FXIIa with the attached heavy chain, retains the capacity to bind to the negatively charged activation surfaces. Cleavage at a second site (site B) outside the disulphide loop results in the release of the carboxy-terminal light-chain ( $\beta$ XIIa) which retains serine proteolytic activity [Cochrane & Griffin, 1982]. The serine proteinase domain bears a high degree of homology to the corresponding regions of the other plasma serine proteinases and the digestive enzymes trypsin and chymotrypsin.

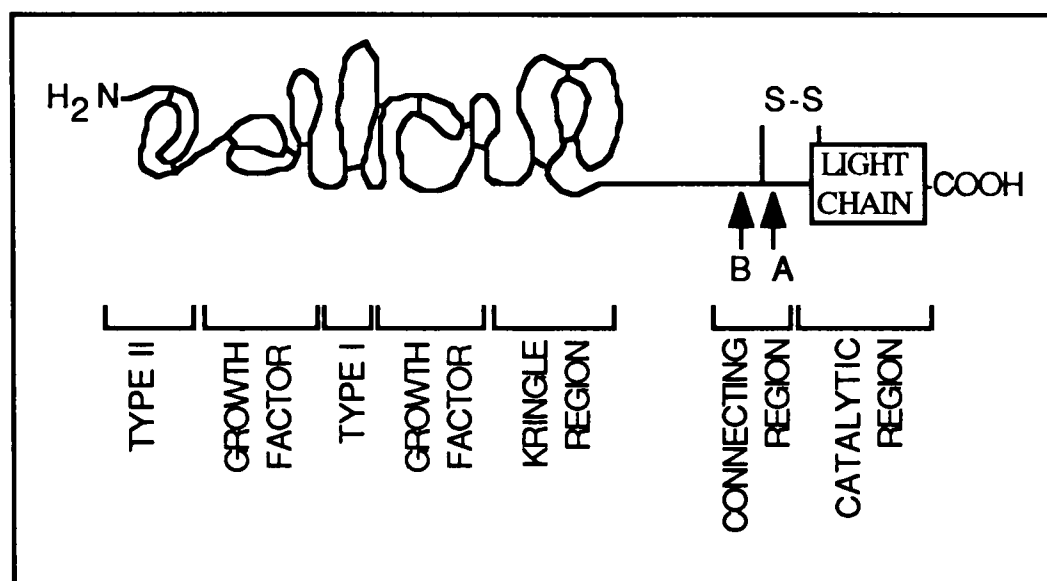


Figure 1.4: Domain Alignment of Factor XII

The exact mechanism of contact activation and the role played by FXII has been subject to intense investigation (see Bouma & Griffin (1986) and Nuijens *et al.* (1989) for references). Continued studies on FXII, which circulates for the major part with prekallikrein (PK; Fletcher Factor) as non-covalently associated bimolecular complexes with HMWK [Nuijens *et al.*, 1989] implicate its importance in a wider sense in that it plays a major role in all the plasma defence systems which additionally include the fibrinolytic, the kallikrein-kinin and the complement systems [Fuhrer *et al.*, 1990].

### 1.3. Activated Hageman Factor Inhibitors

#### 1.3.1. Plasma Inhibitors of the Activated Hageman Factor

Of the plasma proteinase inhibitors, the major inhibitor of both  $\alpha$ - and  $\beta$ -XIIa is the C1-inhibitor (C1-INH) which has been shown to contribute more than 90% to plasma inhibition of  $\alpha$ - and  $\beta$ -XIIa with a second-order rate constant of  $2.2 \times 10^{-5} \text{ M}^{-1} \cdot \text{min}^{-1}$ . Other weak inactivators of FXIIa include  $\alpha_2$ -antiplasmin,  $\alpha_2$ -macroglobulin and antithrombin III [Pixley *et al.*, 1985]. Both the  $\alpha$ - and  $\beta$ -XIIa form 1:1 stoichiometric complexes with all these inhibitors except  $\alpha_2$ -macroglobulin which can inhibit more than one  $\alpha$ -FXIIa molecule [Bouma & Griffin, 1986]. The main physiological roles of these inhibitors remain to be tested - inasmuch as the exact *in vivo* role and importance of FXII and the contact activation system yet remain to be pinpointed.

#### 1.3.2. Plant Inhibitors of the Activated Hageman Factor

Interestingly, a possible means of elucidating the function of FXII and the contact activation system could be provided by the wide variety of plant inhibitors that have been shown to inhibit enzymes of the contact activation system [Fuhrer *et al.*, 1990].

In the course of screening plant materials for serine proteinase inhibitors, several trypsin specific inhibitors have been isolated that, to varying degrees of specificity, inhibit human plasma proteinases. Particularly rich sources of inhibitors are certain plant seeds [Hojima *et al.*, 1980]. This reflects what is possibly a protective function against digestive enzymes of insects and animals. With no obvious physiological connection, several of these inhibitors, in particular, those isolated from cereal and squash seeds

have been found to be potent inhibitors of  $\beta$ XIIa, plasma kallikrein and the other plasma coagulation proteinases. In view of their potential therapeutic applications, such inhibitors are the subject of intense investigation. This is comparable to the wide variety of anti-coagulants and thrombolytic agents found in the saliva of blood-sucking leeches. Hirudin, a 65 amino acid, 7 kDa. polypeptide from the medicinal leech *Hirudo medicinalis* is a potent thrombin inhibitor and Antistasin, a 119 amino acid, 15 kDa. protein from *Haementeria spp.* inhibits FXa ( $K_i=0.4$  nM) [Sawyer, 1991].

### 1.3.3. Sweet Corn Inhibitor

The isolation of sweet corn inhibitor (SCI; popcorn inhibitor, corn hageman factor inhibitor) from the seeds of the high lysine corn cultivar, opaque-2 was first reported by Swartz & co-workers (1977) and estimated to have a molecular weight of 12,500 Da. Subsequent studies showed that, of the coagulation proteinases, **only** FXIIa was inhibited by SCI [Hojima *et al.*, 1980; Ratnoff & Moneme, 1981].

Mahoney & co-workers (1983) determined the primary sequence of SCI as having 112 amino acid residues with a calculated molecular weight of 12,028 Da. By cleavage on trypsin-agarose, the reactive-site peptide bond was identified as Arg<sub>36</sub>-Leu<sub>37</sub>. Extensive sequencing work (unpublished) done on SCI from the seeds of several strains of opaque-2 corn in the labs of M.P. Esnouf (Department of Clinical Biochemistry, University of Oxford) failed to entirely reproduce the sequence of SCI as reported by Mahoney & co-workers (1983). Most significantly, the corrected sequence of SCI provided by M.P. Esnouf (unpublished; see Figure 1.5) did not have the Pro-Arg repeat occupying the sites P<sub>4</sub>-P<sub>1</sub> immediately adjacent to the reactive-site. Barley trypsin inhibitor (BTI), a second representative of this new cereal inhibitor family which shares 54% sequence identity with SCI also has only one Pro-Arg unit adjacent to the reactive-site. Chong & Reeck (1987) attributed the



specificity of SCI to the Pro-Arg repeat in that the Pro residue provided an extra degree of conformational rigidity. This Pro-Arg repeat reported by Mahoney and co-workers (1983) could have been due to a technical error [Kodo, 1988].

The corrected sequence of SCI is illustrated in Figure 1.5. It consists of 110 amino acid residues and a calculated molecular weight of 11,812 Da. The reactive-site peptide is between Arg<sub>34</sub> and Leu<sub>35</sub>. Similar to most serine proteinase inhibitors, SCI has a high Cys content. All ten Cys residues are involved in disulphide pairings (see Figure 1.5). A feature of SCI is its three Trp residues (one in BTI) - Trp is usually absent in the published sequences of plant proteinase inhibitors [Mahoney *et al.*, 1983]. Preliminary structure analysis of SCI by far ultraviolet circular dichroism spectroscopy [Mahoney *et al.*, 1983] indicated that it differs in secondary structure from common proteinase inhibitors of known crystal structure in that it consists of more  $\alpha$ -helical- (about 40%) than  $\beta$ -sheet structure (about 20%). For example, the smaller ( $\approx$ 60 residues), disulphide rich class pancreatic trypsin inhibitor (PTI) (Kunitz) consists of 68%  $\beta$ -sheet and 23%  $\alpha$ -helices and the pancreatic secretory trypsin inhibitor (PSTI) (Kazal), equal  $\alpha$ -helical and  $\beta$ -sheet structure [Read & James, 1986]. The larger ( $\approx$ 110 residues) *Streptomyces* subtilisin inhibitor (SSI) and plasminostreptin (PS) have more extensive anti-parallel  $\beta$ -sheet structure (five strands), with two helices packing on one side of the sheet [Read & James, 1986]. The existence of an extensive hydrophobic core structure in these proteins is precluded due to their relative small sizes (60-110 residues) whereas, the abundance of disulphide pairings could provide the necessary structural stability to the protein [Read & James, 1986].

1.3.4.      **The Putative Reactive-Site Loop of SCI:  
Contribution of Primary Sequence to FXIIa  
Specificity?**

Work by Lei & Reeck (1987) showed that the interaction of SCI with  $\beta$ XIIa followed the standard mechanism of Laskowski Jr. & Kato (1980) and that both trypsin and  $\beta$ XIIa interact with the same site on SCI.

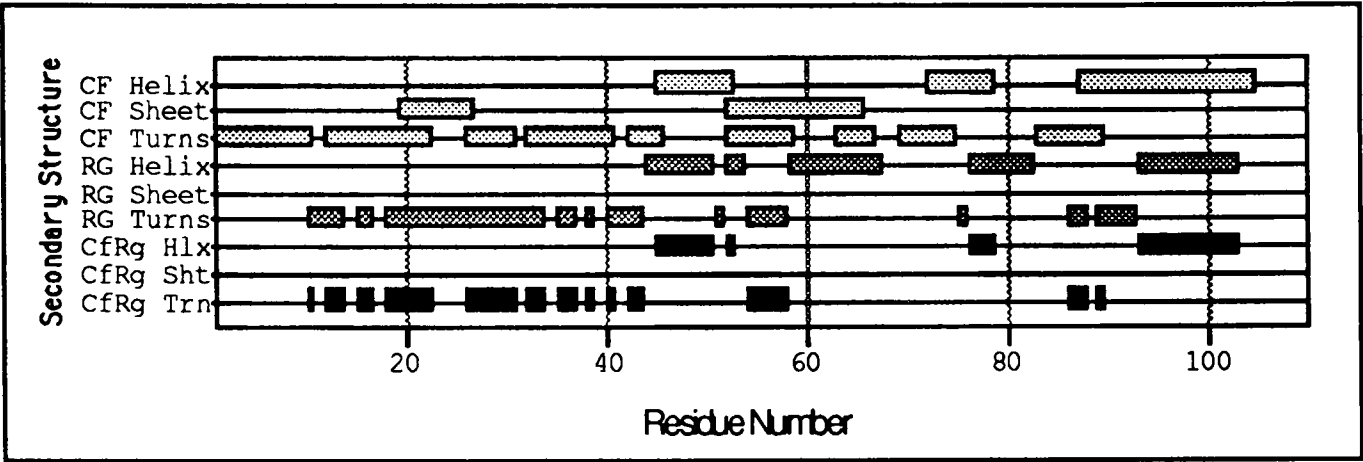


Figure 1.6: Secondary Structure Prediction of SCI

Figure 1.6 illustrates the predicted secondary structure that the different domains of the SCI sequence would probably adopt. Application of the methods according to Chou and Fasman (CF) [Chou & Fasman, 1978] and Robson and Garnier (RG) [Garnier *et al.*, 1978] was carried out with the help of a commercially available software application (MacVector, IBI Ltd., Hemel Hempstead, Herts). Both methods predict the predominance of  $\alpha$ -helical structure and, importantly, the regions flanking the reactive-site, domains of up to 8 residues on either side, are strongly predicted to adopt a  $\beta$ -turn conformation. This is shown by the combination of the results from the above methods (CfRg; Figure 1.6). Keeping in mind the limitations of these prediction methods, one could presume this region to correspond to the reactive-site loop. As previously described, the reactive-site loop binds tightly to the active-site of the cognate proteinase. In this section, the primary sequences of the reactive-site loops of some of the known inhibitors of  $\beta$ XIIa (Table 1.1) will be compared with the sequence of this putative reactive-site

|                 |                      |                      |                      |                      |                                                                                   |                       |                       |                       |                       |                       |                       |                       |                       |                       |                        |                                |  |  |
|-----------------|----------------------|----------------------|----------------------|----------------------|-----------------------------------------------------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|--------------------------------|--|--|
|                 |                      |                      |                      |                      |  |                       |                       |                       |                       |                       |                       |                       |                       |                       |                        |                                |  |  |
|                 | <b>P<sub>5</sub></b> | <b>P<sub>4</sub></b> | <b>P<sub>3</sub></b> | <b>P<sub>2</sub></b> | <b>P<sub>1</sub></b>                                                              | <b>P'<sub>1</sub></b> | <b>P'<sub>2</sub></b> | <b>P'<sub>3</sub></b> | <b>P'<sub>4</sub></b> | <b>P'<sub>5</sub></b> | <b>P'<sub>6</sub></b> | <b>P'<sub>7</sub></b> | <b>P'<sub>8</sub></b> | <b>P'<sub>9</sub></b> | <b>P'<sub>10</sub></b> |                                |  |  |
| <b>SCI</b>      | Gly                  | Ile                  | Gly                  | Pro                  | <b>Arg</b>                                                                        | Leu                   | Pro                   | Trp                   | Pro                   | Glu                   | Leu                   | Lys                   | Arg                   | Arg                   | Cys                    | <b>Cereal<br/>Inhibitors</b>   |  |  |
| <b>BTI</b>      | His                  | Gln                  | Gly                  | Pro                  | <b>Arg</b>                                                                        | Leu                   | Leu                   | Thr                   | Ser                   | Asp                   | Met                   | Lys                   | Arg                   | Arg                   | Cys                    |                                |  |  |
| <b>C1-INH</b>   | Ile                  | Ser                  | Val                  | Ala                  | <b>Arg</b>                                                                        | Thr                   | Leu                   | Leu                   | Val                   | Phe                   | Glu                   | Val                   | Gln                   | Gln                   | Pro                    | <b>Plasma<br/>Serpins</b>      |  |  |
| <b>AT-P</b>     | Gln                  | Ala                  | Ile                  | Pro                  | <b>Arg</b>                                                                        | Ser                   | Ile                   | Pro                   | Pro                   | Glu                   | Val                   | Lys                   | Phe                   | Asn                   | Lys                    |                                |  |  |
| <b>CMTI-I</b>   | Arg                  | Val                  | Cys                  | Pro                  | <b>Arg</b>                                                                        | Ile                   | Leu                   | Met                   | Glu                   | Cys                   | Lys                   | Lys                   | Asp                   | Ser                   | Asp                    | <b>Squash<br/>Inhibitors</b>   |  |  |
| <b>CMTI-III</b> | Arg                  | Val                  | Cys                  | Pro                  | <b>Arg</b>                                                                        | Ile                   | Leu                   | Met                   | Lys                   | Cys                   | Lys                   | Lys                   | Asp                   | Ser                   | Asp                    |                                |  |  |
| <b>CPTI-II</b>  | Arg                  | Val                  | Cys                  | Pro                  | <b>Lys</b>                                                                        | Ile                   | Leu                   | Met                   | Glu                   | Cys                   | Lys                   | Lys                   | Asp                   | Ser                   | Asp                    |                                |  |  |
| <b>CPTI-III</b> | Arg                  | Val                  | Cys                  | Pro                  | <b>Lys</b>                                                                        | Ile                   | Leu                   | Met                   | Glu                   | Cys                   | Lys                   | Lys                   | Asp                   | Ser                   | Asp                    |                                |  |  |
| <b>BDTI-II</b>  | Arg                  | Gly                  | Cys                  | Pro                  | <b>Arg</b>                                                                        | Ile                   | Leu                   | Met                   | Arg                   | Cys                   | Lys                   | Arg                   | Asp                   | Ser                   | Asp                    | <b>Potato I<br/>Inhibitors</b> |  |  |
| <b>CMTI-V</b>   | Thr                  | Pro                  | Val                  | Thr                  | <b>Lys</b>                                                                        | Asp                   | Phe                   | Arg                   | Cys                   | Asn                   | Arg                   | Val                   | Arg                   | Ile                   | Trp                    |                                |  |  |
| <b>BPTI</b>     | Thr                  | Gly                  | Pro                  | Cys                  | <b>Lys</b>                                                                        | Ala                   | Arg                   | Ile                   | Ile                   | Arg                   | Tyr                   | Phe                   | Tyr                   | Asn                   | Ala                    | <b>PTI<br/>(Kunitz)</b>        |  |  |
|                 | <b>P<sub>5</sub></b> | <b>P<sub>4</sub></b> | <b>P<sub>3</sub></b> | <b>P<sub>2</sub></b> | <b>P<sub>1</sub></b>                                                              | <b>P'<sub>1</sub></b> | <b>P'<sub>2</sub></b> | <b>P'<sub>3</sub></b> | <b>P'<sub>4</sub></b> | <b>P'<sub>5</sub></b> | <b>P'<sub>6</sub></b> | <b>P'<sub>7</sub></b> | <b>P'<sub>8</sub></b> | <b>P'<sub>9</sub></b> | <b>P'<sub>10</sub></b> |                                |  |  |

**Table 1.1**  
 Alignment of Reactive-Site Loop Primary Sequences of β Xlla Inhibitors  
 (Arrow indicates reactive-site; BPTI is included for comparison)

loop of SCI in the hope that amino acid residues responsible for the specificity of its interaction with  $\beta$ XIIa might be revealed.

The reactive-site residue,  $P_1$ , is the primary determinant of specificity of the inhibitor. Inhibitors with Arg at  $P_1$ , which includes SCI, together with those with Lys at  $P_1$  tend to inhibit trypsin and trypsin-like enzymes [Laskowski Jr. & Kato, 1980]. The importance of Arg at  $P_1$  for specificity towards  $\beta$ XIIa is illustrated by  $\alpha_1$ -antitrypsin Pittsburgh (AT-P), a variant of the plasma inhibitor,  $\alpha_1$ -antitrypsin ( $\alpha_1$ -AT) that resulted from a point mutation resulting in the Met<sub>358</sub> at  $P_1$  being replaced by Arg [Scott *et al.*, 1985]. This defect causes the molecule to have greatly diminished activity towards its cognate enzyme, neutrophil elastase, but greatly increased activity towards thrombin and the Arg-specific enzymes of the contact system including  $\beta$ FXIIa, kallikrein and FXIa [Scott *et al.*, 1985].  $\alpha_1$ -AT predictably showed no discernible inhibitory activity towards  $\beta$ FXIIa, whereas, AT-P exhibited a second-order inactivation rate constant of  $2.5 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  [Scott *et al.*, 1985]. This represented a nearly 9-fold increase over C1-INH ( $3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ ), the natural plasma inhibitor of  $\beta$ FXIIa [Pixley *et al.*, 1985]. However, a further recombinant mutant,  $\alpha_1$ -AT Ala<sub>357</sub> Arg<sub>358</sub>, with the additional  $P_2$  Pro<sub>357</sub>  $\rightarrow$  Ala mutation (as in C1-INH) was less efficient ( $9.2 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ ) than AT-P, but yet, better than C1-INH [Schapira *et al.*, 1987]. This confirms the positive contributory influence of a Pro residue at the  $P_2$ ,  $P'_3$  and  $P'_4$  positions towards inhibition.

Table 1.1 also lists inhibitors of the squash family, a group of extremely small (27-30 residues) but potent protein proteinase inhibitors isolated from various cucurbitaceae plants [Hojima *et al.*, 1982; Wieczorek *et al.*, 1985; Wynn & Laskowski Jr., 1990] including squash (*Cucurbita maxima*), summer squash (*Cucurbita pepo*), cucumber (*Cucumis sativus*) and red bryony (*Bryonica elioica*). Wieczorek and co-workers (1985) isolated and identified 3 iso-

inhibitors from *C. maxima* (Mr.≈3000) and termed them CMTI-I, CMTI-III and CMTI-IV. They have 3 disulphide bridges and are believed to be the smallest known globular trypsin inhibitors. The X-ray crystal structure of CMTI-I confirms the presence of an N-terminal reactive-site loop in this family [Bode *et al.*, 1989]. A larger squash trypsin inhibitor, CMTI-V (Mr 7106), was isolated by Krishnamoorthi & co-workers (1990) and was shown to have a high degree of homology to the Potato I inhibitor family [Laskowski Jr. & Kato, 1980]. CMTI-V inhibited  $\beta$ XIIa but not kallikrein.

Comparison of the reactive-site sequences of the  $\beta$ XIIa inhibitors listed in table 1.1 shows no striking sequence motifs that are common to all the inhibitors, notwithstanding the expected sequence homologies that exist within the individual families. The P<sub>1</sub> position can be occupied by either Arg or Lys, and the P<sub>2</sub> residue is usually a Pro. As a general rule, residues P<sub>2</sub>-P<sub>4</sub> and P'<sub>1</sub>-P'<sub>3</sub> are hydrophobic and neutral, consistent with their burial in the enzyme-inhibitor interface. Unlike all the other inhibitors compared in Table 1.1, in SCI, P' hydrophobic residues extend right through to P'<sub>6</sub>, their internalisation possibly contributing to the binding free energy. The importance of hydrophobicity as a major factor in stabilising enzyme-inhibitor complexes was shown by Janin & Chotia (1976). They determined the hydrophobic contribution towards the free energy of dissociation by calculating in the known X-ray crystal structures of trypsin complexed with STI and BPTI, the protein surface area buried upon binding. Of greatest importance are the residues that make up the surface areas that are buried upon enzyme-inhibitor binding [Janin & Chotia, 1976]. A hydrophobic, non-polar residue at P'<sub>2</sub> could probably be of importance as trypsin inhibitors with charged residues at this site (e.g. BPTI and STI with Arg [Read & James, 1986]) show no activity towards  $\beta$ XIIa. Ramos (1991) in studies with a computer model of  $\beta$ XIIa based on the known crystal structure of trypsin, suggested that Tyr<sub>39</sub> in the trypsin active-site responsible for hydrogen bonding to the P'<sub>4</sub> Ile

of BPTI is replaced by a Gly in  $\beta$ XIIa. Ramos (1991) suggests this to be the reason why BPTI does not bind to  $\beta$ XIIa in that there are no sites for hydrogen bonding to the P'<sub>4</sub> Ile of BPTI.

The prevalence of Pro residues in the SCI loop (3) as compared to BTI (1) is possibly the main contributory factor towards the greater binding efficiency of the former towards  $\beta$ XIIa; the dissociation constant ( $K_i$ ) for SCI with  $\beta$ XIIa ( $2.1 \times 10^{-9}$  M) being 2 orders of magnitude higher than that for BTI ( $1.1 \times 10^{-7}$  M). This positive effect was seen in AT-P over C1-INH (see above). Furthermore, the hydrophobic residues Trp and Pro at P'<sub>3</sub> and P'<sub>4</sub>, respectively in SCI are replaced in BTI by the polar residues, Thr and Ser, respectively\*.

Wynn & Laskowski Jr. (1990) compared the inhibition of  $\beta$ XIIa by the various squash family inhibitors. Within the family, the most powerful are CMTI-III and BDTI-II (*Bryonia elioica* trypsin inhibitor). Wynn & Laskowski Jr. (1990) attributed this to their positive P'<sub>4</sub> residues, Lys and Arg, respectively. CMTI-I, having the identical sequence of CMTI-III but for the positive Arg P'<sub>4</sub> substituted by the negatively charged Asp residue had a smaller (1/64th) association equilibrium constant ( $K_a$ ). CMTI-V, however, seems to flaunt any tentative rules one could have made of the primary sequence determinants for  $\beta$ FXIIa specificity. Unexpectedly, unlike the other inhibitors listed, CMTI-V, being one of the most potent inhibitors, has charged polar residues in the immediate vicinity of the reactive-site, that is, Asp at P'<sub>1</sub> and Arg at P'<sub>3</sub>. In this case, the loss of hydrophobic interaction could possibly be countered by ionic interaction and hydrogen bonding.

As to the exact nature of interaction between  $\beta$ XIIa and its inhibitors, there is no X-ray structure information available. Ramos (1991) reports an

---

\* On the hydrophobicity scale as suggested by Rose and co-workers (1985), residue hydrophobicity is correlated with the average area buried upon folding, calculated for 4410 residues taken from monomers of 23 X-ray elucidated proteins.

improved three-dimensional model of  $\beta$ XIIa which was built based on sequence homology with the well-known pancreatic serine proteinases, bovine trypsin, chymotrypsin and elastase. The energy-minimised three-dimensional model was determined, based on the known X-ray crystal structure of bovine pancreatic trypsin. Further information of the interactions of  $\beta$ XIIa with its inhibitors could, with caution, be obtained by computer-aided studies of putative complexes between this  $\beta$ XIIa model and specific inhibitors of known crystal structure. To date the crystal structure of only one of the listed inhibitors of  $\beta$ XIIa, namely, CMTI-I, has been determined [Bode *et al.*, 1989]. The options in this direction for obtaining conclusive clues remain limited, as long as the exact crystal structure of  $\beta$ XIIa remains to be determined.

However, with the advent of recombinant DNA technology, one has a very powerful tool by which one can, having established suitable microbial over-expression of the active, recombinant SCI (recSCI), obtain with relative ease and rapidity targeted variants of the inhibitor by the various mutagenetic methods available. Selected reactive-site loop mutations on SCI involving single and multiple residue substitutions based on the information gleaned from the study of the other  $\beta$ XIIa inhibitors and from molecular modeling could give useful insight into the contribution of the individual amino acid residues to the strength and specificity of inhibition of  $\beta$ XIIa by SCI. From the information obtained it should then be possible to further optimise the interaction and in a long-term perspective develop chimaeric inhibitor polypeptides for potential clinical application.

The aim of this thesis is to develop a suitable microbial system for the overexpression of recSCI as a system to provide a ready source of the wild-type and mutant inhibitors.

## **Chapter 2**

# **Expression of a Truncated SCI gene in *Saccharomyces cerevisiae*.**

## **2. The Expression of a Truncated SCI Gene in *Saccharomyces cerevisiae***

### **2.1. Introduction**

The work described in this chapter was a continuation of that described by Y. Kodo (1988) in which he designed and synthesised the SCI gene by automated oligonucleotide synthesis. The gene was designed using yeast-preferred codons [Ikemura & Ozeki, 1982; Sharp & Li, 1987]. Several restriction endonuclease recognition sites were incorporated along its length (Figure 2.1) so as to provide a convenient means for subsequent mutagenesis experiments. The synthetic gene was inserted into a *Saccharomyces cerevisiae* (hereafter, "yeast") expression/secretion plasmid that was previously used to successfully secrete up to 0.5 mg/L medium of active, correctly folded hen egg-white lysozyme (S.K.Field, unpublished results).

#### **2.1.1. Heterologous Expression in *Saccharomyces cerevisiae***

The subject of heterologous expression in yeast has been extensively reviewed [Kingsman *et al.*, 1985; Valenzuela, 1988]. A prerequisite for establishing a yeast expression system is the construction of a suitable expression vector containing the recombinant DNA sequence and those homologous DNA sequences necessary for the proper maintenance and propagation of the vector in the host cells. For the expression of the recombinant protein, sequences such as promoters, enhancers, terminators and upstream regulatory sequences are required [Valenzuela, 1988]. Selection



of a suitable auxotrophic yeast host strain enables one to screen the successful insertion of the expression vector bearing the defective gene as a marker. The *leu2-d* gene expressing 3-isopropylmalate dehydrogenase is commonly used as a selectable marker in *leu2* minus strains [Erhart & Hollenberg, 1983].

### 2.1.2. The 2-Micron Circle

Most heterologous expression vectors for yeast are based on the **2-micron circle**. This extra-chromosomal plasmid is unique amongst eukaryotes in that it is of high copy-number and replicates autonomously. It is a double-stranded, circular plasmid of 6318 b.p. present in most yeast strains at a density of 20-100 copies per haploid genome [Broach, 1982]. Autonomous replication of the 2 $\mu$  circle is conferred by its origin of replication (*ori*), an approximately 350 b.p. region on the plasmid. The *REP1* and *REP2* genes and the *REP3* locus on the 2 $\mu$  circle are responsible for maintaining high copy-number [Broach, 1982].

The vector used for the expression of SCI in yeast was based on the vector pMA3a [Dobson *et al.*, 1982] (Appendix I). This is a **shuttle-vector** in that it contains a combination of DNA elements derived from both *E.coli* and yeast, facilitating replication and selection in both organisms. One can therefore carry out all recombinant and preparative procedures using the powerful *E.coli* technology [Kingsman *et al.*, 1985]. Essentially, pMA3a (Appendix I) consists of the *E.coli* pBR322 plasmid [Bolivar *et al.*, 1977] into which are incorporated the 2 $\mu$  circle *ori*, the *REP3* locus and the *leu2-d* selectable marker. Insertion of pMA3a into a yeast strain containing an endogenous 2 $\mu$ -plasmid (a [*cir*<sup>+</sup>] strain) enables it to replicate (copy number 50-150) using the replication proteins encoded by *REP1* and *REP2*. It is therefore stably maintained by the host strain [Kingsman *et al.*, 1985].

### 2.1.3. The Expression Cassette

For expression of a heterologous protein, a so-called "expression cassette" is inserted into the pMA3a vector. This expression cassette contains the heterologous gene as well as the appropriate homologous yeast regulatory sequences required for the correct, in-frame recognition and transcription of the heterologous gene. These include an upstream promoter region and downstream transcription termination signals. The latter prevent the formation of long transcripts that are unstable and result in low expression levels [Kingsman *et al.*, 1985]. Several yeast promoters have been successfully used in heterologous gene expression. These usually include promoters of genes of proteins that are expressed in relative abundance such as glycolytic enzymes like PGK (phosphoglycerate kinase) [Dobson *et al.*, 1982] or GAPDH (glyceraldehyde 3-phosphate dehydrogenase) [Hitzeman *et al.*, 1983; Valenzuela, 1988]. Yeast promoters are less well defined and are considerably larger, spanning at least 500 b.p. as compared to the 70-100 b.p. of *E.coli* promoters (discussed in Chapter 3). Although greater variation exists in yeast promoters, several common motifs have been noted [see Kingsman *et al.*, 1985].

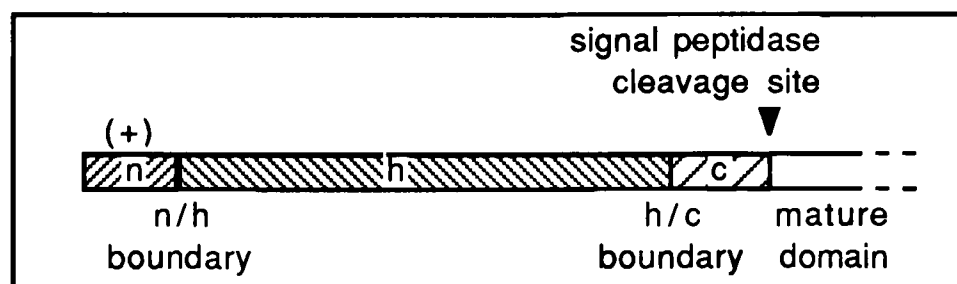
### 2.1.4. Regulation of Heterologous Expression

The potentially deleterious effect of the expressed recombinant protein on the host cells can be circumvented by mechanisms enabling one to, at varying degrees of stringency, repress recombinant protein production during fermentation. This facilitates the healthy growth of the transformed yeast. To date, the most efficient means of regulating expression is at the transcription level by the use of controllable homologous promoters. The most commonly used systems utilised include the PGK-, PHO5- or the MF $\alpha$ 1 promoters. The PGK promoter [Hitzeman *et al.*, 1983] expresses at greatly reduced levels in the presence of non-glycolytic carbon sources (e.g. acetate) and at normal levels in

the presence of glucose. The PHO5 (acid phosphatase) promoter [Kálmán *et al.*, 1990] is repressed at high phosphate levels. The promoter of the yeast MF $\alpha$ 1 gene, a major structural gene for the mating pheromone  $\alpha$ -factor, can be regulated by temperature change in a temperature sensitive *SIR3* strain of yeast [Brake *et al.*, 1984].

### 2.1.5. Protein Secretion in Yeast

Proteins destined to be secreted into the non-cytosolic spaces of both eukaryotes and prokaryotes are initially synthesised at the ribosome as precursors with an additional N-terminal extension known as the **signal peptide** (Figure 2.2). The signal sequence is the only intragenic information that is truly essential for protein secretion and is sufficient in some cases to allow the export of an otherwise cytoplasmic protein [Bieker & Silhavy, 1990]. Signal peptides vary from 15 to 50 amino acids in length and are found in both prokaryotes and eukaryotes. They show little identity in primary structure but can be divided into three structurally and functionally distinct domains: a basic N-terminal region (n), a central hydrophobic region (h) and a more polar C-terminal region (c) defining the cleavage site [Nagohara *et al.*, 1988].



**Figure 2.2:** Structure of eukaryotic signal peptides:  
The n, h and c regions typically basic,  
hydrophobic and polar residues, respectively.

In eukaryotes, the nascent protein is vectorially threaded across the hydrophobic rER membrane into the ER lumen. The signal peptide is cleaved by a signal peptidase on the luminal side of the rER membrane even before the completion of translation. (for a detailed review, see Notwehr & Gordon

(1990)). After translocation into the ER, the proteins destined to be secreted into the extracellular space move through the intra-cellular organelles of the secretion pathway (ER, golgi- and secretory vesicles) by sequential budding and fusion events.

The secretion of heterologous proteins in yeast has been accomplished in several cases by the expression of a DNA sequence encoding the protein fused with an N-terminal signal peptide. Signal peptides from various sources, both prokaryotic and eukaryotic, have been successfully used (see Table 2.1 for a few examples).

### 2.1.6. The $\alpha$ -Factor

Yeast is capable of existing in three distinct cell types: two haploid types designated *a* and  $\alpha$  which fuse to form an *a*/ $\alpha$  diploid cell type [Julius *et al.*, 1984]. The process of *a* and  $\alpha$  mating is mediated by the reciprocal exchange of diffusible peptide pheromones called *a*- and  $\alpha$ -factor, respectively.

The  $\alpha$ -factor pheromone released by  $\alpha$ -cells into the medium is a short peptide 13 residues long, expressed as a precursor, 165 residues long. This precursor can be divided into three parts (Figure 2.3):

- a) A pre-region consisting of a hydrophobic 19 residue peptide,
- b) A 66 residue hydrophilic pro-segment containing four consensus glycosylation sites
- and,
- c) A C-terminal region consisting of tandem repeats of the mature  $\alpha$ -factor. The mature  $\alpha$ -factor repeats are separated from the prepro region and each other by spacer peptides [Kurjan & Herskowitz, 1982] (Figure 2.3).

The pre-region signal peptide mediates the entry of the precursor prepro- $\alpha$ -factor (pp $\alpha$ F) into the secretory pathway by translocation into the

| Heterologous Protein                                         | Signal Sequence                                    | Reference                      |
|--------------------------------------------------------------|----------------------------------------------------|--------------------------------|
| Human Lysozyme                                               | Chicken Lysozyme                                   | Jigami <i>et al.</i> , 1986    |
| Plasminogen Activator Inhibitor Type I (PAI-I)<br>Echistatin | <i>S. cerevisiae</i><br>$\alpha$ -galactosidase    | Hofmann & Schultz, 1991        |
| Human Insulin-Like Growth Factor (hIGFI)                     | <i>S. cerevisiae</i> $\alpha$ -Factor Pre-Sequence | Steube <i>et al.</i> , 1991    |
| Human Epidermal Growth Factor (hEGF)                         | <i>S. cerevisiae</i> $\alpha$ -Factor Pre-Sequence | Brake <i>et al.</i> , 1984     |
| <i>Bacillus amyloliquifaciens</i> $\alpha$ -Amylase          | <i>B. amyloliquifaciens</i> $\alpha$ -Amylase      | Ruohonen, <i>et al.</i> , 1987 |
| Human Interferon (IFN $\alpha$ 1, IFN $\alpha$ 2)            | Human IFN                                          | Hitzemann <i>et al.</i> , 1983 |

**Table 2.1:** Examples of Secretion of Recombinant Proteins in Yeast

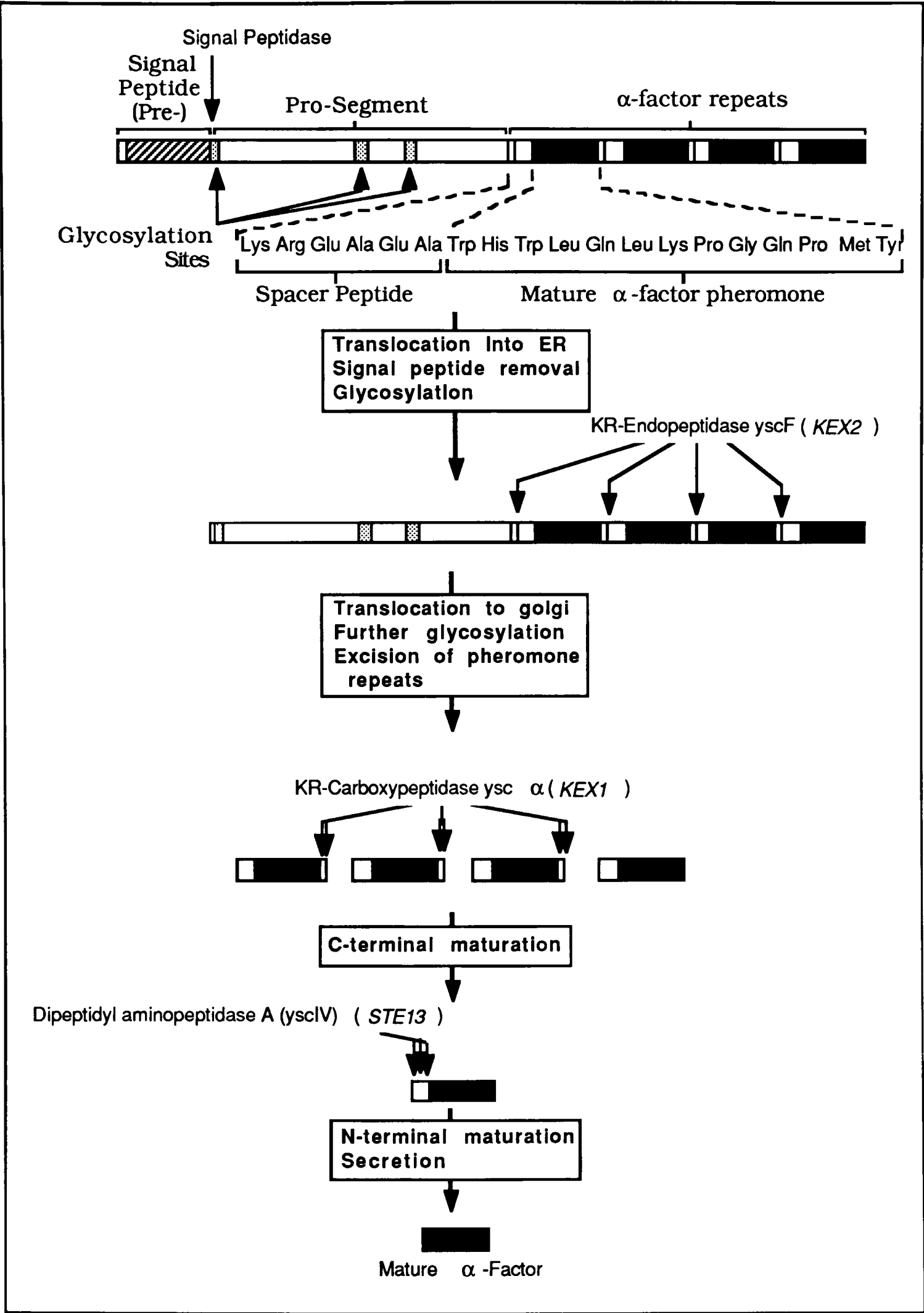


Figure 2.3: Proteolytic Maturation of Yeast Alpha-Factor

rER after which, it is cleaved by a signal peptidase. The pro-sequence is extensively glycosylated both in the ER lumen and the golgi apparatus.

The golgi associated maturation of the ppαF involves the sequential proteolytic cleavage of peptide bonds within the spacer regions. Three proteases have been shown to be involved in this processing step which leads to the formation and secretion of the mature α-factor peptide. The *KEX2* encoded KR-endoropeptidase yscF cleaves the peptide bond on the carbonyl side of the -Lys-Arg- residues. A *KEX1* encoded KR-carboxypeptidase yscα is responsible for the subsequent removal of basic C-terminal Lys or Arg residues. The *STE13* encoded dipeptidyl aminopeptidase A (yscIV) then cleaves off the remaining N-terminal -Glu-Ala- residues to yield the mature α-factor [Fuller *et al.*, 1988].

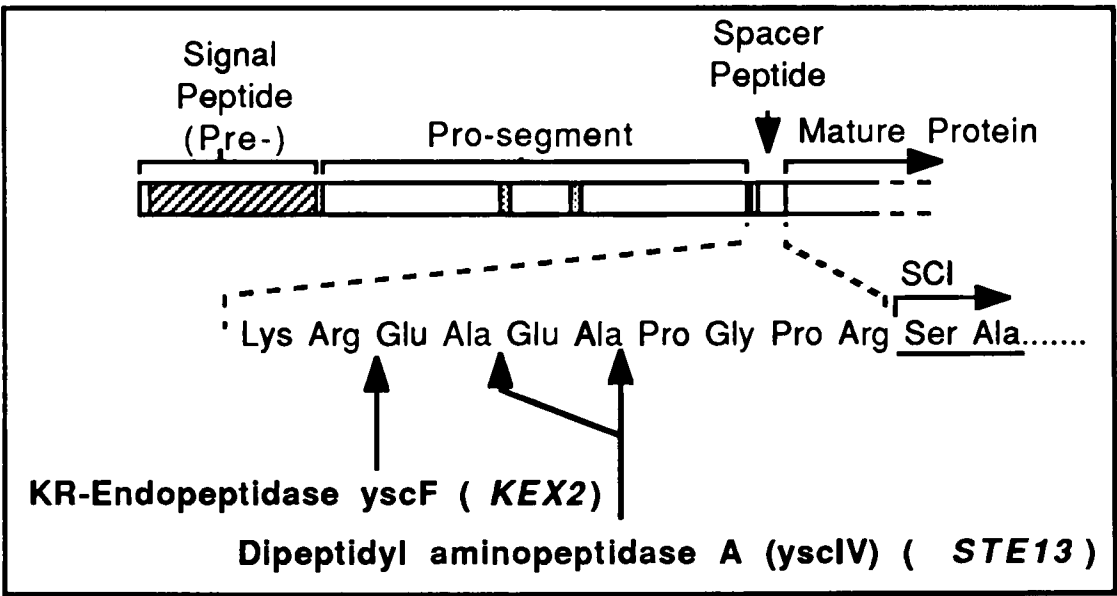


Figure 2.4: Post-translational proteolytic removal of the α-prepro sequence.

Secretion of recombinant proteins have been obtained by the in-frame fusion of the heterologous gene downstream to the prepro region of the yeast α-factor [Brake *et al.*, 1984; Neeper *et al.*, 1990; Steube *et al.*, 1991] and expressing under the control of various yeast promoters, including the MFα1 promoter [Inokuchi *et al.*, 1987]. Post-translational processing of the precursor protein would require cleavage by only the KR-endoropeptidase yscF at the -Pro-Arg- site and the dipeptidyl aminopeptidase A yscIV after the -Glu-Ala- repeat (Figure 2.4).

## 2.2. Expression of SCI in Yeast.

Kodo (1988) in his study of the expression of SCI described the construction of two variants of yeast expression vectors, pMASC-1 and pMASC-20 $\alpha$  under the control of the yeast PGK- and MF $\alpha$ 1 promoters, respectively. In both plasmids, the SCI gene, fused in-frame to the 5' terminal  $\alpha$ -prepro sequence was inserted downstream to the respective promoters with PGK 3' sequences as transcription terminators. With media and conditions designed for both unregulated and regulated expression, Kodo (1988) reported the failure to obtain secretion of any protein as was ascertained by SDS-PAGE and western blotting analysis of the culture medium. Yet, western blot analysis of both the periplasmic and cellular extracts indicated the intracellular accumulation of discernible quantities of recSCI [Kodo, 1988]. This included a correctly processed form which co-migrated with the native SCI on SDS-PAGE and two additional forms, one slightly smaller and the second slightly larger which possibly indicates incorrect or incomplete post-translation processing of the precursor protein. The hydrophobic nature and poor solubility properties of SCI [M.P. Esnouf; personal communication] gave one reason to believe that the nascent recombinant protein underwent self-aggregation or associated with ER- or golgi membranes and related structures - thereby preventing completion of the secretion pathway. Furthermore, premature folding could have hindered translocation. Electron-microscopic examination of the yeast cells indicated such a localisation of the recombinant protein (Y.Kodo; personal communication). The intracellular location of multiple forms of the recombinant protein could be due to the accumulation at more than one location [Steube *et al.*, 1991].

Initially, it was therefore decided to replace the SCI gene in pMASC-20 $\alpha$  with a truncated SCI gene (tSCI) encoding only the first 56 N-terminal amino acid residues of SCI (Figure 2.1). This segment of SCI incorporates the putative reactive-site loop and 7 of the 10 Cys residues. The convenient modular design of the synthetic SCI gene within the MF $\alpha$ 1 promoter/prepro sequence expression cassette facilitated the rapid and convenient truncation of the SCI gene. With tSCI, we hoped to overcome the barriers encountered by SCI along the secretion pathway by removing a major part of the protein's hydrophobic structural framework - but retaining adequate residues where one could reasonably hope that the peptide will fold so as to provide the reactive-site loop with adequate conformational support to obtain a functional inhibitor. Cyclic peptides, 6-15 residues long which mimic the reactive-site structures of several trypsin inhibitors bearing only the reactive-site loop sequence, cyclised by a disulphide bond between two Cys residues at either end have been shown to inhibit their cognate enzymes, yet, not as efficiently as the parent inhibitor (for references, see Powers & Harper, 1986).  $K_i$  values for the cyclic mimics have been found to be several orders of magnitude lower and the peptides degraded fast ( $t_{1/2} \approx 30$  min.) [Powers & Harper, 1986]. The high Pro content of the SCI reactive-site loop would possibly preserve its  $\beta$ -turn conformation in tSCI.

### **2.2.1. Construction of pUCSCR- $\alpha$ - an Expression Cassette for a Truncated SCI Gene**

Figure 2.5 illustrates the construction of the expression cassette for tSCI (see Chapter 5 for experimental details). This required the relevant 3'-truncation of the SCI gene contained in the plasmid pUSC-1 $\alpha$  (Appendix II) which was previously constructed by Y. Kodo (1988). The plasmid pUSC-1 $\alpha$ , consisted of a 1565 b.p. *Eco*RI fragment made up of the synthetic SCI gene cloned as an  $\alpha$ -factor prepro-SCI fusion gene, downstream to the yeast MF $\alpha$ 1

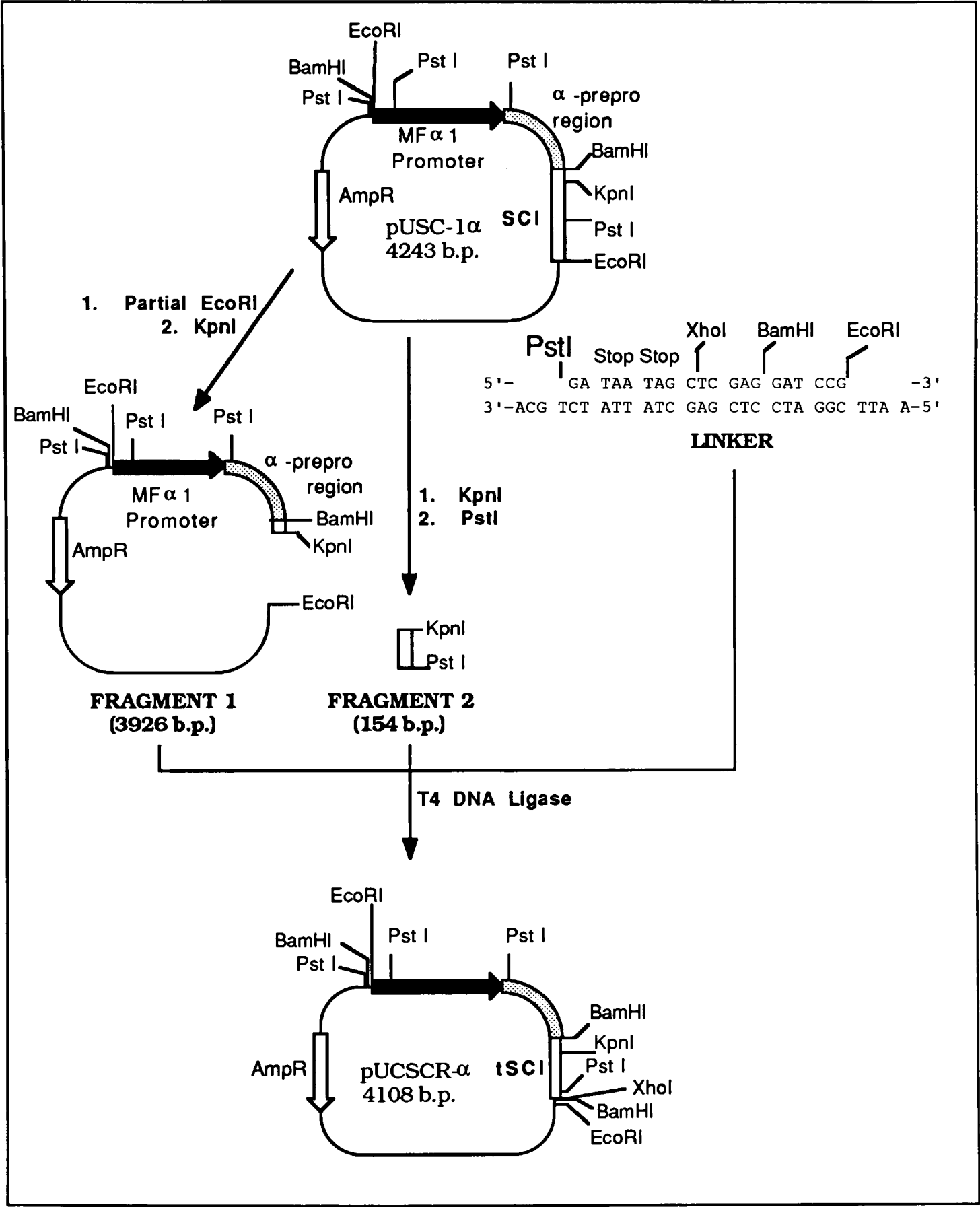


Figure 2.5: A Yeast Expression Cassette for a truncated SCI gene: Construction of pUCSCR- $\alpha$

promoter which was inserted into the *EcoRI* site of the *E.coli* plasmid pUC8 [Vieira & Messing, 1982]. The 1565 b.p. *EcoRI* fragment was essentially the expression cassette for the  $\alpha$ -prepro-SCI precursor without the required 3'-terminal transcription termination sequences. All modifications required for the construction of the corresponding  $\alpha$ -prepro-tSCI expression cassette were carried out with sequences obtained from pUSC-1 $\alpha$ .

Apart from the incomplete C-terminal Arg codon, all of the gene encoding the tSCI peptide is contained in the 163 b.p. *Bam*HI/*Pst*I fragment of the SCI gene contained in pUSC-1 $\alpha$ . Figure 2.5 illustrates procedure adopted in replacing the redundant *Pst*I/*Eco*RI fragment encoding the C-terminal half of SCI with a suitably designed, 24 b.p. *Pst*I/*Eco*RI adapter. The adapter was designed so as to complete the codon encoding the C-terminal Arg of tSCI, directly followed by two translation termination (stop) codons. A *Xho*I and a *Bam*HI site were incorporated into the non-coding segment of the adapter. The unique *Xho*I site facilitated screening for the successful insertion of the adapter into subsequent constructs and the *Bam*HI site for the correct insertion of the expression cassette into the yeast shuttle vector pMA3a.

The complementary oligonucleotide strands of the adapter were synthesised on an Applied Biosystems 380B DNA synthesiser. After deprotection, they were purified by denaturing urea-PAGE (see Chapter 5) and annealed to yield the adapter.

For the completion of the tSCI expression cassette, the following fragments were prepared from pUSC-1 $\alpha$  (Figure 2.5):

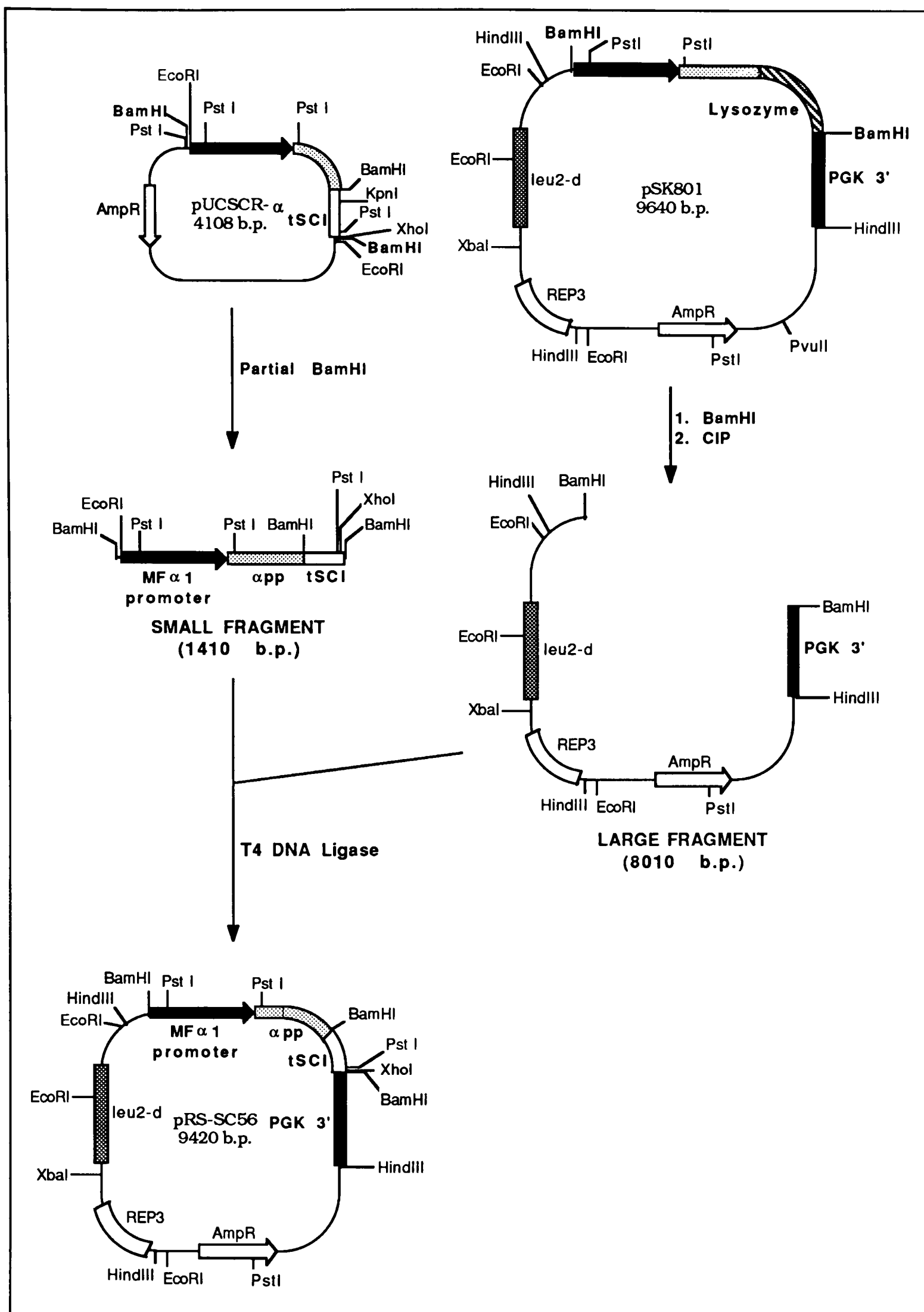
Fragment 1: the 3926 b.p. *Eco*RI/*Kpn*I fragment incorporating the entire pUC8 *Eco*RI fragment, the MF $\alpha$ 1 promoter, the  $\alpha$ -prepro sequence and the first three complete SCI codons, and,

Fragment 2: the 154 b.p. *KpnI/PstI* fragment which included the remaining tSCI coding sequence.

To complete the plasmid pUCSCR- $\alpha$ , the purified fragments 1 and 2 and the *PstI/EcoRI* adapter were mixed in an approximate molar ratio of 1:10:100, respectively and ligated overnight at 4°C with T4 DNA ligase. The different restriction endonucleases used to prepare the various fragments facilitated the correct formation of pUCSCR $\alpha$  due to correct orientation of fragments during ligation. The resulting ligation mix was used to transform competent *E.coli* MC1061 cells which were plated on an LB-agar plate containing ampicillin. Overnight incubation at 37°C resulted in about 300 colonies. Restriction endonuclease screening was carried out on plasmid mini-preparations of 24 single, well-separated colonies, obtained by the rapid boiling method. The unique *XhoI* recognition site in the non-coding region of the linker used to construct the expression cassette enabled one to rapidly screen mini-preps of the positive transformants obtained. Plasmid preparations of four of the transformants were linearised by *XhoI* digestion to yield a single band indicating the successful insertion of the linker. Extensive screening of the plasmids obtained by multiple restriction endonuclease digests and double-stranded sequencing (data not shown) confirmed the correct construction of the plasmid pUCSCR- $\alpha$ . In this plasmid, the expression cassette for the  $\alpha$ -prepro-tSCI fusion cloned downstream to the MF $\alpha$ 1 promoter, but without the necessary 3'-termination sequences, was contained in the 1410 b.p. *BamHI* fragment (Appendix III).

### **2.2.2. Construction of pRS-SC56 - an Expression Vector for a Truncated SCI Gene**

The completion of the tSCI expression vector is illustrated in figure 2.6 (for details, Chapter 5). The necessary pMA3a shuttle vector sequences



**Figure 2.6:** A Yeast Expression Plasmid for a truncated SCI gene:  
Construction of pRS-SC56

together with the 3'-transcription termination sequences of the yeast PGK gene were obtained as the 8 k.b. *Bam*HI fragment from the plasmid pSK801. This plasmid was constructed for the successful MF $\alpha$ 1 promoter mediated secretion of up to 0.5 mg/L active hen egg-white lysozyme into the medium both unregulated in the yeast strain MD40-4C and inducible in the *MAT $\alpha$ SIR3* mutant, JRY188 (S.K. Field; unpublished results).

*Bam*HI digestion of pSK801 resulted in the excision of the lysozyme gene (387 b.p.) together with the  $\alpha$ -prepro sequences and the entire MF $\alpha$ 1 promoter. The remaining, large (8 k.b.) fragment which included the entire pMA3a and the PGK 3'-termination sequences was isolated and purified. The 5'-phosphate groups of the fragment obtained were removed by treatment with calf intestinal alkaline phosphatase (CIP). This minimised the preferred tendency for the fragment to circularise during ligation which would prevent the insertion of the tSCI expression cassette.

The tSCI expression cassette was obtained by the partial *Bam*HI digestion of pUSCR- $\alpha$ . Of the various fragments obtained, the 1410 b.p. fragment, corresponding to the tSCI expression cassette was isolated. The plasmid pRS-SC56 was completed by the ligation of this fragment to the large 8k.b.*Bam*HI fragment of pSK801. The resulting ligation mix was used to transform *E.coli* MC1061 cells, which were then plated on LB-agar containing ampicillin. After overnight incubation at 37°C, about 20 ampicillin resistant colonies were identified.

Plasmid "minipreps" of successful transformants were extracted by the rapid boiling method (see Chapter 5). In the first instance, insertion of the *Bam*HI expression cassette in pRS-SC56 was confirmed by the incorporation of a new *Xho*I site. Yet, the expression cassette can be inserted in two possible orientations. Correct orientation of the insert was confirmed by mixed restriction endonuclease cuts. Double-stranded DNA sequencing of the

ligation junctions of the tSCI expression cassette and the shuttle vector confirmed correct insertion (data not shown). One of these plasmid preparations, pRS-SC56 (Appendix IV), the completed yeast expression vector for the  $\alpha$ -prepro-tSCI fusion, was used to transform the yeast strains JRY188, MD40-4C and KHY2.

### 2.2.3. Transformation of Yeast Spheroplasts

Spheroplasts of the yeast strains JRY188, MD40-4C and KHY2 were prepared according to the procedure of Hinnen *et al.* (1978) by the enzymatic removal of the cell wall using Zymolyase 20T, a mixture of  $\beta$ -glucanases from *Arthrobacter luteus* (Seikagaku Kogyo Co. Ltd., Tokyo, Japan) (N.B. the incorporation of 1M Sorbitol in all reagents ensured osmotic support for the fragile spheroplasts). To facilitate the uptake of episomal DNA, the spheroplasts were pre-treated with 10mM calcium chloride and following the addition of the relevant DNA, uptake induced by the addition of polyethylene glycol ( $M_r$  4000) to the incubation mix. A subsequent method of transforming yeast by electroporation described by Becker & Guarente (1991) provided higher transformation efficiencies.

Aliquots of the cells were grown on minimal agar medium in the absence of leucine (Leu<sup>-</sup>). The MD40-4C and KHY2 yeast strains were incubated at 30°C whereas, the JRY188 strain was kept at the elevated temperature of 37°C so as to repress expression of the recombinant protein. Well isolated colonies of leucine prototrophs that appeared after about 36h incubation were selected for expression studies.

### 2.2.4. Expression Studies

Transformants of the *MAT $\alpha$ SIR3* strain, JRY188 were used for investigating the temperature-induced regulation of expression of tSCI with

KHY2 transformants for reference. For comparison, transformants the yeast strain MD40-4C were used for constitutive expression.

Quantities of 5ml of SDAA selective medium (Leu<sup>-</sup>) were incubated with well isolated colonies of transformants of each strain. To repress expression of the recombinant protein, JRY188[pRS-SC56] was grown at 37°C. MD40-4C[pRS-SC56] and KHY2[pRS-SC56] transformants were grown at 30°C. After overnight growth, the starter cultures were diluted 1/10th, 1/50th and 1/100th into 25ml each of fresh SDAA medium in 100 ml conical flasks.

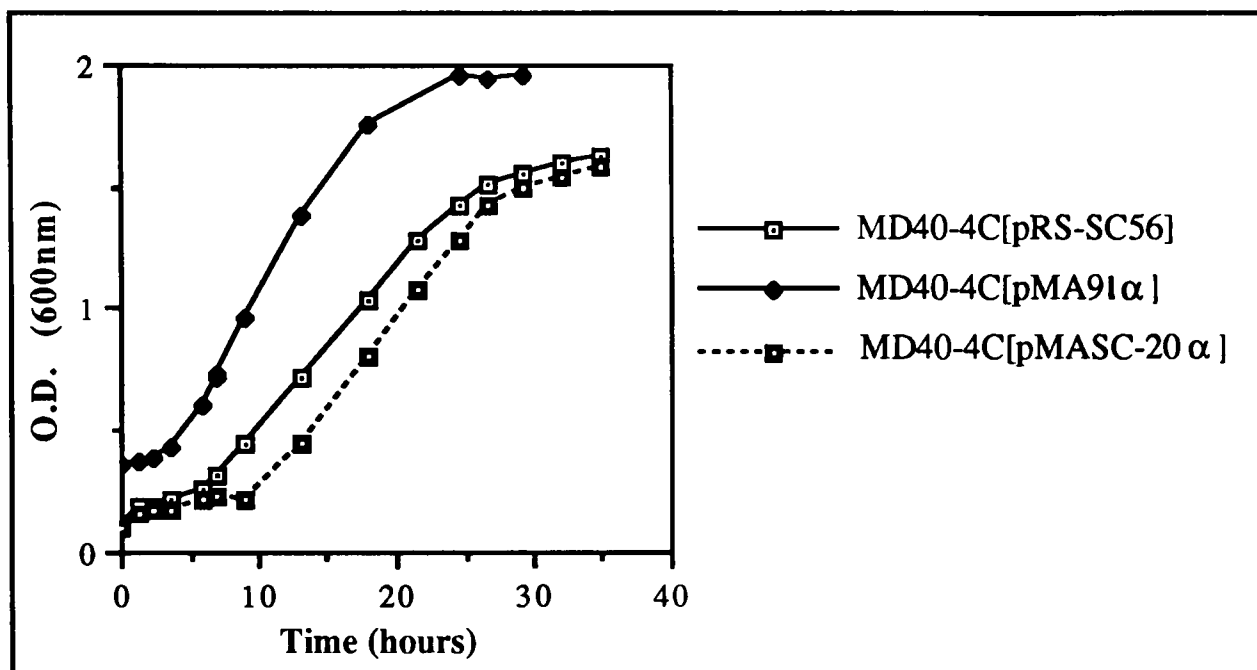
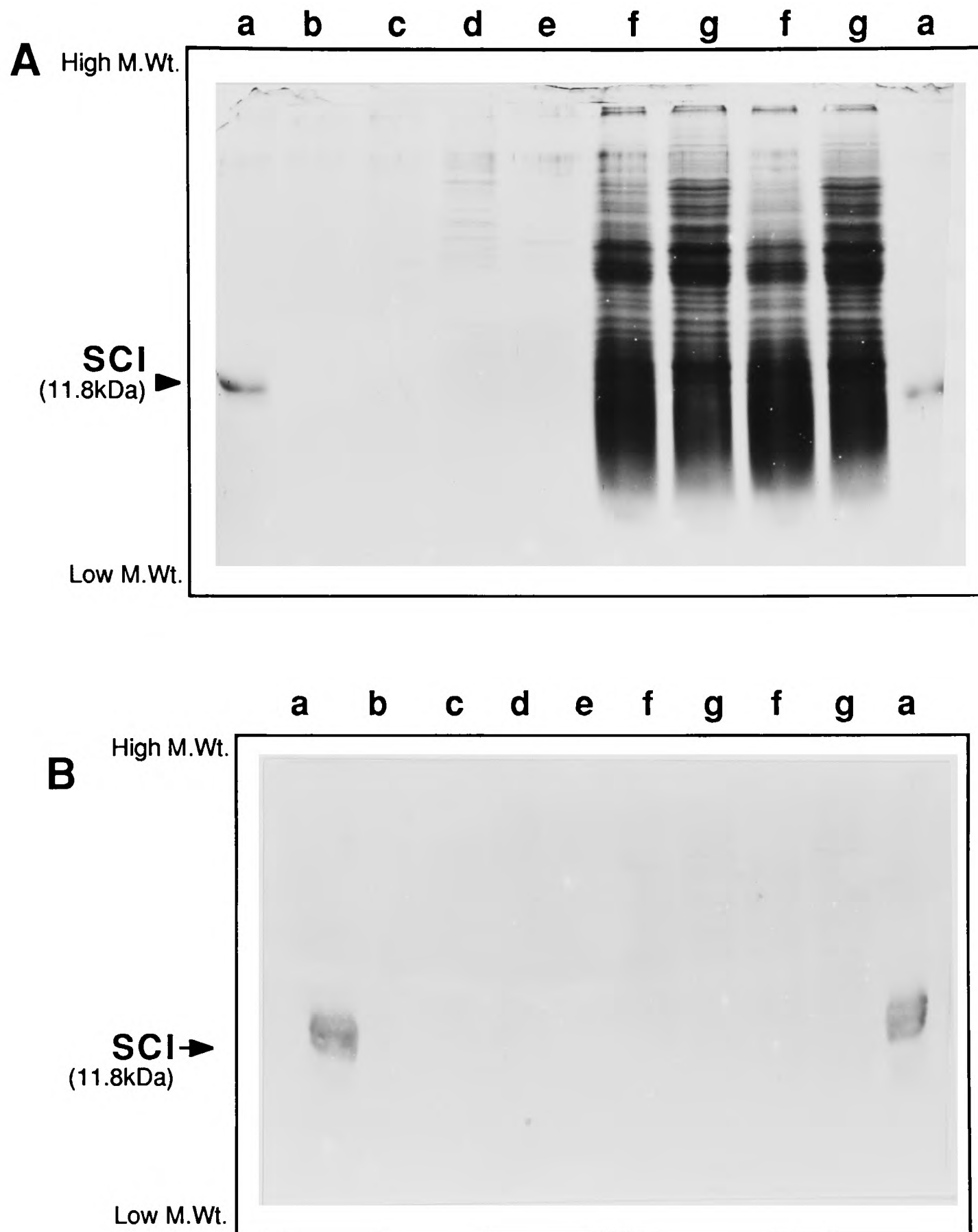


Figure 2.7:  
Growth curves of yeast MD40-4C transformants

For constitutive expression, the diluted MD40-4C[pRS-SC56] culture was grown at 30°C for 4 days with vigorous rotation (300rpm). The growth rate was spectrophotometrically monitored by measuring the increase of the turbidity of the medium (O.D. at 550-600nm). Figure 2.7 illustrates the typical growth curves obtained. For comparison, the growth curves of MD40-4C transformed both with pMASC-20α (dotted line), the same expression plasmid, but with the entire SCI gene [Kodo, 1988] and pMA91α, a pMA3a derived plasmid [Dobson *et al.*, 1982] without a heterologous gene insert were determined. The transformant without a heterologous insert showed healthier growth rates (figure 2.7) with a shorter generation time and a higher saturation level ( $\approx 2.0$ ) as compared to those with the tSCI and SCI inserts,

MD40-4C[pRS-SC56] and MD40-4C[pMASC-20 $\alpha$ ], respectively ( $\approx 1.6$ ). This indicated the deleterious effect of the presence of both the SCI and tSCI genes - which however, did not prove lethal to the cells. No substantial difference could be seen between cells containing pRS-SC56 and pMASC-20 $\alpha$  expressing tSCI and recSCI, respectively. Routine examination of the growth media for the presence of trypsin and  $\beta$ XIIa inhibitory activity failed to indicate successful secretion of any active inhibitor. After 4 days induction, aliquots of the culture were removed and the yeast cells separated from the medium by centrifugation at 13000g for 5 min. Protein extracts of both the media and the periplasmic and cytoplasmic fractions of the cells were analysed by SDS-PAGE and western blotting using a polyclonal anti-SCI antibody, obtained, with thanks, from M.P. Esnouf (Nuffield Department of Clinical Biochemistry, University of Oxford), failed to reveal any detectable levels of expressed tSCI. Figure 2.8 illustrates the SDS-PAGE gel obtained, and western blot profiles of different extracts. Note the complete absence of any detectable SCI antigenic activity in any of the obtained extracts.

For induced expression, JRY188[pRS-SC56], JRY188[pMASC-20 $\alpha$ ] and, as reference, KHY2[pRS-SC56], were initially grown overnight in SDAA at the repressive temperature of 37°C to OD<sub>600</sub> $\approx 1.3$ . After dilution of the cultures in fresh SDAA (see above) the expression of tSCI was induced by reducing the incubation temperature to 24°C and incubating for a further 4 days. For reference, the growth behaviour of the same transformants was monitored at the continued repressive temperature of 37°C. As controls, untransformed cells of both strains and transformants with pMA91 $\alpha$  (that is, without a heterologous gene insert) were grown under the same conditions (results not shown). Aliquots of the media were removed at regular intervals to monitor the growth rate of the cultures and the presence of secreted trypsin and  $\beta$ XIIa activity. Figure 2.9 shows the growth curves of the cultures of the JRY188 and KHY2 transformants. Contrary to the expectation that growth at 24°C, with



**Figure 2.8:** Constitutive Expression of tSCI in Yeast: Profiles of intracellular, periplasmic and secreted proteins (reduced) from yeast transformants on 15% SDS-polyacrylamide gels.

A: Coomassie Stain  
B: Western Blot

Lanes a: Authentic SCI  
Lane b: MD40-4C[pRS-SC56] ] medium  
Lane c: KHY2[pRS-SC56]  
Lane d: MD40-4C[pRS-SC56] ] periplasm  
Lane e: KHY2[pRS-SC56]  
Lanes f: MD40-4C[pRS-SC56] ] cytoplasm  
Lanes g: KHY2[pRS-SC56]

the induced expression of the potentially deleterious recombinant proteins would be slower than at the repressive temperature of 37°C, JRY188 transformants grew at a greatly enhanced rate at the former temperature, leading to higher saturation levels (Figure 2.9). A similar difference was seen with the controls and the KHY2 transformants. This difference was probably due to the effect of temperature on the organisms in that the preferred temperature for healthier growth was the lower 24°C. The growth rate comparison therefore gave no indication of the appearance of any potentially deleterious recombinant proteins under permissive conditions.

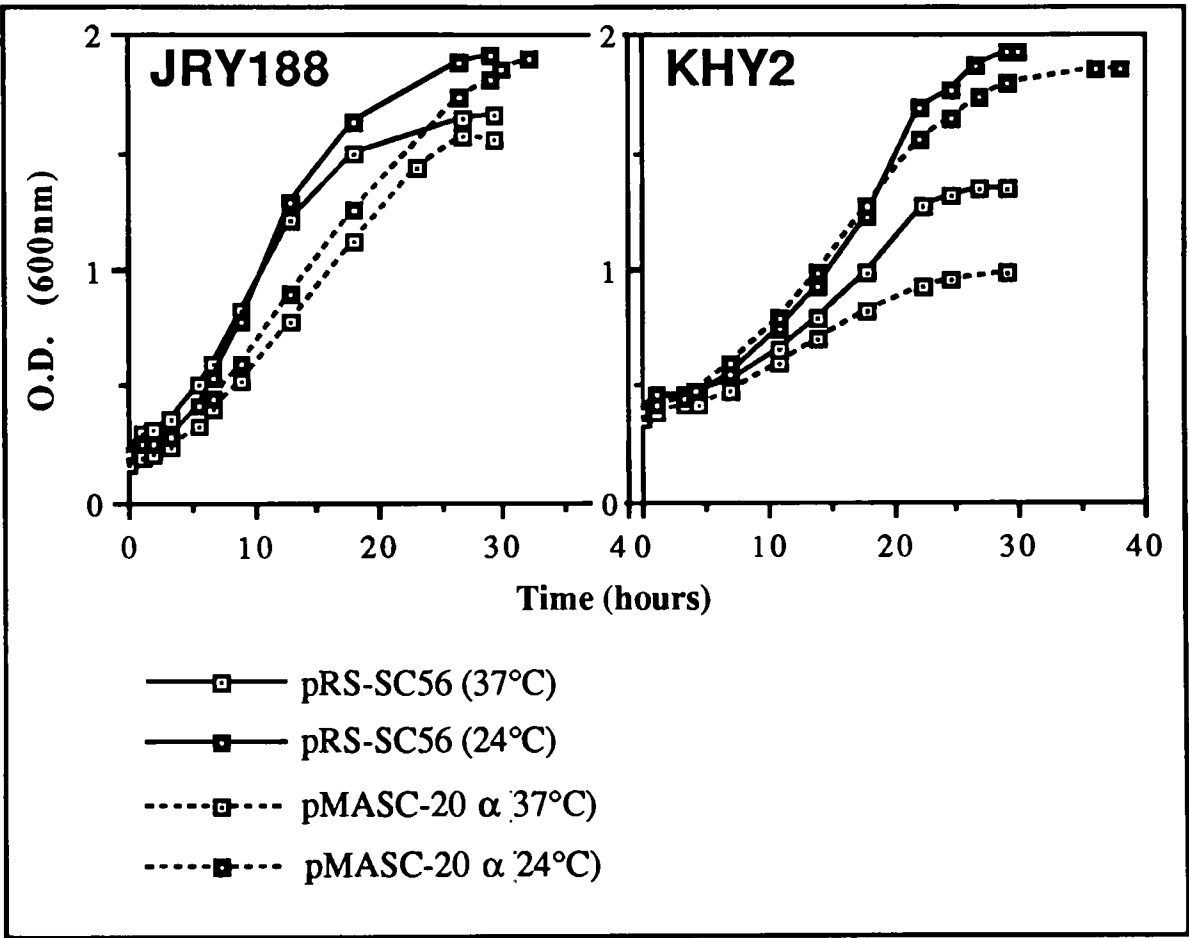
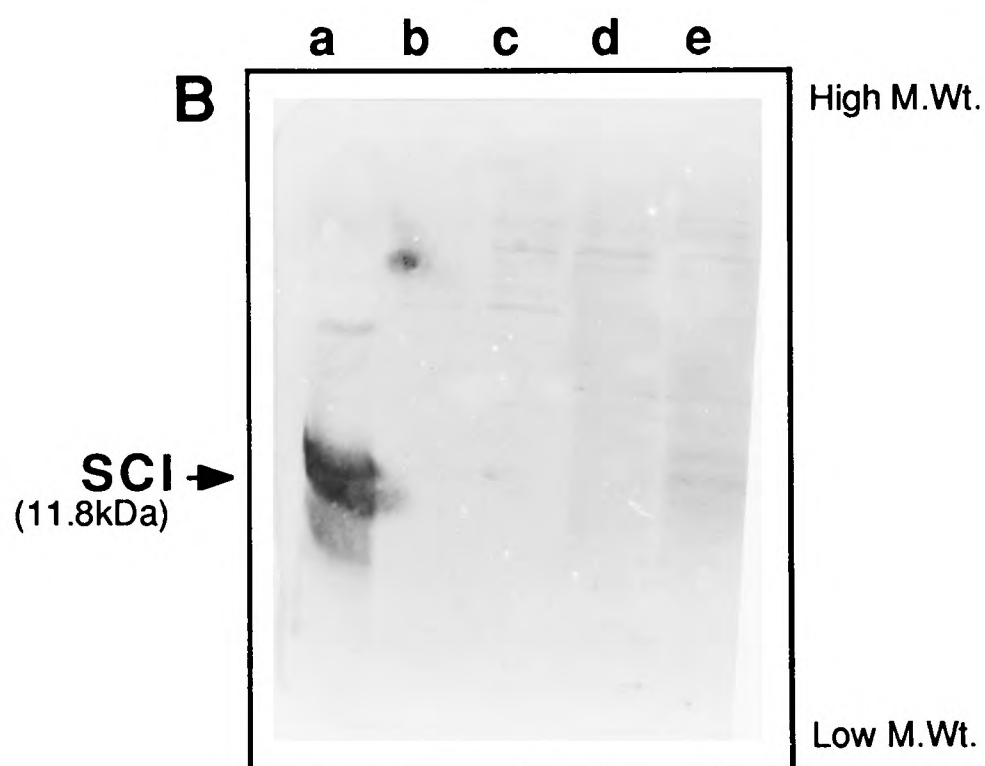
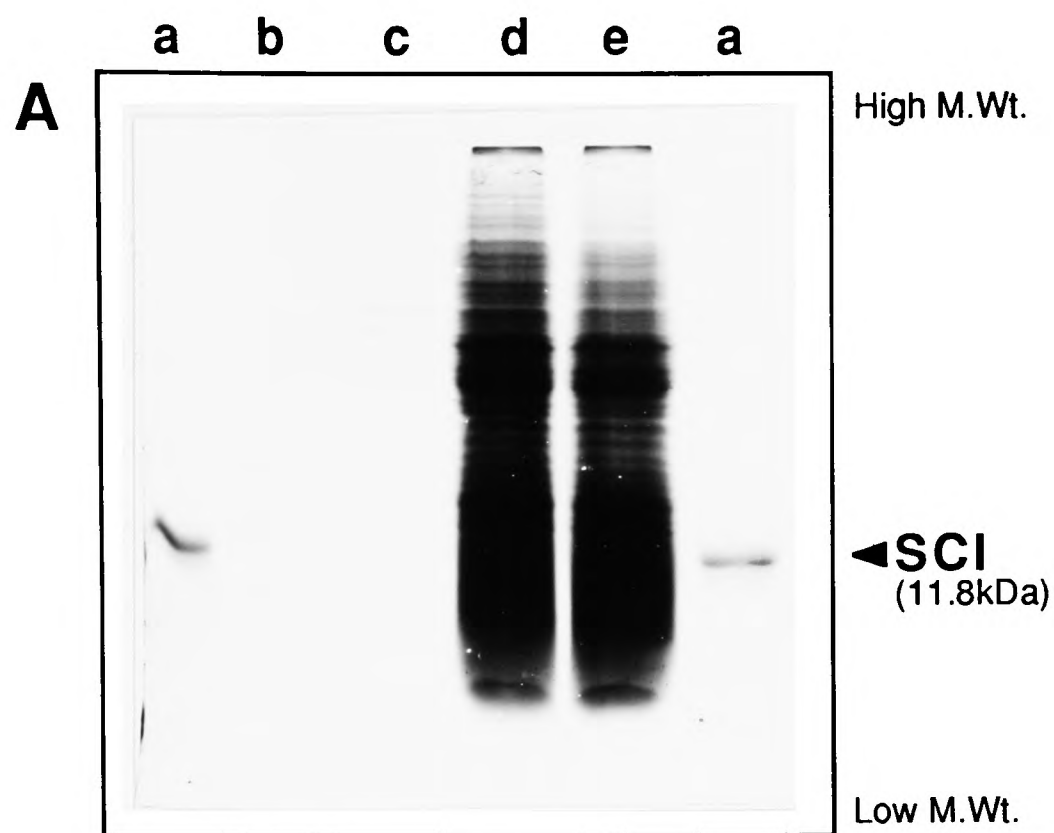


Figure 2.9: Induced Expression of tSCI: Growth curves of yeast JRY188 transformants. (the yeast KHY2 transformants were grown as reference)

After the completion of incubation, the resulting yeast cells were separated from the growth media by centrifugation. Both were analysed by SDS-PAGE and Western blotting for the presence of expressed recombinant SCI. Figure 2.10 illustrates the results obtained. The cytoplasmic extract of JRY188[pMASC-20α] showed trace levels of SCI antigenic activity (panel B; lane e), note the presence of two faint bands of about the same molecular



**Figure 2.10:** Induced Expression of tSCI in Yeast: Profiles of intracellular, periplasmic and secreted proteins (reduced) from yeast transformants on 15% SDS-polacrylamide gel

A: Coomassie Stain

B: Western Blot

Lanes a: Authentic SCI

Lane b: JRY188[pRS-SC56]; medium

Lane c: JRY188[pRS-SC56]; periplasm

Lane d: JRY188[pRS-SC56]; cytoplasm

Lane e: JRY188[pMASC-20 $\alpha$  ]; cytoplasm

weight of authentic SCI, these possibly correspond to traces of the mature SCI and the unprocessed  $\alpha$ -prepro-SCI precursor. Yet again, no trace of any expressed SCI antigenic activity could be detected in the other protein extracts of any of the cellular fractions inasmuch as no trypsin or  $\beta$ XIIa inhibitory activity was present in the growth medium. Even on 10-times concentration of larger volumes of the media by lyophilisation in water after salt removal on a PD-10 column (Pharmacia), tSCI could not be detected.

## 2.3. Discussion

Kodo (1988) observed the intracellular accumulation of recSCI forms, both of higher and lower molecular weight than the native SCI. From this, one had reason to attribute the lack of expression and the detectable accumulation of recSCI within the cells of transformants to the hydrophobic properties of SCI and therefore, its poor solubility. It has been also known that the overproduction of secreted proteins can lead to their mislocalisation, in that it never enters the correct pathway for secretion which results in the accumulation of the unprocessed form of the protein [Rothman *et al.*, 1986].

In the attempt to obtain secretion of the truncated SCI (tSCI), the possibilities for pinpointing the cause of failure were limited to a minimum as there was no discernible expression of the tSCI gene at any level. The absence of expression could have been due to a complex of reasons such as degradation of the expression products, both the mRNA at the transcription stage, and the nascent polypeptide after translation. Further investigation of the expression system could be carried out both at transcription stage, by determining the corresponding levels of tSCI mRNA within the cell and after translation by a closer analysis of the cell's protein extracts.

Yeast, an eukaryote, was initially chosen in preference to *E.coli* as, common to higher eukaryotes, it possesses those intra-cellular membraneous organelles such as the endoplasmic reticulum (ER) and the golgi bodies, that are responsible for the correct post-translational processing and subsequent secretion of nascent proteins that are destined to be secreted into the extra-cellular medium. The redox conditions and enzymes within the lumen of the ER and golgi bodies catalyse the correct disulphide pairings of Cys residues [Kingsman *et al.*, 1985] leading to the secretion of the correctly folded recSCI. However, all attempts to obtain functional secretion of both recSCI and tSCI failed. It was postulated that the hydrophobic nature of the recombinant protein caused it to adhere to the amphiphilic intracellular lipid membranes. With this, it was concluded that this system and the previous systems developed by Kodo (1988) were not suitable for the development of a functional secretion system for obtaining workable quantities SCI. There remain various other yeast and eukaryotic expression systems that could be assessed. Yet, within the framework of this project, it was important to establish a more rapid, straightforward and efficient expression system for SCI as the more immediate priority was to undertake mutagenic studies in view of obtaining information on the structure-function relationships of SCI.

## 2.4. **Addendum**

Recent work by [Steube *et al.*, 1991] of the pp $\alpha$ F-leader-directed secretion of recombinant human-insulin-like growth factor I (IGFI) under the control of a yeast glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter postulated an undefined Golgi or post-Golgi bottleneck representing a major obstacle in the secretion of the recombinant protein. IGFI is a small polypeptide 73 residues long with 6 disulphide-paired Cys residues. Common to most expression systems using the  $\alpha$ -factor-leader fusions, secretion seemed to be limiting, leading to the intracellular accumulation of greater

amounts of unprocessed forms of the protein [for references, see Steube *et al.*, 1991]. It was postulated that correctly matured and folded protein precursors are successfully released into the medium, whereas, the majority of unfolded protein accumulates in a Golgi vesicle. A possible chaperone-like function can be attributed to the  $\alpha$ -prepro-sequence in that aggregation of the nascent protein occurred only after the former is removed by KR-endopeptidase yscF [Steube *et al.*, 1991]. The polypeptide IGFI shares common properties with SCI in that it is small and disulphide-rich. However, the use of two different yeast strains, AB110 and H449 transformed with the same expression plasmid for IGFI showed markedly varied levels of secretion of the recombinant protein. Only the strain AB110, a cross between the yeasts *S.cerevisiae* and *S.carlsbergensis* secreted significant amounts of mature IGFI into the extracellular medium. The strain H449 secreted only trace amounts, the majority of the protein being accumulated within the cell. Neepers & co-workers (1990) reported the successful secretion of active recombinant tick anticoagulant peptide (TAP) (from *Ornithodoros moubata*) in yeast, having expressed it as a pp $\alpha$ -TAP fusion under the transcription control of the yeast *GAL10* promoter. TAP, a small (60 residue), disulphide-rich (6 Cys) protein is a highly selective inhibitor of FXa.

## **Chapter 3**

# **Expression of SCI in *Escherichia coli*.**

## 3. Expression of SCI in *Escherichia coli*

### 3.1. Introduction

Following the failed attempts to obtain functional expression of SCI in yeast, it was decided to revert to expression in the prokaryote *Escherichia coli* (*E.coli*). *E.coli*, being the best studied living organism, has featured as a standard host organism for the cloning and manipulation of genetic material and the over-expression of foreign proteins. Rapid growth in inexpensive media and easy manipulation, coupled to the wealth of detailed knowledge available have made *E.coli* the most widely used host in gene manipulation studies. Recently, expression vectors of ever-increasing refinement and sophistication for maximising heterologous gene expression have been developed and *E.coli* host strains obtained for minimising the turnover of gene products [Das, 1990]. It is now commonplace to obtain expression of heterologous proteins to levels of 25% and more of the total bacterial protein [Marston, 1986]. This greatly facilitates the accelerated purification of a protein.

Expression of eukaryotic proteins in *E.coli* is hampered by the major disadvantage of the lack of suitable post-translational processing mechanisms. In most cases, high level and rapid expression of the recombinant protein, coupled with the reductive environment of the cytoplasm results in the recombinant protein being expressed as an insoluble, non-native and, therefore, inactive form [Marston, 1986]. Very often, the insoluble recombinant proteins associate to form large protein aggregates called inclusion bodies. The formation of inclusion bodies are usually undesirable

as the insoluble protein can only be solubilised in detergent, strong denaturant or at extreme pH values [Wetzel *et al.*, 1991]. Therefore, in addition to the isolation and purification of the recombinant protein, one has to devise a suitable solubilisation and re-folding strategy so as to obtain the correctly folded, functional protein. The lack of cost-effective methods for resolubilisation and folding can often render an otherwise efficient expression system valueless.

Inclusion body formation depends on both the expressed recombinant protein and the homologous expression signals which control the production of the protein. Different proteins or even mutants of the same protein under control of the same homologous expression signals show different tendencies to form inclusion granules [Wetzel *et al.*, 1991]. The same variability has been noted in the case of the same recombinant protein under the control of different homologous promoter sequences. For example, isopenicillin N synthase (IPNS) under the control of the  $\lambda$ -phage  $P_L$  promoter is expressed as insoluble inclusion bodies to levels of up to 20% of the total bacterial protein whereas, under the control of the *trc* promoter, an inducible *trp/lac* fusion promoter, it is expressed active and soluble to levels up to 50% (J.D. Sutherland; personal communication). However, this variability is not necessarily a function of only the proteins, promoters and *E.coli* strains used. Growth conditions, including nutrients and growth temperature have a marked influence on inclusion body formation [Bowden *et al.*, 1991].

With the availability of a wide variety of expression vectors for *E.coli* that facilitated relatively rapid cloning and efficient expression of heterologous proteins in *E.coli*, the development of a suitable expression system for the over-production of functional recSCI was considered reasonably feasible. This chapter describes the various attempts to establish such an *E.coli* expression system and concludes with the detailed description

of successful studies with an expression vector based on the  $\lambda$ -phage  $P_L$  promoter.

### **3.1.1. Expression of Heterologous Proteins in *Escherichia coli* - Expression Vectors.**

Synthesis of a functional protein depends upon transcription of the appropriate gene, efficient translation of the mRNA transcript and, in many cases, post-translational processing and compartmentalisation of the nascent polypeptide. The universality of the genetic code permits the recognition and expression of a foreign gene that has been correctly introduced into the host organism.

Insertion of the foreign gene into *E.coli* is greatly facilitated by the existence of a wide variety of expression plasmids that have been derived for the stable maintenance of the gene in *E.coli*. Most *E.coli* expression vectors are derived from the high copy-number plasmid, pBR322. This versatile and widely used synthetic plasmid was constructed to contain the replication elements of pMB1, a ColE1-like plasmid [Bolivar *et al.*, 1977].

The foreign gene, together with the necessary homologous regulatory sequences are incorporated into such plasmids for efficient maintenance and expression in *E.coli*. These regulatory sequences include:

- 1) a powerful upstream promoter recognised by the *E.coli* RNA polymerase,
- 2) an mRNA ribosomal binding site which includes the start codon (usually, AUG) and Shine-Dalgarno sequence (SD sequence), and,
- 3) transcription termination sequences.

Secretion of the recombinant protein can be achieved by the in-frame fusion of its gene to a sequence encoding an N-terminal signal peptide which

as in yeast, directs the nascent protein across the *E.coli* cytoplasmic membrane into the periplasmic space.

### 3.1.2. *E. coli* Promoters

The excess of 3000 genes in the *E.coli* chromosome are often found in groups encoding functionally related polypeptides. These groups which are usually expressed separately as individual units of transcription are called operons. For each operon there is a separate sequence called the promoter which signals the site for transcription initiation. This is brought about by providing a recognition site for interaction between the DNA template and the RNA polymerase holoenzyme. The latter is responsible for the process of transcription, where mRNA is produced under the direction of the DNA template. In addition to the promoter, the operon also contains DNA sequences that signal transcription termination.

The DNA sequence composition of the promoter directly determines the rate of transcription by dictating the rate at which stable initiation complexes are formed with the  $\sigma$ -subunit of the RNA polymerase holoenzyme. This could be attributed to varied needs of the cell for different groups of protein produced. There is, in fact, great variation in the amount of different proteins produced in the cell at any given time. For example, ribosomal proteins may each be present in excess of 10,000 copies per cell whereas certain regulatory proteins occur at less than 10 copies [Voet & Voet, 1990].

Sequence comparison of 112 different *E.coli* promoters were compiled [Reznikoff & McClure, 1986]. The consensus sequence obtained by this compilation revealed two conserved regions in a domain of about 40 b.p. upstream, that is, on the 5' side of the transcription initiation site. The most conserved region, also known as the Pribnow box is found at the -10 region,

that is, 10 b.p. upstream from the transcription initiation site. The consensus non-template sequence is the hexameric TATAAT. Similarly, a second conserved region of consensus sequence TCTTGACAT is found around the -35 region. The spacing between the -35 and -10 regions plays an important role in promoter function. The most frequently occurring spacing which is also best for promoter function is given as  $17 \pm 1$  base pairs [Reznikoff & McClure, 1986].

### 3.1.3. Regulation of Transcription

To facilitate adaptation to environmental changes such as availability of nutrients and change in temperature, production of polypeptides and enzymes can be varied with changes in the cell's conditions. In particular, many genes that encode catabolic functions are expressed only when the cells are presented with the relevant pathway's substrate, and many genes that encode anabolic functions are expressed at lower levels when the cells are presented with the relevant pathway's end-product [Reznikoff & McClure, 1986]. This is an example of inducible or adaptive enzymes as compared to constitutive enzymes which are synthesised at a relatively constant rate.

Inducible expression systems are particularly useful in the expression of heterologous proteins. In most cases, overexpression is deleterious to the host cells due to the excess utilisation of resources essential for healthy cell growth and the potential toxicity of the expressed protein. Such effects can be circumvented using inducible expression in that healthy cell growth is assured in the absence of recombinant protein expression. Expression can subsequently be induced once optimal biomass has been achieved.

### 3.1.4. The Bacteriophage Lambda P<sub>L</sub> Promoter

The leftward promoter of the coliphage  $\lambda$  is one of the most active *in vivo* [Das, 1990]. The phage  $\lambda$  is the most extensively characterised complex

virus with respect to its molecular biology. The intricate genetic regulatory mechanisms involved in selecting and maintaining its life cycle are thought to be analogous to systems which underlie many cellular processes [Ptashne, 1987]. The phage  $\lambda$  which adsorbs to the *E.coli* outer membrane, injects its linear DNA into the host cell. Being an example of a temperate phage as opposed to virulent phages, the phage  $\lambda$  can undergo two possible life cycles in the host cell, the lytic or the lysogenic life cycles. In the lytic mode, the phage replicates rapidly, resulting in the release of new phage particles as a result of host cell lysis. In the lysogenic cycle, the phage DNA incorporates into the host chromosome at a pre-determined site. The resulting phage is now defined as the prophage and the host cell, a lysogen. The choice of life cycles depend on environmental conditions. Ptashne (1987) deals with the intricate genetic regulatory mechanisms responsible for the switch between the lytic and lysogenic modes and *vice versa*. The  $P_L$  promoter and its contiguous operator sequence ( $O_{L1}$ ) and, in particular, the right-ward promoter,  $P_R$  and operators with the  $\lambda$  repressor play key roles in the maintenance of the lysogenic mode. Similar to the *lac* repressor, the  $cI$  repressor inhibits transcription by binding to the  $P_L$  promoter operator ( $O_{L1}$ ) sequence. This prevents binding of the RNA polymerase and initiation of transcription. The  $P_L$  promoter and operator,  $O_{L1}$  in conjunction with the  $cI$  repressor have found widespread application in the expression of heterologous proteins in *E.coli*. Plasmids containing heterologous gene sequence cloned downstream to the  $P_L$  promoter can be introduced into *E. coli*  $\lambda$  lysogens expressing the  $cI$  repressor. Expression of the foreign protein is initiated by treating cultures with SOS-inducing agents such as nalidixic acid [Das, 1990]. This results in changes in *RecA*, a host protein responsible for DNA repair and recombination in *E.coli* which cause it to become a highly specific proteinase that cleaves the  $cI$  repressor, thereby activating the  $P_L$  promoter and recombinant gene expression [Ptashne, 1987]. However, a

temperature-sensitive allele of *ci*, namely *ci857* has found more widespread use as a repressor of  $P_L$  transcription [Bernard & Helinski, 1979; Milman, 1987]. At low temperatures (27°C), the *ci857* repressor is active and efficiently represses  $P_L$  function by binding to the  $\lambda O_L$  operator [Bernard & Helinski, 1979]. At higher temperatures (42°C), the *ci857* repressor is deactivated. Induction of transcription can therefore be achieved by shifting the growth temperature from 27°C to 42°C. Some widely used lysogenic strains of *E.coli* (e.g. AR68 [Watt *et al.*, 1985]) contain this temperature-sensitive allele. This work describes the use of a  $P_L$  expression plasmid for SCI in which the *ci857* gene is co-resident on the expression plasmid. This enables one to insert the expression plasmid even into non-lysogenic host strains. Furthermore, increased gene dosage due to the high copy-number of the plasmids used ensures a more efficient inhibition of transcription due to higher concentrations of the repressor.

### 3.1.5. Translation Initiation

The translation of mRNA into protein in *E.coli* involves the ribosomal binding-site the Shine-Dalgarno (SD) sequence and the translation start codon, usually AUG. The initiation of translation in *E.coli* essentially involves the formation of the initiation complex between the 70S ribosome, the initiator tRNA recognising the start codon (fMet.tRNA) bound together with the mRNA [Stormo, 1986]. The fMet.tRNA carries an unusual form of Met in that it is N-formylated as N-formyl methionine (fMet). Some *E.coli* proteins carry these N-terminal fMet residues. However, most are post-translationally modified by deformylation and subsequent removal of the N-terminal Met by deformylases [Adams, 1968; Livingston & Leder, 1969] and methionine aminopeptidase (MAP), respectively [Ben-Bassat *et al.*, 1987; Hirel *et al.*, 1989].

Initial interaction involves the base-pairing of the purine-rich SD sequence, an mRNA region about 10 nucleotides upstream to the start codon with a pyrimidine-rich complementary region on the 3' end of the 16S ribosomal RNA (rRNA) molecule. The rRNA molecule together with several factors that facilitate translation initiation are found in the 30S subunit of the ribosome.

Investigation of several mRNA SD sequences led to an optimal consensus sequence required to interact with the 16S RNA 3' sequence ACCUCCU. Efficient initiation requires complementarity to at least four of these nucleotides [Stormo, 1986]. Furthermore, an optimal spacing of 9 nucleotides, consisting mostly of A and U, between the SD-sequence and the initiation codon AUG was suggested. Further optimisation of translation efficiency can be brought about by modifying the nucleotide sequence of the spacer region between the SD sequence and the start codon. The SD-sequence and start codon should not be involved in any secondary structures and not overlap with any RNA protein binding-sites thereby occluding ribosome access [Das, 1990]. In efficient translation, the SD sequence and the initiation codon have been shown to be situated on the accessible loop of a hairpin structure, that is, ribosomes may recognise structural signals in addition to the linear mRNA nucleotide sequence [Tessier *et al.*, 1986].

The presence of strong terminators at the end of a cloned gene is thought to increase mRNA stability in that excessively long, unstable transcripts are not produced.

Following the basic approach of expression of a foreign gene in *E.coli*, literature reports the rapidly growing numbers of successful protein expression experiments. With the aid of the vast amounts of knowledge gathered on the detailed aspects of *E.coli* expression and its optimisation, one

can with relative confidence, adopt these basic strategies and develop a suitable expression system for the foreign protein in question. Yet, each new protein poses new problems as the complexity of interactions of the new protein with the host systems have so-far prevented one from extracting any hard-and-fast rules that would universally apply to all expression studies. A rather more heuristic approach has to be adopted in the quest for an optimal expression system for any one heterologous protein. Yet, the wealth of knowledge that already exists, together with the plethora of "tailor-made" expression vectors incorporating different, tried combinations of the regulatory sequences necessary for efficient heterologous expression, make the task of establishing the ideal combination much easier.

## 3.2. Cytoplasmic Expression of SCI in *E.coli*

### 3.2.1. Inducible Expression of SCI; Construction of pRS101

The plasmid pRS101 was constructed by replacing the sequences encoding the gene of *Cephalosporium acremonium* isopenicillin N synthase (IPNS) in the plasmid pIT353 [Aplin *et al.*, 1990]. The plasmid pIT353 (Appendix V), obtained, with thanks, from J.D. Sutherland (Dyson Perrins Laboratory, University of Oxford) has been in routine use to overproduce IPNS under the control of the  $\lambda$ -phage  $P_L$  promoter to levels of up to 25% of the total bacterial protein. The recombinant IPNS is produced as insoluble inclusion bodies and has to be solubilised in 7M urea prior to purification [Aplin *et al.*, 1990]. Sequences necessary for the stable maintenance of the plasmid in *E.coli* together with the gene encoding tetracycline resistance ( $Tet_R$ ) are derived from the plasmid pBR322 [Bolivar *et al.*, 1977]. The expression of the IPNS is under the control of a  $P_L$  promoter and a  $cI857$  repressor, both co-resident on the plasmid. Due to the commercial origin of pIT353, details of additional termination and anti-termination sequences on the plasmid could not be obtained. *E.coli* host cells transformed with pIT353 were grown at 27°C in the absence IPNS expression. To induce expression, the incubation temperature is increased to 42°C.

The construction of pRS101, the expression vector for SCI, required the in-frame replacement of the IPNS gene sequence with the SCI synthetic gene. This was facilitated by an *NcoI* restriction site (C/CATGG) at the promoter/gene junction that contained the ATG translation initiation codon. Figure 3.1 illustrates the construction of pRS101 (for experimental details, see Chapter 5).

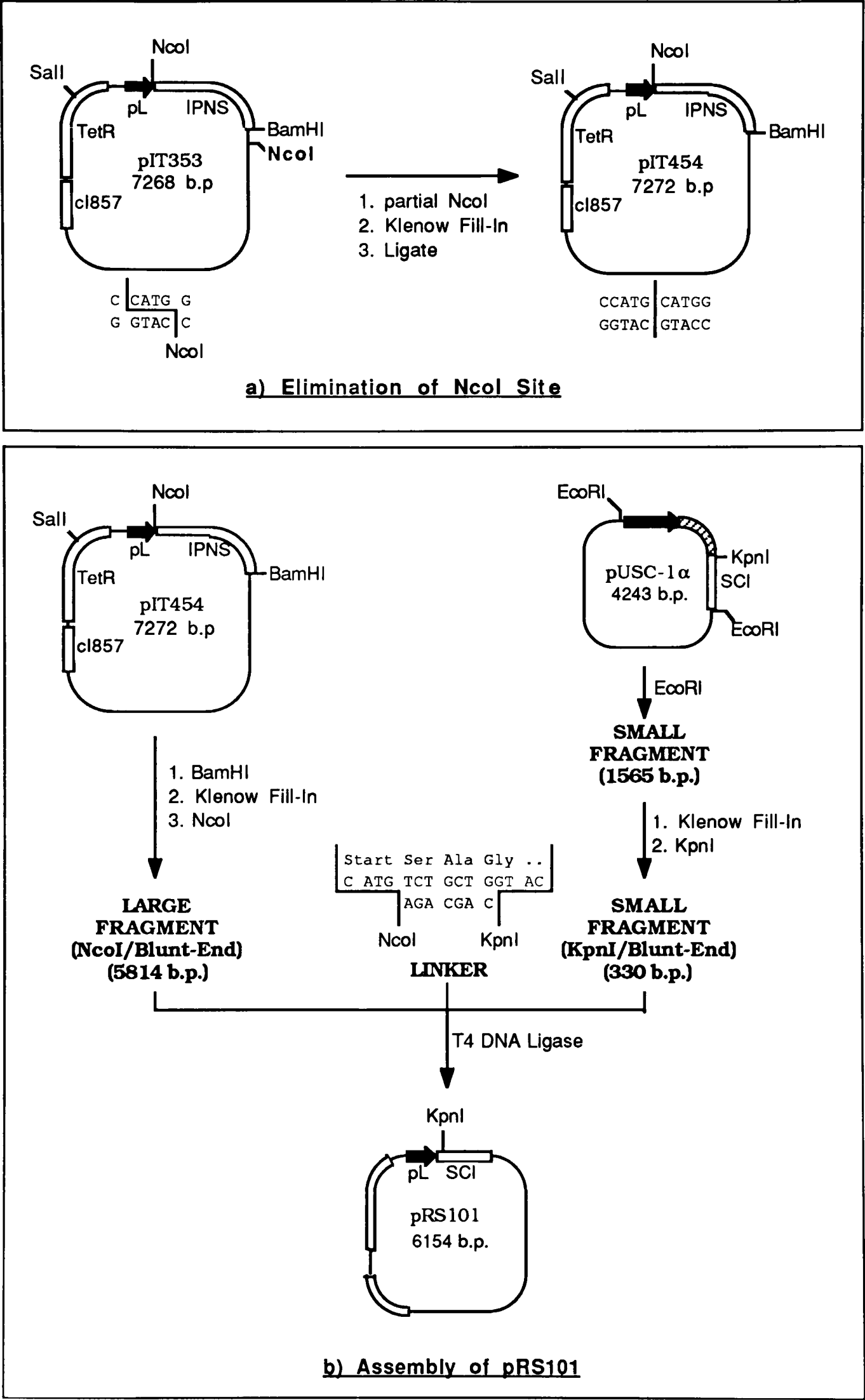


Figure 3.1: Construction of pRS101

In pIT353 (Appendix V), the IPNS gene can be excised as a 1458 b.p. *NcoI/BamHI* fragment. However, prior to removal of the IPNS gene, a second *NcoI* site located about 200 b.p downstream to the *BamHI* site was deleted, thereby, preventing loss of the 200 b.p. *BamHI/NcoI* fragment. Elimination of this *NcoI* site was brought about by "filling-in" the cohesive overhangs of the site with DNA polymerase I (Klenow fragment). Re-ligation resulted in the loss of the *NcoI* recognition site (Figure 3.1). The total ligation mix was then used to transform competent *E.coli* cells which were spread on LB-agar plates containing tetracycline. (N.B. To ensure effective repression of deleterious recombinant gene expression, heat-shock for DNA uptake, and growth of the transformed cells had to be carried out at 27°C). After incubation for 24h. at 27°C, about 40 colonies expressing tetracycline resistance were counted. Plasmid "minipreps" were obtained from these colonies. Plasmids (8 colonies) with a successfully eliminated *NcoI* site were identified as those that yielded a single fragment of 7272 b.p. upon digestion with *NcoI*. In the unmodified pIT353, an *NcoI* cut resulted in two distinct bands (5633 b.p. and 1635 b.p.). As to which of the two *NcoI* site had been eliminated, those plasmids with the correct elimination were identified by a *SalI* cut followed by an *NcoI* cut. In both cases the *SalI* cut yielded two fragments (5858 b.p. and 1414 b.p.). Those with the desired elimination could then be identified after the following *NcoI* cut in that the smaller *SalI* fragment (1414 b.p.) was cleaved into 2 fragments (1266 b.p. and 148 b.p.) (data not shown). Fortuitously, just one of the prepared plasmids had the required elimination. This modified plasmid, pIT454, was used for all subsequent cloning experiments.

For the removal of the IPNS gene from pIT454 (Figure 3.1), the plasmid was first linearised by incubation with *BamHI*. The cohesive overhangs resulting from the *BamHI* cut were filled-in with DNA polymerase I (Klenow

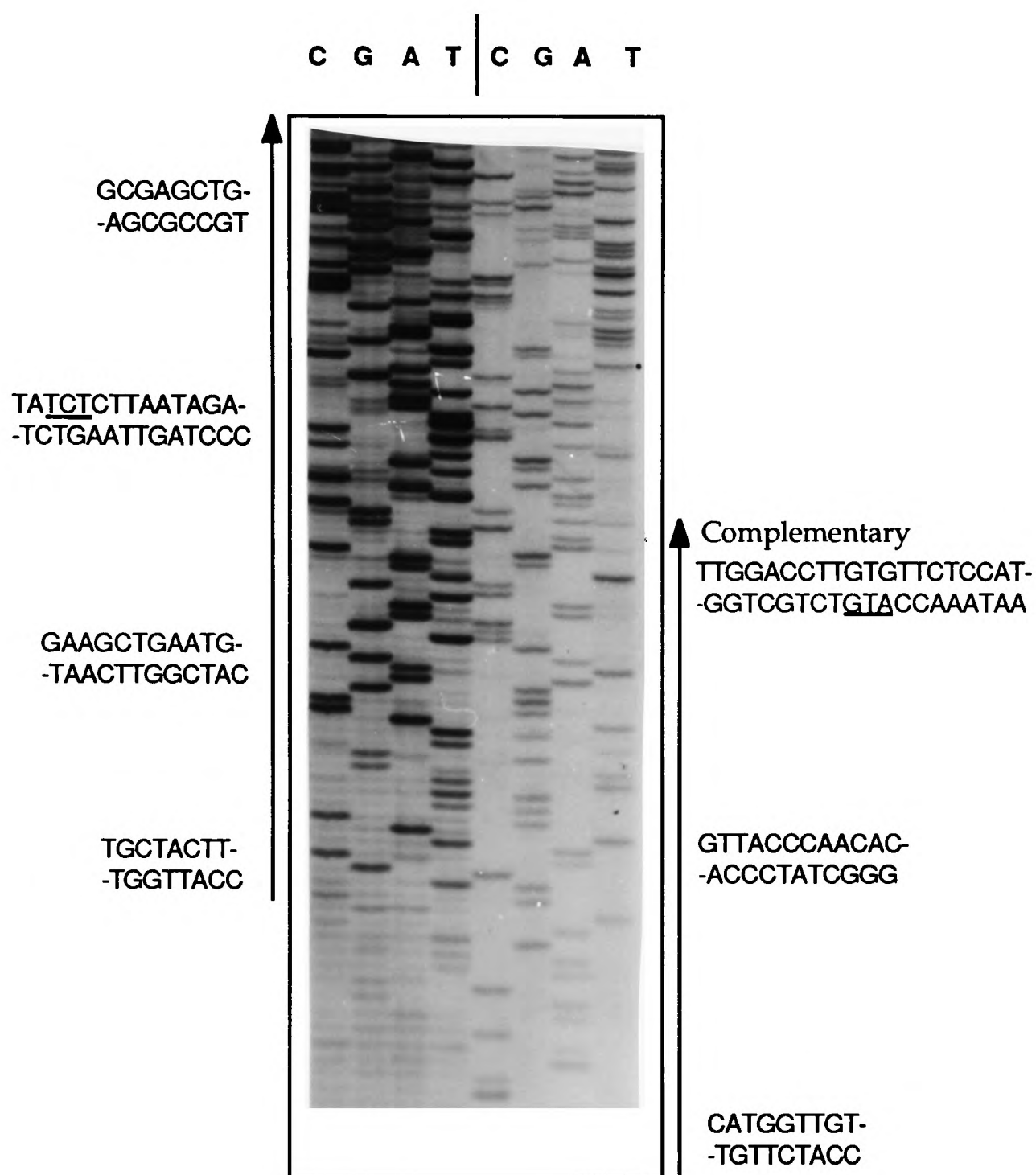
fragment). The removal of the IPNS gene was completed by digesting with the resulting blunt-ended fragment with *Nco*I. The resulting digest consisted of two fragments. The larger 5814 b.p. fragment (Fragment I) was isolated and purified.

The DNA sequences encoding the SCI gene were excised from the plasmid pUSC-1 $\alpha$  as the 330 b.p. *Kpn*I/*Eco*RI fragment (Fragment II). In order to facilitate correct, in-frame insertion of the SCI gene into the above expression vector, it has to be adapted, firstly, by filling-in the *Eco*RI sticky-end with DNA polymerase I (Klenow fragment) so as to make it compatible with the pIT454 blunt-end and, secondly, an *Nco*I/*Kpn*I linker was synthesised that incorporated the ATG start codon and the codons encoding the first three amino acids of SCI (Figure 3.1).

The expression plasmid pRS101 was finally completed by ligating the fragments I and II and the *Nco*I/*Kpn*I linker. The resulting ligation mix was used to transform competent *E.coli* JM109 cells which were smeared on LB-agar plates containing tetracycline and incubated for 24h at 27°C. About 100 positive transformants were counted. Transformants containing the correctly assembled plasmid (8 of the first 24 screened) were identified by restriction endonuclease screening of plasmid "minipreps". Extensive restriction endonuclease screening and double-stranded sequence analysis (Figure 3.2) of the new construct, pRS101, confirmed the correct insertion and orientation of the SCI gene in the expression plasmid. This expression plasmid, pRS101 (Appendix VI) was used for subsequent expression studies.

### **3.2.2. Expression Studies Using The SCI Expression Plasmid pRS101**

In the first instance, *E.coli* JM109 cells were transformed with pRS101. An overnight inoculum in 5ml SOB medium containing tetracycline was



**Figure 3.2:** Double-Stranded Sequence Analysis of pRS101:  
Autoradiograph of 6% Urea-Polyacrylamide Gel

Left Half: Sequence of the 3'- junction of the SCI gene  
Right Half: Sequence of the 5'- junction of the SCI gene

The plasmid was sequenced using the primers YK22-15  
and YK22-8, respectively [Kodo, 1988] (see Chapter 5)

diluted 1/10th in 25ml fresh SOB medium containing tetracycline and incubated at 27°C with vigorous shaking (300 rpm). The growth curve of the culture was monitored spectrophotometrically by measuring the turbidity of the medium ( $OD_{600}$ ) at regular intervals. Figure 3.3 illustrates a growth curve obtained. Typically, after a lag-phase of about 1h, where the cells adapt to the fresh growth medium and begin expression of antibiotic resistance, the cells begin to grow rapidly (log-phase) for about 6h. Thereafter, exhaustion of nutrients and the increased acidity of the medium cause the slowing and eventual halting of cell growth, resulting in a stationary phase, which was followed by gradual cell death and lysis.

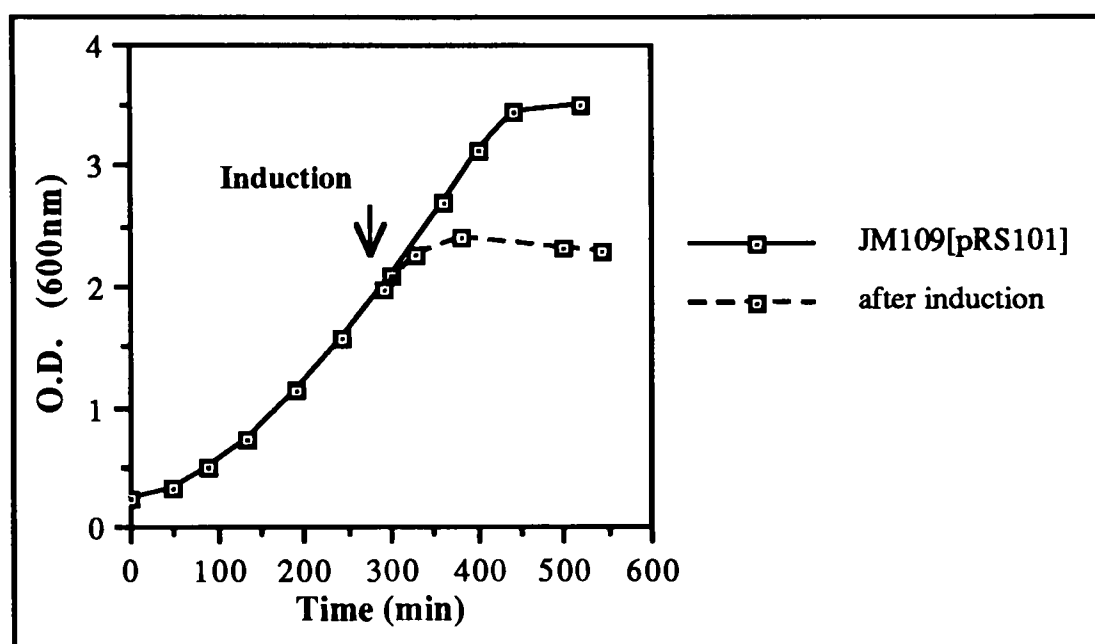


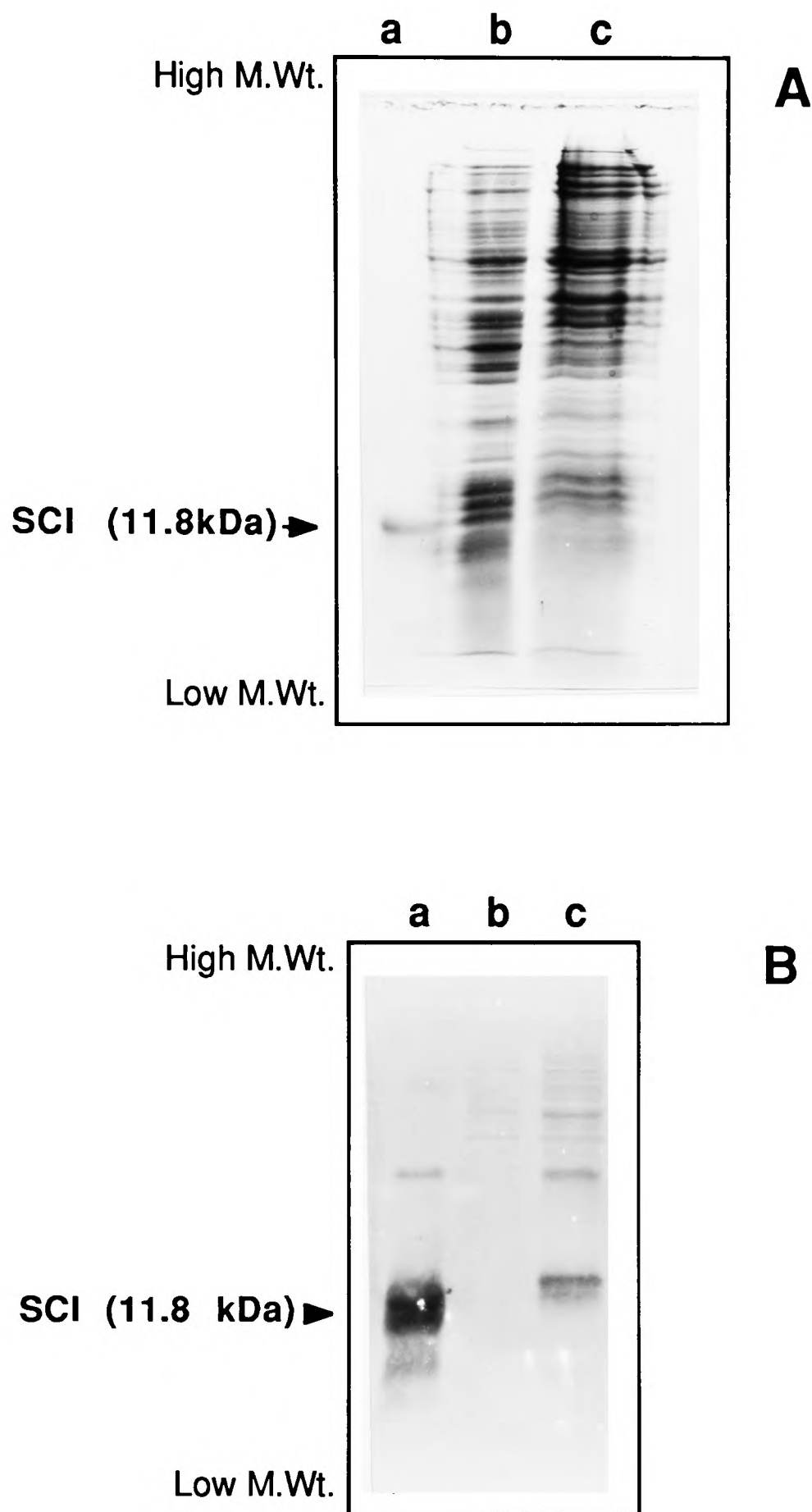
Figure 3.3: Expression of SCI in *E.coli*  
Growth of *E.coli* transformants.

For the induction of expression, the incubation temperature of transformants grown at 27°C in SOB medium that were in mid-log phase of growth ( $OD_{600} \approx 2.0$ ) was increased to 42°C. Incubation was continued overnight. Aliquots of the culture were removed at regular intervals to monitor the ensuing growth-rate and analyse cell protein extracts for the presence of recSCI.

Figure 3.3 (dotted line) shows the progress of cell growth after temperature-induction. It can be observed that the cells initially continue to

grow at a similar rate, but then, very rapidly stop growing. The following decrease in cell density indicated accelerated cell death and lysis which in turn, indicated the expression of activity toxic to the cell. Similar behaviour was not noted in cultures of untransformed JM109 cells where increase in temperature caused an increase in growth rate (data not shown).

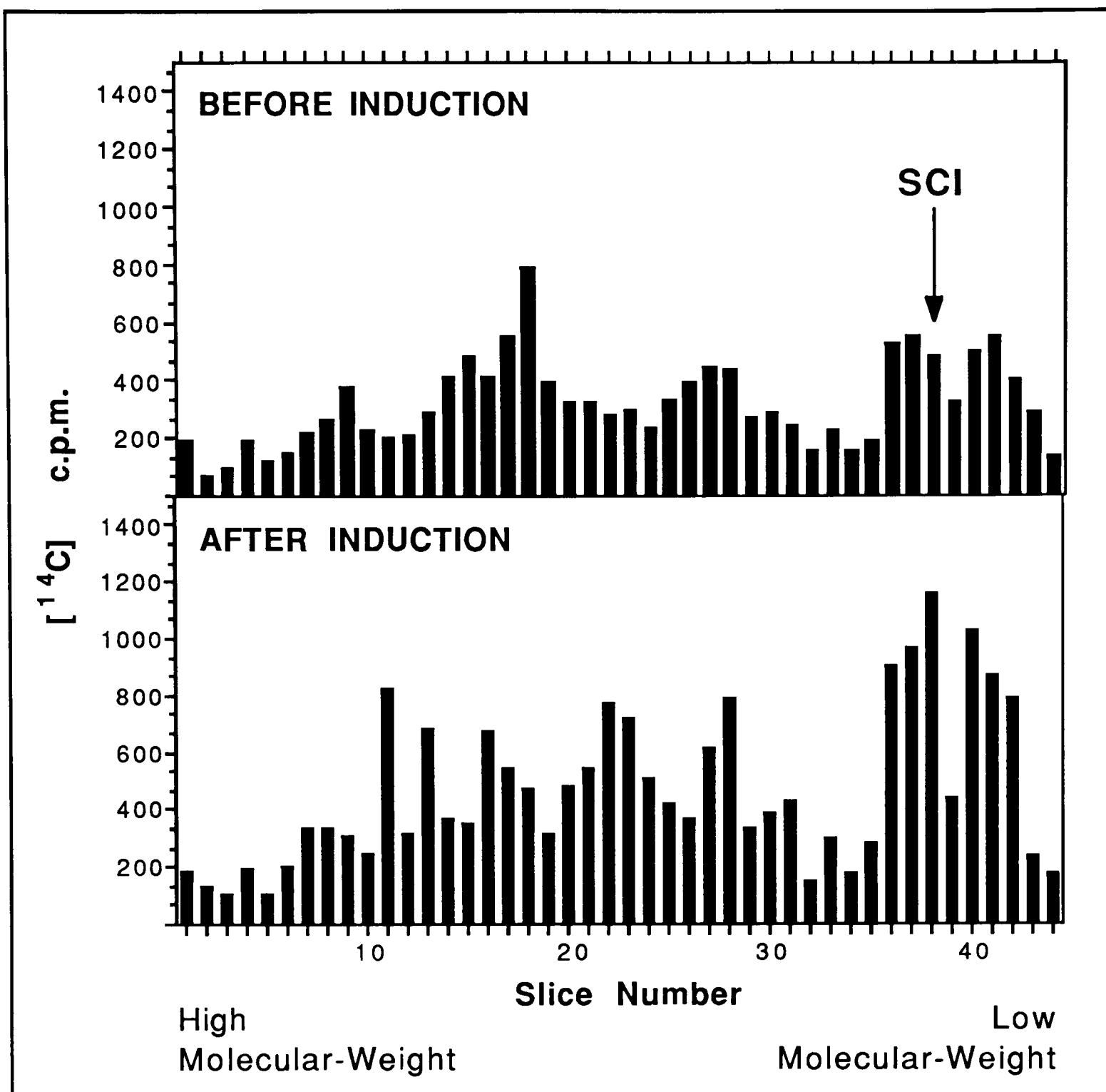
Total protein extracts of the cells were analysed by SDS-PAGE and Western blots for the presence of recSCI. Protein stains of the SDS-PAGE separated proteins using Coomassie Blue R and the more sensitive silver stain failed to reveal the appearance of any band corresponding in molecular weight to SCI. This indicated the failure to obtain any substantial overproduction of recSCI as the major protein component of the recombinant cells. Yet, Western blot analysis of the separated proteins using an anti-SCI polyclonal antibody confirmed, for the first time, that detectable quantities of recSCI were being produced after induction, reaching maximum levels after overnight incubation at 42°C. Figure 3.4 illustrates the results of SDS-PAGE and Western blotting analysis: Panel A illustrates Coomassie Blue R stained protein extracts separated by 15% SDS-PAGE and panel B shows the corresponding western blot. Note that in the Coomassie Blue stained gel the total protein extract of the JM109[pRS101] cells after overnight induction (lane c) shows no discernible difference to the corresponding uninduced cells (lane b) - in that no identifiable protein corresponding to the molecular weight of the native SCI (lane a) had been formed. However, in the corresponding western blot (panel B), the anti-SCI polyclonal antibody clearly recognises the presence of recSCI. Yet, the molecular weight of the recSCI appeared slightly larger than that of the authentic SCI. This discrepancy persisted in all subsequent expression systems, but disappeared on purification of the recombinant SCI. It could be attributed to tight association with a cellular component [Darby & Creighton, 1990] or, post-translational modification of the protein and/or anomalous electrophoretic behaviour of recSCI.



**Figure 3.4:** Cytoplasmic Expression of SCI in *E. coli*  
 Profiles of total cell protein extracts (reduced) from  
*E. coli* transformants on 15% SDS-polyacrylamide gel.  
 A: Coomassie Stain  
 B: Western Blot  
 Lane a: Authentic SCI  
 Lane b: JM109[pRS101]; without induction  
 Lane c: JM109[pRS101]; after induction

An estimate of the expression level was obtained by the method described by Remaut & co-workers (1987). An overnight inoculum of JM109[pRS101] was diluted to a cell density of  $2 \times 10^8$ /ml in fresh LB containing 5 $\mu$ Ci/ml of U-<sup>14</sup>C labelled protein hydrolysate (Amersham) and grown at 27°C. At the mid-log phase of cell growth, an aliquot (50 $\mu$ l) of the culture was removed and then expression induced, by increasing the incubation temperature to 42°C. After overnight induction, a further aliquot (50 $\mu$ l) was removed and both analysed for total cell protein extract on SDS-PAGE using a 9-18% gradient gel. After fixing the protein by Coomassie Blue R staining, each lane was cut into slices approximately 1mm thick. The radioactivity in each sample was counted using a scintillation counter. The obtained profiles are illustrated in Figure 3.5. Note the relative increase in radioactivity in fractions 36-39 when compared to the profile of the remaining region corresponding to the higher molecular weight proteins. This region corresponds to the bands of the molecular weight of SCI. Control growths using untransformed JM109 and JM109[pIT353] showed no parallel increase. From this experiment, the expression level of recSCI was estimated at about 2% of the total cell protein. Though encouraging, this expression level of recSCI was disappointingly low and featured no improvement, quantitatively or qualitatively, on the levels previously obtained by Kodo (1988) using a yeast expression system. The same *E.coli* expression system had been used to successfully obtain up to 25% isopenicillin N synthase (IPNS) [Aplin *et al.*, 1990].

With the aim of improving the level of expression, the SCI expression vector, pRS101 was used to transform various available strains of *E.coli* including MC1061, CJ236, TG1, NM522 and JM105. Although marginal strain-dependent differences in growth rates were noted, SDS-PAGE and Western blot analysis using a polyclonal anti-SCI antibody (a kind gift from M.P.



**Figure 3.5:** Radioactive profile of <sup>14</sup>C-labelled protein extracts of *Escherichia coli* JM109[pRS101].

An overnight inoculum of *E.coli* JM109[pRS101] was diluted to a cell density of  $2 \times 10^8$  cells/ml in fresh LB-medium containing 5 $\mu$ Ci/ml of U-<sup>14</sup>C labelled protein hydrolysate. The cells were grown to mid-log phase (OD<sub>600</sub>~2) at 27°C and expression induced by increasing the incubation temperature to 42°C. The total protein extracts of the cells from 50 $\mu$ l aliquots of the media were separated by SDS-PAGE on a 9-18% gradient gel. After fixing the proteins by Coomassie blue staining, each lane was cut into 1mm slices. The individual slices were dissolved in organic tissue solubiliser (TCS; Amersham) and the radioactivity in each slice in organic counting scintillant (OCS; Amersham) monitored by scintillation counting.

Esnouf, Nuffield Department of Clinical Biochemistry, University of Oxford) showed no notable improvements in expression level (data not shown). JM109[pRS101] transformants were grown and expression induced in different rich-media, including L-broth and SOB medium at different pH values ranging from 6.5 to 8.0. Yet again, though differences in growth rates were noted, the faster rates in more basic conditions (Figure 3.6), the high expression levels expected of this system remained elusive as no improvement in expression was noted. Analysis of cell lysates failed to locate trypsin and  $\beta$ XIIa inhibitory activity. This indicated that the low levels of recSCI produced were not in the native conformation. In spite of the optimised transcriptional and translational signals in pRS101, only low levels of expression of recSCI were obtained. This could be due to several reasons including the toxicity of the gene product, induction of heat-shock response and unfavourable codon usage.

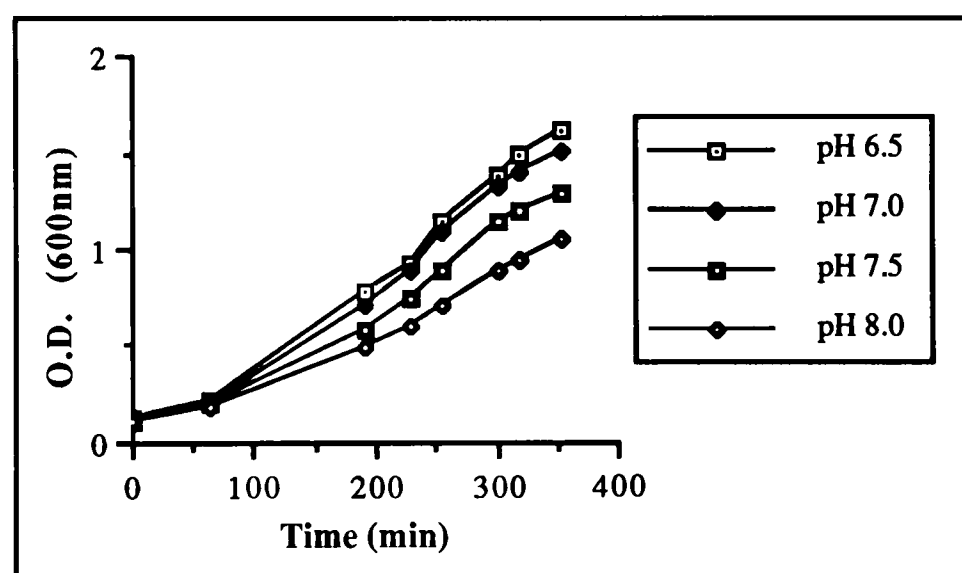


Figure 3.6: Comparison of growth rates of *E.coli* JM109[pRS101] in SOB medium at different pH values.

### 3.2.3. The *dnaY* Gene Product

Brinkmann & co-workers (1989) reported on the correlation between the availability of the *E.coli dnaY* gene product and the level of expression of various recombinant eukaryotic genes together with its effect on the efficiency of DNA replication and cell growth. The *dnaY* gene was shown to produce a 77 nucleotide transcript which corresponded to a minor tRNA that recognises

the rare arginine AGA and AGG codons [Garcia *et al.*, 1986]. Sharp & Li (1987) set up the codon adaptation index, in which they compared the codon usage in genes of several highly expressed proteins in *E.coli* and *S.cerevisiae*. This was based on the knowledge that there was a clear positive correlation between the degree of codon bias and level of gene expression in both organisms. With the codon adaptation index, the most efficiently translated codon of the alternative synonymous codons for an amino acid can be determined and the relative extent to which the other synonymous codons are disadvantageous determined [Sharp & Li, 1987]. This, so-called major codon bias (Appendix VIII) has been shown to be directly correlated to the level of cognate tRNA present in the cell [Ikemura & Ozeki, 1982; Kurland, 1991]. In the results compiled by Ikemura & Ozeki (1982) and Sharp & Li (1987), yeast shows a definite preference for the Arg codon AGA, for which reason the synthetic SCI gene was designed using this codon [Kodo, 1988]. However, in *E.coli*, CGU is the preferred Arg codon. The 8 Arg residues of SCI in its synthetic gene were all encoded by the rare yeast-preferred codon, AGA (Figure 2.1). This is 7.3% of the total codon content of the SCI gene. Brinkmann & co-workers (1989) showed that translated regions containing more than 3% rare Arg (AGA and AGG) codons were expressed at levels of less than 5% of the total bacterial protein, showing reductions to expression levels below 1% in genes with 5% rare Arg codons. Furthermore, high rare-codon content caused decreased cell growth and DNA replication [Brinkmann *et al.*, 1989]. The incorporation of the *dnaY* gene product into the expression system resulted in the dramatic improvement of expression of these proteins to levels over 30% of the total bacterial protein and also improved cell viability and plasmid stability [Brinkmann *et al.*, 1989].

Noting the high rare-codon content in the SCI gene (7.3%), it was decided to supplement the JM109[pRS101] expression system with the *dnaY* gene product. This was achieved by co-transforming JM109[pRS101] with the

plasmid pUBS520. This plasmid, a kind gift from Dr. P. Buckel (Department of Genetics, Boehringer Mannheim GmbH, Penzberg, Germany), contains the *E.coli dnaY* gene together with the *E.coli lacI<sup>q</sup>* gene and a kanamycin resistance gene [Brinkmann *et al.*, 1989]. Successful double-transformants were selected for by growing the resultant transformants on media containing both tetracycline and kanamycin for phenotypic selection.

An overnight inoculum of a single, well separated colony grown at 27°C in SOB (pH 7.5) containing tetracycline and kanamycin was diluted 1/10th in fresh growth medium and incubated at 27°C with vigorous shaking (300rpm). Figure 3.7 illustrates the growth curve at 27°C of JM109[pRS101+pUBS520] compared with that of JM109/pRS101. The growth behaviour of both transformants is essentially similar.

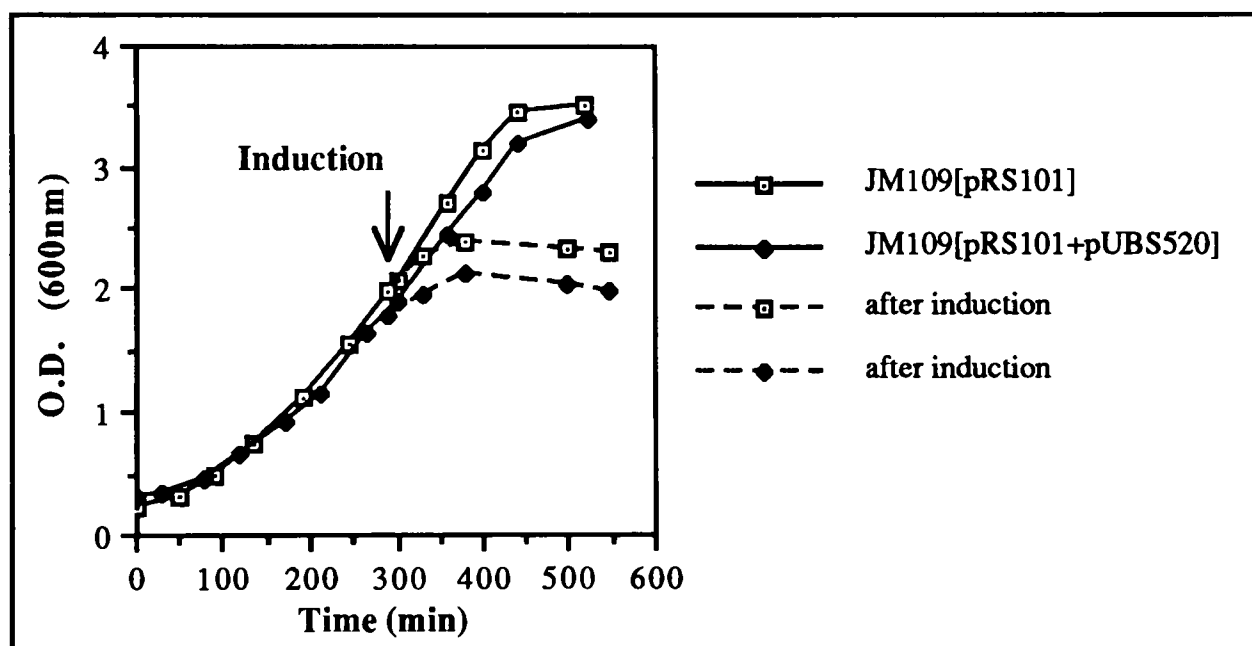


Figure 3.7: Growth Curves of *E.coli* JM109[pRS101] in the presence and absence of the *dnaY* gene product (pUBS520)

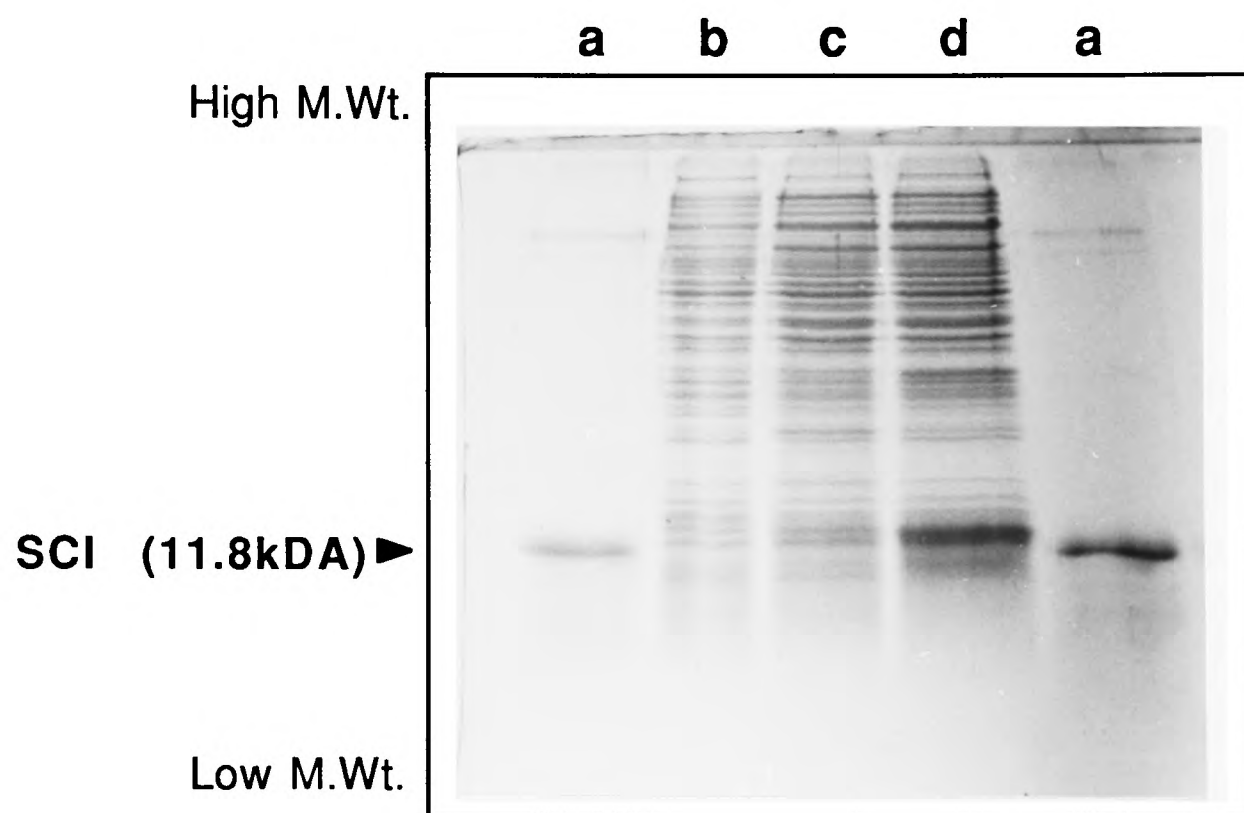
For the comparison of expression rates, induction of expression was achieved by increasing the incubation temperature of cells in mid-log phase ( $OD_{600} \approx 2.0$ ) of growth from 27°C to 42°C. The cells were incubated overnight at this permissive temperature and aliquots of the growth media removed to monitor the growth curve of the cultures and analyse the cells for protein

expression. Both JM109[pRS101] and JM109[pRS101+pUBS520] showed similar growth behaviour (Figure 3.7).

On comparison of the total protein extracts of the *E.coli* transformants by SDS-PAGE (Figure 3.8), it was noted that the expression level of recSCI was dramatically enhanced upon incorporation of the *dnaY* gene product (in the JM109[pRS101+pUBS520] double-transformants). The level of expression of recSCI, estimated at around 10% of the total bacterial protein resulted, for the first time, in a clearly visible band of approximately the required molecular weight on a Coomassie Blue stained protein gel (Figure 3.8). This was confirmed by a western blot, using the anti-SCI polyclonal antibody (result not shown).

#### **3.2.4. Purification and Characterisation of recSCI Inclusion Bodies**

It was important to determine the morphology of the expressed recSCI in the cytoplasm in order to devise a suitable strategy for its isolation and purification. As was previously mentioned, over-expressed heterologous eukaryotic proteins in *E.coli* aggregate in many cases to form dense, insoluble inclusion bodies [Marston, 1987]. In the majority of cases, the eukaryotic protein is incorrectly folded, and therefore, inactive. The inclusion bodies formed as a result of the aggregation of recombinant proteins usually contain a few contaminating homologous proteins that inevitably co-aggregate. Yet, the major component of the inclusion bodies is the recombinant protein itself. The compactness and high density of the inclusion bodies (shown by phase-contrast microscopy), facilitate their isolation from the other bacterial cell lysate components by sedimentation with low-speed centrifugation [Marston, 1987].



**Figure 3.8:** Expression of SCI in *E.coli*  
 Profiles of the total cell protein extracts (reduced)  
 of *E.coli* transformants on Coomassie stained  
 15% SDS-polyacrylamide gel.

Lanes a: Authentic SCI  
 Lane b: JM109[pRS101+pUBS520]; without induction  
 Lane c: JM109[pRS101]; induced  
 Lane d: JM109[pRS101+pUBS520]; induced

For the analysis of the recSCI obtained, a 3.5-litre culture of JM109[pRS101+pUBS520] in SOB (pH 7.5) was prepared:

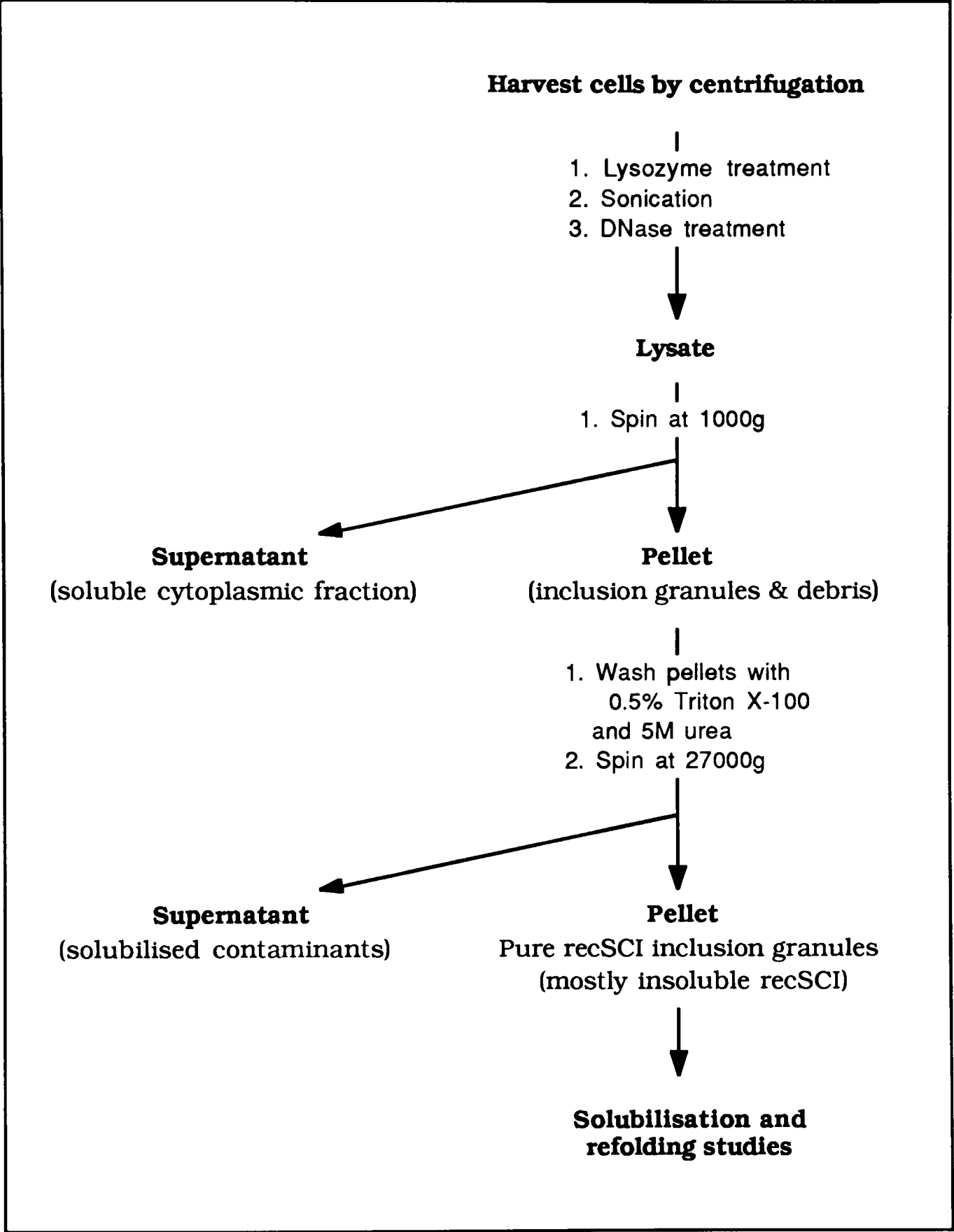
An overnight culture grown from a single colony of JM109[pRS101+pUBS520] in 350ml SOB (pH 7.5) containing tetracycline and kanamycin was diluted 1/10th in 3.5 litres fresh medium contained in 3 4-litre conical flasks and incubated with shaking (250 rpm) at 27°C for 4h until mid-log phase of growth was reached ( $OD_{600} \approx 2.0$ ). At this point, expression was induced by adding an equal volume of medium heated to 57°C and increasing the incubator temperature to 42°C. Incubation was continued overnight. Cells were harvested by centrifugation for 20min at 10,000g at 4°C, washed in ice-cold lysis buffer (50mM Tris-HCl pH 8.0, 1mM EDTA & 100mM sodium chloride) and re-centrifuged. All manipulations at this point and beyond were carried out at 4°C (unless otherwise stated) so as to minimise the proteolytic degradation of the recombinant protein. A total of 23g wet-weight of cells was obtained.

Disruption and fractionation of the cells were carried out as described by Marston (1986). A quantity of 6g wet-weight of the cells were resuspended in 35ml lysis buffer containing 1M sucrose. To reduce proteolytic activity, phenylmethylsulphonyl fluoride (PMSF) was added to the suspension to a final concentration of 0.15mM. The suspension was homogenised using a tissue homogeniser and lysozyme added to a final concentration of 1mg/ml. The completion of spheroplast formation was monitored by diluting a small portion of the suspension 50-fold in water. Spheroplast formation was taken to be complete when the optical density at 450nm fell 80-85% within 10 sec. The suspension was incubated for 30 min. on ice with occasional stirring and then sonicated on ice with two 15 sec bursts. The successful disruption of the spheroplasts was indicated by the very viscous consistency of the resulting suspension. This was due to leakage of *E.coli* chromosomal DNA. A quantity

of 200µl of DNaseI (1mg/ml in lysis buffer) was added and the suspension incubated on a 37°C water bath for 10 min. after which time it was no longer viscous. The cell lysate obtained was used for further cell fractionation and recSCI purification experiments.

Whether the SCI was in the inclusion bodies was tested by the centrifugation of small aliquots of the lysate at various speeds (150g-11,500g at 4°C) for 10 min. The recSCI was found to be predominantly sequestered in inclusion bodies as, even with speeds as low as 500g, most of the detectable SCI antigenic activity was located in the pellets as was ascertained by SDS-PAGE and Western blots. Due to the high density of the inclusion bodies the low-speed centrifugation step facilitated their effective separation from the soluble proteins in the cell lysates. This was carried out by centrifuging the lysate at 1000g for 10 min.

The pellet obtained (2.7g wet-weight) was resuspended in 5ml lysis buffer. Aliquots of 100µl were transferred into 1.5 ml tubes and centrifuged in a microfuge at 13000g for 15min. The pellets were washed with 1ml each of lysis buffer, containing either 0.5% (v/v) of the non-ionic detergent Triton X-100, or urea at concentrations between 0.5M and 7M, or both urea (0.5M-7M) and 0.5% Triton X-100 and immediately centrifuged for 10 min at 27,000g at 4°C. The pellets and supernatants were analysed by SDS-PAGE so as to determine the optimal urea concentration that, together with the detergent Triton X-100, could be used to remove the contaminating proteins and lipid sub-structures that are more loosely associated with the inclusion bodies. A combination 0.5% Triton X-100 and 5M urea produces a remarkably high degree of purity of recSCI in the precipitated inclusion bodies. Table 3.1 summarises the procedure for the isolation and purification of the recSCI inclusion bodies.

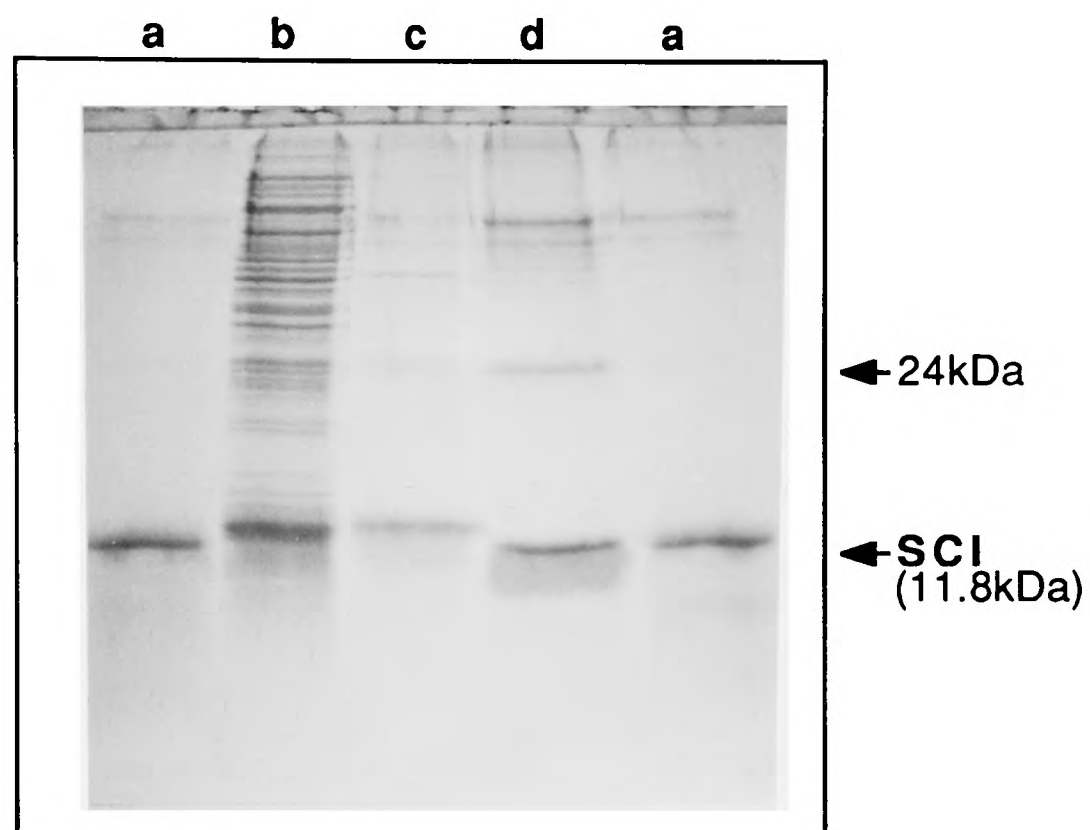


**Table 3.1:** Isolation and Purification of recombinant SCI Inclusion Bodies

Figure 3.9 illustrates the SDS-PAGE results of protein extracts at the various stages of inclusion body purification. Note the discrepancy in size of the recSCI in the total bacterial extract (lane b) and the unwashed inclusion bodies (lane c) as compared with that of the native SCI (lane a). The slightly slower migration could possibly be due to the tight association of the recSCI with some component of the *E.coli* cell [Ellis, 1990; Darby & Creighton, 1990]. This association seemed to be broken on washing the inclusion bodies in 5M urea and 0.5% Triton X-100 as the recSCI in the obtained inclusion bodies showed no perceivable difference in molecular weight (lane c). The purity of the obtained recSCI inclusion bodies was deemed adequate for preliminary solubilisation and refolding studies. The higher molecular components in the purified recSCI (lane d), of which traces were also apparent in authentic SCI (lane a) are possibly the dimeric ( $\approx 24$  kDa) and polymeric forms of the protein.

For solubilisation of the recSCI inclusion bodies the pellet (1g) was resuspended in 5ml 20mM Tris-HCl pH 7.8 containing 1mM DTT, 5mM EDTA and 6M guanidine hydrochloride (GuHCl). The suspension was gently stirred (avoiding foaming) for 12h at 4°C. A clear suspension indicated that the recSCI inclusion bodies were successfully solubilised. Any undissolved cell debris was removed by centrifugation of the suspension for 10min at 10,000g at 4°C.

Noting the high protein concentration ( $>50$ mg/ml) in the suspension and the poor solubility of SCI ( $<1$ mg/ml), aliquots of the suspension were diluted to a concentration of 20-50 $\mu$ g/ml in the above buffer prior to the removal of GuHCl during refolding studies.



**Figure 3.9:** Purification of recombinant SCI Inclusion Bodies: Protein samples (reduced) were resolved by 15% SDS-PAGE and stained with Coomassie blue.

Lanes a: Authentic SCI  
 Lane b: Total bacterial protein extract of JM109[pRS101+pUBS520] (after induction)  
 Lane c: Inclusion bodies (crude)  
 Lane d: Inclusion bodies (washed)

### 3.2.5. Attempted Refolding of Solubilised recSCI to Yield the Biologically Active Inhibitor

The very nature of inclusion bodies, with the recombinant protein aggregated in insoluble and, in most cases the non-native, inactive conformation requires the two stage process of unfolding and solubilisation of the inclusion bodies followed by the refolding of the protein under selective and controlled conditions. The conditions required for the effective *in vitro* refolding of the solubilised protein vary greatly with each system and each protein used. Marston (1987) has compiled examples of published protocols of refolding procedures used for various recombinant proteins solubilised from inclusion bodies. These carefully selected “environmental factors” required for the optimal refolding of proteins include temperature, protein concentration, redox buffer and labilising agents [Buchner & Rudolph, 1991]. The listed protocols were used as a guideline for developing a suitable refolding strategy for the obtained recSCI.

Solubilisation of the inclusion bodies was brought about with a combination of high concentrations of GuHCl, and the thiol agent dithiothreitol (DTT) which reduces the covalent inter- and intramolecular disulphide bonds. The correct pairing of the ten Cys residues in SCI may be important for the correct folding of the recombinant protein. This requires the controlled and gradual removal of both GuHCl and DTT, enabling correct disulphide pairings and therefore correct folding of the protein. Work by M.P. Esnouf (personal communication) on the authentic SCI extracted from sweet corn showed that following unfolding of the protein in 6M GuHCl and 0.1M DTT, upto 60% of the initial total inhibitory activity was recovered on slow removal of the reagents by dialysis.

Aliquots of the solubilised recSCI inclusion bodies were used in an attempt to re-enact this refolding procedure, in which, 1ml suspension (see

above) containing about 1mg protein in 20mM Tris-HCl pH 7.8 containing 1mM DTT, 5mM EDTA and 6M GuHCl was degassed by vacuum evacuation and sealed in a dialysis tube (Spectra/Por; molecular weight cutoff 3500). The GuHCl and DTT concentrations were reduced, step-wise and gradually, by immersing the dialysis sample in 4ml, degassed renaturation buffer (20mM Tris-HCl pH 8.0, containing 50mM sodium chloride and 5mM EDTA) and successively replacing the latter buffer at 6h intervals over a period of 24h. This procedure failed to yield any trace of trypsin or  $\beta$ XIIa inhibitory activity from the recSCI.

Other non-enzymatic methods summarised by Marston (1987) and Buchner & Rudolph (1991) and variations thereof were utilised for the attempted renaturation of recSCI solubilised from inclusion bodies. This included the creation of "oxido shuffling" conditions that encourage the breaking of the incorrect, intrinsically labile disulphide bonds and promote the formation of the correct, energetically stable native disulphide pairings [Buchner & Rudolph, 1991]. These oxido shuffling conditions are commonly obtained by adding the low-molecular weight thiol compound, glutathione to the renaturation buffer [Buchner & Rudolph, 1991]. In several published procedures for the refolding of proteins similar optimal ratios of the reduced glutathione (GSH) to oxidised glutathione (GSSG) (e.g. 2mM GSH and 0.2mM GSSG) in the renaturation buffer have been determined [Saxena & Wetlaufer, 1970; Buchner & Rudolph, 1991]. The following, basic procedure was adopted:

A 1ml aliquot of recSCI inclusion bodies solubilised in 50mM Tris-HCl pH 8.0 containing 1mM EDTA, 0.3M DTT and 6M GuHCl at a concentration of 20-50 $\mu$ g/ml protein was dialysed at 4°C for several days in renaturation buffer containing 5mM GSH and 0.5mM GSSG. Aliquots were removed at regular intervals to check for trypsin and  $\beta$ XIIa inhibition. No inhibitory activity was detected at any point during the dialysis. Analysis of the protein

obtained by SDS-PAGE confirmed that it had not been degraded and was indistinguishable from the authentic SCI.

Several variations of the above procedure also failed to yield native SCI. These included changes in the ratios and concentrations of GSH (2-10mM) and GSSG (0.2mM-1mM) and in further experiments GSH was replaced by DTT as the reducing agent (3mM DTT and 4mM GSSG) or the preferential destabilisation of incorrectly folded protein was induced by the addition, as "labilising agent", of 0.5M L-arginine to the renaturation buffer [Orsini & Goldenberg, 1978].

A further method attempted was as suggested by Hoess & co-workers (1988) where they successfully solubilized insoluble recombinant proteins in bacterial lysates to yield the active proteins by treatment of the lysates with ion exchange resin: A quantity of 10mg wet weight of recSCI inclusion bodies were resuspended in 0.5ml renaturation buffer containing 1mM DTT, 1mM PMSF and 10% (v/v) glycerol and mixed with 0.5ml S-Sepharose Fast-Flow resin (Pharmacia) equilibrated in the same buffer and shaken at 4°C for 2h. The resin together with the bound protein was separated from the buffer by centrifugation for 5 min. in a microfuge at 13000g. The protein was eluted by resuspending the resin in 0.5ml of the same buffer, containing 0.5M sodium chloride. As ascertained by SDS-PAGE, solubilisation of the inclusion bodies was successful in that the elution buffer contained recSCI as the major component. But, yet again, it failed to yield any traces of inhibitory activity.

The repeated failure to obtain even traces of the active native form of recSCI by the standard solubilisation and refolding procedures could have been due to several reasons. The features of the final protein product are a function and result of the expression of the protein and the methods used to recover the active, soluble protein [Marston & Hartley, 1990]. One or more of

the following reasons could result in the failure to obtain the native, functional form of the protein being produced:

1. A modified N- and/or C-terminus
2. The lack of post-translational modification
3. Chemical modifications resulting from solubilisation conditions (e.g. residual cyanates in urea can react with amino groups to form carbamylated derivatives, high pH can cause deamidation, low pH can cause Met oxidation and the conversion of cysteine to cysteic acid)
4. The lack of conformational authenticity of the refolded molecule [Marston & Hartley, 1990].

It was difficult to ascertain the exact reason for the failure to obtain the native, correctly folded recSCI. Double-stranded DNA sequencing of the entire length of the SCI gene in pRS101 together with the gene junctions on both the 5'- and 3'-end (Figure 3.2) confirmed that the sequence of the gene was correct and had not undergone any *in vivo* mutations.

A sample of the purified inclusion bodies (0.5mg/ml) was submitted for N-terminal sequencing by automated Edman degradation. Only very small traces of this relatively large concentration of the recSCI were shown to have the correct N-terminal sequence, that is, Ser-Ala-Gly-Thr-Ser- confirming the presence of recSCI (A.R. Willis; personal communication). The majority of the protein appeared to be blocked at the N-terminal amino group. Such a blockage could have two origins, firstly, the high-level of expression of the recombinant polypeptide can lead to the rapid formation of insoluble and inaccessible inclusion bodies, too fast for enzymatic deformylation [Livingston & Leder, 1969] and subsequent removal of the N-terminal Met [Ben-Bassat *et al.*, 1987; Yasueda *et al.*, 1990] (see section 3.1.5). Secondly, the chemical derivatisation of the N-terminal amino group during the extraction of the inclusion bodies (e.g. residual cyanates in urea used in

washing the inclusion bodies can react with amino groups to form carbamylated derivatives [Marston & Hartley, 1990]).

Further investigation of the recSCI would be necessary in order to determine the precise nature of its N-terminus and whether an incorrectly processed N-terminal Met or any other modifications can be implicated in the failure of the obtained recSCI to fold into the native and active form. It is also plausible, but less probable that the recSCI together with other inclusion body components preferentially adopted an abnormally stable configuration by inter- or intramolecular interactions or “concealed” disulphide linkages that were not disrupted by the conventional methods described above.

### 3.2.6. Low-Temperature Expression

The formation of inclusion bodies is a general problem associated with induction at high temperatures and may involve the redistribution of induced proteins into insoluble fractions [Yatvin *et al.*, 1986] due, in some cases, to perturbations in the protein folding pathway at high temperatures [Lowman & Bina, 1990]. The growth temperature has been shown to drastically affect the formation of a variety of inclusion bodies by a variety of polypeptides expressed in *E.coli* [Bowden & Georgiou, 1990]. However, the direct *in vivo* correlation of temperature to the tendency for polypeptides to aggregate cannot be made as temperature also has a pleiotropic effect on the cell physiology - processes such as translation and the interaction of the partially folded polypeptide with chaperonins show a strong temperature dependence [Bowden & Georgiou, 1990].

The induction of expression of recSCI was attempted at lower temperatures in the hope that a soluble form of the protein may be produced and therefore facilitating post-translational processing of the N-terminal fMet. Lowman & Bina (1990) showed that optimal expression of recombinant

proteins utilising the  $\lambda P_L$  promoter and the heat-sensitive cI857 repressor was obtained at by incubation for 6h. at temperatures as low as 36°C. Furthermore, Poindexter & Gayle III (1991), utilising the same promoter and repressor obtained enhanced expression levels of recombinant proteins at 30°C by shifting the pH of the growth medium of transformants that had entered the stationary phase of growth to the basic value of 9. This caused the SOS induced deactivation of cI857 and resulted in the continued growth of the cells with high level accumulation of the recombinant protein [Poindexter & Gayle III, 1991]. The cells continued to grow until the culture medium again attained a pH of about 6.5. This characteristic acidification of the culture medium causes inhibited cell growth and the cells attaining stationary phase even with an abundance of nutrients. The maintenance of the pH of the growth medium at 9 by the addition of NaOH resulted in the further continuation of cell growth to OD<sub>600</sub> values as high as 20 as compared to the values of around 4 obtained by temperature induction [Poindexter & Gayle III, 1991].

The effect of temperatures between 35 and 42°C on the induction of recSCI expression in JM109[pRS101+pUBS520] was studied adopting the same procedure described in section 3.2.3. in which overnight cultures of the transformants in SOB were diluted 1/10th in SOB and grown at 27°C. At the mid-log phase of growth, the cultures were transferred to incubators at temperatures of 36, 38, 40 and 42°C. After overnight incubation, aliquots of the cells were analysed by SDS-PAGE and Western blotting for recSCI. At the induction temperatures of 38°C and below, though expression at trace levels was detected, no marked improvement in expression was achieved, as was determined by Western blotting. Maximum levels of expression were observed at the induction temperature of 42°C. The low temperature induction as suggested by Lowman & Bina (1990) was therefore not suitable for the overexpression of recSCI.

For induction by alkaline shift [Poindexter & Gayle III, 1991], JM109[pRS101+pUBS520] was grown overnight at 27°C in 25ml supplemented superbrotth containing M9 minimal salts and 1% (w/v) glucose. At stationary phase ( $O.D_{600} \approx 7.0$ ), the pH of the resulting culture was 6.5. For induction of expression, the pH was shifted to 9.0 by the addition of 390  $\mu$ l 10N sodium hydroxide. Contrary to the results obtained by Poindexter & Gayle III (1991) no further growth of the cells was observed and analysis of protein by SDS-PAGE and western blotting of extracts of cells following overnight growth at 27°C and pH 9 indicated that no detectable levels of recSCI were produced.

In the low-temperature induction methods described by Lowman & Bina (1990) and Poindexter & Gayle III (1991), the cI857 repressor was provided by lysogenic strains of *E.coli* containing the cI857 gene. This is unlike the SCI expression system used here where the cI857 gene is co-resident on the expression plasmid. The higher gene dosage of cI857 provided by the latter could possibly have been the reason that recSCI expression was not induced as induction conditions were not sufficient for the complete inactivation of the higher levels of cI857 produced.

As a further attempt to obtain recSCI expression at low temperatures, the plasmid pRS201 (4570 b.p.) was constructed (Appendix VII). In this plasmid, transcription of the heterologous gene was under the chemically inducible control of the powerful *trc* promoter ( $P_{trc}$ ) [Russell & Bennett, 1982; de Boer *et al.*, 1983], a fusion promoter consisting of the operator and -35 regions of the inducible lactose operon (*lac*) promoter and the Pribnow box (-10 region) and Shine-Dalgarno (SD) sequence derived from the powerful tryptophan (*trp*) promoter. The plasmid contained a chloramphenicol acetyltransferase (CAT) gene which facilitated phenotypic selection of transformants on media containing the antibiotic, chloramphenicol. The *lac* operator region facilitated inducible expression whereby expression is

effectively switched off during growth in minimal media. Expression was initiated by the addition of the inducer, isopropylthiogalactoside (IPTG). This plasmid was successfully used to express very high levels (over 30% of the total bacterial protein) of soluble IPNS [R.J. Heath; personal communication]. We therefore had reason to believe that similar results could be obtained with *recSCI*.

The construction of the expression plasmid pRS201 (Figure 3.10) was essentially the same as the construction of pRS101. An *NcoI* site containing the ATG initiation codon directly downstream to the *trc* promoter and SD sequence facilitated the correct, in-frame insertion of the SCI gene. The IPNS in the plasmid pRH101 was excised as the 1458 b.p. *NcoI/BamHI* fragment. The sequences for the SCI gene were obtained as the 330 b.p. *KpnI/BglII* fragment from pUSC-1 $\alpha$ . Together with the same *NcoI/KpnI* adapter used in the construction of pRS101, the SCI gene was ligated with the 4120 b.p. *NcoI/BamHI* fragment of pRH101 (cleavage with both *BamHI* and *BglII* create compatible single-stranded overhangs). The correct insertion of the SCI gene in pRS201 (4570 b.p.) was confirmed by extensive screening with mixed restriction endonuclease digests.

For expression studies, several *E.coli* strains were transformed with pRS201 and double-transformed with both pRS201 and pUBS520. The transformants were grown at 37°C in M9 minimal medium containing casamino acids (M9CA) and antibiotics (see chapter 5) required for phenotypic selection (chloramphenicol for the single transformants and chloramphenicol and kanamycin for the double-transformants). Expression was induced at the mid-log phase of growth by the addition of IPTG to a final concentration of 5mM and incubation continued overnight. The investigation of protein content of cell extracts by SDS-PAGE and Western blotting unfortunately failed to detect any production of *recSCI*.

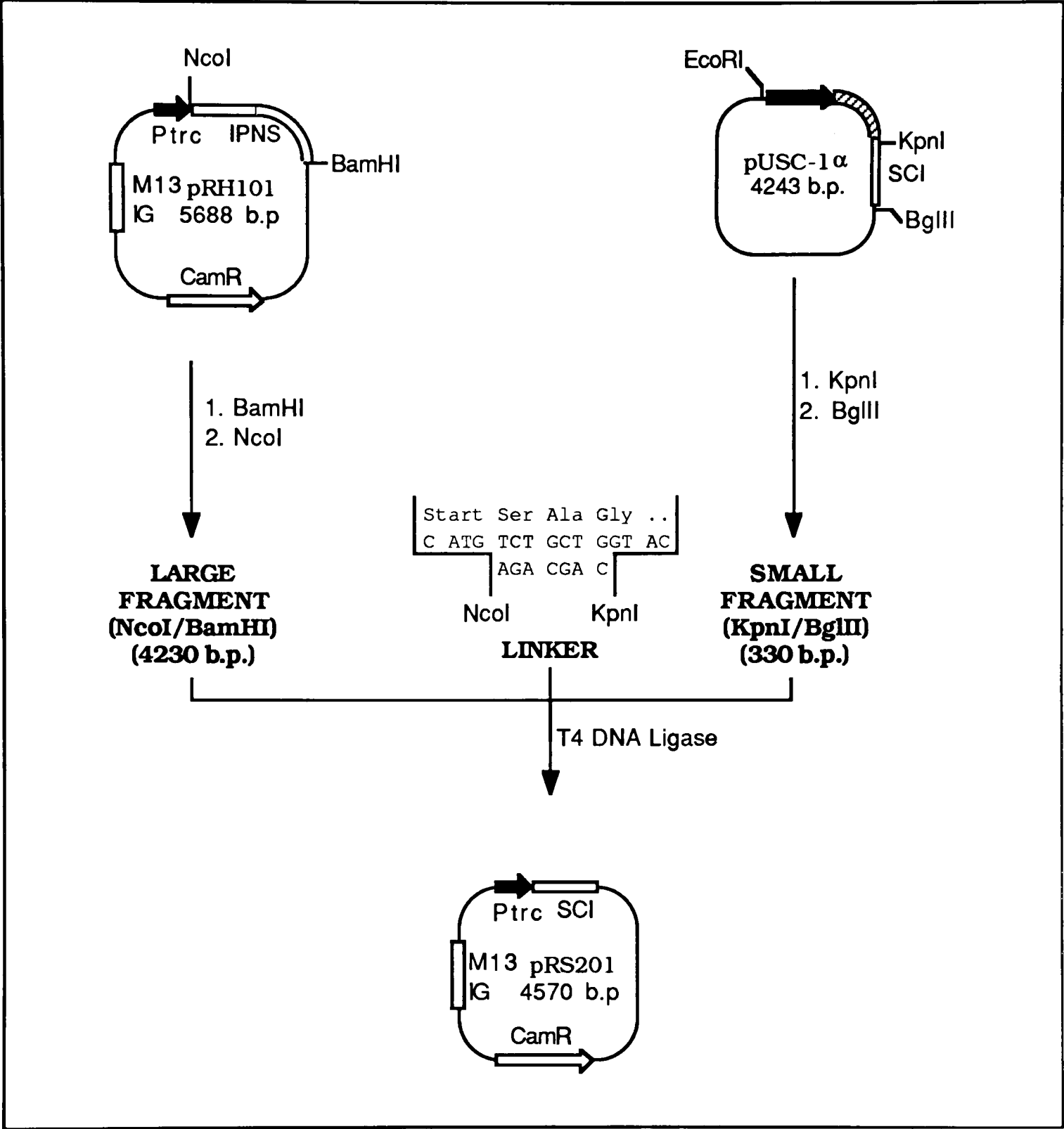


Figure 3.10: Construction of pRS201

However, in a parallel project Quaderi (1991) described the successful expression of recSCI at 37°C using the bacteriophage T7 polymerase/promoter system developed by Tabor & Richardson (1985). In it, transcription of the SCI gene was under the control of the T7 promoter which was recognised exclusively by the T7 RNA polymerase. This system utilised a two-plasmid system in which the first plasmid based on the pT7-7 described by Tabor & Richardson (1985) carried the SCI gene cloned downstream to the T7 promoter. The second construction, a specialised M13 phage vector mGP1-2 carried the gene for the T7 RNA polymerase and was introduced into the cell at the time of induction by infection [Tabor & Richardson, 1987]. This induction by transfection therefore does not require high temperatures. The morphology of the recSCI produced in this system remains to be determined. Yet, recombinant protein production utilising this system has been shown to share characteristics with the  $\lambda P_L$  promoter system in that most proteins have the propensity to be produced as insoluble inclusion bodies (Stan Tabor; personal communication).

### 3.3. Periplasmic Expression of SCI in *E.coli*.

#### 3.3.1. Introduction

The secretion in *E.coli* of a cloned gene product across the cytoplasmic membrane can be achieved by fusing an appropriate signal peptide to the N-terminal end of the gene product. Although prokaryotes lack the intracellular mechanisms involved in the eukaryotic secretion pathway (Section 2.1.5.), the basic mechanisms of secretion remain essentially conserved (Figure 3.11) (see Bieker & Silhavy (1990) for a review).

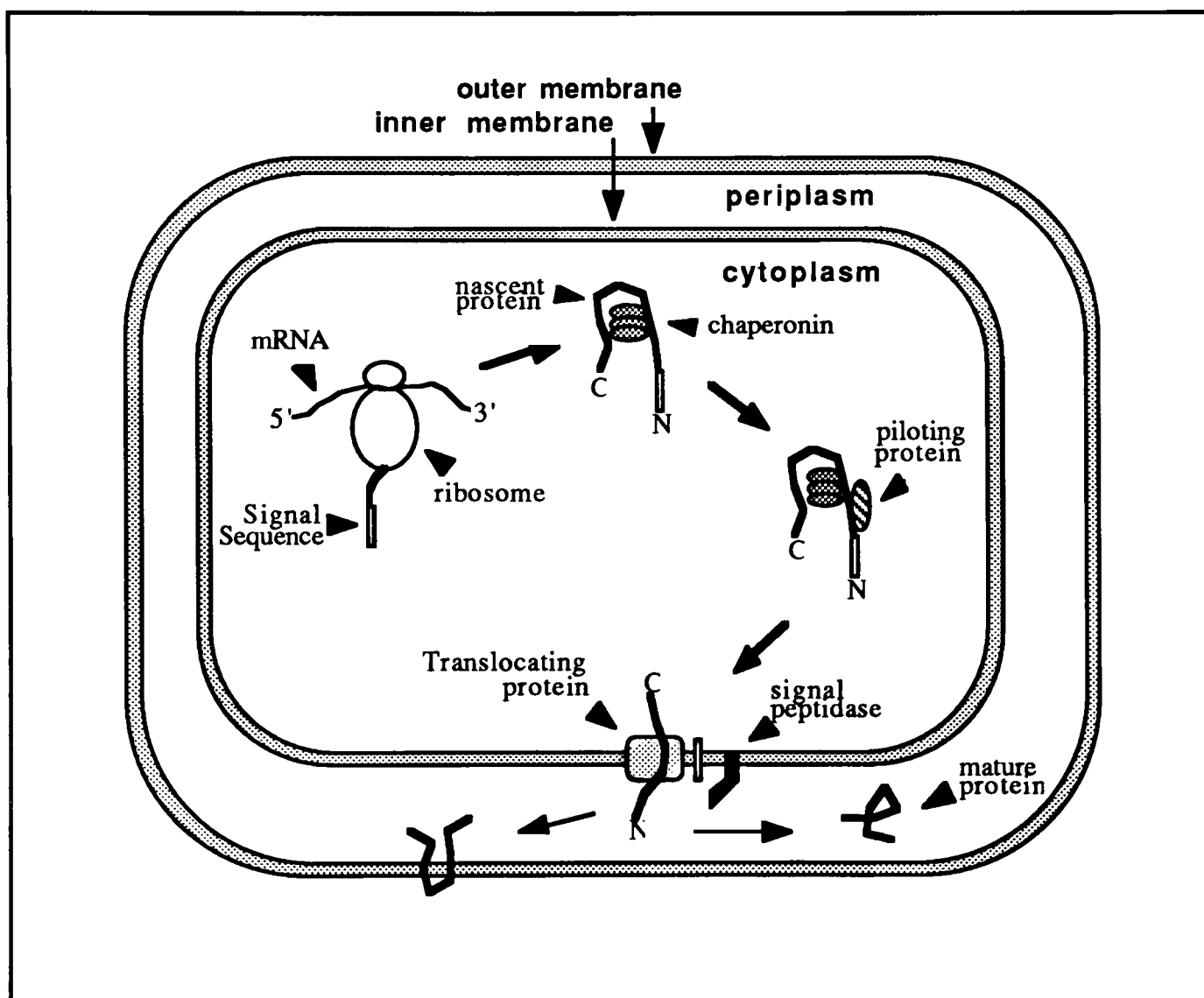


Figure 3.11: Secretion of Proteins in *E.coli*

*E.coli*, a gram-negative bacteria contains two surface membranes, the inner cytoplasmic membrane and the outer membrane. The inner membrane is the main permeability barrier between the cytoplasm and the external environment - the outer membrane is permeable to many chemicals and external agents. Between the two membranes is the periplasm. Protein

secretion involves the signal peptide mediated translocation of the nascent polypeptides across the inner cytoplasmic membrane into the periplasmic space where on the periplasmic face of the cytoplasmic membrane, the signal peptide is cleaved by the signal peptidase, resulting in the mature protein [Kusters *et al.*, 1991] (Figure 3.11). The only intragenic information that is truly essential for protein secretion is the signal sequence [Bieker & Silhavy, 1990].

In the periplasm, the secreted polypeptides are exposed to conditions considerably different to those in the cytoplasm. Due to the permeability of the outer membrane, these conditions depend on those of the extra-cellular medium and may be more favourable for correct disulphide bonding than the highly reductive conditions in the cytoplasm [Gatenby *et al.*, 1990]. Furthermore, the cleavage of the signal peptide yields the correct N-terminally processed form of the expressed polypeptide. The potential toxicity of the recombinant gene product could lead to the activation of heat-shock and SOS responses [VanBogelen *et al.*, 1987] which include the activation of degradatory proteolytic activity [Goff & Goldberg, 1985]. Proteolytic activity is considered to be much less in the periplasmic space than in the cytoplasm leading to a stabler gene product. Table 3.2 shows a few published examples of eukaryotic proteins that have been secreted in *E.coli* in active form, using homologous signal sequences. These include structurally similar serine proteinase inhibitors which were secreted, correctly folded and active.

### **3.3.2. Construction of pRS-om101; a Secretion Vector for SCI**

The secretion vector pRS-om101 was constructed by replacing the SCI gene in pRS101 by a gene encoding a fusion of the signal sequence of the *E.coli* OmpA protein and SCI. The OmpA protein is a major component of the outer membrane of *E.coli* and is the most abundant protein in the cells [Ghrayeb & Inouye, 1984]. It is expressed in the cytoplasm as a secretion

| Protein                                                                     | Promoter | Signal Sequence | Reference                      |
|-----------------------------------------------------------------------------|----------|-----------------|--------------------------------|
| Bovine Pancreatic Trypsin Inhibitor (BPTI)                                  | lpp      | OmpA            | Goldenberg, 1988               |
| Human Pancreatic Secretory Inhibitor                                        | trp      | PhoA            | Kanamori <i>et al.</i> , 1988  |
| Human Chymotrypsin inhibitor-2                                              | lpp      | OmpA            | Longstaff <i>et al.</i> , 1990 |
| Human $\beta$ -amyloid precursor protein kunitz proteinase inhibitor domain | trp      | PhoA            | Schilling <i>et al.</i> , 1991 |
| Human Kringle-2-Serine Protease                                             | tac      | PhoA            | Obukowicz <i>et al.</i> , 1990 |
| Porcine Phospholipase A2                                                    | T7       | OmpA            | Deng <i>et al.</i> , 1990      |

Table 3.2: Secretion of heterologous proteins in *E.coli*

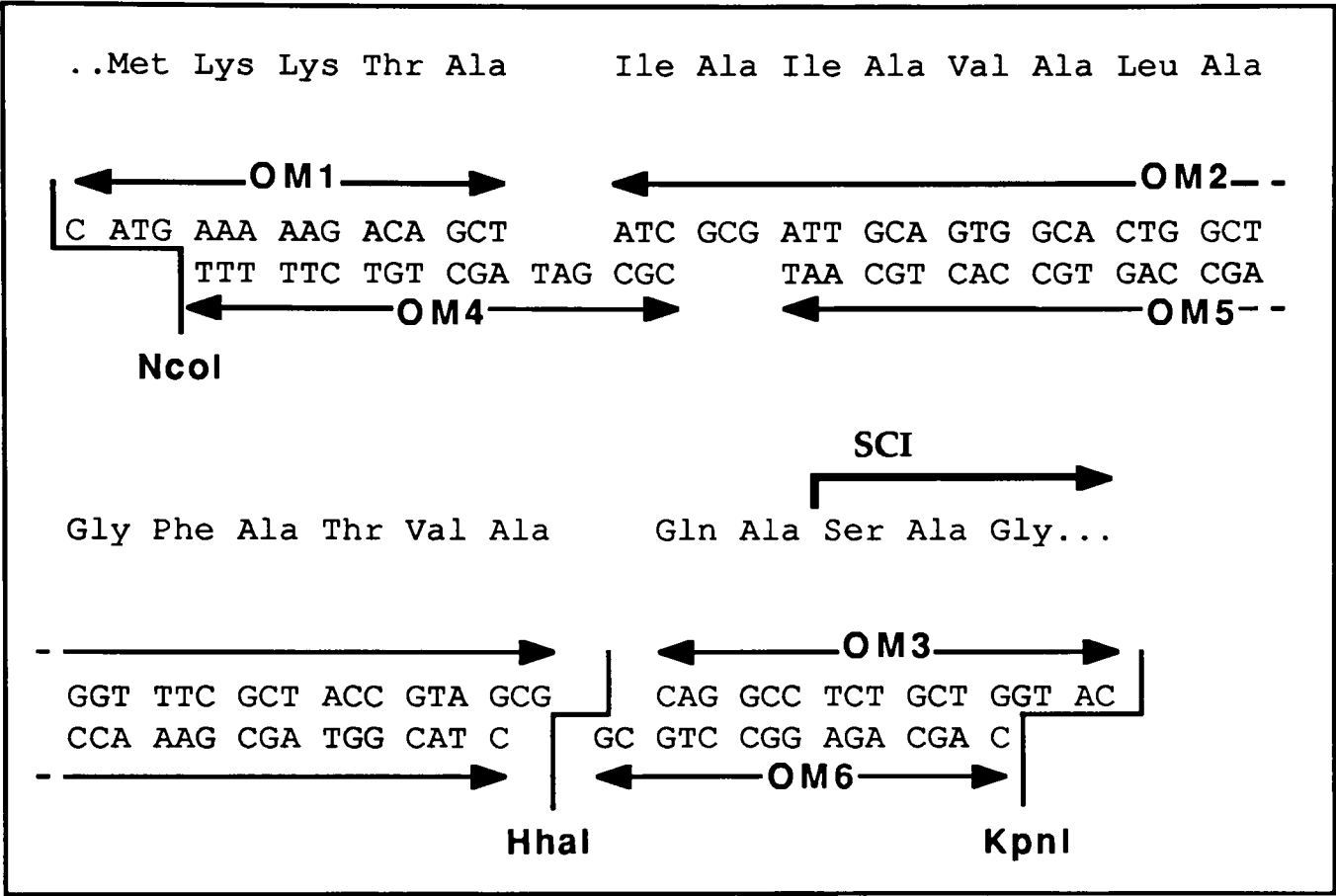
precursor with an additional 20 amino acid N-terminal signal peptide. Upon translocation into the periplasm, the signal peptide is cleaved and the mature protein assembled in the outer membrane [Ghrayeb & Inouye, 1984]. The amino acid sequence of the signal peptide determined by Movva & co-workers (1980) is as follows:

**-21                                          -10                                          -1**

**Met-Lys-Lys-Thr-Ala-Ile-Ala-Ile-Ala-Val-Ala-Leu-Ala-Gly-Phe-Ala-Thr-Val-Ala-Gln-Ala-**

It complies with the general rules for a signal peptide (section 2.1.5) in that it has basic N-terminal residues followed by an extended hydrophobic region between the basic N-terminal region and the cleavage site with a Gly residue almost in the middle of the hydrophobic portion. The amino acid at the cleavage site (-1) is an Ala residue.

The plasmid pRS-om101 was constructed by ligating the 330 b.p. *KpnI/BglII* SCI gene fragment from pUSC-1 $\alpha$  to a 71 b.p. synthetic *NcoI/KpnI* adapter bearing the DNA sequence encoding the OmpA signal peptide. The signal peptide adapter was designed as illustrated in figure 3.12, using the same codons as determined by [Movva *et al.*, 1980]. An *NcoI* site incorporating the ATG start codon was added to the 5'-end so as to ensure correct insertion of the fused OmpA signal-SCI gene into the pIT454 derived expression vector. The 3'-end of the adapter incorporated the region encoding the junction of the signal sequence and the SCI and included the missing 5' region of the SCI gene up to the *KpnI* site, which incorporates the first three complete codons. For synthesis, the adapter was divided into three blocks, each consisting of complementary strands with protruding, unpaired cohesive ends, which ensured the correct orientation of each block during ligation. The oligonucleotides labelled OM1 to OM6 were designed so as to minimise any internal self-complementarity and homology between the cohesive ends of the same oligonucleotides so that only the desired hybridisations occur.



**Figure 3.12: Construction of the OmpA Signal Peptide Adapter for SCI**

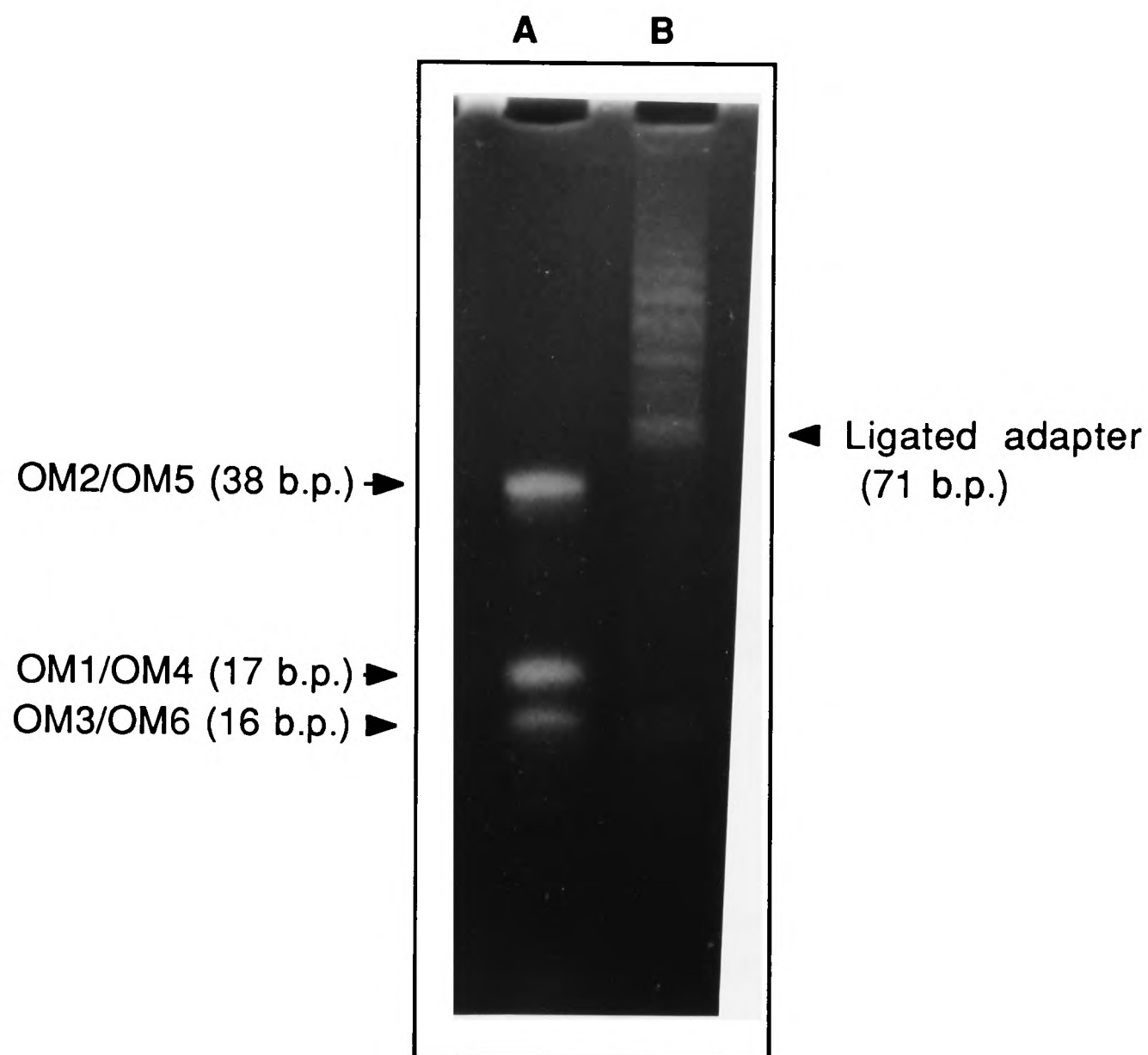
The NcoI/KpnI adapter was divided into 3 blocks.  
The oligonucleotides synthesised for each block are designated OM1 to OM6.

Synthesis of the oligonucleotides was performed by automated DNA synthesis.

The oligonucleotides were deprotected and purified as described in Chapter 5. All the oligonucleotides excepting OM1 and OM6, which were the 5'-terminal ones were phosphorylated at the 5'-ends using T4 polynucleotide kinase in the presence of ATP. The unphosphorylated 5'-terminal oligonucleotides were to prevent concatamerisation in the subsequent ligation steps.

For the assembly of the adapter, equimolar portions of the complementary oligonucleotides of each block were pooled and annealed. The duplexes obtained were mixed and ligated. The completion of the ligation was monitored on a 20% non-denaturing polyacrylamide gel (Figure 3.13). The complete adapter obtained was used without separation from the mixture of longer concatamers seen in figure 3.13. The latter were formed as a result of ligation between the 5'-phosphorylated blocks. Only the correctly formed adapter, with its *NcoI/KpnI* cohesive termini would ligate with the compatible termini of the SCI *KpnI/BglII* fragment and the pIT454 *NcoI/BamHI* fragment to form the expression vector pRS-om101.

Figure 3.14 illustrates the completion of the construction of pRS-om101 where the 5814 b.p. *NcoI/BamHI* fragment was mixed together with the 330 b.p. *KpnI/BglII* of pUSC-1 $\alpha$  bearing the SCI gene sequence and the above adapter with the molar ratio of 1:10:500, respectively and ligated overnight at 4°C with 0.5U T4 DNA ligase in the presence of 2mM HCC and 8mM spermidine. The resulting ligation mix was used to transform competent *E.coli* JM109 cells which were then smeared on LB-agar plates containing tetracycline and incubated at 27°C for 24h. About 40 positive transformants were counted. Restriction endonuclease screening of "minipreps" of the plasmids of the individual colonies, showed that 4 colonies contained the



**Figure 3.13: Assembly of the 71 b.p. NcoI/KpnI OmpA Signal Adapter for SCI**

All the oligonucleotides except OM1 and OM6 were phosphorylated. The oligonucleotide duplex blocks were assembled by annealing equimolar amounts (100pmol each) of the constituent oligonucleotides. The annealed duplexes were mixed and ligated with T4 DNA Ligase. The mixtures were separated on 20% non-denaturing PAGE. The DNA was stained with ethidium bromide and the DNA located in UV light.

Lane A: Before Ligation  
Lane B: After Ligation.

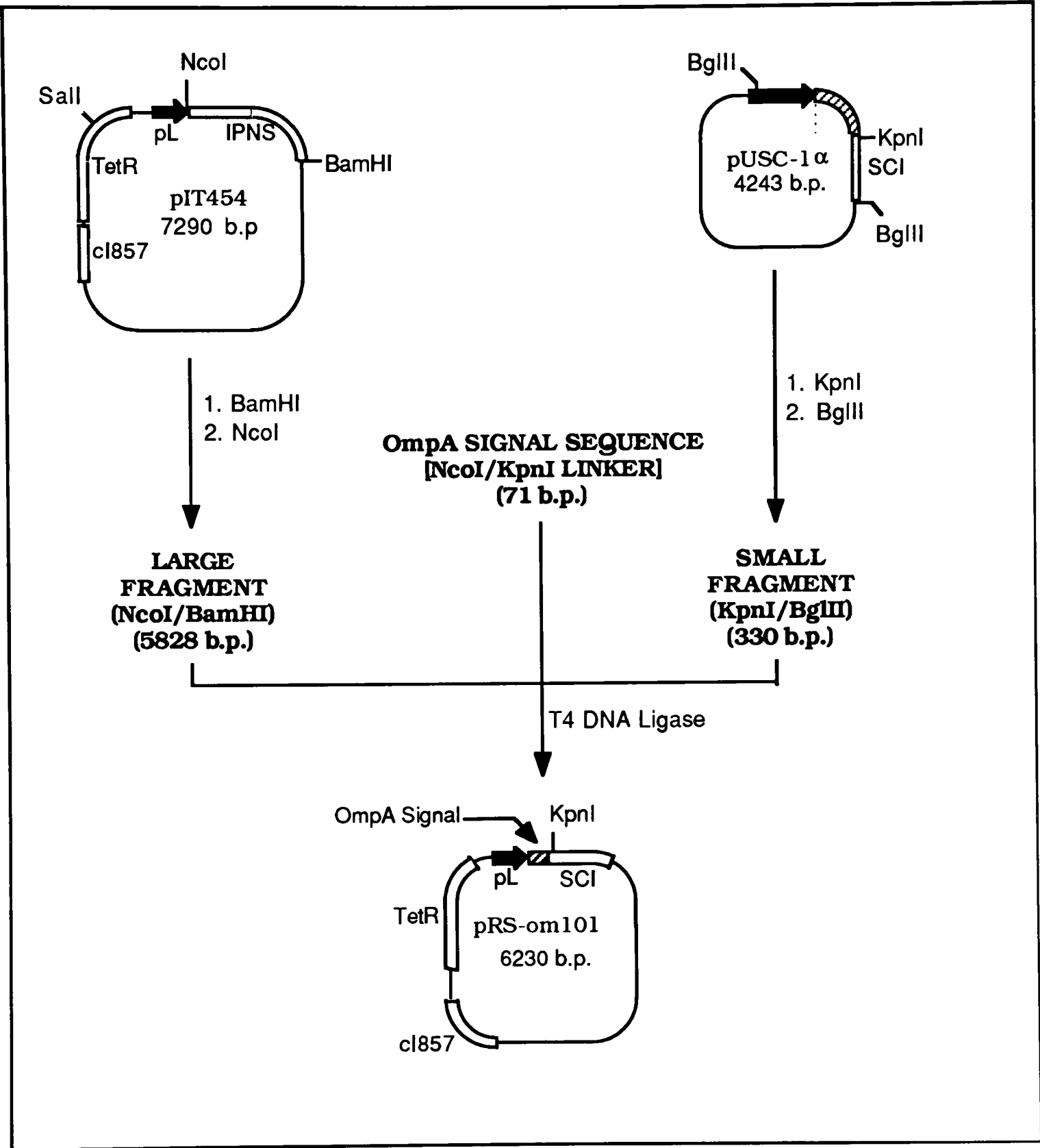
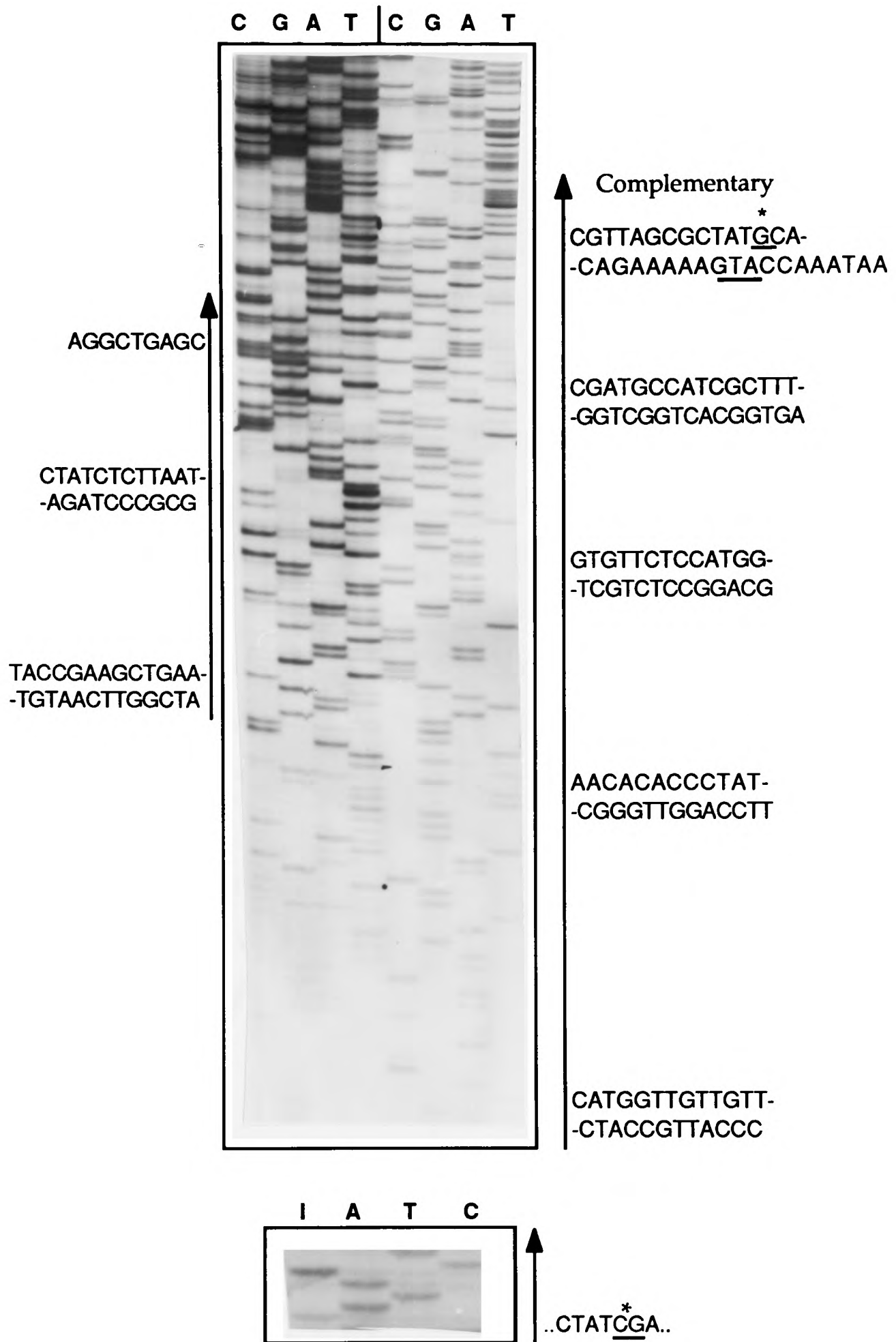


Figure 3.14: Construction of pRS-om101



**Figure 3.15:** Double-Stranded Sequence Analysis of pRS-om101: Autoradiograph of 6% Urea-Polyacrylamide Gel

Left Half: Sequence of the 3'-junction of the SCI gene  
 Right Half: Sequence of the 5'-junction of the SCI gene

The plasmid was sequenced using the primers YK22-15 and YK22-8, respectively [Kodo, 1988] (see Chapter 5).

(The erroneous GC swap in the sequence of the OmpA signal peptide (\*) was confirmed to be a GC compression. The correct sequence (inset) was read by replacing dGTP with dITP).

correctly assembled plasmid. This was confirmed by double-stranded DNA sequencing of the constructs (Figure 3.15).

3.3.3 Expression Studies Using the Periplasmic Expression Vector pRS-om101

Preliminary expression studies with JM109[pRS-om101] were carried out in SOB medium as described for JM109[pRS101] in section 3.2.2. The growth behaviour of JM109[pRS-om101] at 27°C was essentially identical to that of JM109[pRS101] (Figure 3.16). However, after induction of protein expression by increasing the incubation temperature to 42°C, JM109[pRS-om101] showed a considerably more rapid retardation in cell growth and ensuing death and lysis than JM109[pRS101] (Figure 3.16).

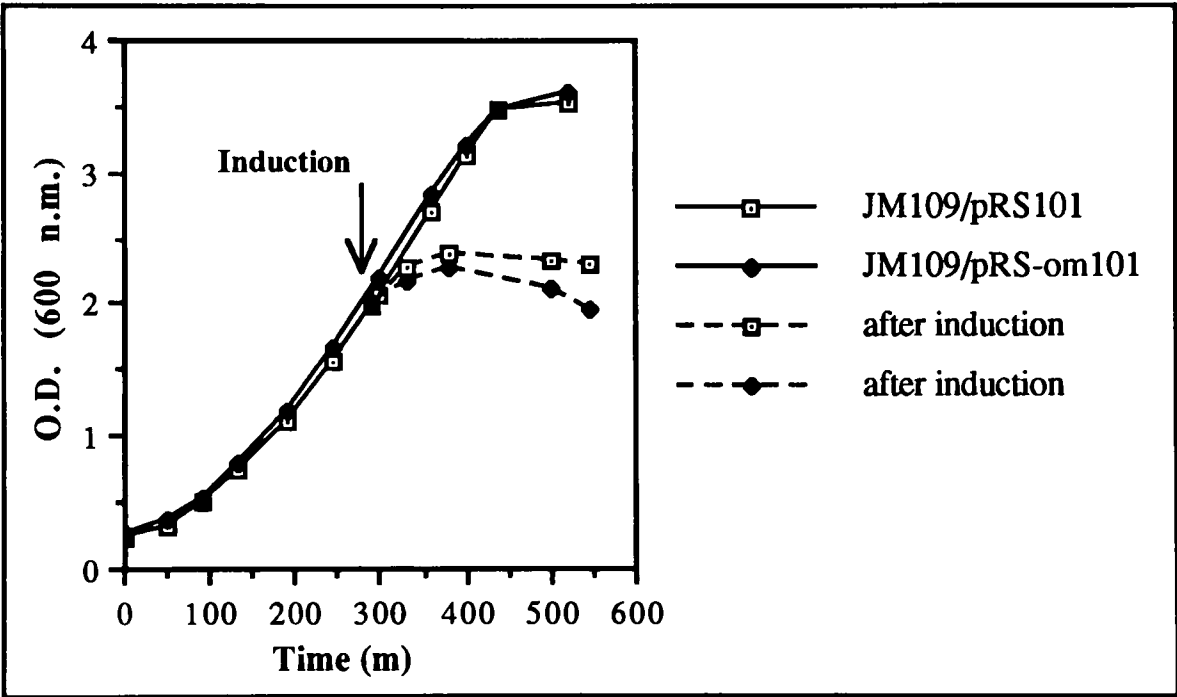
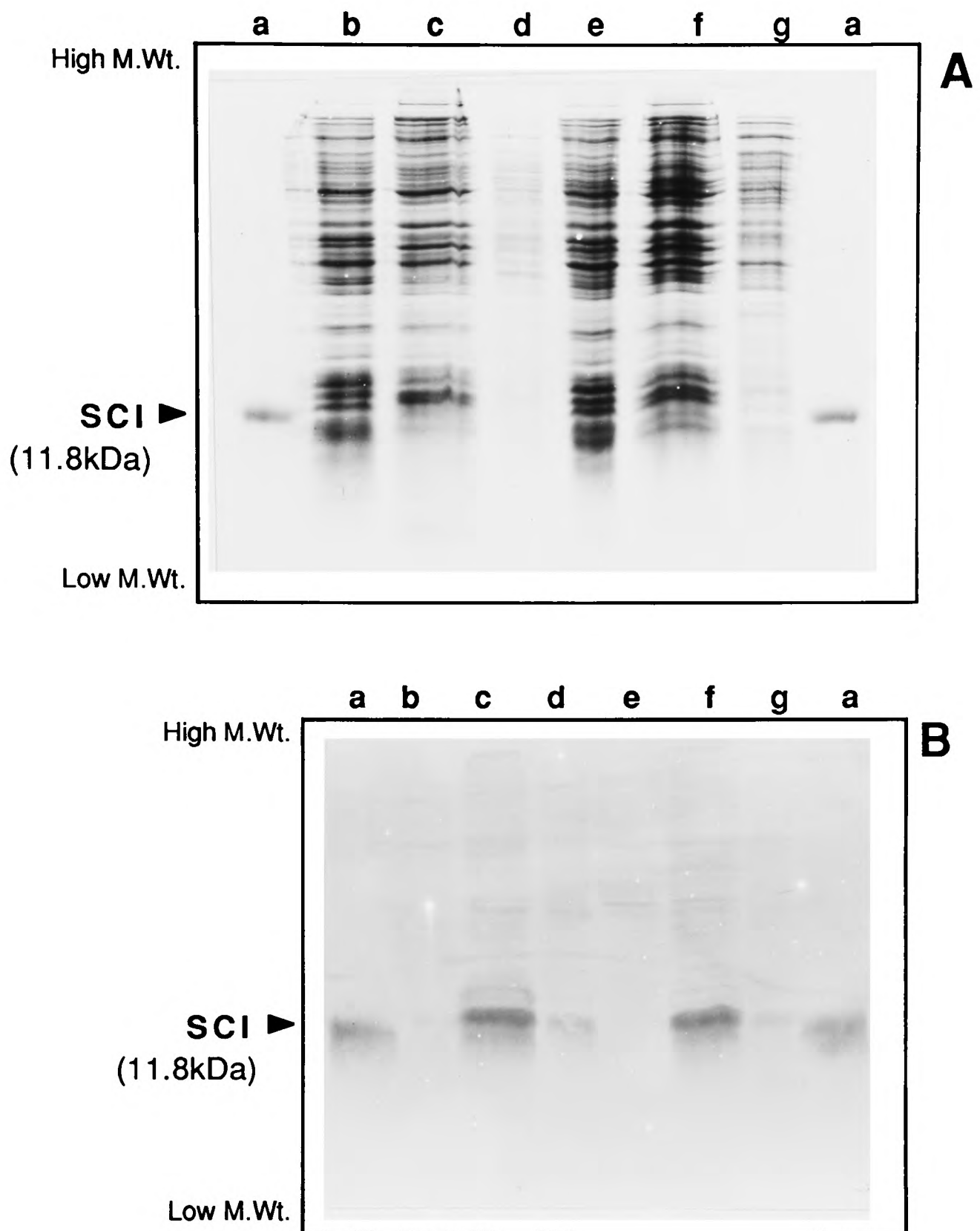


Figure 3.16: Periplasmic Expression of SCI in *E.coli*: Comparison of Growth Rates of *E.coli* JM109 transformants.

Initial comparison of the protein contents by SDS-PAGE and Western blotting of total cell extracts before induction and after overnight induction at 42°C showed that on induction, relatively high levels of recSCI were being successfully produced (Figure 3.17). The expression level was estimated at about 10% of the total bacterial protein, which was substantially higher than those levels obtained in the JM109[pRS101] cytoplasmic expression system in



**Figure 3.17:** Periplasmic Expression of SCI in *E.coli*

Profiles of the total cell protein (reduced) and protein extracts (reduced) of growth media of *E.coli* transformants on 15% SDS-polyacrylamide gel.

A: Coomassie stain

B: Western blot

Lanes a: Authentic SCI

Lane b: JM109[pRS-om101] total cell extract; without induction

Lane c: JM109[pRS-om101] total cell extract; after induction

Lane d: JM109[pRS-om101] growth medium

Lane e: JE5505[pRS-om101] total cell extract; without induction

Lane f: JE5505[pRS-om101] total cell extract; after induction

Lane g: JE5505[pRS-om101] growth medium

the absence of the *dnaY* gene product (pUBS520) and were comparable to those obtained in cytoplasmic expression with JM109[pRS101+pUBS520] (Section 3.2.3). The incorporation of the *dnaY* gene product by co-transforming the plasmid pUBS520 into the periplasmic expression system resulted in only a marginal improvement of the expression level (data not shown). This enhancement was not comparable to those obtained in the cytoplasmic expression systems using pRS101. This was a somewhat puzzling anomaly, as it was the same gene for SCI, with its high rare Arg codon content that was used in the construction of both the cytoplasmic and periplasmic expression vectors. The improved expression with the latter could be attributed to several factors that compensate for the negative effect of the high rare Arg codon content of the SCI gene. These may include the stabilisation of the mRNA transcript and improved translation due to optimised ribosomal binding (section 3.1.5), both resulting from the presence of sequences encoding the highly expressed homogeneous OmpA signal sequence directly downstream to the SD-sequence. It is evident that the secondary structure adopted by the mRNA transcript is an important factor in determining its *in vivo* stability and the efficiency of translation.[Das, 1990].

Post-translational stabilisation of the nascent fusion protein due to the presence of the homogeneous signal peptide could also result in the improvement of expression levels. Though the exact mechanisms involved in the secretion of proteins in *E.coli* remain unclear, several accessory proteins have been implicated in the cytoplasmic recognition of the signal peptide and the subsequent retention of the nascent protein in a secretion-competent conformation. This includes protein SecB, a secretion-specific 'chaperonin' [Bieker & Silhavy, 1990]. Chaperonins, which also include the *E.coli* heat-shock protein GroEL, belong to the larger class of proteins known as molecular chaperones and are thought to stabilise transient folding intermediates of nascent proteins by providing a surface to which the latter

can attach, thus preventing them from aggregating as a result of unfavourable hydrophobic interactions [Gatenby *et al.*, 1990]. Cytoplasmic proteins are thereby enabled to fold correctly and nascent periplasmic proteins destined for secretion retain a stable export-competent conformation unhindered by unnecessary intermolecular interactions [Bieker & Silhavy, 1990]. This additional stabilisation due to the presence of the signal peptide may, in addition, protect the recombinant SCI from proteolytic degradation.

With the hope of optimising the periplasmic expression level of recSCI, several strains of *E.coli* were transformed with pRS-om101 (see section 3.2.2). Apart from marginal strain- and pH-dependent variations in growth rates, no substantial improvement in expression levels were obtained. In addition, JE5505, a murein-lipoprotein deficient (*lpo*<sup>-</sup>) strain of *E.coli* [Suzuki *et al.*, 1978] was transformed with pRS-om101. This strain carries a 'leaky' outer membrane permitting periplasmic proteins to diffuse into the external medium. It has been previously shown to efficiently release secreted active recombinant pancreatic secretory trypsin inhibitor (STI) at about 50µg/ml [Kanamori *et al.*, 1988] and the Kunitz proteinase-inhibitor domain of the  $\beta$ -amyloid precursor protein ( $\beta$ APP-KPI) at levels of 1µg/ml [Schilling *et al.*, 1991] into the medium. (*E.coli* JE5505 was obtained, with thanks, from Dr. Akiko Nishimura, National Institute of Genetics, Mishima, Japan).

The transformants JE5505[pRS-om101] and JE5505[pRS-om101+pUBS520] were obtained by electroporation (Chapter 5). Growth and induction of expression was carried out in SOB-medium as previously described for JM109[pRS101] in section 3.2.2. The growth curves obtained (Figure 3.18) showed slightly, but not substantially slower rates than the JM109 transformants. After induction, the growth media, regularly monitored for both trypsin and  $\beta$ XIIa inhibitory activity, failed to indicate the excretion of active recSCI into the medium. Comparative analysis of the

protein content by SDS-PAGE, western blotting and the determination of protein concentrations (Chapter 5) of the growth media of JM109[pRS-om101] and JE5505[pRS-om101] was carried out after induction.

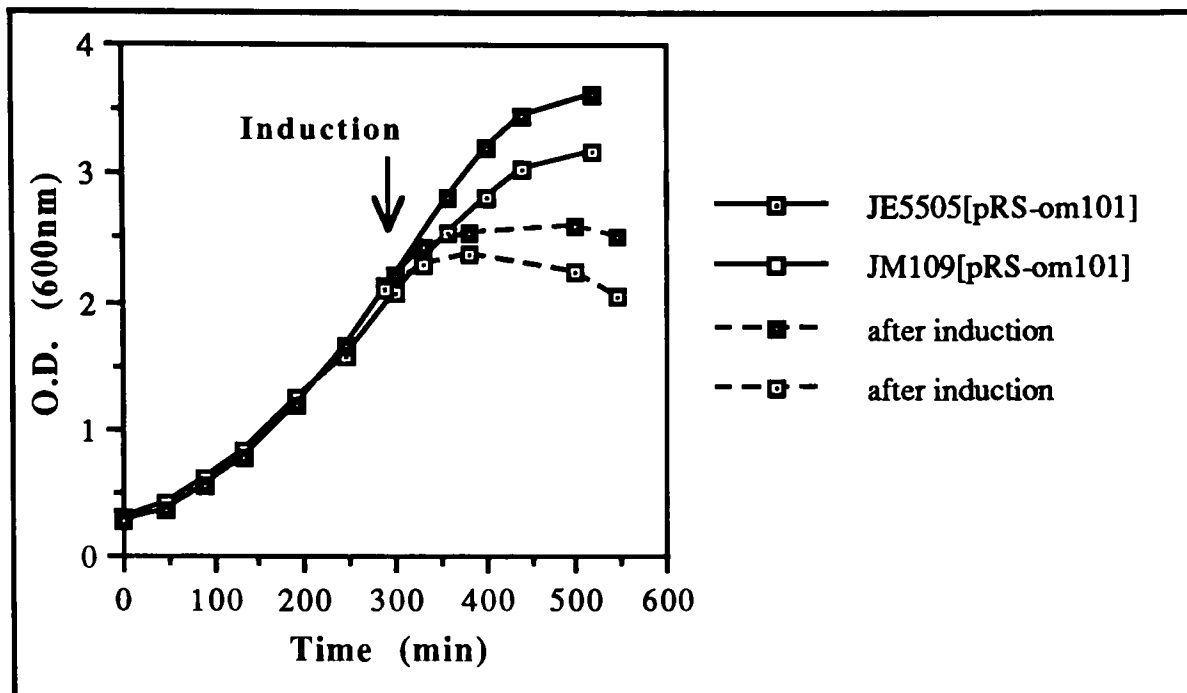


Figure 3.18: Periplasmic Expression of SCI  
Growth of transformants of the *lpo*<sup>-</sup> strain, JE5505

After induction for 6h, whereby maximum expression levels were obtained, the media of both JM109[pRS-om101] and JE5505[pRS-om101] showed high concentrations of protein ( $\approx 0.1$  and  $0.3\text{mg/ml}$ , respectively). The SDS-PAGE and Western blotting results (Figure 3.17) showed that in both strains, the cytoplasmic and media protein profiles were essentially the same. This indicated that the cells had undergone lysis resulting in the release of soluble cytoplasmic proteins, which included recSCI, into the medium. No such leakage of cytoplasmic protein was observed in cytoplasmic expression of recSCI using pRS101 (data not shown).

This obviously indicated that secreted recSCI had an adverse effect on the structural integrity of the host cells. In yeast secretion systems, intracellular accumulation of the recSCI was shown by electron microscopy to be localised at the intra-cellular membranes that could have originated from the endoplasmic reticulum or golgi bodies and secretory vesicles [Kodo, 1988]. Similarly, it is perceivable that recSCI, with its highly hydrophobic nature,

contains a stop-transfer sequence which blocks the translocation process across the inner membrane or will promote insertion within the inner membrane [Duffaud *et al.*, 1987] - the high level accumulation causing the eventual break up of the cells due to the destruction of the structural integrity of cytoplasmic membrane and its function as an effective barrier.

Due to the spontaneous lysis of the cells secreting recSCI, it was unfortunately not possible to determine and analyse the existence and location of soluble, secreted recSCI in the periplasm or medium as all fractions were heavily contaminated with cytoplasmic proteins.

### **3.3.4 Purification of Secreted recSCI**

#### **3.3.4.1. Extraction and Fractionation**

The final location and morphology of a secreted recombinant protein in *E.coli* depends on the expression system and the conditions used. At low levels of expression, secreted proteins are usually found in soluble form within the periplasmic space [Bowden *et al.*, 1991]. This facilitates their release together with the other periplasmic proteins by osmotic shock [Neu & Heppel, 1965] or their excretion as was the case when using the *E.coli* JE5505 strain (see above). However, at elevated expression rates, the formation of inclusion bodies has been observed with several proteins [Ikemura *et al.*, 1987; Obukowicz *et al.*, 1988]. Furthermore, Bowden & Georgiou (1991) in studies of the secretion of  $\beta$ -lactamase showed that the choice of signal sequences affected the maturation and aggregation of the secreted protein.

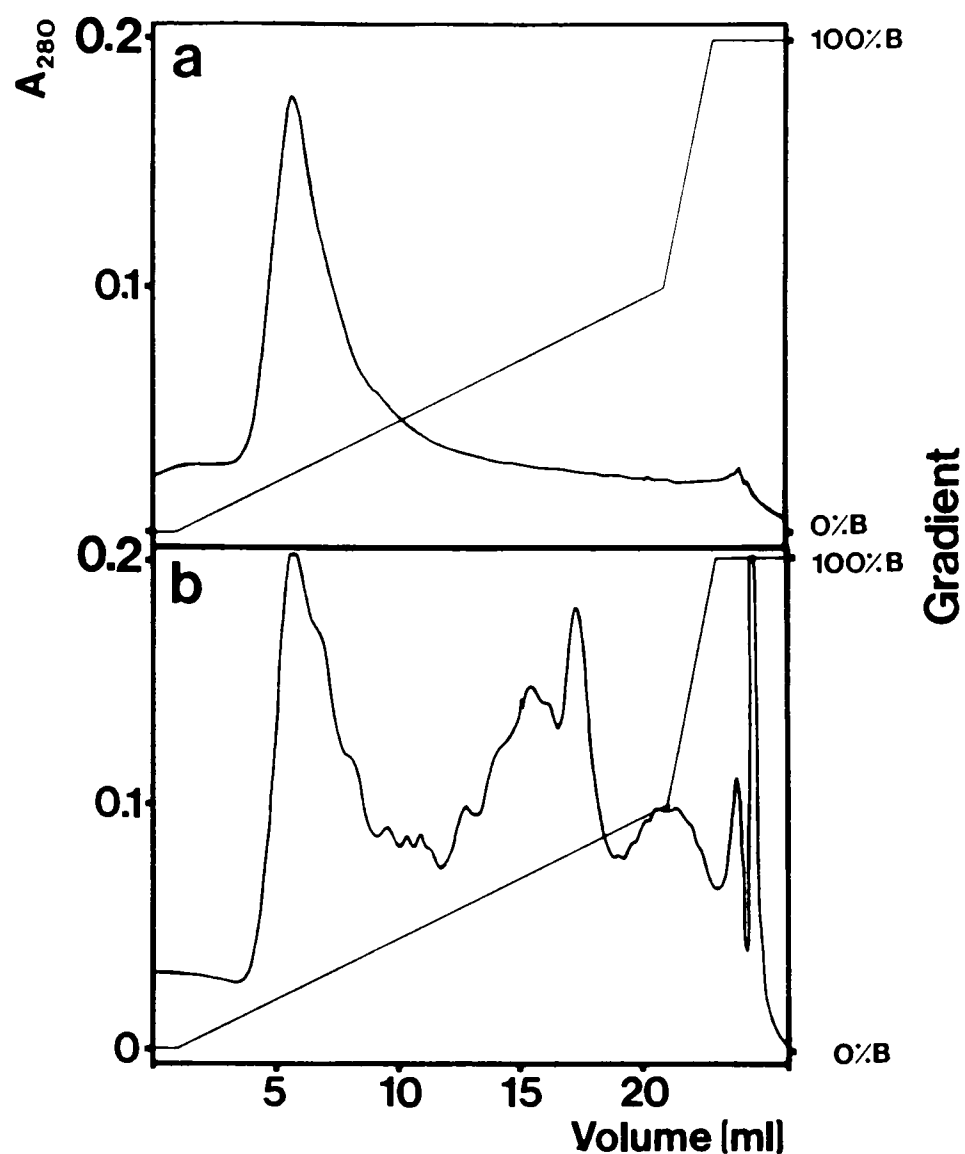
For the analysis of the secreted recSCI, a 4-litre culture of JE5505[pRS-om101] in superbroth containing tetracycline was prepared as described in section 3.2.4. After overnight induction, the cells were harvested by centrifugation for 20min. at 10000g at 4°C (MSE J2-21). The supernatant was retained for subsequent investigation. A total of 12g wet-weight of cells was

obtained. The cells were resuspended in 130ml 50mM Tris-HCl pH 8.0, containing 5mM EDTA, 0.15mM PMSF and homogenised, using a tissue homogeniser. Lysozyme was added to a final concentration of 1mg/ml and the suspension incubated for 30min at 4°C with gentle stirring and sonicated on ice with three 20 sec bursts. After DNaseI treatment of the suspension at 25°C, it was centrifuged for 20 min at 10000g. Both the supernatant, consisting of the remaining soluble cell components and the pellet were separated and analysed by SDS-PAGE. The recSCI was seen to be a major component of the insoluble pellet fraction whereas only trace quantities could be detected in the soluble fraction (data not shown).

The pellet obtained (5.6g wet-weight) was resuspended in 40 ml 50mM Tris-HCl pH 8.0 containing 50mM NaCl, 7M Urea and 0.1M DTT and incubated overnight at 4°C with gentle stirring. Effective solubilisation of the pellet was indicated by the clear suspension obtained. The residual cell debris was removed by centrifuging the suspension at 4°C at 10000g for 20 min.

#### **3.3.4.2. Ion-Exchange Chromatography (IEC)**

Analytical work was carried out on an FPLC system (Pharmacia) using a MonoQ (HR 5/5) column. The column was equilibrated in buffer A (50mM Tris-HCl pH 8.0 containing 50mM NaCl, 0.5mM EDTA, 0.5mM DTT and 5M urea). For calibration, 100µl of a solution of authentic SCI in buffer A (1mg/ml) was loaded onto the column at a flow-rate of 0.5ml/min. After being loaded, the column was washed in 3 times the column volume of buffer A and eluted at a flow-rate of 0.5ml/min with a linear sodium chloride concentration gradient from 50mM NaCl (buffer A) to a final concentration of 1M NaCl in the same buffer. Initially, the gradient was increased over 20min from 50mM NaCl to 0.5M NaCl and thereafter increased to 1M NaCl in 2min. The elution profile of pure, authentic SCI is illustrated in Figure 3.19 (panel a) in which it is seen to elute as a single peak at about 14% B. The MonoQ



|                         |                                                                                                                                                                         |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Gel:</b>             | MonoQ (HR 5/5)                                                                                                                                                          |
| <b>Samples:</b>         | a) 100µl authentic SCI (≈0.5 mg/ml) in Buffer A<br>b) 100µl Solubilised recSCI Inclusion Bodies (1mg/ml protein)<br>in 50mM Tris-HCl pH8, 50mM NaCl, 0.1M DTT & 7M Urea |
| <b>Bed Dimensions:</b>  | 0.5 x 5cm                                                                                                                                                               |
| <b>Buffer A:</b>        | 50mM Tris-HCl pH 8.0, 50mM NaCl, 0.5mM EDTA, 0.5M DTT<br>& 5M Urea.                                                                                                     |
| <b>Buffer B:</b>        | 50mM Tris-HCl pH 8.0, 1M NaCl, 0.5mM EDTA, 0.5M DTT<br>& 5M Urea.                                                                                                       |
| <b>Gradient Volume:</b> | 11ml                                                                                                                                                                    |
| <b>Flow Rate:</b>       | 0.5ml/min.                                                                                                                                                              |
| <b>Time:</b>            | 28 min.                                                                                                                                                                 |

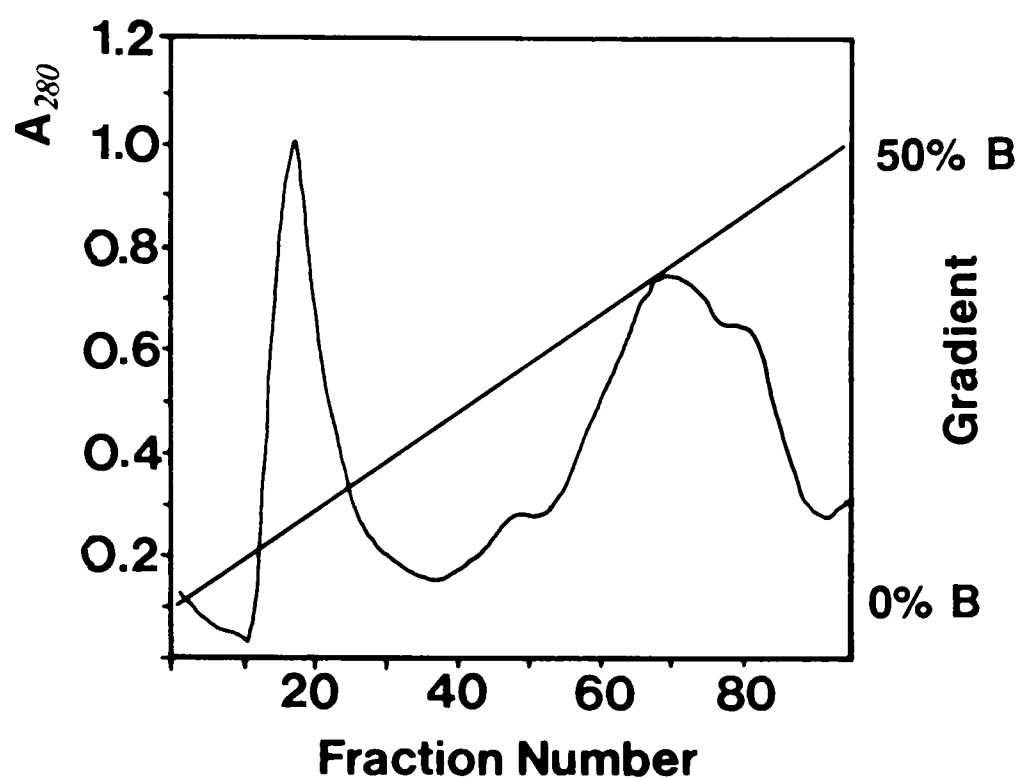
**Figure 3.19: Analytical Ion-Exchange Purification of Solubilised recSCI Inclusion Bodies.**

column was regenerated by washing it in 4 bed volumes of buffer containing 1M NaCl after which it was re-equilibrated in buffer A.

Similarly, 100µl of the clarified urea extract of the cell pellet was loaded onto the MonoQ column that was previously equilibrated with buffer A. After washing the column with 3 times the column volume of buffer A, the loaded proteins were eluted using the same gradient programme as above. Figure 3.19 (panel b) illustrates the obtained profile. The peak of the major component of the proteins eluted coincided with that of the authentic SCI. SDS-PAGE and western blot analysis of the peak fractions confirmed the presence of recSCI (data not shown).

Large-scale preparation of the recSCI contained in the urea suspension was carried out on a Q-Sepharose Fast-Flow (Pharmacia) column. Under the same elution conditions a similar profile to that obtained by FPLC on a MonoQ column (see above) was expected:

A 65 ml Q-Sepharose Fast-Flow column (inner diameter 2.6cm) was equilibrated in buffer A. The total quantity of 40ml of the clarified urea suspension of the cell pellet (section 3.3.4.1) estimated to contain about 1.2mg/ml of recSCI was loaded onto the column at a flow-rate of 3.5 ml/min. After washing the column with 6 times its volume of buffer A (see above), elution was carried out with a gradient from buffer A (50mM NaCl) to the same buffer containing 0.5M NaCl at the same flow-rate (3.5ml/min) and a total gradient volume of 500ml. Ninety six fractions of 5.2ml each were collected and the relative protein content of each fraction determined by measuring the OD<sub>280</sub>. Figure 3.20 illustrates the elution profile obtained. Note that the profile is essentially similar to that obtained over the same gradient range by FPLC on a MonoQ column. The fractions 12 to 36 were analysed by SDS-PAGE. The major component of the peak fraction was shown to be recSCI consisting of at least 60% of the total protein (0.6mg/ml) (see Figure



|                         |                                                                                                                                      |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| <b>Gel:</b>             | Q Sepharose Fast Flow                                                                                                                |
| <b>Sample:</b>          | 40ml Solubilised recSCI Inclusion Bodies ( $\approx 1.2\text{mg/ml}$ protein)<br>in 50mM Tris-HCl pH8, 50mM NaCl, 0.1M DTT & 7M Urea |
| <b>Bed Dimensions:</b>  | 65ml (2.6cm diameter)                                                                                                                |
| <b>Buffer A:</b>        | 50mM Tris-HCl pH 8.0, 50mM NaCl, 0.5mM EDTA, 0.5M DTT<br>& 5M Urea.                                                                  |
| <b>Buffer B:</b>        | 50mM Tris-HCl pH 8.0, 1M NaCl, 0.5mM EDTA, 0.5M DTT<br>& 5M Urea.                                                                    |
| <b>Gradient Volume:</b> | 500ml                                                                                                                                |
| <b>Flow Rate:</b>       | 3.5ml/min.                                                                                                                           |
| <b>Time:</b>            | 2h 30min.                                                                                                                            |

**Figure 3.20: Scaled-Up Preparative Ion-Exchange Purification of Solubilised recSCI Inclusion Bodies.**

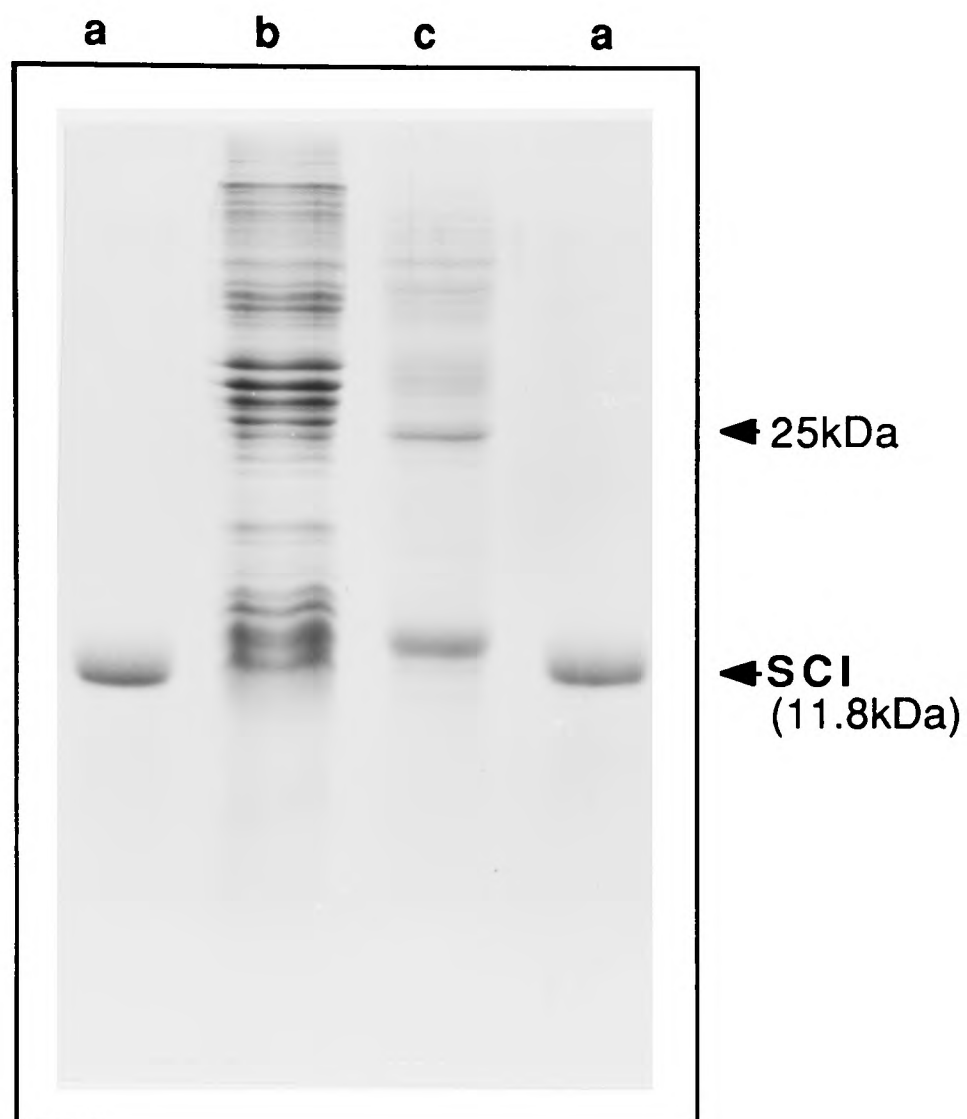
3.21). The urea in the samples was removed by dialysis in 50mM Tris-HCl pH 8.0 buffer containing 50mM NaCl, 0.5mM DTT and 0.5mM EDTA. No trypsin or  $\beta$ XIIa inhibitory activity was detected in the fractions obtained, indicating the failure of the recSCI to refold into its native form.

**3.3.4.3. Analytical Hydrophobic Interaction Chromatography (HIC)**

This was carried out on a Waters HPLC system (Millipore) using a Vydac C<sub>4</sub> reversed-phase column (particle size 5 $\mu$ m, pore diameter 300Å; The Separations Group, Hesperia CA). The column was equilibrated with 0.1% trifluoroacetic acid (TFA) in water (solvent 1) at a flow-rate of 1ml/min. To calibrate the system, 250 $\mu$ l of pure, authentic SCI (0.3mg/ml) was injected and eluted with the following gradient at a flow-rate of 1ml/min:

| Time (min) | Solvent 1<br>(0.1% TFA) | Solvent 2<br>(0.09% TFA in 80% Acetonitrile) |
|------------|-------------------------|----------------------------------------------|
| 0          | 70                      | 30                                           |
| 2          | 70                      | 30                                           |
| 52         | 60                      | 40                                           |
| 60         | 0                       | 100                                          |
| 62         | 0                       | 100                                          |
| 63         | 70                      | 30                                           |
| 80         | 70                      | 30                                           |

The obtained profile for pure, authentic SCI is illustrated in figure 3.22 (panel A). The multiple peaks are characteristic for the protein isolated from its natural source in that it has been shown to exist in several molecular forms including cleaved, partially cleaved and intact sequences of different forms of SCI - at least two other sequence variants of SCI with different solubility properties have been identified [M.P. Esnouf; personal communication] (see Chapter 4).

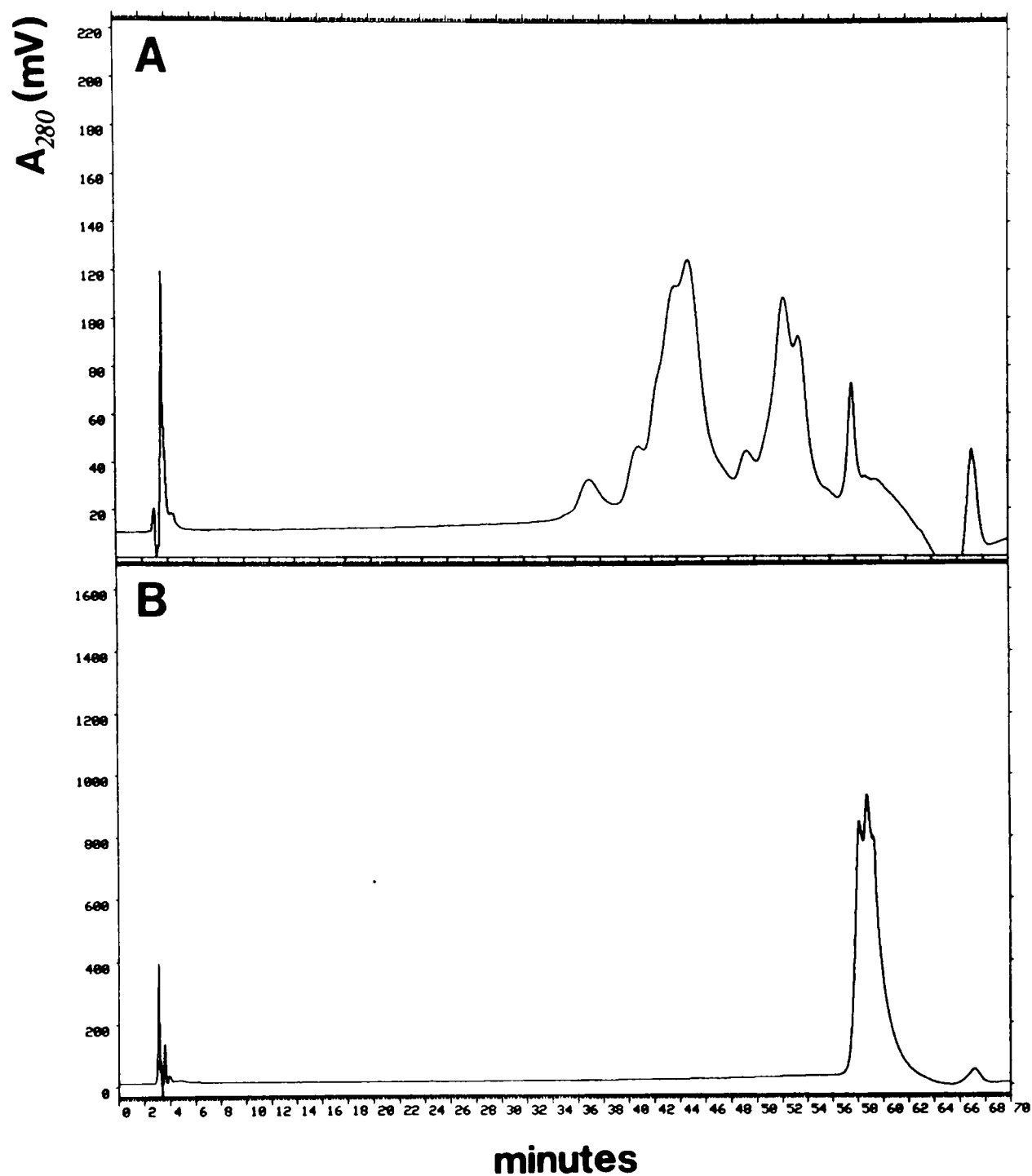


**Figure 3.21:**

Partial Purification of Solubilised recSCI Inclusion Bodies by Ion-Exchange Chromatography on a Q-Sepharose Fast-Flow Resin.

Profiles of protein extracts (reduced) on a Coomassie stained, 15% SDS-polyacrylamide gel.

- Lanes a: Authentic SCI
- Lane b: Crude recSCI Inclusion Body Fraction
- Lane c: Ion-exchange purified rec SCI solubilised from inclusion bodies.



*Sample:* A: 300 $\mu$ l Authentic SCI (0.3mg/ml)  
B: 250 $\mu$ l Q-Sepharose IEC peak fraction of recSCI ( $\approx$ 0.6mg/ml)  
*Column:* Vydac C4 Reversed Phase (particle size 5 $\mu$ m, pore diameter 300 $\text{\AA}$ )  
*Flow Rate:* 1ml/min.  
*Solvent 1:* 0.1% TFA  
*Solvent 2:* 0.09% TFA in 80% Acetonitrile  
*Gradient:* See Section 3.3.4.3.

**Figure 3.22: Hydrophobic Interaction Chromatography of recSCI showing (A) Authentic SCI and, (B) recombinant SCI (IEC fraction)**

A quantity of 250µl of the Q-sepharose IEC peak fraction containing an estimated 0.6mg/ml crude recSCI (section 3.3.4.2) was similarly injected onto the HIC column and eluted under the same conditions, with the same gradient. The elution profile obtained is illustrated in figure 3.22 (panel B). The peak obtained from the crude recSCI mixture did not coincide with any of the major peaks of the authentic SCI - it overlapped with the last minor peak of authentic SCI.

### **3.3.5. Characterisation of recSCI**

#### **3.3.5.1. N-Terminal Sequence**

The N-terminal sequence of the HIC purified peak fraction of the secreted recSCI (100µl) was determined by automated Edman degradation (Tony Willis, MRC Immunochemistry Laboratory, University of Oxford). The result obtained after 15 cycles (Appendix IX) confirmed that the major fraction of the purified protein had the correct N-terminal sequence:

**Ser-Ala-Gly-Thr-Ser-(Cys)-Val-Pro-Gly-Trp-Ala-Ile-Pro-His-Asn-**

This showed that the N-terminal signal peptide had been completely removed to yield the correctly processed SCI with the first 15 N-terminal amino acids in correct sequence. However, this was associated with about the same level of a truncated recSCI with the N-terminal sequence:

**Gly-Thr-Ser-(Cys)-Val-Pro-Gly-Trp-Ala-Ile-Pro-His-Asn-**

in which the N-terminal Ser and Ala residues had additionally been removed. This was probably due the additional signal peptidase recognition of the Ala residue next to the polar Ser residue. This arrangement closely resembles the -Gln-Ala recognition sequence at the signal peptidase cleavage

site of the OmpA signal peptide (see section 3.3.2). The heterogeneity of the recSCI is indicated by the split peak in the HIC profile (Figure 3.22).

The N-terminal sequences also indicated that the two forms of recSCI are tightly associated with at least one further peptide component (N-terminal sequence: Met-Ser-Phe-Arg....).

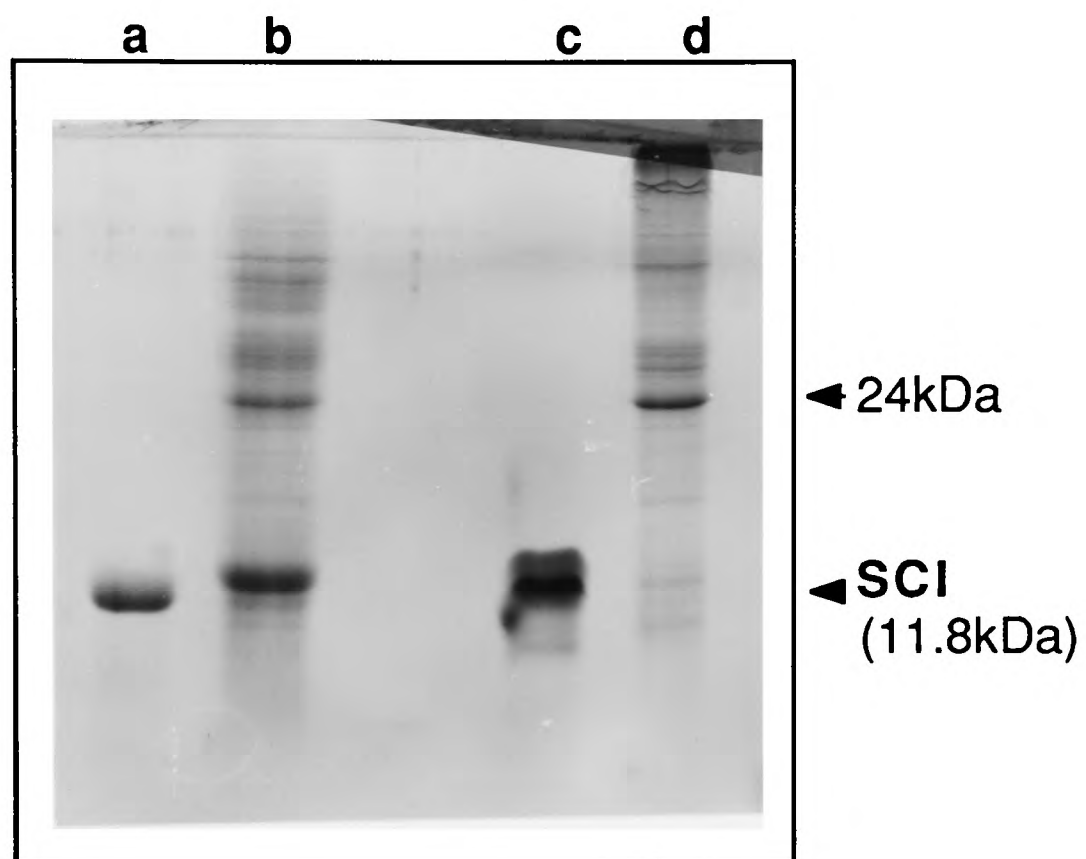
### **3.3.5.2. Non-Reducing SDS-PAGE**

The IEC purified recSCI peak fraction was subject to SDS-PAGE both under reducing and non-reducing conditions (see Chapter 5).

Under non-reducing conditions, the recSCI clearly migrated on SDS-PAGE as a higher molecular weight form ( $\approx 24000$  Da) (Figure 3.23). This molecular weight, which is about double that of recSCI (11812 Da.) indicates that the recSCI had associated into dimers possibly by intermolecular covalent disulphide linkage. The formation of non-native isomers of recombinant proteins obtained from inclusion bodies that result from the formation of intra- and intermolecular disulphide bonding is a common problem associated with inclusion body purification [Tsuji *et al.*, 1987]. The absence of both trypsin and  $\beta$ XIIa inhibitory activity in both the IEC and HIC fractions confirmed the absence of the native, correctly folded isomer of recSCI.

### **3.3.5.3. Molecular Weight Analysis by Electrospray Mass Spectrometry**

Electrospray mass spectrometry (ES-MS) provides a rapid and sensitive method for the molecular weight determination of proteins. This is achieved by the electrospray ionisation into the atmospheric pressure ion source of a quadrupole mass spectrometer [Aplin *et al.*, 1990]. A series of multiply charged ions of the protein enable one to calculate the molecular weights of the protein [Edmonds & Smith, 1990].



**Figure 3.23:** SDS-Polyacrylamide Gel Electrophoretic Analysis of Ion-Exchange Purified, solubilised recSCI Inclusion Bodies

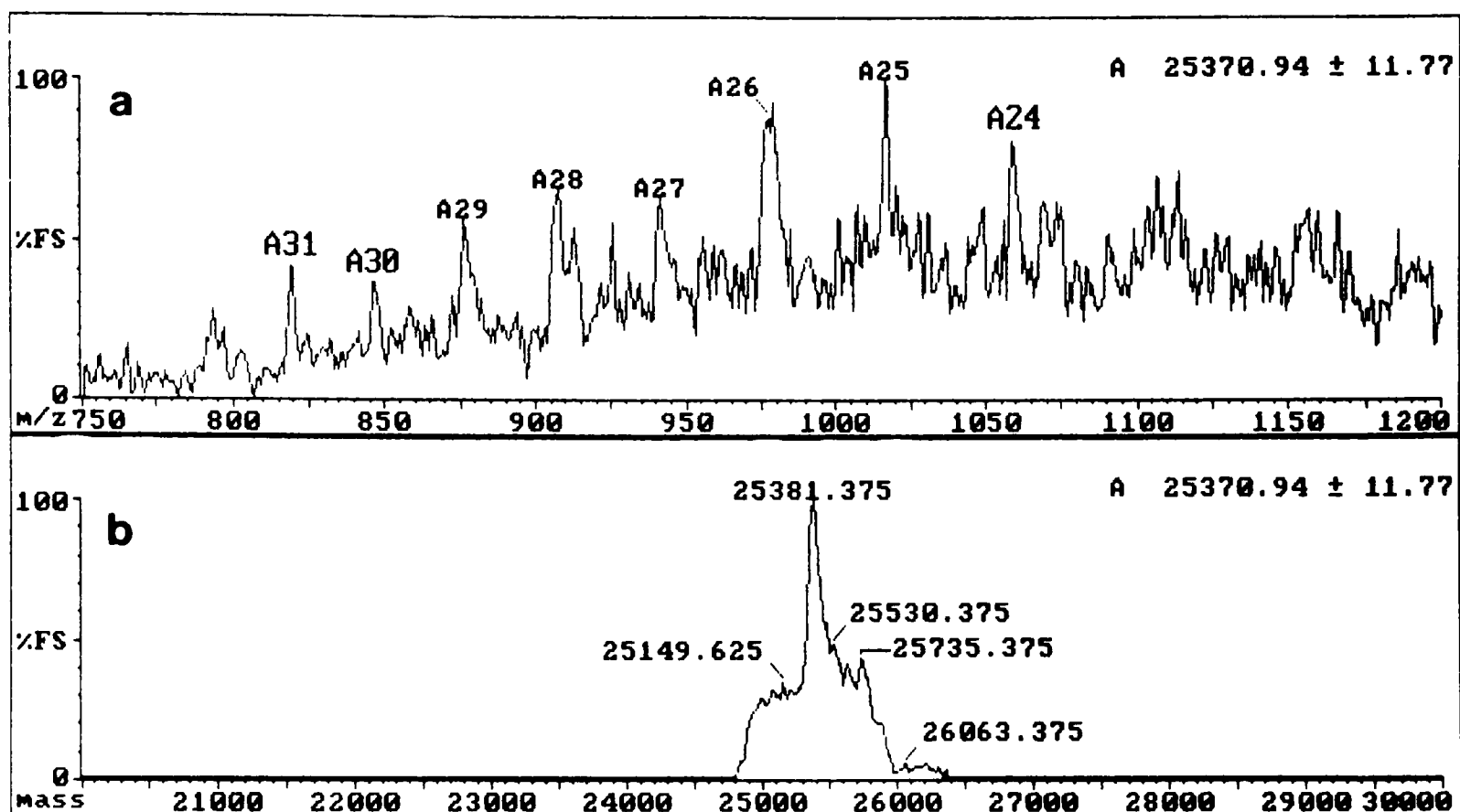
Profiles of protein, Coomassie stained on an SDS-polyacrylamide gel

Lane a: Authentic SCI  
Lane b: rec SCI ] reduced with mercaptoethanol

Lane c: authentic SCI  
Lane d: recSCI ] non-reduced

Prior to analysis, the recSCI in the IEC peak fraction was desalted and buffer exchanged to water using a PD-10 (Pharmacia) column. The resulting solution was concentrated on a SpeedVac concentrator to an estimated 0.4 mg/ml protein. An equal volume of methanol and 1% acetic acid was added to the sample, giving a protein concentration of about 20 pmol/ $\mu$ l. The sample solution (20 $\mu$ l) was injected into the electrospray system of a VG BIO Q quadrupole mass spectrometer calibrated with myoglobin ( $M_w$  16950.5) and scanned over a mass/charge range of 750 to 1200.

The electrospray mass spectrum of the protein sample is illustrated in figure 3.24 (panel a). The major component of the sample appears as a series of multiple protonated ions  $(M + H_n)^{n+}$  ranging from 24<sup>+</sup> and 31<sup>+</sup> indicating a molecular weight of 25370.9 Da. with a standard deviation of 11.8 (panel b). A second scan run together with myoglobin gave a molecular weight of 25369.8 Da. (standard deviation = 8.52). Myoglobin was used as an internal standard and ran as a series giving a calculated molecular weight of 16949.2 Da. (standard deviation = 5.5). This coincides well with the calculated molecular weight of myoglobin (16950.5 Da.). Coupled with the results obtained from N-terminal sequencing and non-reducing SDS-PAGE, the ES-MS result confirms that in both fractions, the purified recSCI was a tightly associated dimeric aggregate. Yet, the molecular weights around 25370 Da. obtained by ES-MS are 1745 Da. higher than 23624 Da., the actual molecular weight of a dimer consisting of only two subunits of SCI. As indicated by the N-terminal sequence information, this additional weight could be due to an additional peptide that has tightly co-aggregated or cross-linked with the dimer, the molecular weight of 1745 Da. being the average for a 14-15 residue peptide.



**Figure 3.24:** Electrospray Mass Spectrometry of recombinant SCI showing (a) the original spectrum indicating the relative intensities (%FS) of multiply protonated ions ( $24^+$  through  $31^+$ ) and, (b) the transformed true molecular mass spectrum.

### 3.3.6. Refolding

The stability of the recSCI dimer is reflected by the residual dimeric aggregate present (Figure 3.23; lane b) even after boiling the protein for 10 min. in SDS-PAGE sample buffer containing the reducing agent, mercaptoethanol (see Chapter 5). The tendency to covalently aggregate into dimers is not a property inherent to the authentic SCI, as was shown in the SDS-PAGE profile of the protein under non-reducing conditions (Figure 3.24; lane c), whereas, the recSCI obtained from the IEC purification step was almost exclusively associated into dimers (Figure 3.23; lane d). This was confirmed by the ES-MS results.

In ongoing work, the attempt is being made to obtain the active, correctly folded recSCI. Yet, prior to any refolding experiments it is of vital importance to dissociate the dimeric recSCI aggregate by reduction and unfolding and isolate the recSCI sub-units from the additional peptide component that is intimately associated with the dimer. Darby & Creighton (1990) in work on recombinant BPTI failed to obtain the correctly folded protein from what was the authentic and apparently pure, reduced and unfolded recombinant protein. This was in spite of established refolding procedures that were routinely used for the quantitative refolding of normal BPTI. They suggest that the tight binding of a cellular component to the recombinant protein during inclusion body formation could be the cause, in the first place, of the tendency to form inclusion bodies and the subsequent tendency during refolding for the protein to form intermolecular mixed disulphides instead of correct intermolecular disulphides [Darby & Creighton, 1990].

In preliminary work, with the aim of isolating the recSCI monomer, the crude protein contained in the IEC peak fraction (section 3.3.4.2) of a sample of the protein, unfolded and reduced in 6M GuHCl containing an

excess of DTT was separated on a C<sub>4</sub> reverse-phase column using an aqueous acetonitrile gradient containing 0.1% TFA (see section 3.3.4.3). The HIC profile showed a distinct peak that coincided exactly with the major peak that corresponded to the unfolded, reduced authentic SCI (data not shown). It could with reasonable certainty be presumed that this peak indeed corresponds to the separated monomer of recSCI. In work that is at present in progress, the pooled peak fractions of multiple runs under the same conditions will be subject to an additional HIC purification step prior to characterisation and refolding of the protein. The acidic conditions of the eluents will retain the protein in unfolded form, until subsequent refolding experiments can be performed.

### 3.4. Discussion

The work described in this chapter culminated in the successful cloning and overexpression of the synthetic SCI gene in *E.coli* and the partial purification and characterisation of the recombinant gene product. Short of the initial aim of obtaining selected reactive-site mutants for investigating the structure-function relationships of SCI and its very narrow specificity, the expression system developed provides one with the firm basis upon which to implement future strategies.

With the vector for cytoplasmic expression which was constructed by the in-frame insertion of the SCI gene directly downstream to the  $\lambda P_L$  promoter, Shine-Dalgarno (SD) sequence and translation initiation codon (ATG), uncharacteristically low levels of recSCI were expressed. These were estimated at 3% or less of the total bacterial protein which is an order of magnitude lower than the levels typically obtained with the same system. Due to the poor protein staining capacity of SCI, these low levels could only be detected by an anti-SCI polyclonal antibody. Failure to achieve high level expression of eukaryotic proteins in *E.coli* despite the optimisation of both transcriptional and translational signals have been reported previously (for references, see Brinkmann *et al.*, 1989). The complexity of the *in vivo* interactions at gene level and of the recombinant gene transcripts and nascent protein with the intracellular environment prevents the formulation of any generalisations that would help one predict the outcome of the attempts to express any one recombinant protein. Low productivity could have been due several possible reasons including inefficient translation initiation because of mRNA secondary structures [Tessier *et al.*, 1986; Yasueda *et al.*, 1990] or the induction of the heat shock response as a result of the toxicity of the gene

product, resulting in the proteolytic degradation of the protein [Goff & Goldberg, 1985]. Supplementation with the *dnaY* gene product (tRNA<sup>Arg</sup>(AGA/AGG)) by way of the plasmid pUBS520 [Brinkmann *et al.*, 1989] led 4- to 5-fold improvement in cytoplasmic expression of SCI using the same expression system. However, enrichment with the *dnaY* gene product in the SCI expression system did not result in any marked improvements in cell viability similar to those reported by Brinkmann & co-workers (1989). The potential improvement in cell viability was possibly negated by the toxic effects of the increased levels of recSCI.

The isolation from the bacterial cell lysates of the recSCI obtained by cytoplasmic expression was facilitated by the formation of dense inclusion bodies. The protein was purified to near homogeneity. Yet, the failure to characterise the obtained recSCI by automated N-terminal sequencing, using the Edman degradation method, was attributed to the presence of either an unprocessed N-terminal N-formyl methionine or the *in vivo* - or *in vitro* derivatisation of the N-terminal residue during subsequent isolation steps (e.g. N-acetylation) [Geisow & Green, 1990]. An N-terminal Met is a characteristic feature of several *E.coli* proteins [Hirel *et al.*, 1989] and often remains unprocessed in overexpressed recombinant proteins [Yasueda *et al.*, 1990]. However, Hirel & co-workers (1989) showed that the effectivity of *in vivo* N-terminal excision in *E.coli* decreased with the increase in maximal side-chain length of the penultimate amino acid. In particular, an N-terminal Met next to a penultimate Ser, as in the expressed recSCI, was cleaved in all studied cases [Hirel *et al.*, 1989; Dalbøge *et al.*, 1990]. It is therefore plausible that the high rate of production of recSCI leading to its accumulation as insoluble inclusion bodies is too fast for enzymatic deformylation and removal of the N-terminal Met [Yasueda *et al.*, 1990].

Extensive efforts to refold the solubilised recSCI inclusion bodies to obtain the functional inhibitor failed. Noting the relative ease and straightforwardness of refolding authentic SCI denatured in 6M guanidine hydrochloride (GuHCl) and 0.1M DTT, where over 60% inhibitory activity was regenerated after refolding by gradual removal of the GuHCl and DTT by dialysis [M.P. Esnouf; personal communication]. As was previously discussed, this failure to correctly fold could have been due to modification of the recombinant protein as compared to the authentic SCI, including the presence of the un- or misprocessed N-terminal residue. On the other hand, the presence of unusually stable intra- and intermolecular interactions that were not disrupted by the conditions used for denaturation could have prevented the formation of the native conformation [Darby & Creighton, 1990].

Without any further attempt to investigate the cause of the failure to obtain correctly folded, active recSCI, it was decided to construct a secretion vector in which the SCI gene was fused in-frame, downstream to a sequence encoding the signal peptide of OmpA, a major outer membrane protein of *E.coli* [Movva *et al.*, 1980]. With the signal peptide mediated secretion of SCI, the recombinant protein is translocated across the inner membrane of the cell into the periplasmic space whereby the signal peptide is cleaved by a signal peptidase, yielding a gene product with a correctly processed N-terminus. The conditions in the periplasm which, due to the permeability of the outer membrane, are directly influenced by those of the external medium, may be more favourable for the correct formation of disulphide bonds as compared to the highly reducing environment of the bacterial cytoplasm. In several reported cases, the particular application of OmpA mediated secretion has led to the successful, direct yield of active eukaryotic proteins that are similar, both in structure and function to SCI.

Expression of the fused gene under the control of the  $\lambda P_L$  promoter resulted in high levels of expression of the recombinant product estimated at around 10% of the total bacterial protein even without the supplemented *dnaY* gene product. The precise localisation of the recombinant protein could not be determined due to rapid death and spontaneous lysis of the cells, which resulted in extensive leakage of cytoplasmic protein. Fractionation of cell lysates into soluble and insoluble proteins showed that, by far, the major fraction of the recSCI was associated with the insoluble fraction, which indicated that the recSCI was yet again sequestered in insoluble inclusion bodies, but this time in the periplasmic space. [Bowden & Georgiou, 1990] in studies of OmpA signal sequence mediated secretion of  $\beta$ -lactamase showed that the mature protein was similarly sequestered as inclusion bodies in the periplasm.

As a first fractionation step, the recSCI inclusion bodies, together with the insoluble fraction of the cell lysate were solubilised with 7M urea and 0.1M DTT and subject to ion-exchange chromatography (IEC) on Q-sepharose resin. The effective binding of recSCI to the resin facilitated the removal of a large fraction of contaminating proteins and cellular components. To prevent the aggregation of protein on the IEC column, all wash and elution procedures were carried out in the presence of 5M urea.

The removal of urea at this stage, resulted in the preferred aggregation of the recSCI into stable, non-native dimers as was ascertained by non-reducing SDS-PAGE and electrospray mass spectrometry (ES-MS). N-terminal sequencing of the obtained dimer confirmed the presence of the mature protein, together with a mis-processed version in which both the N-terminal Ser and Ala had been cleaved. Both N-terminal sequencing and ES-MS indicated the intimate association with the dimer of a short 15-residue peptide. Coupled with the high concentration of the recSCI present on

removal of the denaturant, this peptide could be implicated in the incorrect folding and aggregation of the recombinant protein. [Darby & Creighton, 1990] recently suggested that such tight complexes of the recombinant protein with a cell component impedes correct folding. Such problems could result from incorrect interactions occurring during protein synthesis due to the exposure of interactive surfaces of the nascent protein to cellular surfaces [Ellis, 1990]. Ellis (1990) suggests that yield of the active protein could be improved by the co-expression of relevant molecular chaperones in the heterologous system or the incorporation of the chaperone *in vitro* during refolding of the protein.

Purification recSCI by hydrophobic interaction chromatography (HIC) failed to separate the monomeric subunits of the dimer. Yet, preliminary separation by HIC of the unfolded and reduced sample indicated the successful separation of the monomeric subunit. In ongoing work, the separated protein in the peak fraction will be isolated and characterised. The acidity of the HIC eluent (0.1% TFA) will retain the recSCI in unfolded state, preventing re-oxidation of the reduced Cys-SH residues. The purified recombinant protein that is presumably contained in the peak fraction will be initially characterised following which, refolding experiments will be undertaken.

Subject to the data obtained from these ongoing experiments, the encouraging results obtained indicate the successful culmination of this work with the establishment of a functional *E.coli* expression system for SCI. With the progressive optimisation of the system at the expression and isolation and purification stages, its useful application in the production of site-directed mutants of SCI could provide one with the invaluable tool for the study of the structure-function basis of the interaction of SCI with  $\beta$ XIIa.

## **Chapter 4**

# **Future Directions**

## 4. Future Directions

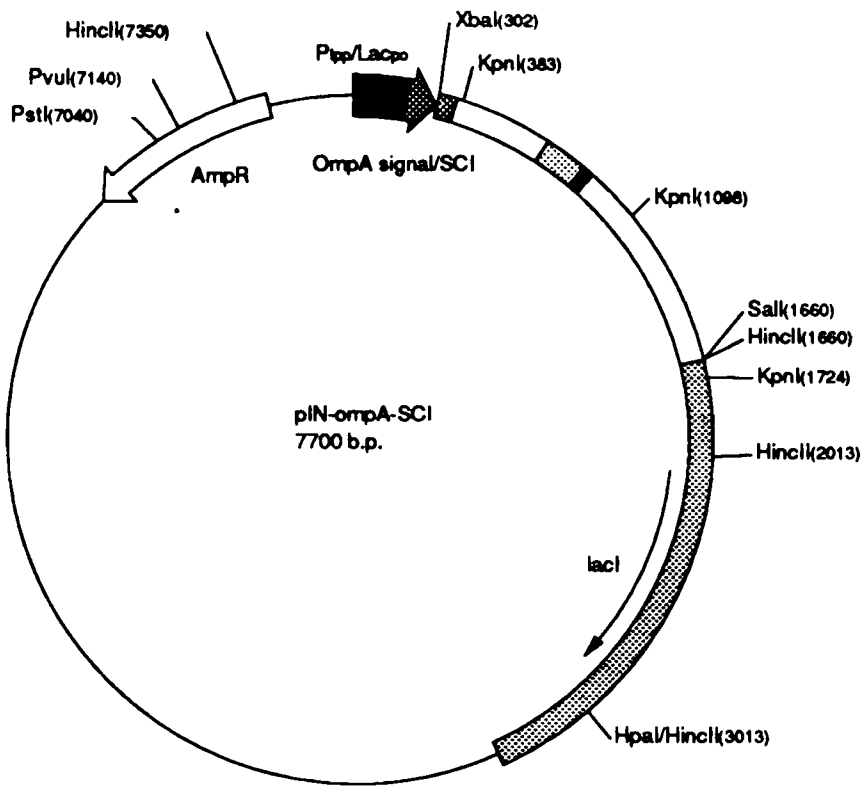
The desirability of establishing a low temperature expression system for SCI was reiterated during the course of this work in that even with periplasmic expression the protein product aggregated to form inclusion bodies. In the case of periplasmic expression system, the hydrophobic nature of recSCI may have caused it to aggregate in the cell membranes thereby causing extensive lysis of the cell.

As was discussed in section 3.2.6, no direct *in vivo* correlation can be made between temperature and the tendency to form inclusion bodies [Bowden & Georgiou, 1990]. Yet, Haase-Pettingell & King (1988) in work on the *Salmonella* phage P22 tailspike protein, showed that increase in temperature decreased the ability of nascent proteins to enter the correct folding pathway, thereby causing the accumulation of a non-native state. They suggested that this has a bearing on the formation of inclusion bodies - where, compounded by the unsuitable environment provided by the bacterial cytoplasm, perturbations to the folding pathway occur at high temperatures. This results in the tight association of non-native or partially folded forms of the protein leading to aggregation and the formation of inclusion bodies.

Reports of the successful use of secretion vectors for the functional production of eukaryotic proteins, including serine proteinase inhibitors that are structurally akin to SCI (see Table 3.2), almost exclusively make use of constant temperature inducible promoters such as the *lpp*, *tac* and *trp* promoters, which in conjunction with the *lac* operator are chemically inducible, by the addition of IPTG to cell cultures grown in minimal media.

With these promoters, expression has been induced at temperatures as low as 30°C. Lower temperatures would also prove more amenable towards the healthy growth of the *E.coli* thereby retaining the structural integrity of the cells in order to make maximal utilisation of the compartmentalisation of the secreted proteins into the periplasm.

However, of the low temperature expression vectors constructed so far, with the exception of the phage T7 promoter based expression vector [Quaderi, 1991] (page74), none have expressed even traces of the recombinant SCI. There remain several possible combinations to be tried on SCI that have been successfully used for the secretion of proteinase inhibitors and related eukaryotic proteins.



**Figure 4.1:** pIN-ompA-SCI; an *E.coli* Secretion Vector

Preliminary preparations have been undertaken in the attempt to clone the SCI gene into the secretion vector pIN-III-ompA3 [Ghrayeb *et al.*, 1984] (Figure 5.1) in which an OmpA signal-SCI fusion will be expressed

under the transcriptional control of the powerful *lpp* promoter. A contiguous *lac* operator provides one with the advantage of inducible expression. This plasmid has been successfully used in the production of various proteins including active chymotrypsin inhibitor 2 [Longstaff, *et al.* 1990], BPTI [Goldenberg, 1988] and related proteins and therefore would be well suited for expressing SCI.

In addition, the T7 promoter based SCI expression vector described by Quaderi (1991), could be adapted to secrete SCI by cloning the gene coding for the OmpA signal-SCI fusion.

Microbial expression of the recombinant protein has proved very difficult due to the very poor solubility (<1mg/ml) and highly hydrophobic properties of SCI.

In methods developed by M.P. Esnouf and co-workers (Nuffield Department of Clinical Biochemistry, University of Oxford), the authentic SCI was isolated from the seeds of various strains of opaque-2 corn by large scale acetone extraction of the seed paste. The hydrophobic environment of the corn seed with its high oil content could account for the fact that the purified SCI has very poor solubility characteristics in aqueous buffers. Furthermore, high-resolution separation of the purified protein (Figure 3.22) shows that it was far from homogeneous in that it consisted of several forms. These include the "short form" as described in this work and two "long forms" consisting of the following additional 9 residue C-terminal extensions that follow the C-terminal Ser residue (Figure 1.5):

-Gly-Val-Ala-Glu-Cys-Pro-Pro-Ile-Leu  
and -Gln-Val-Ala-Glu-Cys-Pro-Pro-Ile-Leu

In addition, the Cys<sub>21</sub>, was found to be replaced by an Arg residue in both long forms [M.P. Esnouf; personal communication]. Subsequent investigation of the "long form" with the additional C-terminal extension and the internal

Cys/Arg substitution have shown that it is at least 10-fold more soluble than the short form described in this work. However, all forms have the identical inhibitory selectivity. The gene for the long form can be obtained by relatively straightforward changes on the synthetic SCI gene - a single base change would be required to change the Cys<sub>21</sub> codon (TGT) to an Arg codon (CGT). The additional C-terminal extension can be incorporated by adding a suitable synthetic adapter at the *Bst*EII site near the C-terminus. The "long form" of SCI may have less tendency to aggregate due to its better solubility.

With the high levels of correctly processed recSCI obtained using the periplasmic expression system described in this work, the imminent establishment and optimisation of a purification and refolding procedure would confirm its usefulness in following up the initial aims of creating site-directed mutants to study the basis of the structure-function relationships of SCI.

## **Chapter 5**

# **Materials and Methods**

## 5. Materials and Methods

### 5.1. Instrumentation

Automated oligonucleotide synthesis was carried out by Mrs. Val Cooper (Dyson Perrins Laboratory, University of Oxford) on an Applied Biosystems DNA synthesiser 380B. Vacuum concentration of samples was carried out on a Savant SpeedVac concentrator. Lyophilisation of samples was performed on a Virtis FreezeMobile 5SL freeze drier. The pH of aqueous solutions was measured with a Radiometer PHM84 pH meter and solutions containing Tris-HCl buffer with a Russell CD660 and special Tris electrode together with a temperature probe. Spectrophotometric measurements were performed by a Pye Unicam SP8-100 UV/VIS spectrophotometer. Centrifugation of small samples were carried out in an MSE Microcentrifuge and larger samples in a refrigerated MSE bench-top centrifuge or a Beckman J2-21 centrifuge. Electroporation was carried out in a Bio-Rad Gene Pulser. Amino-terminal sequencing of protein samples was carried out by Mr. Tony Willis (Laboratory of Molecular Biophysics, University of Oxford) on an Applied Biosystems 470A protein sequencer. Scintillation counting was carried out using an LKB Wallac 1215 Rackbeta Liquid Scintillation Counter. Computer analysis of gene design was performed using the MacVector program (IBI Ltd., Hemel Hempstead, Herts.) for the Apple MacintoshII computer. Electrospray mass spectrometry of proteins was carried out by Dr. Robin Aplin (Dyson Perrins Laboratory, University of Oxford) on a VG BIO Q quadrupole mass spectrometer. Ethidium bromide stained DNA was detected in ultraviolet light on an UV-Transilluminator (UVP Inc., San Gabriel, CA,

U.S.A.). Sonication of samples was carried out in an an MSE Soniprep 150 Ultrasonicator using a 4mm probe.

## **5.2. Materials**

### **5.2.1. Water**

All water used was distilled and deionised by a Milli Q3 water purifying system (Millipore, Watford, Herts.) and sterilised by autoclaving for 20min. at 120°C, 15lb in<sup>-2</sup>.

### **5.2.2. Chemicals, Biochemicals, Enzymes and Proteins**

All chemicals used were of molecular-biological or analytical grade, or the highest purity grade available and were obtained from Sigma (Poole, Dorset), Aldrich (Gillingham, Dorset), BDH (Atherstone, Warwicks.). Labelled adenosine 5'- $\alpha$ -[<sup>35</sup>S]-thiotriphosphate (10mCi/ml) ([ $\alpha$ -<sup>35</sup>S]-dATP) and U-<sup>14</sup>C protein hydrolysate (50 $\mu$ Ci/ml) and organic counting scintillant (OCS), tissue solubiliser (TCS) and X-ray film (Hyperfilm- $\beta$ Max) were obtained from Amersham International plc (Aylesbury, Bucks.). For dialysis of protein solutions, Spectrum Spectra/Por dialysis tubing (M.Wt. cutoff 3500) was used (Pierce & Warriner, Chester).

Agarose, low-melting-point agarose and ultrapure urea were obtained from Gibco BRL Ltd. (Uxbridge, Middx.). Acrylamide and N,N'-methylene-bisacrylamide were obtained from Pharmacia (Milton Keynes, Bucks.).

The Sequenase Version 2.0 sequencing kit was obtained from Cambridge Bioscience (Cambridge). Restriction endonucleases and other DNA modification enzymes, including T4 DNA ligase, DNA polymerase I (Klenow fragment), calf-intestinal phosphatase, T4 polynucleotide kinase

were purchased from Northumbria Biologicals Ltd. (Cramlington, Northumberland), Boehringer Mannheim (Lewes, Sussex), New England Biolabs (Bishop's Stortford, Herts) and Pharmacia.

Zymolyase-20T a mixture of  $\beta$ -glucanases was obtained from Seikagaku Kogyo Corp. (Tokyo, Japan).

DNaseI and RNaseA were obtained from Boehringer Mannheim. RNaseA was dissolved in 50mM sodium acetate (pH 5.0) containing 15% glycerol at a concentration of 5mg/ml, and boiled for 10 min to inactivate any contaminating DNase, and then stored in aliquots at -20°C. Lysozyme and bovine serum albumin were obtained from Sigma.

For the preparation of microbiological growth media, Bacto-tryptone, Bacto-yeast extract, Bacto-peptone, Bacto-agar, yeast nitrogen base without amino acids and casamino acids were obtained from Difco Laboratories Ltd. (East Molesey, Surrey). Biochemical grade amino acids were obtained from BDH. Antibiotics were obtained from Sigma and Boehringer Mannheim.

Authentic sweet corn inhibitor and anti-SCI goat antiserum (lyophilisate) were kind gifts from Dr. M.P. Esnouf (Nuffield Department of Clinical Biochemistry, University of Oxford). Horse-radish-peroxidase-conjugated anti-goat rabbit antibody was obtained from Dakopatts Ltd. (High Wycombe, Bucks.). Tween-20 and the horse-radish-peroxidase substrate, diaminobenzidine (DAB) were obtained from Bio-Rad Laboratories Ltd. (Hemel Hempstead, Herts.).

The chromogenic substrate S2302 was obtained from KabiVitrum Ltd. (Uxbridge, Middx.).

### 5.2.3. Phenol

Liquefied phenol in water (80% (w/w) A.R. from FSA Laboratory Supplies, Loughborough, Leics.) was used to extract proteins from preparations of DNA. It was stored under nitrogen at 4°C. Prior to use, the phenol was equilibrated by extraction first with 1M Tris-HCl (pH 8.0) then in 0.1M Tris-HCl (pH 8.0). The aqueous phase was removed and 8-hydroxyquinoline (0.2%(w/v)) and  $\beta$ -mercaptoethanol (0.2%(v/v)) added to the equilibrated phenol and the phenol finally mixed with 0.1 volume of T.E. Buffer (pH 8.0). Phenol/chloroform/isoamyl alcohol (PCI) was prepared by mixing equilibrated phenol, chloroform and isoamyl alcohol in the ratio 24:24:1.

## 5.3. DNA Preparation and Analysis

Most standard molecular biology protocols were carried out as described in Ausubel *et al.*, (1988) and Sambrook *et al.*, (1989).

### 5.3.1. Buffers

TE Buffers:

pH 7.5

10mM Tris-HCl (pH 7.5)

1mM EDTA (pH 8.0)

pH 8.0

10mM Tris-HCl (pH 8.0)

1mM EDTA (pH 8.0)

Commonly used DNA electrophoresis buffers:

| Buffer             | Working Solution                                | Concentrated stock (per litre)                                                |
|--------------------|-------------------------------------------------|-------------------------------------------------------------------------------|
| Tris-acetate (TAE) | 0.04M Tris acetate<br>1mM EDTA                  | 50x: 242g Tris base<br>57.1ml glacial acetic acid<br>100ml 0.5M EDTA (pH 8.0) |
| Tris-borate (TBE)  | 89mM Tris-borate<br>89mM boric acid<br>2mM EDTA | 5x: 54g Tris base<br>27.5g boric acid<br>20 ml 0.5M EDTA (pH 8.0)             |

5.3.2. Enzyme Reactions

Restriction endonuclease digests ligations using T4 DNA ligase were performed as recommended by the suppliers. The correct reaction conditions were obtained by using the reaction buffer concentrates supplied with the enzymes.

For restriction digests, the quantity of enzyme (in Units) required for a digest was calculated from the following formula:

$$\text{Amount (Units)/}\mu\text{g plasmid} = \frac{48000}{x} \times \frac{y}{z}$$

Where, x is the size (in b.p.) of the plasmid to be digested, y is the number of restriction sites to be cleaved, and z is the number of cleavage sites on the λ-phage plasmid. Usually, quantities 10-fold higher than the calculated value were used at 37°C for 30min to ensure complete digestion. For partial digests, 1/10th of the calculated amount was used, with reduced incubation times (5-15 min) at 37°C. For multiple restriction digests with different enzymes, depending on the buffer requirements of each enzyme, simultaneous or successive incubations were carried out.

De-phosphorylation of the 5'-ends of DNA fragments was carried out using calf intestinal phosphatase (CIP) in 10mM Tris-HCl (pH 8.2) containing

1mM  $\text{ZnCl}_2$  and 1mM  $\text{MgCl}_2$  at the recommended concentrations and conditions. Phosphorylation of oligonucleotides was carried out using T4 polynucleotide kinase in the presence of 1mM ATP in 10mM Tris-HCl (pH7.5) containing 10mM  $\text{MgCl}_2$ , 5mM DTT and 5mg/ml BSA.

Ligations were carried out using T4 DNA Ligase at the conditions recommended by the supplier. Hexaminecobalt (III) chloride (HCC) to a final concentration of 2mM was added to the reaction mix so as to promote blunt-end, intra-molecular ligation [Rusche & Howard-Flanders, 1985] (one minute after the addition of the ligase in the presence of HCC, potassium chloride was added to the reaction mix to a final concentration of 30mM. This was to prevent multiple insertions [Murray, 1986]).

### **5.3.3. Large Scale Preparation of Plasmid DNA from *E.coli* - Alkaline-Lysis Method**

This procedure was according to Birnboim & Doly, (1979) with a few modifications:

*E.coli* cells bearing the required plasmid were grown overnight to stationary phase in 100ml LB-medium supplemented with the appropriate antibiotic. The obtained cells were harvested by centrifugation at 10,000g at 4°C. The obtained cell pellet was resuspended in 10ml QPP1 (25mM Tris-HCl pH 8, 10mM EDTA & 50mM Glucose) containing 2mg/ml lysozyme and incubated at R.T. for 5min. A quantity of 20ml of QPP2 (0.5N NaOH in 1% SDS) was added, mixed well and incubated at R.T. for 5min. To the mixture 15ml of QPP3 (16.8g KOH pellets and 18ml glacial acetic acid in 62ml water) was added, mixed well and incubated for 5min on ice. The mixture was centrifuged for 5min at 4500rpm (MSE table-top centrifuge). The supernatant obtained was transferred to fresh tubes and 25ml isopropanol added and mixed well. After incubation for 15min on ice, the mixture was centrifuged

for 10min at 4500rpm. The supernatant was discarded and the pellet resuspended in 500µl T.E. buffer (pH 7.5) and transferred to a microfuge tube. Any insoluble material was removed by centrifugation. A quantity of 25µl of RNase stock solution (see above) was added and the mixture incubated at 37°C for 30min. The solution was then repeatedly extracted with equal volumes of PCI until the white interface disappears. To the extracted aqueous phase 0.6x volume of 5M ammonium acetate and 1x volume of isopropanol were added, mixed well and incubated for 15min at R.T. The plasmid DNA was precipitated by centrifugation in a microcentrifuge for 10min at 13000g. The obtained pellet was rinsed in ice-cold 70% ethanol, dried in a vacuum centrifuge (Speedvac) and resuspended in 100µl T.E. buffer (pH 7.5). Depending on the plasmid and *E.coli* strain used, yields of 1-2µg plasmid were obtained.

#### **5.3.4. Modified Alkaline-Lysis Method for the Purification of *E.coli* Plasmid DNA Suitable For Double-Stranded Sequencing**

This plasmid "miniprep" was carried out as described by Johnson, (1990). As a further purification step, the plasmid suspensions finally obtained were subjected to PEG precipitation.

A quantity of 4 ml of overnight *E.coli* culture in SOB medium were spun down, portionwise, in 2 ml microfuge tubes for 5 minutes at 14,000g. The supernatant was discarded and the cell pellet resuspended in 50µl ice-cold Tris-Glucose-EDTA solution (50mM Tris-HCl pH 8.0, 50mM Glucose & 10mM EDTA pH 8.0) by vigorous vortexing. A 50 µl quantity of lysozyme solution (8 mg/ml in ice-cold Tris-Glucose-EDTA) was added and the mixture vortexed rapidly for 5 sec. After 5 min. on ice, 200µl of NaOH-SDS solution (0.2N NaOH & 1% SDS) at RT was added, the tubes vortexed and incubated on ice for 5 min. At this stage, a viscous but clear suspension was obtained. A 150 µl

quantity of KAc-formic acid solution (3M potassium acetate & 1.8M formic acid) at RT was then added. The tubes were vortexed and incubated on ice for 15 min. with occasional shaking and inversion. The tubes were then spun at RT for 30 min. at 13000g and the resulting supernatant transferred to another tube. To the supernatant was added 300 µl of PCI, the tubes vortexed vigorously and spun at RT for 5 min. at 14,000g. The aqueous supernatant was transferred to a new tube to which were added 900 µl of 95% ethanol at RT. The tubes were vortexed incubated at RT for 5 min, during which time the tubes were inverted repeatedly and then spun at RT at 14,000g for 5 min. The supernatant was discarded, the pellet dried and resuspended in 50 µl TE pH 8.0 at RT. At this point, the pellet takes several minutes to dissolve. To the resulting suspension 50µl 5M ammonium acetate was added, mixed well by vortexing and placed on ice for 20 min. Following this, the suspension was spun for 3 min. and the supernatant transferred to a new tube and mixed with 200 µl isopropanol at R.T. After 20 min. at RT, the tubes were spun and the obtained plasmid pellet washed with ice-cold 70% ethanol, dried and resuspended in 50 µl TE pH 8.0.

### **5.3.5. Small-Scale Preparation of *E.coli* Plasmid DNA ("Minipreps") - Boiling Method**

This procedure described by Holmes & Quigley, (1981) was applied for the rapid preparation of small quantities of plasmid from *E.coli*:

Quantities of 1.5ml of overnight cultures of single, well-isolated colonies of *E.coli* were transferred to microfuge tubes and centrifuged at RT for 5 min at 13000g. The supernatant was discarded and the cell pellet resuspended in 350µl lysis buffer (50mM Tris-HCl pH 8.0, 50mM EDTA pH 8.0, 0.5% Triton X-100 (v/v) and 8% (w/v) sucrose) containing 1mg/ml lysozyme. The sample tubes were placed in boiling water for exactly 40sec. The samples were then centrifuged for 30min at R.T. at 13000g. The obtained loose

gelatinous pellet was carefully removed with a toothpick. The supernatant was extracted several times with equal volumes of PCI, until no white interface was seen between the upper aqueous phase and the lower phenol phase. The plasmid DNA which was contained in the lower, aqueous phase was precipitated by adding 0.6x volume of 5M ammonium acetate solution and 1x volume of isopropanol. After good mixing, the suspension was incubated at RT and then spun at room temperature at 13000g for 10min. The obtained pellet was washed in chilled 70% ethanol, dried and resuspended in 50µl TE buffer (pH 7.5). A quantity of 3µl DNase free RNase was added and the samples incubated for 30min at 37°C.

#### **5.3.6. PEG Precipitation**

To 50 µl plasmid solution 25 µl of PEG-NaCl solution (20% (w/v) PEG 8000 & 2.5M NaCl) was added. The mixture was mixed well by vortexing and left on ice for 20 min. and spun at RT for 5 min. The resulting supernatant was discarded and the DNA pellet washed in ice-cold 70% ethanol, dried and resuspended in 50 µl TE 8.0. The purity and DNA concentrations were estimated by electrophoresis on a 1% Agarose gel in TBE Buffer.

#### **5.3.7. Ethanol Precipitation**

To aqueous samples of DNA, 1/10th volume of 3M sodium acetate (pH 6.0) and 2.5 volumes of chilled ethanol were added and mixed well. The mixture was left for 30min, at -20°C and then spun in a microcentrifuge (14000g) for 30min. The supernatant was discarded and the pellet carefully washed in chilled 70% (v/v) ethanol, dried and resuspended in an appropriate volume of TE buffer. Analytical grade ethanol was always used in this procedure.

**5.3.8. Analytical Gel Electrophoresis of DNA on Agarose Gels**

Gels were prepared and run in two types of gel apparatus, the BRL H3 gel electrophoresis tank and the BRL baby-gel H6 tanks (Gibco BRL Ltd.) where the gel dimensions were 15 x 12 x 0.5cm and 7 x 5 x 0.5cm, respectively.

Agarose gels were prepared by melting in a microwave oven, the appropriate amount of powdered agarose in 1 x TBE buffer containing 0.5µg/ml ethidium bromide. After cooling to about 50°C, the melted gel was poured into a gel former and allowed to set at RT for about 30min. The concentrations of agarose, depending on the size of DNA fragments to be separated, are listed below:

| <u>Duplex Length</u> | <u>Agarose Concentration (%;w/v)</u> |
|----------------------|--------------------------------------|
| ≈150-500             | 2.5                                  |
| ≈300-1000            | 2.0                                  |
| ≈500-2500            | 1.5                                  |
| ≈700-4000            | 1.0                                  |
| ≈1000-9000           | 0.5                                  |

-----

For each expected band to be separated, 10-100ng of sample in sample buffer (25% (w/v) Ficoll 400 (Pharmacia) and 0.2% (w/v) bromophenol blue) was heated to 65°C for 10min and immediately loaded into the gel sample wells. Electrophoresis was carried out in 1 x TBE buffer at a voltage gradient of 15V/cm. Ethidium bromide stained DNA was detected over an UV transilluminator (245nm).

**5.3.9. Preparative Gel Electrophoresis of DNA in Low-Melting-Point Agarose Gels**

The method routinely used for satisfactory yields of DNA restriction endonuclease fragments using low-melting-point agarose (LMP agarose) was first described by Weislander, (1979). LMP agarose is a derivative of agarose

MRC IMMUNOCHEMISTRY UNIT      SEQUENCING DATA SHEET  
Applied Biosystems 470A/120A/473

|                                           |                        |                         |                 |                |
|-------------------------------------------|------------------------|-------------------------|-----------------|----------------|
| Initial yield                             | Run no.<br>91124H      | Std no.<br>h124         | Date<br>4.9.91  | Stepwise yield |
|                                           |                        | Multimethod.<br>rsci152 |                 |                |
|                                           | Name<br>Rohan da Silva |                         |                 |                |
| Sample name<br>Recomb. SCI Fr. 15 hplc #2 |                        |                         | Amount<br>100µl |                |

| No. | Yield 1 | AA residue(s) |        |   |        |    | No. | Amnt.<br>picomol | AA residue(s) |   |   |   |   |
|-----|---------|---------------|--------|---|--------|----|-----|------------------|---------------|---|---|---|---|
|     |         | 1             | yield2 | 2 | yield3 | 3  |     |                  | 1             | 2 | 3 | 4 | 5 |
| 1   | 36.1    | S             | 36.1   | G | 24.3   | M  | 26  |                  |               |   |   |   |   |
| 2   | 30.0    | A             | 21.7   | T | 7.1    | S  | 27  |                  |               |   |   |   |   |
| 3   | 20.9    | G             | 11.1   | S | 10.4   | F  | 28  |                  |               |   |   |   |   |
| 4   | 21.1    | T             |        | - |        | RV | 29  |                  |               |   |   |   |   |
| 5   | 11.5    | S             | 23.5   | V |        |    | 30  |                  |               |   |   |   |   |
| 6   | -       | -             | 15.6   | P |        |    | 31  |                  |               |   |   |   |   |
| 7   | 23.4    | V             | 19.8   | G |        | AL | 32  |                  |               |   |   |   |   |
| 8   | 17.8    | P             | 8.7    | W |        | D  | 33  |                  |               |   |   |   |   |
| 9   | 15.6    | G             | 20.7   | A |        |    | 34  |                  |               |   |   |   |   |
| 10  | 8.6     | W             | 16.2   | I |        | V  | 35  |                  |               |   |   |   |   |
| 11  | 22.7    | A             | 12.9   | P |        | N  | 36  |                  |               |   |   |   |   |
| 12  | 13.3    | I             | 5.2    | H |        |    | 37  |                  |               |   |   |   |   |
| 13  | 17.7    | P             | 12.5   | N |        |    | 38  |                  |               |   |   |   |   |
| 14  | 6.2     | H             |        | - |        |    | 39  |                  |               |   |   |   |   |
| 15  | 14.5    | N             | 10.5   | L |        |    | 40  |                  |               |   |   |   |   |
| 16  |         |               |        |   |        |    | 41  |                  |               |   |   |   |   |
| 17  |         |               |        |   |        |    | 42  |                  |               |   |   |   |   |
| 18  |         |               |        |   |        |    | 43  |                  |               |   |   |   |   |
| 19  |         |               |        |   |        |    | 44  |                  |               |   |   |   |   |
| 20  |         |               |        |   |        |    | 45  |                  |               |   |   |   |   |
| 21  |         |               |        |   |        |    | 46  |                  |               |   |   |   |   |
| 22  |         |               |        |   |        |    | 47  |                  |               |   |   |   |   |
| 23  |         |               |        |   |        |    | 48  |                  |               |   |   |   |   |
| 24  |         |               |        |   |        |    | 49  |                  |               |   |   |   |   |
| 25  |         |               |        |   |        |    | 50  |                  |               |   |   |   |   |

Notes:                      Yields of peptides are in picomoles.

Appendix IX:      N-Terminal Sequence Analysis of recSCI

spatial arrangement as in chymotrypsin. The catalytic triad of chymotrypsin was retained in its known spatial configuration having discarded the remaining protein structure of chymotrypsin. The supporting framework of three amphipathic helices was constructed so as to hold the three residues in their correct configuration. A fourth helix-forming chain with an N-terminal Glu residue was added so as to stabilise the entire structure and provide backbone hydrogen bonds to form the oxyanion hole. All four helices were covalently linked together at their C-termini through the side-chain amine groups of Orn and Lys residues. The correctly assembled chymohelizyme-1 was shown to measurably hydrolyse selected chymotrypsin substrates.

The substrate acetyltyrosine ester was hydrolysed by chymohelizyme-1 at the rate of catalysis  $k_{\text{cat}}$  of  $0.042 \text{ s}^{-1}$ , a value  $\approx 10^5$  times greater than spontaneous hydrolysis and about 0.03% that of chymotrypsin. It was inactivated by treatment with PMSF, an irreversible inhibitor (inactivator) of chymotrypsin and serine proteinases. This is possibly the first instance in which present knowledge and methods in the prediction and design of peptide structures have been used to successfully model and construct a synthetic polypeptide with catalytic properties.

#### **1.1.4. Inhibitors of Serine Proteinases**

Although proteinases are a physiological necessity, they are also potentially a great hazard to the organism. Uncontrolled proteolysis could lead to the destruction of essential protein components or the premature activation of zymogens. Examples of disease states involving serine proteinases include inflammation, rheumatoid arthritis, pulmonary emphysema and cerebral and coronary infarction resulting from thromboses. The importance of this hazard is reflected by the fact that nearly 10% of plasma proteins are proteinase inhibitors [Powers & Harper, 1986] constituting by weight, the third largest group of proteins in human plasma after albumin

For the analysis of the recSCI obtained, a 3.5-litre culture of JM109[pRS101+pUBS520] in SOB (pH 7.5) was prepared:

An overnight culture grown from a single colony of JM109[pRS101+pUBS520] in 350ml SOB (pH 7.5) containing tetracycline and kanamycin was diluted 1/10th in 3.5 litres fresh medium contained in 3 4-litre conical flasks and incubated with shaking (250 rpm) at 27°C for 4h until mid-log phase of growth was reached ( $OD_{600} \approx 2.0$ ). At this point, expression was induced by adding an equal volume of medium heated to 57°C and increasing the incubator temperature to 42°C. Incubation was continued overnight. Cells were harvested by centrifugation for 20min at 10,000g at 4°C, washed in ice-cold lysis buffer (50mM Tris-HCl pH 8.0, 1mM EDTA & 100mM sodium chloride) and re-centrifuged. All manipulations at this point and beyond were carried out at 4°C (unless otherwise stated) so as to minimise the proteolytic degradation of the recombinant protein. A total of 23g wet-weight of cells was obtained.

Disruption and fractionation of the cells were carried out as described by Marston (1986). A quantity of 6g wet-weight of the cells were resuspended in 35ml lysis buffer containing 1M sucrose. To reduce proteolytic activity, phenylmethylsulphonyl fluoride (PMSF) was added to the suspension to a final concentration of 0.15mM. The suspension was homogenised using a tissue homogeniser and lysozyme added to a final concentration of 1mg/ml. The completion of spheroplast formation was monitored by diluting a small portion of the suspension 50-fold in water. Spheroplast formation was taken to be complete when the optical density at 450nm fell 80-85% within 10 sec. The suspension was incubated for 30 min. on ice with occasional stirring and then sonicated on ice with two 15 sec bursts. The successful disruption of the spheroplasts was indicated by the very viscous consistency of the resulting suspension. This was due to leakage of *E.coli* chromosomal DNA. A quantity

which melts at 60-65°C, well below the melting-point of DNA duplexes. It also contains less polysulphated polysaccharides than standard agarose. They inhibit many of the DNA enzymes (e.g. DNA polymerase I). Gels (7 x 5 x 0.5cm) were prepared as described above (section 4.3.3.) with the exception of replacing 1 x TBE with 1 x TAE buffer. The agarose was melted by boiling, cooled to 37°C and poured and set at 4°C. To obtain optimal resolution and prevent the gel melting by overheating, electrophoresis was carried out at a voltage gradient of 5V/cm at 4°C. The desired ethidium bromide stained DNA fragment, was cut out of the gel. The obtained gel slice was transferred to a 1.5ml microfuge tube and melted by placing it for 10min at 65°C. Approximately 3x volumes of TE (pH 7.5) buffer was added and the mixture re-heated to 65°C and mixed well. After cooling to RT, an equal volume of phenol was added and the suspension kept for 15min. at RT with occasional vigorous shaking and then centrifuged for 5min in a microfuge at 14000g. The upper aqueous phase containing the extracted DNA was separated from the lower phenol phase, taking care not to disturb the white material at the interface. To optimise yield, the phenol was re-extracted with T.E. buffer and the aqueous phase pooled with that of the previous extraction. The obtained aqueous sample was extracted twice with PCI and the DNA precipitated with ethanol and resuspended in an appropriate volume of TE buffer.

#### **5.3.10. Purification of Synthetic Oligonucleotides by Denaturing Urea-Polyacrylamide Gel Electrophoresis**

Oligonucleotides were synthesised on an Applied Biosystems 380B DNA synthesiser. For most applications the synthetic oligonucleotide has sufficient purity but in certain cases such as in the use of the oligonucleotide as a primer, the invariable presence of traces of truncated, incompletely synthesised oligonucleotide could be disadvantageous. Separation and

purification of the deprotected oligonucleotide was carried out by urea/polacrylamide gel electrophoresis.

The crude, protected synthetic oligonucleotide which was supplied in about 1ml of a saturated ammonia solution was deprotected by incubating the sample in an airtight vial for 8 hours at 55°C. The sample was then vacuum dried in a Savant® Speedvac concentrator, the dried pellet resuspended in 400µl water and ethanol precipitated and the obtained, dried pellet resuspended in 100µl water. Five microlitres were removed to determine the concentration of the oligonucleotide by measuring against water, the value of OD<sub>260</sub>\*. The remaining solution was dried again in a Savant® Speedvac concentrator, the pellet resuspended in 10µl water. This solution can then be stored at -20°C.

Five microlitres of the above oligonucleotide solution was mixed with 15µl deionised formamide and loaded onto a pre-heated 5% denaturing urea/polacrylamide gel (10µl oligonucleotide/formamide mix per well). The vertical slab gel electrophoresis apparatus used for this purpose was obtained from Koch-Light Ltd. (Haverhill, Suffolk, U.K.), facilitating the assembly of gels 18cm x 36cm using 1mm spacers. Prior to assembly, the glass plates were washed thoroughly with detergent, dried and swabbed clean with ethanol. For convenient handling of the gel after electrophoresis, both the notched and the

---

\*The concentration of a DNA solution can be determined spectrophotometrically by applying the following equation:

$$\text{Conc. } (\mu\text{M}) = \frac{\text{OD}_{260}}{\epsilon_{260}} \times 10^6$$

where OD<sub>260</sub> is the absorbance of the DNA solution at 260nm and  $\epsilon_{260}$  extinction coefficient (cm<sup>-1</sup> M<sup>-1</sup>) of the given DNA molecule. The total extinction coefficient of an oligonucleotide is calculated by summing the contribution of the individual deoxynucleotide bases. The  $\epsilon_{260}$  (cm<sup>-1</sup> M<sup>-1</sup>) of the single bases are: A: 15400, G: 11700, T: 8800 and 7300. The average molecular weight of a deoxynucleotide base is 330.

long plates were treated on their inside surfaces with methyl silane as described in section 4.4.6. This ensures that the gel will not stick to either plate, facilitating separation after electrophoresis. After treatment, the plates were assembled as instructed, using 1mm spacers.

Depending on the length of the oligonucleotide to be purified, gels of 4% or 6% were prepared using Sequagel™ Concentrate (National Diagnostics, Nashville N.J., U.S.A.) which consisted of 25% (19:1) acrylamide/bisacrylamide and 8.3M Urea. The concentrate was diluted in 1xTBE buffer containing 7M Urea, to make up a 100ml solution of the required percentage. Quantities of 600µl 10% (w/v) ammonium persulphate and 100µl TEMED were then added, mixed thoroughly and then poured and left for at least an hour for complete polymerisation.

Prior to loading, the gel was pre-heated for 30min. at 1100V (25-27mA) in 1xTBE following which the oligonucleotide/formamide samples were preheated for 2min. at 65°C, loaded (10µl per well) and run at the same settings until the dark-blue bromophenol blue dye front ran about two-thirds of the length of the gel.

The gel cassette was then disassembled and the gel carefully transferred onto a plastic cling-film. The oligonucleotide bands, located by UV shadowing on thin-layer chromatography plates were cut out. For the elution of the separated oligonucleotide, the gel slice was sealed in a dialysis tubing together with 3ml of TE buffer (pH 7.5) and dialysed overnight in distilled water. The resulting solution was removed from the dialysis tubing and an aliquot of 30µl removed for determining the DNA concentration and purity. Measurement of OD<sub>260</sub> indicates the concentration of DNA (see above) and the ratio OD<sub>260</sub>/OD<sub>280</sub> indicates the purity of the preparation (values of 1.7 and over are acceptable). The remaining solution was evaporated to dryness

in a Savant® Speedvac concentrator and resuspended in water to give a final concentration of 10pmoles/ $\mu$ l.

## **5.4. DNA Sequence Analysis**

### **5.4.1. Denaturing of Double-Stranded Plasmids**

Prior to priming and sequencing, it is necessary to denature the double-stranded plasmids.

To 45  $\mu$ l purified plasmid, added 5  $\mu$ l 2M NaOH, incubated 30 min. at 37°C following which, the mixture was neutralised by adding 5  $\mu$ l of 3M ammonium acetate at pH 5.4. The DNA was then precipitated by adding 200  $\mu$ l ice-cold ethanol, incubation at -20°C for 20 min. and spinning at 14,000g at 4°C for 30 min. The resulting pellet was then washed with 70% ice-cold ethanol, dried and resuspended in 10  $\mu$ l water. A yield of >1  $\mu$ g of plasmid was estimated at this stage.

### **5.4.2. Sequencing Reactions**

All sequencing reactions were carried out as described in the instruction booklet accompanying the commercially obtained sequencing kit ["Sequenase® Version 2.0: Step-By-Step Protocols for DNA Sequencing With Sequenase® Version 2.0", (2nd Edition), 1989, United States Biochemical Corporation, Cleveland, Ohio, U.S.A.]. This describes a greatly improved method of DNA sequencing by the chain-termination method, initially developed by Sanger *et al.*, (1977), using Sequenase® Version 2.0, a genetic variant of bacteriophage T7 DNA polymerase [Tabor & Richardson, 1987]. The compositions of all the reaction mixes provided in the kit were listed therein. All sequencing reactions were run in 1.5 ml microfuge tubes.

**5.4.3. Annealing Templates and Primers**

For the sequencing of all plasmids containing the complete SCI gene sequence, the two oligonucleotides, YK22-8 and YK22-15 were used. These oligonucleotides were initially used in the construction of the synthetic SCI gene [Kodo, 1988] and available, phosphorylated, as 10 pM solutions in water. The sequences of these primers are as follows:

YK22-8 (34-mer): 5'-CAATTCTGGCCATGGCAATCTTGGTCTTGGACCG-3'  
YK22-15 (35-mer): 5'-GGGCCAGACGCTCAATTGGAAGGTCGAGTGGAAGA-3'

These oligonucleotides were selected as primers so as to maximise the sequence information obtained. This would include reading the sequence of most of the SCI gene as well the junctions at both ends of the gene so as to ascertain correctness of insertion of the gene relative to the various expression and secretion vectors. Each oligonucleotide was separately annealed to equal portions of each denatured plasmid as follows:

|                                |     |
|--------------------------------|-----|
| Denatured plasmid              | 5µl |
| Primer                         | 4µl |
| Sequenase Reaction Buffer (5x) | 3µl |
| Water                          | 2µl |

The mixture was warmed in a to 65°C for 5 min. and cooled down slowly over a period of 1 hour to R.T.

**5.4.4. Labelling Reaction**

These were carried out as described, where the separate portions of each plasmid annealed to the two primers are mixed as follows:

|                                      |      |
|--------------------------------------|------|
| Plasmid/Primer                       | 10µl |
| DTT 0.1M                             | 1µl  |
| 1/5 diluted labelling mix            | 2µl  |
| [α- <sup>35</sup> S]-dATP (10mCi/ml) | 1µl  |

Diluted Sequenase® Version 2    2µl    (Enzyme stock diluted 1:8 in Sequenase Enzyme Dilution Buffer.)

The mixture was incubated at R.T. for 5 min.

#### 5.4.5. Termination Reactions

Quantities of 2.5µl each of the ddGTP, ddATP, ddTTP and ddCTP Termination Mixes were placed in microfuge tubes labelled G, A, T and C, respectively. The tubes were pre-warmed to 37°C after which 3.5µl of the above labelling mix was added to each of the tubes and mixed well and the incubated for upto 30 min. at 37°C. Following incubation, 4µl Stop Solution was added to each tube and mixed thoroughly. The samples were hereby ready for analysis by denaturing urea-polyacrylamide gel electrophoresis.

#### 5.4.6. Denaturing Gel Electrophoresis

The vertical slab gel electrophoresis apparatus used for this purpose was obtained from Koch-Light Ltd. (Haverhill, Suffolk, U.K.), facilitating the assembly of gels 18cm x 36cm. A "sharkstooth" comb (0.4mm thick) was used for sample loading.

Prior to assembly, the glass plates were washed thoroughly with detergent, dried and swabbed clean with ethanol. For convenient handling of the gel after electrophoresis, the plates were then treated as follows on their inside surfaces with silanising agents.

Long Plate ("sticky"):      The following were mixed:  
                                          15µl γ-(methacryloxy)-propyltrimethoxysilane  
                                          150µl 3% (v/v) Acetic Acid  
                                          in 5ml ethanol  
                                          The plate was smeared evenly with this  
                                          mixture, left to dry for 5 min. and wiped clean  
                                          with ethanol.

Notched Plate ("non-stick"):

One side of plate was treated with 5% (v/v) dichloromethylsilane in chloroform, left to dry for 30min. and wiped with ethanol.

The polymerised acrylamide gel will covalently bind to the long methacryl silane treated plate and be repelled by the methyl silane treated notched plate. After treatment, the glass plates were assembled as instructed, using 0.4mm spacers.

Sequencing gels (3-4%) were prepared using Sequagel™ Concentrate (National Diagnostics, Nashville N.J., U.S.A.) which consisted of 25% (19:1) acrylamide/bisacrylamide and 8.3M Urea. The concentrate was diluted in 1xTBE buffer containing 7M Urea, to make up 40ml solution of the required percentage. Quantities of 240µl 10% (w/v) ammonium persulphate and 40µl TEMED were then added, mixed thoroughly and then poured and left for at least an hour for complete polymerisation.

Prior to loading, the gel was pre-heated for 30min. at 1100V (25-27mA) following which the samples were preheated for 2min. at 65°C, loaded and run at the same settings until the dark-blue bromophenol blue dye front ran off the bottom of the gel.

The gel sandwich was then carefully disassembled and the gel (mounted on the long "sticky" plate) fixed for 15min. in 10% (v/v) acetic acid, dried at 80°C for an hour, cooled to R.T and exposed overnight to suitable X-ray film (Hyperfilm-βMax, Amersham). The film was processed as instructed.

## 5.5. Plasmid Constructions

### 5.5.1. Construction of pUCSCR- $\alpha$

The yeast expression cassette for tSCI was constructed into a pUC8 plasmid as illustrated in figure 2.5.

For the assembly of the *Pst*I/*Eco*RI linker, the following oligonucleotides were synthesised and purified:

RS1-1: 5'-GATAATAGCTCGAGGATCCG-3'

RS1-2: 5'-AATTCGGATCCTCGAGCTATTATCTGCA-3'

Quantities of 100pmoles (10 $\mu$ l) of RS1-1 and RS2-2 were 5'-phosphorylated with polynucleotide kinase. The linker was assembled by annealing the oligonucleotides as follows: Equimolar quantities (50pmoles each) of RS1-1 and RS2-2 were mixed and heated to 90°C for 5min after which the mixture was allowed to cool slowly to room temperature over a period of an hour. Slow cooling was obtained in a water bath.

The following fragments were prepared from pUSC-1 $\alpha$ :

#### Fragment 1 (3926 b.p.):

The plasmid pUSC-1 $\alpha$  (75 $\mu$ g) was incubated for 15 min. at 37°C with 30U of *Eco*RI. This resulted in a mixture of uncut, partially cut and fully cut plasmid. The partially cut plasmid which corresponded in length (4243 b.p.) to pUSC-1 $\alpha$  linearised by a single cut (e.g. *Kpn*I) was isolated. It consisted of a mixture of two fragments of equal length that resulted from *Eco*RI cleavage at either site on the plasmid.

A quantity of 4 $\mu$ g of this 4243 b.p., partial *Eco*RI fragment was incubated for with 80U *Kpn*I. The resultant digest consisted of 4 fragments from the

two variants of the linearised plasmid. Fragment 1, the largest of these bands (3926 b.p.) was isolated.

#### **Fragment 2 (154 b.p.):**

The plasmid pUSC-1 $\alpha$  (75 $\mu$ g) was incubated with 450U *Kpn*I. This single-cut resulted in the linearisation of the plasmid. The fragment (4243 b.p.) was purified by P.C.I. extraction and ethanol precipitation.

A quantity of 16  $\mu$ g of the *Kpn*I linearised pUSC-1 $\alpha$  was then incubated for with 150U of *Pst*I. This resulted in the formation of 5 fragments of which the smallest one (154 b.p.) corresponding to fragment 2, was isolated.

For the completion of pUCSCR- $\alpha$  a quantity of 50 ng of the purified fragment 1 was mixed with an approximately 10-fold molar excess of fragment 2 (20 ng) and a 100-fold molar excess of the *Pst*I/*Eco*RI adapter (35 ng) and ligated 0.5U of T4 DNA ligase.

### **5.5.2. Construction of pRS-SC56**

For the completion of the yeast expression shuttle vector pRS-SC56 for the secretion of tSCI, the expression cassette for tSCI was excised from pUSCR- $\alpha$  as a 1410 b.p. partial *Bam*HI fragment and used to replace the respective lysozyme expression cassette sequences in pSK801 (Figure 2.6):

A quantity of 10 $\mu$ g of pUSCR- $\alpha$  was incubated for 15 min. at 37°C with 48U of *Bam*HI. Of the various fragments obtained, the 1410 b.p. fragment was isolated. This corresponded to the expression cassette for tSCI minus the 3' transcription termination sequences.

A quantity of 3 $\mu$ g of pSK801 was incubated with 20U *Bam*HI. The larger 8010 b.p. fragment was isolated. In order to minimise the preferred tendency

for this fragment to re-circularise during ligation due to the compatible *Bam*HI overlaps, it was dephosphorylated using calf intestinal alkaline phosphatase (CIP). A quantity of 1.5µg fragment was incubated with 0.1U CIP at 37°C for 30min. A further aliquot of 0.1U CIP was added and the mixture incubated at 37°C for 30min. The CIP was deactivated at 68°C in SDS and the DNA fragment purified by two PCI extractions followed by ethanol precipitation.

To complete the assembly of pRS-SC56, quantities of 50 ng. each of the prepared *Bam*HI fragments of pUSCR-α and pSK801 were ligated with 0.5U T4 DNA Ligase.

### 5.5.3. Construction of pRS101

A quantity of 2.4 µg of pIT353 was digested for 15 min. at 37°C with 15 Units of *Nco*I. The resultant partial digest was a mixture of uncut plasmid, nicked plasmid (7268 b.p.) and fully digested plasmid (5633 b.p. and 1635 b.p.) (Figure 3.1). The nicked plasmid fragment (7268 b.p.) consisting of a mixture of fragments cut at either *Nco*I site was isolated and purified. A quantity of 700 ng of the purified fragment was incubated for 30 min. at 37°C with 1.7 U DNA polymerase I (Klenow fragment) in the presence of 20µM dNTPs. The enzyme was then deactivated by heating to 60°C and the DNA purified by PCI extraction and ethanol precipitation. The obtained blunt-ended fragments were re-circularised by ligation overnight at 4°C with 1U T4 DNA Ligase. To encourage blunt-end ligation, a quantity of 2mM HCC was incorporated in the ligation mix. The total ligation mix was then used to transform competent *E.coli* cells which were spread on LB-agar plates containing tetracycline (15 µg/ml).

For the removal of the IPNS gene from pIT454 (see Figure 3.1), a quantity of 10µg of plasmid was incubated with 30 U *Bam*HI. The resulting

linearised plasmid was purified by ethanol precipitation. The cohesive overhangs resulting from the *Bam*HI cut were filled-in with Klenow enzyme. This was successfully achieved by incubating 0.5µg of *Bam*HI nicked pIT454 for 15min. at R.T. with 0.3 U DNA polymerase I (Klenow fragment) in the presence 20µM dNTPs. After purification of the resulting blunt-ended fragment by ethanol precipitation, the removal of the IPNS gene was completed by digesting with 4UN*co*I. The resulting digest consisted of two fragments. The larger 5814 b.p. fragment (Fragment I) was isolated and purified.

The DNA sequences encoding the SCI gene were excised from the plasmid pUSC-1α as the 330 b.p. *Kpn*I/*Eco*RI fragment (Fragment II). In order to facilitate correct, in-frame insertion of the SCI gene into the above expression vector, it has to be adapted, firstly, by filling-in the *Eco*RI sticky-end so as to make it compatible to the pIT454 blunt-end. Secondly, an *Nco*I/*Kpn*I linker was synthesised that incorporated, the ATG start codon and the codons encoding the first three amino acids of SCI (see Figure 3.1):

A quantity of 6µg of pUSC-1α was incubated with 10U *Eco*RI. This separated the 1550 b.p. yeast expression cassette for SCI from the 2770 b.p. pUC8 sequences. The former was isolated and purified and 2.5µg incubated for 15 min. at R.T. with 1U DNA polymerase I (Klenow fragment) in the presence of 20µM dNTPs. The blunt-ended fragment was purified by PCI extraction and ethanol precipitation and subsequently incubated with 5U *Kpn*I. The smaller, 330 b.p fragment consisting of the *Kpn*I/blunt-ended SCI gene fragment was isolated and purified.

The complementary oligonucleotides RS1-1 and RS1-2 were prepared by automated DNA synthesis. Equimolar quantities of the purified oligonucleotides were annealed to yield the *Nco*I/*Kpn*I linker required to insert the SCI gene fragment into the pIT454 vector fragment.

For assembly of pRS101, fragments I and II and the *NcoI/KpnI* linker were mixed at a molar ratio of 1:5:50, respectively and ligated with T4 DNA ligase.

#### 5.5.4. Construction of pRS-om101

The secretion plasmid pRS-om101 was constructed by replacing the IPNS gene in pIT454 with a the SCI gene fused in-frame to a 5'-adapter encoding the OmpA signal sequence [Movva, *et al.*, 1980]. The synthetic adapter was constructed by the assembly of the oligonucleotides OM1-OM6. The oligonucleotides were obtained as 1ml solutions in saturated ammonia. They were deprotected and prepared as described in section 4.3.10.

For the assembly of the adapter, equimolar portions (1nmol each) of the complementary oligonucleotides of each block were pooled and annealed by heating for 5 min. to 90°C and cooling slowly to room temperature, over a period of 30 min. Quantities of 200pmoles each of the obtained duplexes were mixed and ligated overnight at 4°C with 5U T4 DNA ligase in the presence of 1mM ATP and 4mM spermidine. The completion of the ligation was controlled on a 20% non-denaturing polyacrylamide gel (Figure 3.13). The obtained, complete adapter, was used without separation from the mixture of longer concatamers seen in figure 3.13. The latter were formed as a result of ligation between the 5'-phosphorylated blocks. Only the correctly formed adapter, with its *NcoI/KpnI* cohesive termini would ligate with the compatible termini of the SCI *KpnI/BglII* fragment and the pIT454 *NcoI/BamHI* fragment to form the expression vector pRS-om101.

The remaining sequences for SCI were obtained as the 320 b.p. *KpnI/BglII* fragment from the plasmid pUSC-1α: A quantity of 10μg pUSC-1α was linearised by digesting for 30min at 37°C with 500U *KpnI*. The obtained fragment was purified by repeated PCI extractions and ethanol precipitation.

To obtain the required SCI gene sequences as the 330 b.p. *KpnI/BglII* fragment, 5µg of the *KpnI* linearised pUSC-1α was digested with 200U *BglII*. Of the three fragments obtained, the smallest (320 b.p.) was isolated and purified. In order to minimise concatamerisation, the obtained fragment was dephosphorylated: A quantity of ≈3pmoles (0.5µg) of the isolated 320 b.p. *KpnI/BglII* fragment in 20µl 1x CIP buffer was incubated with 2U CIP for 15 min. at 37°C. An additional 2U CIP was added and the mixture incubated for 45 min. at 55°C. The CIP was then deactivated by adding 0.1% SDS and proteinase K and the dephosphorylated fragment purified by phenol extraction and ethanol precipitation.

For the completion of the plasmid pRS-om101, the 5828 b.p. *NcoI/BamHI* fragment of pIT454 (section 5.5.3.) was mixed with the dephosphorylated 320 b.p. *KpnI/BglII* fragment of pUSC-1α and the 72 b.p. OmpA signal sequence adapter in the molar proportions of 1:20:100, respectively and ligated overnight at 4°C with 0.5U T4 DNA ligase in the presence of 2mM HCC. The obtained ligation mix was used to transform competent JM109 cells and the cells smeared on LB plates containing tetracycline. After 24h incubation at 27°C, a total of about 100 positive transformants were seen. Restriction endonuclease screening of plasmid “minipreps” of 24 well separated colonies showed that 4 contained the desired plasmid. Double-stranded sequencing confirmed correct insertion and orientation of the OmpA signal/SCI fusion sequences.

## **5.6. Protein Analysis and Preparation**

### **5.6.1. Preparation of Proteins from Yeast**

Quantities of 20ml of yeast cells grown to stationary phase in SDAA were harvested by centrifugation in 50ml polypropylene tubes at 3500g for 5 min. Secreted proteins which are contained in the supernatant were precipitated by addition of 4 volumes methanol, 1 volume chloroform and 3 volumes water. After mixing well, the mixture was centrifuged at 6500g for 5 min. Taking care not to disturb the protein deposited at the interface between the two phases formed, the upper phase was discarded. The protein was pelleted by adding 3 volumes of methanol and centrifuging at 6500g for 5 min. The obtained pellet was dried and resuspended in 1x SDS-PAGE sample loading buffer (10 $\mu$ l) (see below). The cells were resuspended in 400 $\mu$ l 1.0M sorbitol containing 25mM EDTA (pH 8.0) and 0.86M  $\beta$ -mercaptoethanol and incubated at 30°C for 15 min. The cells were centrifuged as above, resuspended in 1.0M sorbitol containing 0.5mg/ml Zymolyase 20T and incubated a further 15 min. at 30°C. The spheroplasts were centrifuged and the supernatant (containing the periplasmic proteins) transferred to a fresh microfuge tube and 5x SDS-PAGE loading buffer (40 $\mu$ l) added. The remaining spheroplasts were resuspended in 1x SDS-PAGE loading buffer (400 $\mu$ l) and boiled for 20 min. Cell debris was removed by centrifugation. The suspension contained the cytoplasmic proteins of the yeast cells. Quantities of 10 $\mu$ l of each sample were applied for SDS-PAGE.

### **5.6.2. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins**

Protein samples were analysed by polyacrylamide gel electrophoresis according to Laemmli, (1970) using a 15% polyacrylamide separating gel (10 x 6 x 0.3cm) containing 3% cross-linker (N,N'-methylene bisacrylamide) and a 5%

stacking gel. Electrophoresis was carried out in a Mini-Protean apparatus (Bio-Rad Laboratories Ltd., Hemel Hempstead, Herts.). The apparatus was assembled as instructed.

The compositions of the gels and reagents were as follows:

|                             |                                                                                                                                                                       |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Resolving Gel:              | 14.55% (w/v) acrylamide<br>0.45% (w/v) N,N'-methylene bisacrylamide<br>0.25M Tris-HCl pH8.8<br>0.1% (w/v) SDS<br>0.033% (v/v) TEMED<br>0.033% (w/v) APS               |
| Separating Gel:             | 4.85% (w/v) acrylamide<br>0.15% (w/v) N,N'-methylene bisacrylamide<br>0.125M Tris-HCl pH 6.8<br>0.1% (w/v) SDS<br>0.05% (v/v) TEMED<br>0.05% (w/v) APS                |
| Running Buffer:             | 0.025M Tris base<br>0.192M glycine<br>0.1% (w/v) SDS                                                                                                                  |
| Sample Loading Buffer (3x): | 0.675M Tris-HCl pH 6.8<br>10%(v/v) glycerol<br>2% (w/v) SDS<br>5% (v/v) $\beta$ -mercaptoethanol (left out in non-reducing buffers)<br>0.002% (w/v) bromophenol blue. |

Protein samples were prepared by suspending in 1x sample loading buffer. The samples were boiled for 2min. (for total cell protein, the cell pellet was boiled for 15min, and the sample vortexed and centrifuged to get rid of cell debris). The prepared samples were loaded in the sample wells of the gel. The gel was run in running buffer, firstly at 13mA until the bromophenol blue front moved into the separating gel after which the current was increased to 16mA until the bromophenol blue reached the bottom of the gel.

Gradient gels (9-18%) of dimensions 16 x 18 x 0.5cm were prepared and run in a Hoefer SE600 vertical slab gel unit (Hoefer Scientific Instruments,

Newcastle, Staffs.). For counting  $^{14}\text{C}$ -labelled proteins separated by SDS-PAGE (Section 3.2.2), the gel was fixed and stained with Coomassie Blue. Each lane was then cut into 1-1.5mm thin slices. Prior to counting, the gel slices were solubilised with the organic tissue solubiliser (TCS; Amersham). The organic counting scintillation cocktail OCS (Amersham) was used for counting.

Regular protein staining was done with Coomassie Blue R (0.1% (w/v) in methanol:glacial acetic acid:water, 5:1:5 by volume). Destaining to remove background stain on the gel was done in water containing methanol (5%) and acetic acid (7.5%).

### **5.6.3. Western Blotting - Electrophoretic Transfer of Proteins From SDS-Polyacrylamide Gels To Nitrocellulose Membrane**

This was carried out as described by Burke *et al.*, (1982) with various modifications. Polyclonal antibody (about 50% IgG) from goat against SCI was obtained, with thanks, from Dr. M.P. Esnouf (Nuffield Department of Clinical Biochemistry, University of Oxford). Horse-radish-peroxidase-(HRP)-conjugated secondary anti-goat antibody from rabbit (P160; immunoglobulin fraction of antisera) was obtained from Dakopatts Ltd. (High Wycombe, Bucks.).

Following separation of proteins by SDS-PAGE, the obtained gel was soaked in transfer buffer (25mM Tris-HCl pH 8.3-8.5, 192 mM glycine & 20% (v/v) methanol) for about 30 min. This also results in the removal of SDS from the gel.

Electrophoretic transfer was carried out in a mini Trans-Blot® Electrophoretic transfer cell obtained from Bio-Rad Laboratories Ltd. (Hemel Hempstead, Herts.). The assembly of the pre-soaked gel and the nitrocellulose membrane (0.2 micron, Bio-Rad) into the gel holder cassette was carried out

as instructed in the accompanying manual. Electrophoretic transfer was carried out at 4°C for 1h at a constant current of 200mA or overnight at 50mA.

Following transfer, the gel holder cassette was disassembled and the nitrocellulose (NC) membrane washed three times in PBS (10mM sodium phosphate pH 7.4 & 0.15M sodium chloride). To optimise results by ensuring full renaturation of the transferred proteins which bind irreversibly to the membrane, the membrane was incubated overnight at 37°C in PBS. Following incubation, the transfer membrane was blocked by washing the membrane for 30 min. in 1% (w/v) BSA in PBS. This prevents any non-specific binding of both the primary and secondary antibodies.

Prior to antibody binding, the membrane was then washed three times for 5 min. a time with 0.1% (w/v) BSA in PBS. Antibody binding is carried out by incubating the NC membrane in antibody binding buffer (1%(w/v) BSA, 0.05% (v/v) Tween-80, 2M glucose & 20% (v/v) glycerol in PBS) containing 5mg/ml goat anti-SCI antibody. Due to curtailed stocks of the antibody this incubation was carried out in minimal volume (5ml). For optimal antibody binding, the NC membrane was sealed with the antibody solution in a plastic bag. Incubation with shaking was carried out for 1 h at R.T. Following antibody binding, the NC membrane was washed three times for 5 min. each in PBS containing 0.1% (w/v) BSA and then incubated for 2h. at R.T. with shaking, with the secondary peroxidase-conjugated antibody, diluted (1:500) in 20ml PBS containing 0.1% (w/v) BSA and 0.05% (v/v) Tween-80. After incubation, the NC membrane was rinsed once with PBS containing 0.1% Tween-80, then washed twice for 5min. each, with PBS containing 0.05% (v/v) Tween-80 and finally, washed for 5 min. in PBS alone. This washing procedure ensures the complete removal of unbound antibodies.

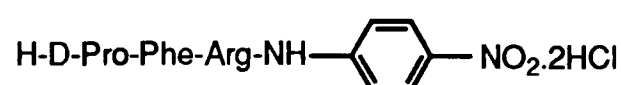
The bound peroxidase was stained in a solution of DAB (10µg/ml) in 0.1M Tris-HCl pH 7.4 containing 0.01% hydrogen peroxide. The stained NC membrane was rinsed six times in distilled water and dried at 55°C.

#### 5.6.4. Determination of Protein Concentrations; Bradford Method

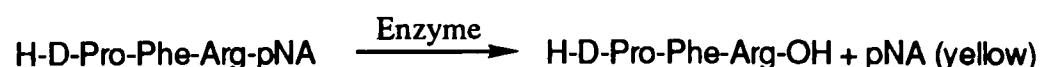
Protein concentrations were determined by the method described by Bradford, (1976). The Bradford reagent was prepared as follows: To a quantity of 100mg Coomassie Blue G-250 dissolved in 50ml 90% (v/v) ethanol, 100ml of 85% (w/v) phosphoric acid was added and water added to a total of 1L. The obtained solution was filtered and stored in the dark, ready to use. After calibration of the reagent with BSA (10-100µg), concentrations of proteins were determined by adding 5ml of the Bradford reagent to 0.1ml of the protein solution and measuring absorbance at 595nm.

#### 5.6.5. Colourimetric Determination of Proteolytic Activity of Trypsin and βXIIa

The proteolytic activities of trypsin and βXIIa and their inhibition were determined, using the chromogenic substrate S2302 (KabiVitrum):



The assay is based on the following reaction which is catalysed by the proteinases:



The rate of formation of the yellow p-nitroanilide measured spectrophotometrically at 405nm is proportional to the proteolytic activity [S2302 Data Sheet; KabiVitrum, Stockholm Sweden].

The S2302, available as a powder was dissolved sterile water to give a stock solution of 4mM. Assays were carried out at 37°C in 0.1M Tris-HCl

containing 0.1M NaCl and 0.1mg/ml BSA. To assay the activity of trypsin or  $\beta$ XIIa, the enzyme, dissolved in water or the above assay buffer is added to a spectrometric cuvette and the assay buffer added to a total volume of 950 $\mu$ l. A quantity of 50 $\mu$ l of the S2302 stock was then added and mixed well, and the rate of change of OD<sub>405</sub> measured against a blank reference containing only 50 $\mu$ l S2302 in a total of 1ml assay buffer. In order to increase the sensitivity of the assay, a series dilution of either proteinase was made and the optimal concentration of proteinase for determining inhibition ascertained. To measure inhibition, the proteinases, trypsin or  $\beta$ XIIa were incubated for 5min. at 37°C with varying volumes and dilutions of the sample to be tested in a total of 950 $\mu$ l. The incubations were carried out in plastic cuvettes after which, 50 $\mu$ l S2302 stock solution was added and mixed and the change of OD<sub>405</sub> measured

## 5.7. Microbiological Techniques

### 5.7.1. Strains

The genotypes of the microbial strains used during the course of this work are listed below:

| Strain                                 | Genotype                                                                                                                                                                                       | Source/Reference                                                  |
|----------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|
| <hr/>                                  |                                                                                                                                                                                                |                                                                   |
| <b><i>Saccharomyces cerevisiae</i></b> |                                                                                                                                                                                                |                                                                   |
| JRY188                                 | <i>Mat<math>\alpha</math>, sir3 -8, leu2 -3, leu2 -112, trp1, ura3 -52, his4</i>                                                                                                               | J. Rine, Berkeley, CA<br>[Rine <i>et al.</i> , 1979]              |
| MD40-4C                                | <i>ura2, trp1, leu2 -3, leu2 -112, his3 -11, his3 -15</i>                                                                                                                                      | S. Kingsman, Oxford<br>[Mellor <i>et al.</i> , 1985]              |
| KHY-2                                  | <i>Sc252[pGAL-10/GAL-4] cir+ a leu2-2 leu2-112 his3 ade1 MEL 1</i>                                                                                                                             | L.D. Schultz,<br>West Point, PA<br>[Schultz <i>et al.</i> , 1987] |
| <b><i>Escherichia coli</i></b>         |                                                                                                                                                                                                |                                                                   |
| JM109                                  | <i>F' traD36 lacI<sup>q</sup><math>\Delta</math>(lacZ)M15 proAB/recA1 endA1 gyrA96 (Nal<sup>r</sup>) thi hsdR17 (r<sub>K</sub>-m<sub>K</sub>+) supE44 relA1 <math>\Delta</math>(lac-proAB)</i> | [Yanish-Perron <i>et al.</i> 1985]                                |
| JM105                                  | <i>F' traD36 lacI<sup>q</sup><math>\Delta</math>(lacZ)M15 proAB/thi rpsL (Str<sup>r</sup>) endA1 sbcB15 (sbcC) hsdR4 (r<sub>K</sub>-m<sub>K</sub>+) <math>\Delta</math>(lac-proAB)</i>         | [Yanish-Perron <i>et al.</i> 1985]                                |
| MC1061                                 | <i>F- araD139 <math>\Delta</math>(ara-leu)7696 <math>\Delta</math>(lac) X74 galU galK hsdR2 (r<sub>K</sub>-m<sub>K</sub>+) mcrB1 rpsL (Str<sup>r</sup>)</i>                                    | [Wertman & <i>et al.</i> , 1986]                                  |
| CJ236                                  | <i>F- dut ung1 thi-1 relA1/pCJ105 (Cm<sup>r</sup>)</i>                                                                                                                                         | [Raleigh <i>et al.</i> , 1989]                                    |
| NM522                                  | <i>F' lacI<sup>q</sup><math>\Delta</math>(lacZ)M15proAB/supE thi <math>\Delta</math>(lac-proAB) <math>\Delta</math>(hsdMS-mcrB)5 (r<sub>K</sub>-m<sub>K</sub>-McrB-)</i>                       | [Gough & Murray, 1983]                                            |
| TG1                                    | <i>K12 F' traD36 lacI<sup>q</sup> <math>\Delta</math>(lacZ)M15 proAB/supE thi hsdD5 <math>\Delta</math>(lac-proAB)</i>                                                                         | [Carter <i>et al.</i> , 1985]                                     |
| JE5505                                 | <i>F- lpo5508 pps his-4 proA2 ArgE3 thi-1 galK2 lacY1 xyl-5 mtl-1 tsx-29</i>                                                                                                                   | Y. Nishimura,<br>Mishima, Japan<br>[Suzuki <i>et al.</i> , 1978]  |
| <hr/>                                  |                                                                                                                                                                                                |                                                                   |

5.7.2.      **Growth of Strains**

*S. cerevisiae* and *E. coli* were grown in liquid media contained in conical flasks of at least 10 times greater volume than that of the culture medium. After inoculation, cultures were grown aerobically by shaking in orbital incubators (250-300rpm). The temperature of incubation depended on the strain and expression of yeast or *E.coli* used. Media for agar plates were prepared in 100ml Duran bottles. Unless otherwise stated, all components were autoclaved together for 20min at 120°C 15lb in<sup>-2</sup> pressure. The agar media were allowed to cool to 60°C and plates poured, allowed to set, and left to dry.

5.7.3.      **Media for Growth of *Escherichia coli***

5.7.3.1.    **Rich Media**

LB (Luria-Bertani)-Medium (per litre):

|                              |     |
|------------------------------|-----|
| Bacto-tryptone               | 10g |
| Bacto-yeast extract          | 5g  |
| Sodium chloride              | 5g  |
| Bacto-agar (for plates only) | 15g |

Superbroth-Medium (per litre):

|                       |     |
|-----------------------|-----|
| Bacto-tryptone        | 32g |
| Bacto-yeast extract   | 20g |
| Sodium chloride       | 5g  |
| Sodium hydroxide (1N) | 5ml |

SOB-Medium (per litre):

|                     |      |
|---------------------|------|
| Bacto-tryptone      | 20g  |
| Bacto-yeast extract | 5g   |
| Sodium chloride     | 0.5g |

Before autoclaving, adjust pH to 7.5 with potassium hydroxide.

5.7.3.2.    **Minimal Media**

M9 Medium (per litre):

|                                  |      |
|----------------------------------|------|
| Na <sub>2</sub> HPO <sub>4</sub> | 6g   |
| KH <sub>2</sub> PO <sub>4</sub>  | 3g   |
| NaCl                             | 0.5g |
| NH <sub>4</sub> Cl               | 1g   |

Adjust pH to 7.4, autoclave, cool to 60°C, and then add:

|                      |       |
|----------------------|-------|
| 1M MgSO <sub>4</sub> | 2ml   |
| 20% (w/v) glucose    | 10ml  |
| 1M CaCl <sub>2</sub> | 0.1ml |

These solutions were made separately and sterilised separately by sterile filtration (glucose) or autoclaving.

**M9CA Medium:**  
Identical to M9 medium with the inclusion of 2.0g/L of casamino acids.

**5.7.3.3. Antibiotics**

The following antibiotic concentrations were used. Stock solutions were prepared, filter sterilised and stored at -20°C:

| Antibiotic      | Stock Solution                                                      | Working Concentration |
|-----------------|---------------------------------------------------------------------|-----------------------|
| Ampicillin      | 25mg/ml of the sodium salt in water                                 | 50µg/ml               |
| Tetracycline    | 12.5mg/ml of tetracycline hydrochloride in ethanol/ water (50% v/v) | 12.5µg/ml             |
| Kanamycin       | 25mg/ml in water                                                    | 50µg/ml               |
| Chloramphenicol | 34mg/ml in 100% ethanol                                             | 30µg/ml               |

-----

**5.7.4. Media for Growth of *Saccharomyces cerevisiae***

|                                |                              |     |
|--------------------------------|------------------------------|-----|
| YPD (rich medium) (per 900ml): | Bacto-yeast extract          | 10g |
|                                | Bacto-peptone                | 20g |
|                                | Bacto-agar (for plates only) | 20g |

After autoclaving, 100ml 20% (w/v) glucose (filter sterilised) was added.

|                                   |                                         |       |
|-----------------------------------|-----------------------------------------|-------|
| SDAA (minimal medium) (per 87ml): | Yeast nitrogen base without amino acids | 0.67g |
|-----------------------------------|-----------------------------------------|-------|

|                                             |                                 |
|---------------------------------------------|---------------------------------|
| The following were added after autoclaving: |                                 |
|                                             | 20% (w/v) glucose 10ml          |
|                                             | 50x amino acids supplement* 2ml |
|                                             | 0.2% Tryptophan 1ml             |

**\*50x amino acids supplement contains the following:**

|                 |       |                 |       |
|-----------------|-------|-----------------|-------|
| Adenine         | 0.1%  | L-Histidine     | 0.1%  |
| Uracil          | 0.1%  | L-Arginine      | 0.1%  |
| L-Methionine    | 0.1%  | L-Tyrosine      | 0.15% |
| L-Lysine        | 0.15% | L-Phenylalanine | 0.25% |
| L-Glutamic Acid | 0.5%  | L-Aspartic Acid | 0.5%  |
| L-Valine        | 0.75% | L-Isoleucine    | 1.5%  |
| L-Threonine     | 1.0%  | L-Serine        | 1.75% |

**SS3% (top agar for transformation) (per 85ml):**

|                                            |       |
|--------------------------------------------|-------|
| Yeast nitrogen base<br>without amino acids | 0.67g |
| D-sorbitol                                 | 18.2g |
| Bacto-agar                                 | 3.0g  |

The following were added after autoclaving:

|                             |      |
|-----------------------------|------|
| 20% (w/v) glucose           | 10ml |
| 50x amino acids supplement* | 2ml  |
| 0.2% tryptophan (100x)      | 2ml  |

**SS2% (bottom agar for transformation) (per 87ml):**

|                                            |       |
|--------------------------------------------|-------|
| Yeast nitrogen base<br>without amino acids | 0.67g |
| D-sorbitol                                 | 18.2g |
| Bacto-agar                                 | 2.0g  |

The following were added after autoclaving:

|                             |      |
|-----------------------------|------|
| 20% (w/v) glucose           | 10ml |
| 50x amino acids supplement* | 2ml  |
| 0.2% Tryptophan (100x)      | 1ml  |

## **5.7.5. Transformation of *Escherichia coli*-Calcium Chloride Method**

### **5.7.5.1. Preparation of Competent *Escherichia coli* Cells**

For the purpose of preparation and storage of competent *E.coli* cells, the following modified calcium chloride treatment of cells in early log-phase of growth gave very high transformation efficiency to all the strains of *E.coli* used regularly during the course of this thesis. These strains, listed in section 5.7.1, included MC1061, JM109 and NM522.

An overnight culture of *E.coli* was obtained by inoculating 5ml of LB-medium with a well isolated plate colony of the respective *E.coli* strain and overnight growth at 37°C and rotational shaking at 250 rpm. The overnight culture was then diluted 100-fold by transferring 2ml of the overnight culture to 200ml fresh LB-medium in a 1 litre conical flask. The diluted cells, initially having an OD<sub>650</sub> of around 0.1 (this value varies depending on the strain of *E.coli* used) were grown to an OD<sub>650</sub> of 0.3. The cells were then harvested by centrifuging the culture at 4000xg for 10 min. at 4°C. The cell pellet was carefully resuspended in half the initial culture volume (100ml) of ice-cold 50mM calcium chloride in 10mM piperazine-N,N'-bis-[2-ethanesulphonic acid] (PIPES) buffer (pH 6.6) containing 15% glycerol. Following a 20 min. incubation on ice, the cells were spun down at 4000g for 10 min. 4°C and carefully resuspended in a tenth of the initial culture volume (20 ml) of the above solution. The resulting cell suspension on ice can be used for transformation after about 40 minutes incubation. Transformation efficiency has been seen to improve following overnight storage at 4°C. Aliquots of 200 µl in sterile 1.5ml. microfuge tubes were rapidly frozen in liquid nitrogen and stored at -80°C.

#### 5.7.5.2. Transformation

The 200  $\mu$ l frozen aliquots of competent *E.coli* were thawed by placing on ice for 20 min. For transformation, a suitable quantity (50-100ng) of plasmid DNA is added and the mixture incubated at 4°C for 30 min. To facilitate DNA uptake in to the cells, they were then heat-shocked for 5 min. at 37°C (when transforming cells with expression plasmids under the control of the heat inducible  $\lambda P_L$  promoter, to avoid potentially lethal expression of recombinant protein, the cells were heat-shocked and grown at 27°C). Following heat-shock, the transformed cells were cooled on ice followed by the addition of 1ml fresh LB-medium. Prior to plating, the cells were grown for 1 hour at 37°C. This enables the cells to reconstitute and express the antibiotic resistance incorporated on the plasmid DNA. The cells are then pelleted by spinning at 6500 r.p.m., resuspended in 100ml LB-medium and plated on to LB-agar plates containing the relevant antibiotic. Incubation of the plates at 37°C results in transformant cell colonies appearing after 12-18 hours.

#### 5.7.6. Transformation of *Escherichia coli*- Electroporation

Certain strains of *E.coli* (e.g. the murein-lipoprotein deficient strain JE5505) proved resistant to the above calcium chloride transformation method. Electroporation, whereby DNA is introduced into *E.coli* by subjecting the cells in the presence of DNA to a very short, intense pulse of current, proved an effective method of transformation.

A quantity of 2 ml of an overnight culture of a single colony of the required strain of *E.coli* in SOB medium was diluted into 200ml of fresh 2xTY medium and incubated at 37°C until the growth culture attained an OD<sub>550</sub> of about 0.5. The medium was then chilled on ice and centrifuged at 4°C at 4500

rpm for 15 min. The obtained cell pellet was then subject to repeated washes with ice-cold 10% (v/v) glycerol, by resuspending in volumes of 200ml, 100ml and 8ml, respectively, followed by centrifugation at 4°C at 4500 rpm. Finally, the washed pellet was resuspended in 200µl of the same solution. Together with 1µl of the required plasmid DNA (10-50ng), aliquots of 20µl of the prepared cell were placed in sterile electroporation cuvettes (Bio-Rad) and subject to electroporation in a Bio-Rad Gene Pulser. The apparatus was preset at the following values: 2.5kV, 200Ω and 25µF (t/c≈4.7). A quantity of 1ml of SOB medium was immediately added to the electroporated mixture and the mixture incubated for 1h. at the required growth temperature (27°C or 37°C). The cells were then plated on SOB agar containing the appropriate antibiotic for the screening of correct transformants and grown 12-24h at a suitable temperature. Consistently efficient transformation frequencies were obtained by this method.

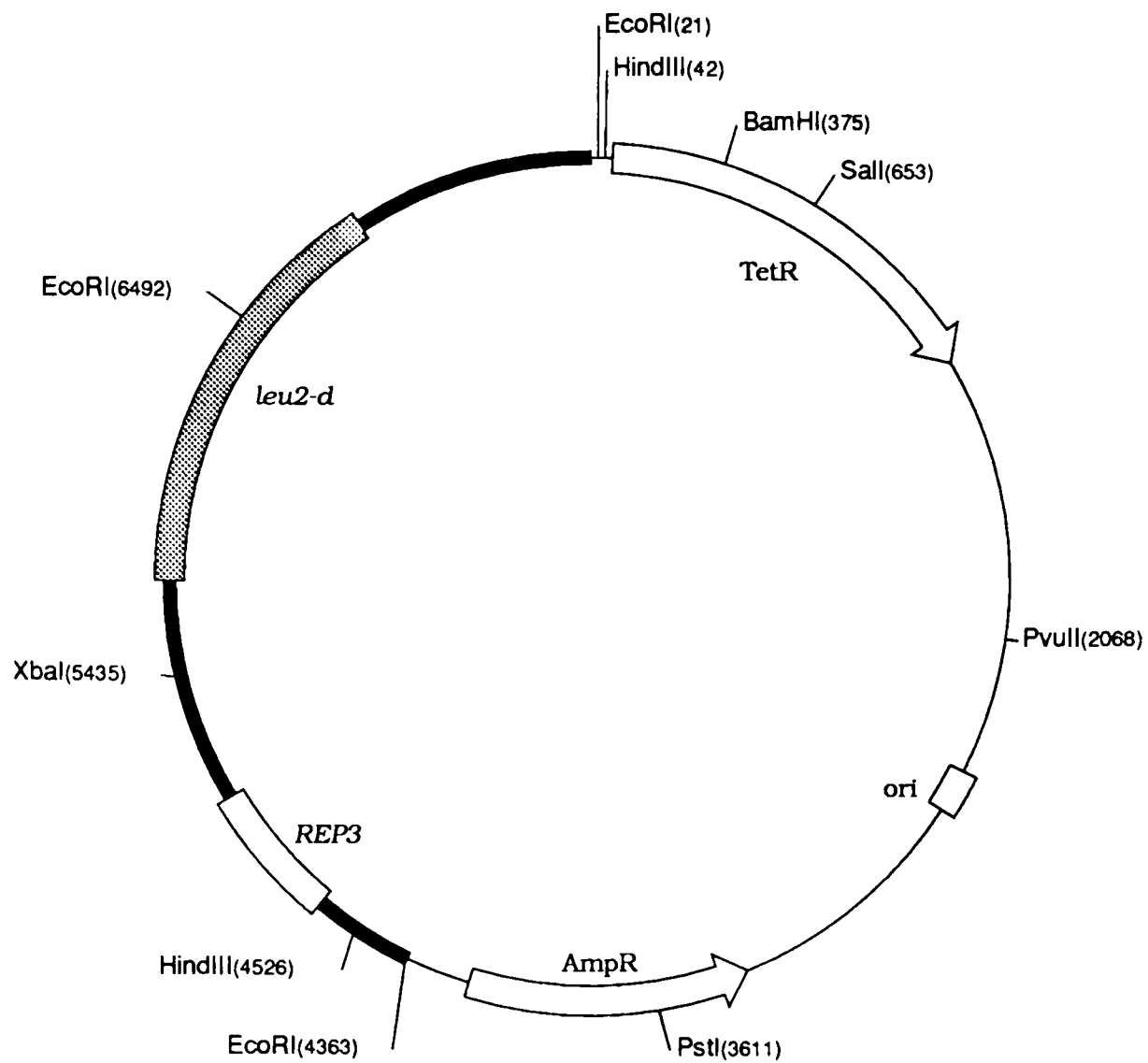
#### **5.7.7. Transformation of Yeast Spheroplasts**

Transformation of yeast cells was performed as described by [Hinnen *et al.*, 1978], with some modifications:

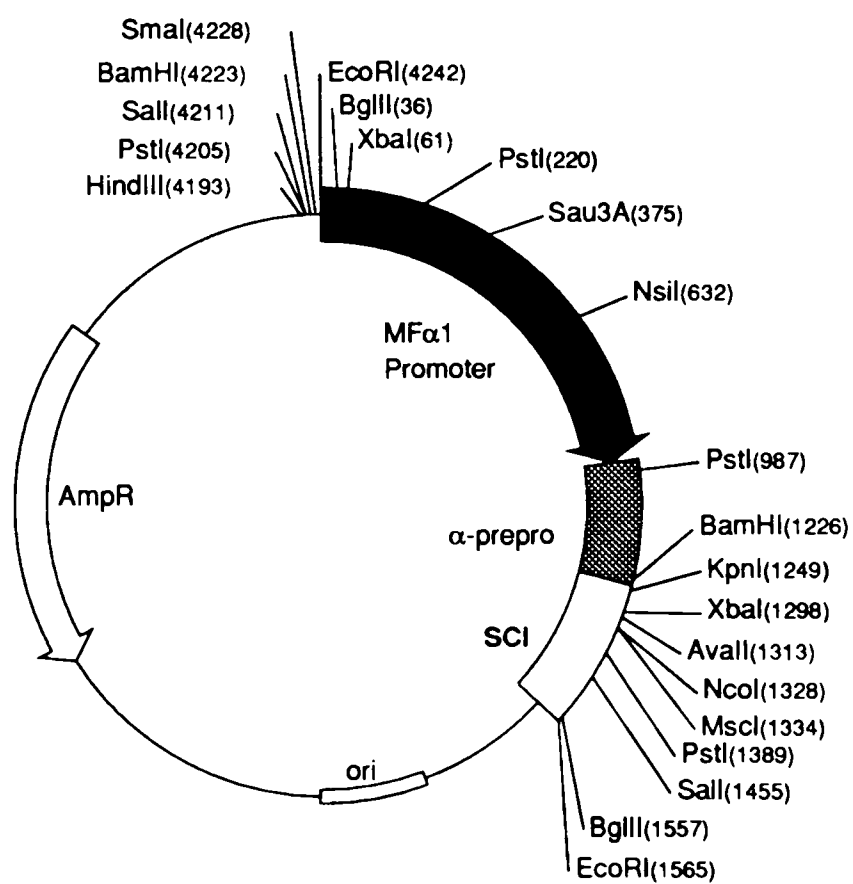
Yeast cultures were grown at 30°C to mid-log phase ( $1-2 \times 10^7$  cells/ml or OD<sub>600</sub> ≈1-2) in YPD and harvested by centrifugation at 4000g for 3 min. The cell pellet was washed once in 20 ml 1M sorbitol (for osmotic support) and then resuspended in 1M sorbitol solution to give a cell density of ≈ $2 \times 10^8$  cells/ml. For the removal of the cell wall, 1/200th the suspension volume of Zymolyase-20T (20mg/ml in 1M Sorbitol) was added and the suspension shaken gently at RT for 30 min. Spheroplast formation was checked after incubation and was judged to be complete if the OD<sub>600</sub> upon dilution of 100-fold into water was 30% of that upon dilution of 100-fold into 1M Sorbitol. The spheroplasts were gently washed three times in 10ml 1M sorbitol to remove the Zymolyase-20T and the cell pellet resuspended in TCS buffer

(10mM Tris-HCl pH 7.5 containing 10mM calcium chloride and 1M sorbitol) to a cell density of  $2 \times 10^8$  cells/ml. Aliquots of 160 $\mu$ l were centrifuged in a microfuge at 6500g. and the cell pellet resuspended in TCS buffer containing 10-500 $\mu$ g/ml of the transforming DNA. Following the incubation of the mixture for 10 min,. at RT, DNA uptake was induced by adding 100 $\mu$ l TCP buffer (20mM Tris-HCl pH 7.5 containing 10mM calcium chloride and 44% polyethylene glycol 4000 (BDH)) and the suspension incubated a further 10min at RT. Finally, aliquots (0.2ml) were combined with 10ml selective regeneration agar (SS3%) and poured onto transformation plates (SS2%). After setting, the plates were incubated at 30°C. After 2 days, positive transformants (leucine prototrophs) were seen. Single, well separated colonies were selected for expression studies.

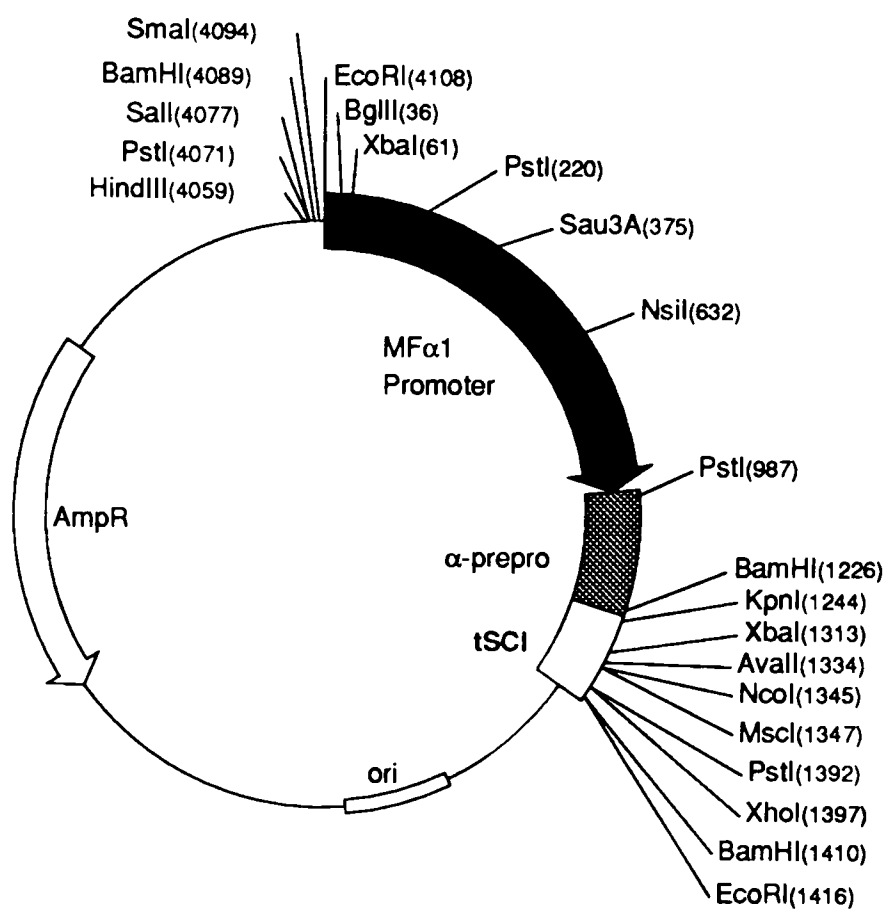
# **Appendices**



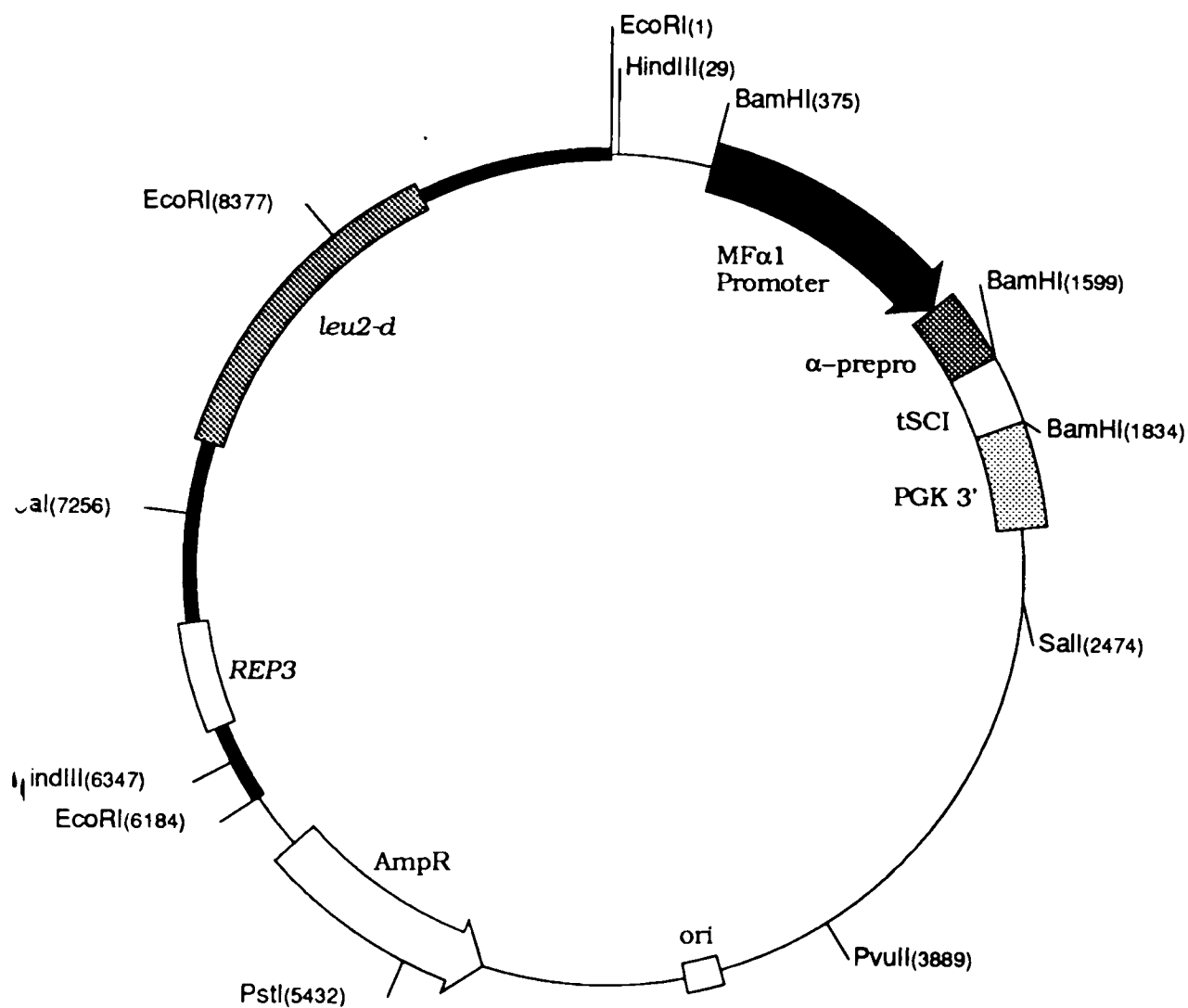
**Appendix I: pMA3a, a Yeast/*E. coli* Shuttle Vector (7620 b.p.)**  
 (This plasmid consists of sequences derived from the yeast 2-micron plasmid (thick line) and the *E. coli* pBR322 plasmid (thin line). Restriction site positions (b.p.) are indicated in parentheses) [from Dobson *et al.*, 1982].



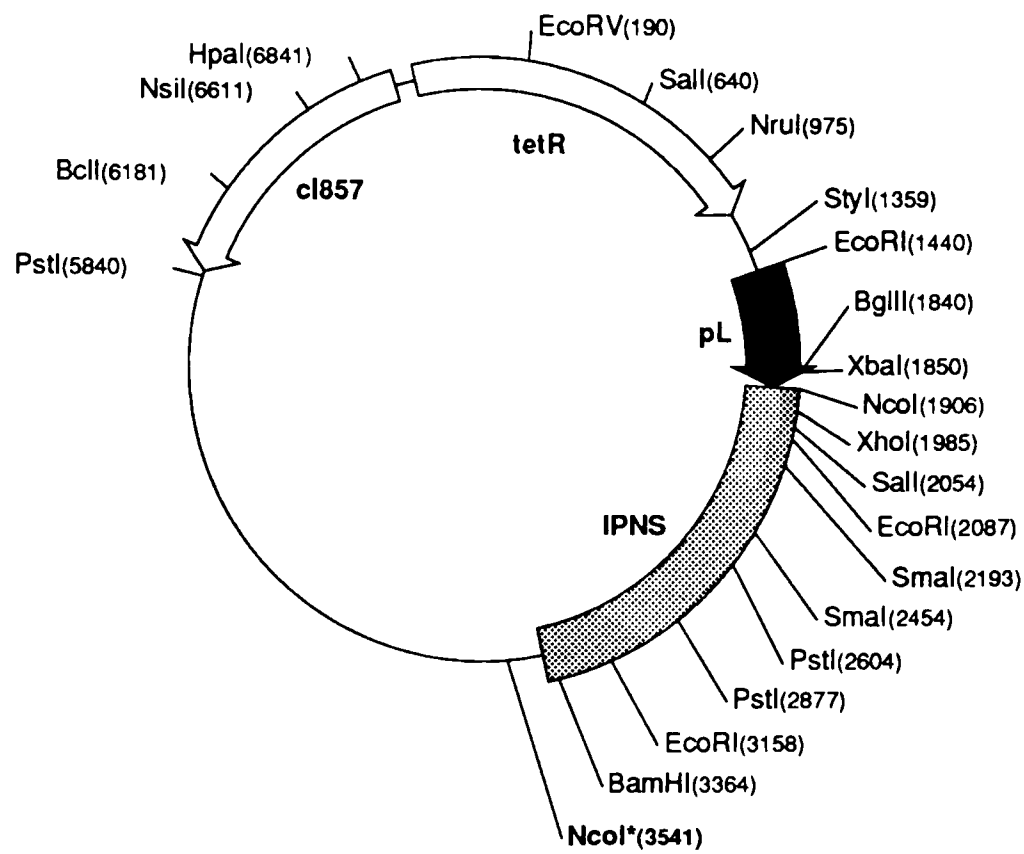
**Appendix II: pUSC-1α (4243 b.p.)** [Kodo, 1988]  
 Restriction endonuclease site-positions in parentheses.  
 Thin line represents sequences derived from pUC8



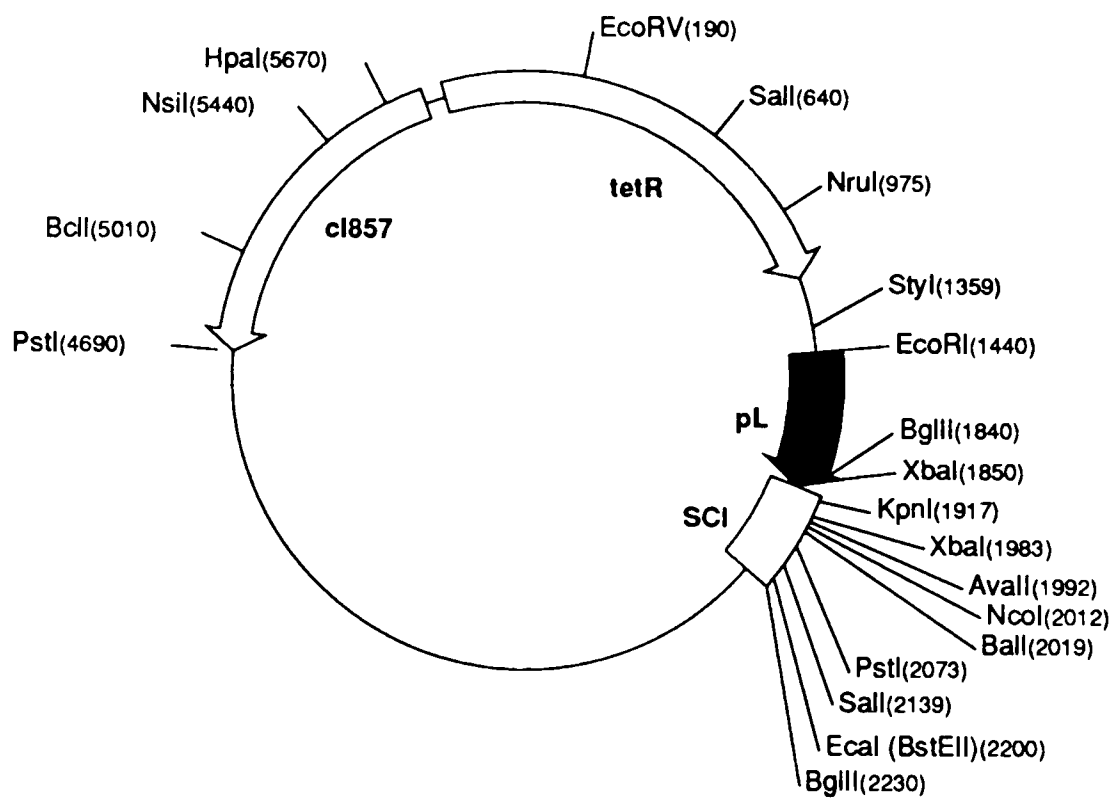
**Appendix III: pUCSCR-α (4100 b.p.)**  
 Restriction endonuclease site positions in parentheses  
 Thin line represents sequences derived from pUC8



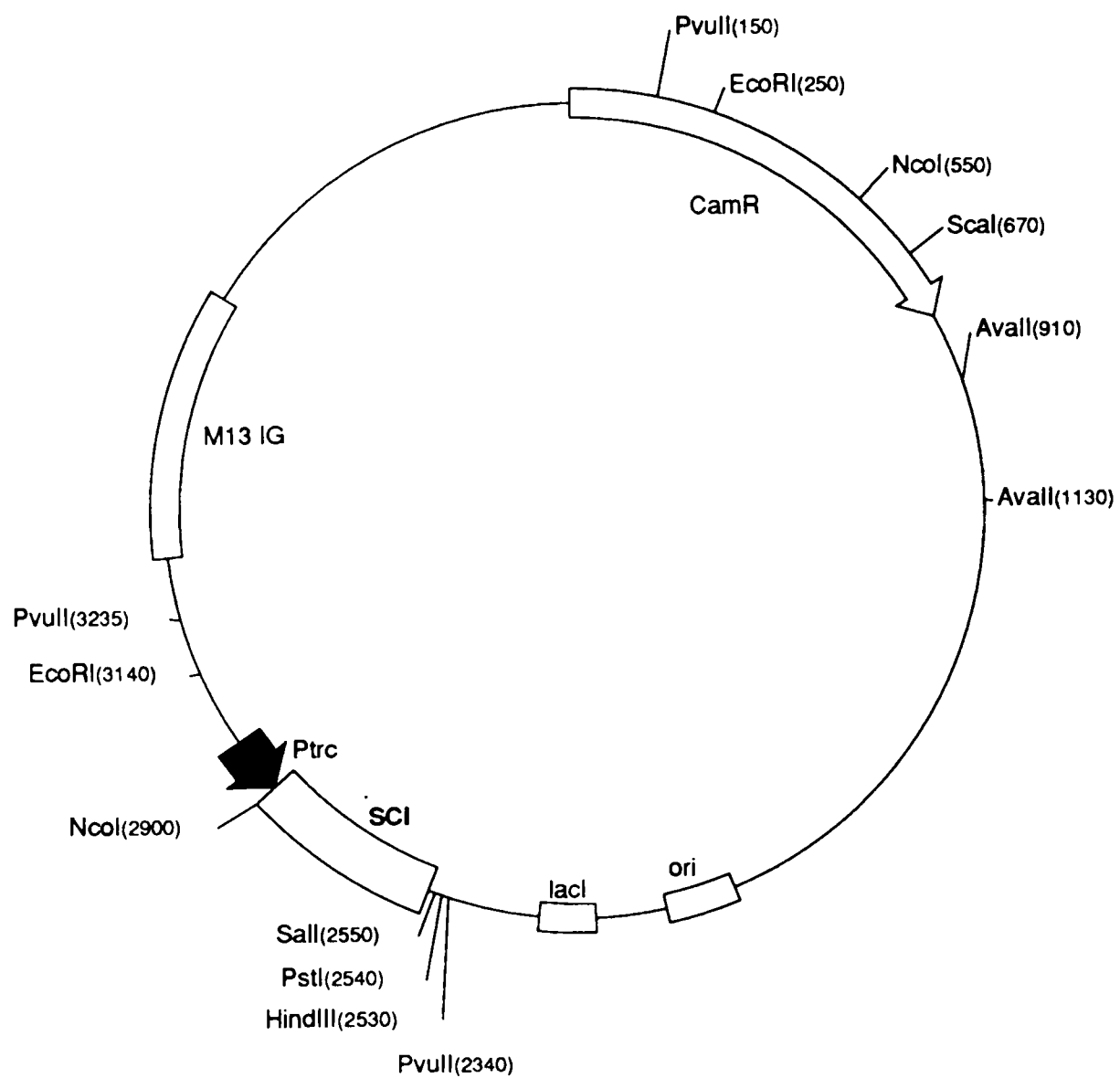
**Appendix IV: pRS-SC56 (9435 b.p.)**  
 A yeast expression vector for truncated SCI (tSCI).  
 (Restriction site positions (b.p.) are indicated in parentheses).



**Appendix V: pIT353/pIT454 (7272 b.p.)**  
 Restriction endonuclease site positions are shown in parentheses  
 \*In the plasmid pIT454, the NcoI site (3541 b.p.) has been deleted.



**Appendix VI: pRS101 (6150 b.p.)**  
 An *E.coli* expression vector for SCI.  
 (Restriction site positions (b.p.) are in parentheses)



**Appendix VII: pRS201 (4570 b.p.)**  
 An *E. coli* expression vector for SCI.  
 Restriction sites (b.p.) are shown in parentheses.

| Amino Acid | <i>E.coli</i> | <i>S.cerevisiae</i> |
|------------|---------------|---------------------|
| Leu        | CUG           | UUG                 |
| Arg        | CGU/CGC       | AGA                 |
| Pro        | CCG           | CCA                 |
| Gln        | CAG           | CAA                 |
| Lys        | AAA           | AAG                 |
| Ala        | GCN           | GCU/GCC             |
| Val        | GUN           | GUU/GUC             |
| Gly        | GGU/GGC       | GGU                 |
| Ser        | UCU/UCC       | UCU/UCC             |
| Thr        | ACU/ACC       | ACU/ACC             |
| Ile        | AUU/AUC       | AUU/AUC             |
| Asn        | AAC           | AAC                 |
| Phe        | UUC           | UUC                 |
| Tyr        | UAC           | UAC                 |
| Glu        | GAA           | GAA                 |
| Cys        | UGU/UGC       | UGU                 |
| His        | CAC           | CAC                 |
| Asp        | GAU/GAC       | GAU/GAC             |
| Met        | AUG           | AUG                 |
| Trp        | UGG           | UGG                 |

**Appendix VIII:** Codon Bias in *Escherichia coli* and *Saccharomyces cerevisiae* [from Ikemura & Ozeki, 1982]  
 (Determined by comparing levels of isoaccepting tRNAs)

MRC IMMUNOCHEMISTRY UNIT      SEQUENCING DATA SHEET  
Applied Biosystems 470A/120A/473

|                            |         |              |            |                |
|----------------------------|---------|--------------|------------|----------------|
| Initial yield              | Run no. | Std no.      | Date       | Stepwise yield |
|                            | 91124H  | h124         | 4.9.91     |                |
|                            |         | Multimethod. |            |                |
|                            |         | rsci152      |            |                |
| Name                       |         |              | Operator   |                |
| Rohan da Silva             |         |              | Tony Wills |                |
| Sample name                |         | Amount       |            |                |
| Recomb. SCI Fr. 15 hplc #2 |         | 100µl        |            |                |

| No. | Yield 1 | AA residue(s) |        |   |        | No. | Amnt. picomol | AA residue(s) |   |   |   |   |
|-----|---------|---------------|--------|---|--------|-----|---------------|---------------|---|---|---|---|
|     |         | 1             | yield2 | 2 | yield3 | 3   |               | 1             | 2 | 3 | 4 | 5 |
| 1   | 36.1    | S             | 36.1   | G | 24.3   | M   | 26            |               |   |   |   |   |
| 2   | 30.0    | A             | 21.7   | T | 7.1    | S   | 27            |               |   |   |   |   |
| 3   | 20.9    | G             | 11.1   | S | 10.4   | F   | 28            |               |   |   |   |   |
| 4   | 21.1    | T             |        | - |        | RV  | 29            |               |   |   |   |   |
| 5   | 11.5    | S             | 23.5   | V |        |     | 30            |               |   |   |   |   |
| 6   | -       |               | 15.6   | P |        |     | 31            |               |   |   |   |   |
| 7   | 23.4    | V             | 19.8   | G |        | AL  | 32            |               |   |   |   |   |
| 8   | 17.8    | P             | 8.7    | W |        | D   | 33            |               |   |   |   |   |
| 9   | 15.6    | G             | 20.7   | A |        |     | 34            |               |   |   |   |   |
| 10  | 8.6     | W             | 16.2   | I |        | V   | 35            |               |   |   |   |   |
| 11  | 22.7    | A             | 12.9   | P |        | N   | 36            |               |   |   |   |   |
| 12  | 13.3    | I             | 5.2    | H |        |     | 37            |               |   |   |   |   |
| 13  | 17.7    | P             | 12.5   | N |        |     | 38            |               |   |   |   |   |
| 14  | 6.2     | H             |        | - |        |     | 39            |               |   |   |   |   |
| 15  | 14.5    | N             | 10.5   | L |        |     | 40            |               |   |   |   |   |
| 16  |         |               |        |   |        |     | 41            |               |   |   |   |   |
| 17  |         |               |        |   |        |     | 42            |               |   |   |   |   |
| 18  |         |               |        |   |        |     | 43            |               |   |   |   |   |
| 19  |         |               |        |   |        |     | 44            |               |   |   |   |   |
| 20  |         |               |        |   |        |     | 45            |               |   |   |   |   |
| 21  |         |               |        |   |        |     | 46            |               |   |   |   |   |
| 22  |         |               |        |   |        |     | 47            |               |   |   |   |   |
| 23  |         |               |        |   |        |     | 48            |               |   |   |   |   |
| 24  |         |               |        |   |        |     | 49            |               |   |   |   |   |
| 25  |         |               |        |   |        |     | 50            |               |   |   |   |   |

Notes:                      Yields of peptides are in picomoles.

Appendix IX:      N-Terminal Sequence Analysis of recSCI

# References

# References

- Adams, J. M. (1968). "On the release of the formyl group from nascent protein." *J. Mol. Biol.* 33: 571-589.
- Aplin, R. T., J. E. Baldwin, Y. Fujishima, C. J. Schofield, B. N. Green & S. A. Jarvis (1990). "Molecular weight analysis of isopenicillin N synthase by electrospray mass spectrometry." *FEBS Lett.* 264(2): 215-217.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith & K. Struhl, (Eds.) (1988). Current Protocols in Molecular Biology. New York, John Wiley & Sons.
- Barrett, A. J. & J. K. McDonald (1986). "Nomenclature: protease, proteinase and peptidase." *Biochem. J.* 237: 935.
- Barrett, A. J. (1986). "An introduction to the proteinases." in Proteinase Inhibitors. (Eds. A. J. Barrett & G. Salvesen), Elsevier Science Publishers B.V. (Biomedical Division), Amsterdam. pp. 3-22.
- Becker, D. M. & L. Guarente (1991). "High-efficiency transformation of yeast by electroporation." *Meth. Enzymol.* 194: 182-187.
- Ben-Bassat, A., K. Bauer, S.-Y. Chang, K. Myambo, A. Boosman & S. Chang (1987). "Processing of the initiation methionine from proteins: Properties of the *Escherichia coli* methionine aminopeptidase and its gene structure." *J. Bacteriol.* 169(2): 751-757.
- Berger, A. & I. Schechter (1970). "Mapping the active site of papain with the aid of peptide substrates and inhibitors." *Phil. Trans. R. Soc. London B*257: 249-264.
- Bergmann, M. & J. S. Fruton (1937). "On proteolytic enzymes. XII. Regarding the specificity of aminopeptidase and carboxypeptidase. A new type of enzyme in the intestinal tract." *J. Biol. Chem.* 117: 189-202.
- Bernard, H.-U. & D. R. Helinski (1979). "Use of the  $\lambda$ -phage promoter  $P_L$  to promote gene expression in hybrid plasmid cloning vehicles." *Meth. Enzymol.* 68: 482-492.

Bieker, K. & T. Silhavy (1990). "The genetics of protein secretion in *E. coli*." *Trends Genet.* 6(10): 329-334.

Birnboim, H. C. & J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." *Nucl. Acids Res.* 7(6): 1513-1523.

Bode, W., H. J. Greyling, R. Huber, J. Otlewski & T. Wilusz (1989). "The refined 2.0 Å X-ray crystal structure of the complex formed between bovine  $\beta$ -trypsin and CMTI-I, a trypsin inhibitor from squash seeds (*Cucurbita maxima*): Topological similarity of the squash seed inhibitors with the carboxypeptidase A inhibitor from potatoes." *FEBS Lett.* 242(2): 285-292.

Bolivar, F., R. L. Rodriguez, P. Y. Greens, M. C. Betlach, H. L. Heynecker, H. W. Boyer, J. H. Crosa & S. Falkow (1977). "Construction and characterization of new cloning vehicles. II. A multipurpose cloning system." *Gene* 2: 95-113.

Bouma, B. N. & J. H. Griffin (1986). "Initiation mechanisms: The contact activation system in plasma." in Blood Coagulation. (Eds. R. F. A. Zwaal & H. C. Hemker), Elsevier Science Publishers B.V. (Biomedical Division), pp. 103-128.

Bowden, G. A. & G. Georgiou (1990). "Folding and aggregation of  $\beta$ -lactamase in the periplasmic space of *Escherichia coli*." *J. Biol. Chem.* 265(28): 16760-16766.

Bowden, G. A., A. M. Paredes & G. Georgiou (1991). "Structure and morphology of protein inclusion bodies in *Escherichia coli*." *Bio/Tech.* 9: 725-730.

Bradford, M. M. (1976). "A rapid sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal. Biochem.* 72: 248-254.

Brake, A. J., J. P. Merryweather, D. G. Coit, U. A. Heberlein, F. R. Masiarz, G. T. Mullenbach, M. S. Urdea, P. Valenzuela & P. J. Barr (1984). "Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*." *Proc. Natl. Acad. Sci. USA* 81: 4642-4646.

Brinkmann, U., R. E. Mattes & P. Buckel (1989). "High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product." *Gene* 85: 109-114.

Broach, J. R. (1982). "The yeast 2 $\mu$  circle." *Cell* 28: 203-204.

Buchner, J. & R. Rudolph (1991). "Renaturation, purification and characterisation of recombinant F<sub>ab</sub>-fragments produced by *Escherichia coli*." *Bio/Tech.* 9(9): 157-162.

Burke, B., G. Griffiths, H. Reggio, D. Louvard & G. Warren (1982). "A monoclonal antibody against a 135-K Golgi membrane protein." *EMBO J.* 1(12): 1621-1628.

Carter, P., H. Bedouelle & G. Winter (1985). "Improved oligonucleotide site-directed mutagenesis using M13 vectors." *Nucl. Acids Res.* 13(12): 4431-4443.

Chong, G. L. & G. R. Reeck (1987). "Interaction of trypsin,  $\beta$ -factor XIIa, and plasma kallikrein with a trypsin inhibitor isolated from barley seeds: a comparison with the corn inhibitor of activated Hageman Factor." *Thrombosis Res.* 48: 211-221.

Chou, P. Y. & G. D. Fasman (1978). "Empirical predictions of protein conformations." *Ann. Rev. Biochem.* 47: 251-276.

Cochrane, C. G. & J. H. Griffin (1982). "The biochemistry and pathophysiology of the contact system of plasma." *Adv. Immunol.* 33: 241-306.

Cool, D. E., C.-J. S. Edgell, G. V. Louie, M. J. Zoller, G. D. Brayer & R. T. A. MacGillivray (1985). "Characterization of human blood coagulation factor XII cDNA: Prediction of the primary structure of factor XII and the tertiary structure of  $\beta$ -factor XIIa." *J. Biol. Chem.* 260: 13666-13676.

Dalbøge, H., S. Bayne & J. Pedersen (1990). "In vivo processing of N-terminal methionine in *E. coli*." *FEBS Lett.* 266(1,2): 1-3.

Darby, N. J. & T. E. Creighton (1990). "Folding Proteins." *Nature* 344: 715-716.

Das, A. (1990). "Overproduction of proteins in *Escherichia coli*: vectors, hosts and strategies." *Meth. Enzymol.* 182: 93-112.

Davie, E. W. & K. Fujikawa (1975). "Basic mechanisms in blood coagulation." *Ann. Rev. Biochem.* 44: 799-829.

Davie, E. W. & O. D. Ratnoff (1964). "Waterfall sequence for intrinsic blood clotting." *Science* 145: 1310-1312.

de Boer, H.A. & L.J. Comstock & M. Vasser (1983). "The tac promoter: a functional hybrid derived from the trp and lac promoters." *Proc. Natl. Acad. Sci. USA* 80: 21-25.

Deng, T., J. P. Noel & M.-D. Tsai (1990). "A novel expression vector for high-level synthesis of foreign proteins in *Escherichia coli*: overproduction of bovine pancreatic phospholipase A<sub>2</sub>." *Gene* 93: 229-234.

Dobson, M. J., M. F. Tuite, N. A. Roberts, A. J. Kingsman, S. M. Kingsman, R. E. Perkins, S. C. Conroy, B. Dunbar & L. A. Fothergill (1982). "Conservation of high efficiency promoter sequences in *Saccharomyces cerevisiae*." *Nucl. Acids Res.* 10(8): 2625-2637.

Duffaud, G. D., P. E. March & M. Inouye (1987). "Expression and secretion of foreign proteins in *Escherichia coli*." *Meth. Enzymol.* 153: 492-507.

Edmonds, C. G. & R. D. Smith (1990). "Electrospray ionization mass spectrometry." *Meth. Enzymol.* 193: 413-431.

Ellis, R. J. (1990). "Impropriety in proteins." *Nature* 345: 484.

Erhart, E. & C. Hollenberg (1983). "The presence of a defective *LEU2* gene on 2μ DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number." *J. Bacteriol.* 156(2): 625-635.

Esnouf, M. P. (1984). "Biochemistry of Blood Coagulation." in Anticoagulants and Myocardial Infarction. (Ed. T. W. Meade), John Wiley & Sons Ltd., pp. 29-61.

Fuhrer, G., M. J. Gallimore, W. Heller & H.-E. Hoffmeister (1990). "F XII." *Blut* 61: 258-266.

Fuller, R. S., R. E. Sterne & J. Thorner (1988). "Enzymes required for yeast prohormone processing." *Ann. Rev. Physiol.* 50: 345-362.

Furie, B. & B. C. Furie (1988). "The Molecular Basis of Blood Coagulation." *Cell* 53: 505-518.

Gailani, D. & G. J. Broze Jr. (1991). "Factor XI activation in a revised model of blood coagulation." *Science* 253: 909-912.

Garcia, G. M., P. K. Mar, D. A. Mullin, J. R. Walker & N. E. Prather (1986). "The *E.coli dnaY* gene encodes an arginine transfer RNA." *Cell* 45: 453-459.

Garnier, J., D. J. Osguthorpe & B. Robson (1978). "Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins." *J. Mol. Biol.* 120: 97-120.

Gatenby, A., P. Viitanen & G. Lorimer (1990). "Chaperonin assisted polypeptide folding and assembly: implications for the production of functional proteins in bacteria." *Trends Biotech.* 8: 354-358.

Geisow, M. J. & B. N. Green (1990). "Rapid detection and identification of post-translational modifications by electrospray mass spectrometry". *BioSpec* 1 (VG Biotech).

Ghrayeb, J. & M. Inouye (1984). "Nine amino acid residues at the NH<sub>2</sub>-terminal of lipoprotein are sufficient for its modification, processing, and localisation in the outer membrane of *Escherichia coli*." *J. Biol. Chem.* 259(1): 463-467.

Ghrayeb, J., H. Kimura, M. Takahara, H. Hsiung, Y. Masui & M. Inouye (1984). "Secretion cloning vectors in *Escherichia coli*" *EMBO J.* 3(10): 2437-2442.

Goff, S. A. & A. L. Goldberg (1985). "Production of abnormal proteins in *E. coli* stimulates transcription of *lon* and other heat shock proteins." *Cell* 41: 587-595.

Goldenberg, D. P. (1988). "Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols." *Biochemistry* 27: 2481-2489.

Gough, J. & N. Murray (1983). "Sequence diversity among related genes for recognition of specific targets in DNA molecules." *J. Mol. Biol.* 166: 1-19.

Grassmann, W. & H. Dyckerhoff (1928). "Über die Proteinase und die Polypeptidase der Hefe." *Hoppe-Seyler's Z. Physiol. Chem.* 179: 41-78.

Haase-Pettingell, C.A. & J. King (1988). "Formation of aggregates from a thermolabile *in vivo* folding intermediate in P22 tailspike maturation." *J. Biol. Chem.* 263 (10): 4977-4983.

Hahn, K. W., W. A. Klis & J. M. Stewart (1990). "Design and synthesis of a peptide having chymotrypsin-like esterase activity." *Science* 248: 1544-1547.

Hartley, B. S. (1960). "Proteolytic enzymes." *Ann. Rev. Biochem.* **29**: 45-72.

Hinnen, A., J. B. Hicks & G. R. Fink (1978). *Proc. Natl. Acad. Sci. USA* **75**: 1929.

Hirel, P.-N., J.-M. Schmitter, P. Dessen, G. Fayat & S. Blanquet (1989). "Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid." *Proc. Natl. Acad. Sci. USA* **86**: 8247-8251.

Hitzeman, R. A., D. W. Leung, L. J. Perry, W. J. Kohr, H. L. Levine & D. V. Goeddel (1983). "Secretion of human interferons by yeast." *Science* **219**: 621-625.

Hoess, A., A. K. Arthur, G. Wanner & E. Fanning (1988). "Recovery of soluble, biologically active recombinant proteins from total bacterial lysates using ion exchange resin." *Bio/Tech.* **6**: 1214-1217.

Hofmann, K. J. & L. D. Schultz (1991). "Mutations of the  $\alpha$ -galactosidase signal peptide which greatly enhance secretion of heterologous proteins by yeast." *Gene* **101**: 105-111.

Hojima, Y., J. V. Pierce & J. J. Pisano (1980). "Plant inhibitors of serine proteinases: Hageman factor fragment, kallikreins, plasmin, thrombin, factor Xa, trypsin and chymotrypsin." *Thrombosis Res.* **20**: 163-171.

Hojima, Y., J. V. Pierce & J. J. Pisano (1982). "Pumpkin seed inhibitor of human factor XIIa (Activated Hageman Factor) and bovine trypsin." *Biochemistry* **21**: 3741-3746.

Holmes, D. S. & M. Quigley (1981). "A rapid boiling method for the preparation of bacterial plasmids." *Anal. Biochem.* **114**: 193-197.

Hubbard, S. J., S. F. Campbell & J. M. Thornton (1991). "Molecular recognition: Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors." *J. Mol. Biol.* **220**: 507-530.

Ikemura, H., H. Takagi & M. Inouye (1987). "Requirement of Pro-sequence for the production of active subtilisin E in *Escherichia coli*." *J. Biol. Chem.* **262**(16): 7859-7864.

Ikemura, T. & H. Ozeki (1982). "Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to isoacceptor contents." *Cold Spring Harbor Symp. Quant. Biol.* **47**: 1087-1097.

Ikemura, T. & H. Ozeki (1982). "Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to isoacceptor contents." *Cold Spring Harbor Symp. Quant. Biol.* 47: 1087-1097.

Inokuchi, K., A. Nakayama & F. Hishinuma (1987). "Identification of sequence elements that confer cell type specific control of MF-alpha-1 expression in *Saccharomyces cerevisiae*." *Mol. Cel. Biol.* 7(9): 3185-3193.

Janin, J. & C. Chotia (1976). "Stability and specificity of protein-protein interactions: The case of the trypsin-trypsin inhibitor complexes." *J. Mol. Biol.* 100: 197-211.

Jigami, Y., M. Muraki, N. Harada & H. Tanaka (1986). "Expression of synthetic human-lysozyme gene in *Saccharomyces cerevisiae*: use of a synthetic chicken-lysozyme signal sequence for secretion and processing." *Gene* 43: 273-279.

Johnson, K. R. (1990). "A Small-Scale Plasmid Preparation Yielding DNA Suitable for Double-Stranded Sequencing and *in vitro* Transcription." *Anal. Biochem.* 190: 170-174.

Julius, D., R. Schekman & J. Thorner (1984). "Glycosylation and processing of prepro-alpha-factor through the yeast secretory pathway." *Cell* 36: 309-318.

Kálmán, M., I. Cserpán, G. Bajszár, A. Dobi, E. Horváth, C. Pázmán & A. Simoncsits (1990). "Synthesis of a gene for human serum albumin and its expression in *Saccharomyces cerevisiae*." *Nucl. Acids Res.* 18(20): 6075-6081.

Kanamori, T., S. Mizushima, Y. Shimizu, H. Morishita, H. Kubota, A. Nii, H. Ogino, Y. Nagase, M. Kisaragi & M. Nobuhara (1988). "Expression and excretion of human pancreatic secretory trypsin inhibitor in lipoprotein-deletion mutant of *Escherichia coli*." *Gene* 66: 295-300.

Kingsman, S. M., A. J. Kingsman, M. J. Dobson, J. Mellor & N. A. Roberts (1985). "Heterologous gene expression in *Saccharomyces cerevisiae*." *Biotech. Genet. Eng. Rev.* 3: 377-416.

Kodo, Y. (1988). D.Phil Thesis University of Oxford.

Krishnamoorthi, R., Y.-X. Gong & M. Richardson (1990). "A new protein inhibitor of trypsin and activated Hageman factor from pumpkin (*Cucurbita maxima*) seeds." *FEBS Lett.* 273(1,2): 163-167.

Kurjan, J. & I. Herskowitz (1982). "Structure of a yeast pheromone gene (MF-alpha): A putative alpha-factor precursor contains four tandem copies of mature alpha-factor." *Cell* 30: 933-943.

Kurland, C. G. (1991). "Codon bias and gene expression." *FEBS Lett.* 285(2): 165-169.

Kusters, R., W. Dowhan & B. de Kruijff (1991). "Negatively charged phospholipids restore prePhoE translocation across phosphatidylglycerol-depleted *Escherichia coli* inner membranes." *J. Biol. Chem.* 266(14): 8659-8662.

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

Laskowski Jr., M. & I. Kato (1980). "Protein inhibitors of proteinases." *Ann. Rev. Biochem.* 49: 593-626.

Lei, M.-G. & G. R. Reeck (1987). "Resynthesis by factor XIIa of the trypsin-cleaved peptide bond in the corn protease inhibitor." *Thromb. Res.* 45: 87-94.

Livingston, D. M. & P. Leder (1969). "Deformylation and protein biosynthesis." *Biochemistry* 8(1): 435-443.

Longstaff, C., A. F. Campbell & A. R. Fersht (1990). "Recombinant chymotrypsin inhibitor 2: Expression, kinetic analysis of inhibition of  $\alpha$ -chymotrypsin and wild-type and mutant subtilisin BPN', and protein engineering to investigate inhibitory specificity and mechanism." *Biochem.* 29: 7339-7347.

Lowman, H. B. & M. Bina (1990). "Temperature-mediated regulation and downstream inducible selection for controlling gene expression from bacteriophage  $\lambda$  P<sub>L</sub> promoter." *Gene* 96(133-136):

MacFarlane, R. G. (1964). "An enzyme cascade in the blood clotting mechanism and its function as a biochemical amplifier." *Nature* 202: 498-499.

Mahoney, W. C., M. A. Hermodson, B. Jones, D. D. Powers, R. S. Corfman & G. R. Reeck (1983). "Amino acid sequence and secondary structure analysis of the corn inhibitor of trypsin and activated Hageman factor." *J. Biol. Chem.* 259(13): 8412-8416.

Marston, F. A. O. & D. L. Hartley (1990). "Solubilization of protein aggregates." *Meth. Enzymol.* 182: 264-276.

Marston, F. A. O. (1986). "The purification of eukaryotic polypeptides synthesized in *Escherichia coli*." *Biochem. J.* 240: 1-12.

Marston, F. A. O. (1987). "The purification of eukaryotic polypeptides expressed in *Escherichia coli*." in DNA Cloning: A Practical Approach. (Ed. D. M. Glover), IRL Press, Oxford. pp. 59-89.

Mellor, J., M. J. Dobson, N. A. Roberts, A. J. Kingsman & S. M. Kingsman (1985). "Factors affecting heterologous gene expression in *Saccharomyces cerevisiae*." *Gene* 33: 215-226.

Milman, G. (1987). "Expression plasmid containing  $\lambda$ -P<sub>L</sub> promoter and cI857 repressor." *Meth. Enzymol.* 153: 482-491.

Movva, N. R., K. Nakamura & M. Inouye (1980). "Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of *Escherichia coli*." *J. Biol. Chem.* 255(1): 27-29.

Murray, J. A. H. (1986). "HCC ligation: rapid and specific construction with blunt ended fragments." *Nucl. Acids Res.* 14(24): 10118.

Nagohara, H., H. Fujisawa & Y. Jigami (1988). "Alterations in the cleavage site of the signal sequence for the secretion of human lysozyme by *Saccharomyces cerevisiae*." *FEBS Lett.* 238(2): 329-332.

Naito, K. & K. Fujikawa (1991). "Activation of human blood coagulation factor XI independent of factor XII." *J. Biol. Chem.* 266(12): 7353-7358.

Neeper, M. P., L. Waxman, D. E. Smith, C. A. Schulman, M. Sardana, R. W. Ellis, L. W. Schaffer, P. K. S. Siegl & G. P. Vlasuk (1990). "Characterization of recombinant tick anticoagulant peptide: A highly selective inhibitor of blood coagulation factor Xa." *J. Biol. Chem.* 265(29): 17746-17752.

Neu, H. C. & L. A. Heppel (1965). "The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts." *J. Biol. Chem.* 240(9): 3685-3692.

Neurath, H. (1984). "Evolution of proteolytic enzymes." *Science* 224: 350-357.

Notwehr, S. F. & J. I. Gordon (1990). "Targeting of proteins into the eukaryotic secretory pathway: Signal peptide structure/function relationships." *BioEssays* 12(10): 479-484.

Nuijens, J. H., C. C. M. Huijbregts, A. J. M. Eerenberg-Belmer, J. C. M. Meijers, B. N. Bouma & C. E. Hack (1989). "Activation of the contact system of coagulation by a monoclonal antibody directed against a neodeterminant in the heavy chain region of human coagulation factor XII (Hageman factor)." *J. Biol. Chem.* 264(22): 12941-12949.

Obukowicz, M. G., M. A. Turner, E. Y. Wong & W. C. Tacon (1988). "Secretion and export of IGF-1 in *Escherichia coli* strain JM101." *Mol. Gen. Genet.* 215: 19-25.

Obukowicz, M. G., M. E. Gustafson, K. D. Junger, R. M. Leimgruber, A. J. Wittwer, T.-C. Wun, T. G. Warren, B. F. Bishop, K. J. Mathis, D. T. McPherson, N. R. Siegel, M. G. Jennings, B. B. Brightwell, J. A. Diaz-Collier, L. D. Bell, C. S. Craik & W. C. Tacon (1990). "Secretion of active kringle-2-serine protease in *Escherichia coli*." *Biochemistry* 29: 9737-9745.

Orsini, G. & M. E. Goldenberg (1978). "The renaturation of reduced chymotrypsinogen A in guanidine HCl: refolding versus aggregation." *J. Biol. Chem.* 253(3453-3458):

Pauling, L. (1946). "Molecular architecture and biological reactions." *Chem. Engin. News* 24: 1375-1377.

Pixley, R. A., L. G. Stumpo, K. Birkmeyer, L. Silver & R. W. Colman (1987). "A monoclonal antibody recognizing an icosapeptide sequence in the heavy chain of human factor XII inhibits surface-catalysed activation." *J. Biol. Chem.* 262(21): 10140-10145.

Pixley, R. A., M. Schapira & R. W. Colman (1985). "The regulation of human factor XIIa by plasma proteinase inhibitors." *J. Biol. Chem.* 260(3): 1723-1729.

Poindexter, K. & R. B. Gayle III (1991). "Induction of recombinant gene expression in *Escherichia coli* using an alkaline pH shift." *Gene* 97: 125-130.

Powers, J. C. & J. W. Harper (1986). "Inhibitors of serine proteinases." in Proteinase Inhibitors. (Eds. A. J. Barrett & G. Salvesen), Elsevier Science Publishers B.V. (Biomedical Division), Amsterdam. pp. 55-152.

Ptashne, M. (1987). A genetic switch; gene control and phage  $\lambda$ . Oxford, Blackwell Scientific Publications.

Quaderi, N. A. (1991). Part II Thesis, University of Oxford.

Raleigh, E. A., K. Lech & R. Brent (1989). in Current Protocols in Molecular Biology. (Eds. F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith & K. Struhl), John Wiley & Sons, New York.

Ramos, M. J. (1991). "Model structure for the human blood coagulation agent  $\beta$ -factor XIIa." *J. Mol. Graphics* 9: 91-93.

Ratnoff, O. D. & J. E. Calopy (1955). "A familial haemorrhagic trait associated with a clot-promoting fraction from plasma." *J. Clin. Invest.* 34: 602-613.

Ratnoff, O. D. & V. Moneme (1981). "Inhibition of ellagic acid-activated Hageman factor (Factor XII) and Hageman Factor fragments by popcorn inhibitor." *Proc. Soc. Exp. Biol. Med.* 166: 297-299.

Read, R. J. & M. N. G. James (1986). "Introduction to the protein inhibitors: X-ray crystallography." in Proteinase Inhibitors. (Eds. A. J. Barrett & G. Salvesen), Elsevier Science Publishers B.V. (Biomedical Division), Amsterdam. pp. 301-336.

Remaut, E., A. Marmenout, G. Simons & W. Fiers (1987). "Expression of heterologous unfused protein in *Escherichia coli*." *Meth. Enzymol.* 153: 416-431.

Reznikoff, W. S. & W. R. McClure (1986). "*E. coli* promoters." in Maximizing Gene Expression. (Ed. W. S. Reznikoff & L. Gold), Butterworths, Boston. pp. 1-33.

Rine, J., J. N. Strathern, J. B. Hicks & L. Herskowitz (1979). "A supressor mating-type locus mutations in *Saccharomyces cerevisiae*: Evidence for identification of cryptic mating-type loci." *Genetics* 93: 877-901.

Rose, G. D., A. R. Geselowitz, G. J. Lesser, R. H. Lee & M. H. Zehfus (1985). "Hydrophobicity of amino acid residues in globular proteins." *Science* 229: 834-838.

Rothman, J. H., C. P. Hunter, L. A. Valls & T. H. Stevens (1986). "Overproduction-induced mislocalization of a yeast vacuolar protein allows isolation of its structural gene." *Proc. Natl. Acad. Sci. USA* 83: 3248-3252.

Ruohonen, L., P. Hackman, P. Lehtovaara, J. K. C. Knowles & S. Keränen (1987). "Efficient secretion of *Bacillus amyloliquifaciens*  $\alpha$ -amylase cells by its own signal peptide from *Saccharomyces cerevisiae* host." *Gene* 59: 161-170.

Rusche, J. R. & P. Howard-Flanders (1985). "Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase." *Nucl. Acids Res.* 13(6): 1997-2008.

Russell, D. R. & G. N. Bennett (1982). "Construction and analysis of in vivo activity *E.coli* promoter hybrids and promoter mutants that alter the -35 to -10 spacing." *Gene* 20: 231-243.

Sambrook, J., E. F. Fritsch & T. Maniatis (1989). Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen & A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." *Proc. Natl. Acad. Sci. USA* 74: 5463.

Sawyer, R. T. (1991). "Thrombolytics and anti-coagulants from leeches." *Bio/Tech.* 9: 513-518.

Saxena, V. P. & D. B. Wetlaufer (1970). "Formation of three-dimensional structure in proteins. I. Rapid non-enzymatic reactivation of reduced lysozyme." *Biochemistry* 9(5015-5021):

Schapira, M., M.-A. Ramus, B. Waeber, H. R. Brunner, S. Jallat, D. Carvallo, C. Roitsch & M. Courtney (1987). "Protection by recombinant  $\alpha_1$ -antitrypsin Ala<sup>357</sup> Arg<sup>358</sup> against arterial hypotension induced by factor XII fragment." *J. Clin. Invest.* 80: 582-585.

Schilling, J., Y. Wang, K. Lau, L. Smith & B. Cordell (1991). "Synthesis and characterisation of the Kunitz protease-inhibitor domain of the  $\beta$ -amyloid precursor protein." *Gene* 98: 225-230.

Schultz, L. D., K. J. Hofmann, L. M. Mylin, D. L. Montgomery, R. W. Ellis & J. E. Hopper (1987). "Regulated overproduction of the GAL4 gene product greatly increases expression from galactose inducible promoters on multi-copy expression vectors in yeast." *Gene* 61: 123-133.

Scott, C. F., R. W. Carrell, C. B. Glaser, F. Kueppers, J. H. Lewis & R. W. Colman (1985). "Alpha-1-antitrypsin-Pittsburgh; A potent inhibitor of human plasma factor XIa, kallikrein and factor XII<sub>f</sub>." *J. Clin. Invest.* 77: 631-634.

Sharp, P. M. & W.-H. Li (1987). "The codon adaptation index - a measure of directional synonymous codon usage bias and its potential application." *Nucl. Acids Res.* 15(3): 1281-1295.

Steube, K., B. Chaudhuri, W. Marki, J. P. Merryweather & J. Heim (1991). "Alpha-factor-leader-directed secretion of recombinant human-insulin-like growth factor I from *Saccharomyces cerevisiae*: Precursor formation and processing in the yeast secretory pathway." *Eur. J. Biochem.* 198: 651-657.

Stormo, G. D. (1986). "Translation Initiation." in Maximizing Gene Expression. (Eds. W. Reznikoff & L. Gold), Butterworths, Boston. pp. 195-224.

Suzuki, H., Y. Nishimura, S. Yasuda, A. Nishimura, M. Yamada & Y. Hirota (1978). "Murein-lipoprotein of *Escherichia coli*: a protein involved in the stabilization of bacterial cell envelope." *Mol. Gen. Genet.* 167: 1-9.

Suzuki, H., Y. Nishimura, S. Yasuda, A. Nishimura, M. Yamada & Y. Hirota (1978). "Murein-lipoprotein of *Escherichia coli*: a protein involved in the stabilization of bacterial cell envelope." *Mol. Gen. Genet.* 167: 1-9.

Swartz, M. J., H. L. Mitchell, D. J. Cox & G. R. Reeck (1977). "Isolation and characterization of trypsin inhibitor from opaque-2 corn seeds." *J. Biol. Chem.* 252: 8105-8107.

Tabor, S. & C. C. Richardson (1985). "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes." *Proc. Natl. Acad. Sci. USA* 82: 1074-1078.

Tabor, S. & C. C. Richardson (1987). "DNA sequence analysis with a modified bacteriophage T7 DNA polymerase." *Proc. Natl. Acad. Sci. USA* 84: 4767-4771.

Tessier, L.-H., S. Jallat, M. Sauvageot, R. G. Crystal & M. Courtney (1986). "RNA structural elements for expression in *Escherichia coli*:  $\alpha_1$ -antitrypsin synthesis using translational control elements based on the cII ribosome-binding site of phage  $\lambda$ ." *FEBS Lett.* 208(2): 183-188.

Travis, J. & G. S. Salvesen (1983). "Human plasma proteinase inhibitors." *Ann. Rev. Biochem.* 2: 655-709.

Tsuji, T., R. Nakagawa, N. Sugimoto & K. Fukuhara (1987). "Characterization of disulfide bonds in recombinant proteins: Reduced human interleukin-2 in inclusion bodies and its oxidative refolding." *Biochemistry* 26: 3129-3134.

Valenzuela, P. (1988). "The production of foreign proteins in *Saccharomyces cerevisiae*." *Arch. Biol. Med. Exp.* 21: 231-240.

VanBogelen, R. A., P. M. Kelley & F. C. Neidhardt (1987). "Differential induction of heat shock, SOS, and oxidation stress regulons and accumulation of nucleotides in *Escherichia coli*." *J. Bacteriol.* 169(1): 26-32.

Voet, D. & J. G. Voet (1990). Biochemistry. New York, John Wiley & Sons.

Warshel, A., G. Naray-Szabo, F. Sussman & J.-K. Hwang (1989). "How do serine proteases really work?" *Biochemistry* 28(9): 3629-3637.

Watt, R. A., A. R. Shatzmann & M. Rosenberg (1985). "Expression and characterization of the human c-myc DNA-binding protein." *Mol. Cell. Biol.* 5: 448-456.

Weislander, L. (1979). "A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels." *Anal. Biochem.* 98(305-309):

Wertman, K. F., A.R. Wyman & D. Botstein, (1986). "Host/Vector Interactions which affect the viability of recombinant phage lambda clones." *Gene* 49: 253-262.

Wetzel, R., L. J. Perry & C. Veilleux (1991). "Mutations in human interferon gamma affecting inclusion body formation identified by a general immunochemical screen." *Bio/Tech.* 9: 731-737.

Wieczorek, M., J. Otlewski, J. Cook, K. Parks, J. Leluk, A. Wilimovska-Pelc, A. Polanowski, T. Wilusz & M. Laskowski (1985). "The squash family of serine proteinase inhibitors. Amino acid sequences and association equilibrium constants of inhibitors from squash, summer squash, zucchini, and cucumber seeds." *Biochem. Biophys. Res. Comm.* 126(2): 646-652.

Wynn, R. & M. Laskowski Jr. (1990). "Inhibition of human  $\beta$ -factor XIIa by squash family serine proteinase inhibitors." *Biochem. Biophys. Res. Comm.* 166(3): 1406-1410.

Yanish-Perron, C., J. Vieira & J. Messing (1985). "Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors." *Gene* 33: 103-119.

Yasueda, H., K. Nagase, A. Hosoda, Y. Akiyama & K. Yamada (1990). "High-level direct expression of semi-synthetic human interleukin-6 in *Escherichia coli* and production of N-terminas Met-free product." *Bio/Tech.* 8: 1036-1040.

Yatvin, M. B., K. M. Smith & F. L. Siegel (1986). "Translocation of nascent non-signal sequence protein in heated *Escherichia coli*." *J. Biol. Chem.* 261(17): 8070-8075.