BACTERIAL EXPRESSION OF HUMAN β-UROGASTRONE FROM A CLONED SYNTHETIC GENE

by

Ian Graham Fotheringham B.Sc.

A thesis submitted for the degree of

Doctor of Philosophy

in the Faculty of Science of the University of Glasgow

Department of Genetics, Glasgow University

and

Departments of Molecular Genetics and Microbiology,

G.D. Searle and Co

February, 1986

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgements.

I would like to acknowledge the encouragement, guidance and advice of all my supervisors during this work. Firstly Prof. David Sherratt for his continuous support and for extending to me (at a moment's notice) the facilities of the Genetics department at Glasgow University. Also my supervisors at Searle (in order of appearance), Dr. Mike Doel, Dr. David Lilley and Dr. Jan-Ingmar Flock. Special thanks to Jan-Ingmar for his advice, enthusiasm and his friendship. Thanks also to Dr. Mark Edwards for turning the occasional blind eye during the past two years.

I would also like to thank Mrs Sally Pheby and especially Mr.Edward "Ted" Cook for my introduction to practical molecular biology. I am also extremely appreciative of the extra-ordinary secretarial skills of Mrs. Maureen Bevan in the final presentation of this work.

Most importantly, I would like to stress my appreciation to my family for their patience and their faith in the eventual completion of this thesis and to Alex for her support and for bravely suffering the final weeks of writing.

A final thanks to all my (now former) colleagues at G.D.Searle for their help and best wishes to them all in their future careers.

To my Grandfather

MR. THOMAS GRAHAM

The following abbreviations were used in this work:

secs	seconds.
mins	minutes.
hrs	hours.
l	litre(s).
ml(s)	millilitre(s).
ul	microlitre(s).
g	gramme(s).
mg(s)	milligramme(s).
ug	micrograms.
ng	nanogramme(s).
M	molar.
mM	millimolar.
uM	micromolar.
bp	base pair(s).
kb	kilobase pairs.
m.wt.	molecular weight.
C	degrees celcius.
V	volts.
W	watts.
uCi	microCurie.
G	centrifugal force equal to gravitational accelaration.
cm	centimetre.
mm	millimetre.

Tris	tris (hydroxymethyl) aminomethane.
EDTA	ethylene diamine tetra-acetic acid.
BSH	beta mercaptoethanol.
DTT	di-thiotreitol.
X-gal	5-bromo, 4-chloro, 3-indolyl B-D galactopyranoside.
PMSF	phenylmethylsulphonyl Fluoride.
ATP	adenosine triphosphate.
CTP	cytosine triphosphate.
GTP	guanosine triphosphate.
TTP	thymidine triphosphate.

DNA	deoxyribonucleic acid.
cDNA	complementary DNA.
RNA	ribonucleic acid.
tRNA	transfer RNA.
RNAse	ribonuclease.
kan	kanamycin.
amp	ampicillin.
SDS	sodium dodecyl sulphate.

CONTENTS

ABSTRACT

This work describes the design, cloning and expression of a synthetic gene encoding a human peptide hormone, β-urogastrone-epidermal growth factor. This is a 53 amino-acid polypeptide incorporating two separate biological activities: i) It inhibits the secretion of gastric HCl and ii) It stimulates growth of epidermal cells both in vivo and in cell culture. Preliminary studies using β-urogastrone isolated from human urine have indicated that the peptide may have potential medical applications in the treatment of gastric ulcers. However its isolation and purification from this source is extremely inefficient and time consuming.

Bacterial expression was therefore investigated as a viable alternative source of the peptide. In addition, total gene synthesis was examined as a suitable alternative to the isolation of the gene from natural sources. Various theoretical and practical concepts of gene design are discussed. The preliminary expression of β-urogastrone in E.coli was carried out, using regulatory regions derived from the host tryptophan (trp), biosynthetic operon. Subsequently a series of vectors were designed to enable expression in an alternative host, B.subtilis. The expression system was incorporated on a dual-replicon plasmid allowing selectable propagation and a study of expression in both B.subtilis and in E.coli. Transcription was initiated from an early promoter derived from a B.subtilis phage and fully characterised here. Translation was achieved using synthetic ribosome binding sites.

Subsequently, the use of site-directed mutagenesis permitted the development of this system to enable the future expression of additional genes.

CHAPTER 1

INTRODUCTION

1.1 Foreword

Genome organisation and the regulation of gene expression in living organisms has been subject in the past ten years, to one of the most intense investigations seen in biological science. The primary reason for this intense activity has been the development of an entirely new technology in molecular biology which has enabled investigators to isolate, manipulate and ultimately determine the nucleotide sequence of large, specifically selected regions of DNA. Using plasmids as vectors, particular genes can be maintained for study and subsequently re-introduced to either their original host or to a different organism.

The vast majority of experiments involving this technology has been performed in the bacterium Escherichia coli for two main reasons. Firstly it is better characterised biochemically and genetically than any other organism, providing many immediate topics for investigation and secondly it has proved the most suitable host for the preparation and manipulation of chromosomal and plasmid DNA. It is therefore not surprising that the greatest degree of knowledge has been attained concerning the structure and function of the E.coli chromosome as well as those of its phages. This has in turn enabled much of the classical biochemical and genetical analysis of E.coli to be elegantly complemented at the molecular level.

However the molecular genetics of many other micro-organisms especially Bacillus subtilis and yeast have also been investigated using these techniques and there is an increasing understanding of the mechanisms they employ in the regulation of gene expression. Similarly a large number of mammalian genes have been isolated and sequenced giving a preliminary understanding of the organisation of the

eukaryotic genome. Evolutionary considerations of genome structure have been greatly enhanced by the comparison of DNA sequence both in closely related and widely divergent organisms and with sequence analysis of transposable elements and introns has come substantial insight to the fluidity of bacterial and eukaryotic genomes.

In short, the vast accumulation of sequence data has permitted an understanding of both the coding sequence and of specific regulatory regions of DNA involved in such processes as genome replication, recombination, transcription and translation at a level of definition not previously attainable.

Perhaps the most exciting aspect of this revelation in the control of gene expression lies in the prospect of specific alteration of the DNA sequence. Whereas in the past geneticists have used the natural occurence of mutation or non-specific mutagens to study biological systems, the potential now exists to study the effects on cellular processes of DNA mutations specifically introduced at any given point in the sequence. Furthermore it is becoming a realistic possibility to specifically insert, delete or combine particular fragments of DNA known to be involved in the control of gene expression, both in an effort to better resolve the function of regulatory regions and to bring about the production of heterologous peptides in any given host. Crucial to this kind of investigation has been the use of chemically synthesised oligomers of DNA. A process not particularly new in itself but whose recent rapid improvements in speed, accuracy and efficiency have turned it into one of the most important tools available to the molecular biologist.

Using the genetic code and having determined the amino-acid sequence of a particular peptide it is now possible to design and build DNA sequences capable of encoding that peptide. Furthermore by considering the regulation of gene expression at the primary sequence level for the chosen host and speculating upon the possible role of DNA and RNA secondary structure it should also be possible to design putative regulatory elements of DNA for the gene in question and study their function in the organism.

The work in this thesis involves many of the current rapid DNA handling techniques and uses specifically designed, chemically synthesised sequences of DNA. Both coding and regulatory sequences have been designed to examine the production of a human hormone, β-urogastrone both in E.coli and in Bacillus subtilis.

1.2 Reasons for cloning the β-urogastrone gene.

The cloning in bacteria of human genes has not only been desirable to examine the organisation of the human genome. It has also enabled the evaluation of the possible therapeutic potential of several hormones and enzymes which previously could not be isolated either in sufficient quantity for study or without extreme difficulty and cost. Obvious examples at present are human insulin (1), human growth hormone (2), and interferon (3) which have been subject to great commercial interest and which can now be synthesised and isolated from bacteria in far greater quantities than from pre-existing sources.

β-urogastrone similarly has considerable therapeutic potential and has emerged as an important regulatory molecule with much current interest being directed into its structure and function (4). The β form is the more abundant of the two forms of urogastrone detected by sensitive radio-immunoassay in humans. Found predominantly in the salivary glands, duodenum and urine (5,6,7), it is a 53 amino-acid single peptide chain of mwt 6.2 kDal incorporating three disulphide bonds (8). The γ form lacks the C-terminal arginine residue but is otherwise identical (8) and shows very similar biological activities.

β-urogastrone is strikingly homologous in structure and function to mouse epidermal growth factor (EGF) isolated from the submaxillary gland (9). Both peptides inhibit gastric acid secretion (10,11) and will similarly stimulate growth of epidermal cells both in vivo (11,8) and in vitro (12). The amino-acid sequence bears such close homology to mouse EGF and to the bulk amino-acid composition obtained for human EGF that it seems likely that β-urogastrone and EGF are one and the same molecule (8). Residues in the C-terminus of the peptide are crucial to the EGF activity as removal of the terminal six residues by limited proteolysis reduces this activity ten fold whilst not affecting the inhibition of gastric acid secretion (13).

Obviously a molecule with activity to temporarily reduce the secretion of acid in the stomach has potential medical applications in the treatment of peptic ulcers, as gastric HCl is an antagonist to the healing of ulcerated tissue. Furthermore the growth factor activity of β-urogastrone may prove to be useful in actually increasing the rate of healing of the lesion. The therapeutic potential of the peptide has indeed shown promising results in human trials which have been carried out using urine derived material (14). However it seems likely that much larger

quantities of β-urogastrone will be required not only if its clinical applications are to be pursued, but also for a substantial basic research effort into its tertiary structure and its biological activities, with particular emphasis on its mechanism of action as a growth factor.

Epidermal growth factor and its cellular receptor have, along with other human growth factors become the subject of intense interest to tumour biologists. This is due to the interaction of the receptor with known tumour promoters such as phorbol esters (15) and very recently with their observed structural homology and proposed functional significance with the products of retroviral oncogenes (16) and retrovirus transformed cells (17). Obviously this facet of the biological function of EGF is of extreme importance and may be of fundamental relevance to the molecular biology of cancer. It is also an aspect of β-urogastrone action which must be closely investigated with respect to its potential clinical applications.

Its conventional isolation however, in yields of less than 1mg per 1000l of urine (8) is clearly an impractical and unattractive source of material for study. This has led to the investigations into the possibility of the molecular cloning of its gene.

Molecular cloning of the genes for other human peptides has recently yielded sufficient material for satisfactory studies to be carried out where conventional methods have proved too arduous. It seems reasonable to imagine that in this case the same approach could lead to sufficient quantities of β-urogastrone becoming available.

1.3 Possible gene cloning strategies

In order to isolate a DNA sequence encoding a particular peptide it is necessary to have both a source of appropriate DNA or RNA and a suitable selection procedure, either for the cloned gene or for the product it encodes. In the case of a eukaryotic gene and in particular a human gene there are usually considerable problems associated with these requirements.

Direct selection of cloned DNA sequences expressing human peptides is usually most easily carried out using specific radiolabelled antisera, as biological assay if at all possible is generally complex and time consuming. Since in this case neither an antiserum nor any form of workable biological assay was available at the time, the alternative remained to isolate β-urogastrone encoding sequences by means of hybridisation to a homologous DNA sequence. Since no natural DNA fragment known to be homologous to the β-urogastrone gene was available to use as a "probe" this would require the synthesis of an oligomer of DNA derived from the known amino-acid sequence of the peptide.This could then be used either to directly probe a human genomic or cDNA library or as a primer to specifically initiate synthesis of a cDNA on the desired messenger RNA.

The drawbacks to these techniques are numerous. Firstly, due to the degeneracy of the genetic code it is usually impossible to find a region of the peptide from which to deduce a probe sequence long enough to be sufficiently stringent yet with only one possible coding sequence. The best that can usually be attained is to determine a region containing one or more amino acids encoded by only one or two possible codons, such as tryptophan or methionine. In this event only very few probes need be constructed to cover all possible DNA coding

sequences. In the case of β-urogastrone the most appropriate region is shown below.

```
residues  (47) LEU.....  -LYS-TRP-TRP-GLU- .....LEU (52)

Base           TTA         AAA TGG TGG GAA         TTA
variations.    C G           G           G         C G
                 C                                   C
                 T                                   T
                           --- 12 mer ----
                           --- 14 mer --------------
```

Four possible DNA sequences can encode this tetrapeptide from reidues 48 to 51. Further elongation of the 12'mer would necessitate building eight 14'mers and so the disproportionate rise in workload during the screening process becomes evident. Hybridisation of such numbers of short sequences to a genomic library of human DNA is extremely impractical as sequences identical and closely homologous to these will occur quite frequently. This would probably require a great deal of analysis with a high background due to the difficulties in achieving optimum stringency. Furthermore, the isolation of eukaryote genomic DNA for genes intended to be expressed in prokaryotes is usually undesirable due to the presence of introns in most eukaryote genes which are not correctly processed in bacteria (18). Finally there is some question over the representative nature of gene libraries and the stability of inserts of particular size or structure (19). It is possible that some sequences are present in low abundance or not at all.

The more attractive proposition is to selectively enrich for clones of the desired sequence. This is achieved by specifically probing or priming cDNA derived from the mRNA of cells known to produce the protein in question. Since there were no appropriate cell lines available for the isolation of β-urogastrone mRNA this would necessitate the extraction of the RNA from fresh tissue. In the case of a human protein however, tissue availability is obviously, with few exceptions such as the placenta, a severe restriction. In addition β-urogastrone is produced in the cells of the small intestine which even when available following surgical removal present considerable obstacles to the extraction of mRNA, since this tissue is tough, difficult to denature quickly and abundant in nucleases. This necessitates prolonged and difficult work at low temperatures.

Consideration of these biological and technical difficulties coupled with the developments occuring in the field of gene synthesis and the various other advantages of a specifically designed gene, led to the serious consideration of total gene synthesis as a realistic alternative to the isolation of the coding sequence from natural sources. Previously, oligonucleotides of synthetic DNA have been used in gene synthesis (20) but recent improvements in the methodology of chemical synthesis and HPLC purification (21) have vastly increased the rate and accuracy of oligonucleotide synthesis and purification. This enabled the 23 oligomers required for the assembly of the β-urogastrone gene, which ranged from 12 to 20 bases in length, to be synthesised within a few weeks.

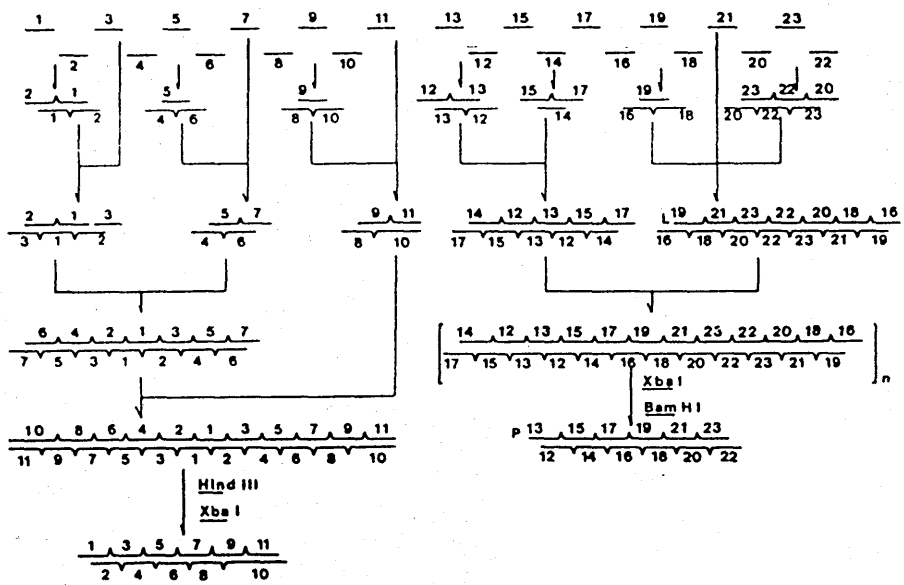


Figure 1. Strategy of gene assembly.

Schematic representation of the stages of gene assembly and structure of the intermediate sub-assemblies.

1.4 Design of the synthetic gene

1.4.1 Codon choice

Obviously the degeneracy of the genetic code permits some degree of flexibility in the choice of nucleotide sequence designed to encode a given peptide. This allows various biological and technical parameters to be considered and compromised upon, in designing the optimum sequence.

E.coli like most organisms for which DNA sequence data are available displays a preference for the use of particular codons in the coding sequence of its genome (22). This reflects the abundance of the corresponding charged tRNA molecules in the cytoplasm (23). The effect of this bias on the efficiency of translation is still unclear but it would seem prudent to incorporate where possible within a synthetic gene designed to be expressed in E.coli, the most frequently used codons of the host or to at least avoid the most infrequently used ones. This minimises the possibility of encountering any translational limitation which may be imposed by the cell. This may be especially relevent to a plasmid borne gene present in multiple copies and expressed at high levels. Codon preference in B.subtilis is less well documented due to the small number of genes yet sequenced. However it appears that there is no gross dissimilarity of codon choice with that of E.coli involving the codons used in the synthetic β-urogastrone gene (24). Very recently it has been suggested that codon bias may be less significant in B.subtilis than in E.coli (25).

1.4.2 Considerations as to the method of gene assembly

I considered that the assembly of the gene would be most easily carried out by means of sequential hybridisation-ligation reactions, as illustrated in Figure 1, with purification of correctly assembled intermediates for subsequent assemblies. This minimises the risk of

undesired hybridisations between non-adjacent fragments which could occur if large numbers of oligomers were mixed. Obviously the greatest yield of each correctly ligated sub-assembly will be achieved by optimising the ratio of correctly to incorrectly hybridised oligomers. Avoidance of incorrectly aligned fragments is best achieved using maximum overlap between oligomers and minimum incorporation of repeated or palindromic sequences. This is obviously most relevant to the initial stages of gene assembly since the purification of intermediates lessens the need for stringent avoidance of potentially homologous regions between distant fragments, as these will have been incorporated into higher molecular weight structures by the later stages of asembly. In order to avoid undesired homology between oligomers, computer programmes were employed which using free energies of base pairing (26) predict the relative stability of correct and undesired hybridisations between specific fragments.

1.4.3 Inclusion of specific restriction endonuclease sites

Using these two major criteria to design a framework sequence for the gene it is then extremely useful to incorporate where possible, specific restriction endonuclease sites both terminally and internally. This facilitates not only cloning and sequencing of the completed gene but analysis of assembly intermediates and following expression permits the replacement of particular fragments by alternative sequences. This in turn allows the simple alteration of the structure of the gene or the peptide.

1.5 Bacterial expression of the synthetic gene

1.5.1 Levels of regulation

The expression in a bacterium of any plasmid-borne gene whether natural or synthetic in origin, is regulated at several levels by the host cell. The efficiency of transcription and translation of the gene, the stability and the copy number of the plasmid and the effects of cellular nucleases and proteases all exert a major influence upon the levels of the particular mRNA and peptide within the cell. Whereas the mechanisms and pathways of specific nucleic acid and protein degradation are only beginning to be elucidated (27,28) the mode of regulation of gene expression has now been examined in great detail. Specifically, the role of the primary structure of DNA on the regulation of mRNA and protein synthesis has been sufficiently elucidated to enable the isolation of sequences of DNA and RNA which regulate transcription and translation both positively and negatively. These sequences can then be combined and modified to form expression systems capable of synthesising high levels of foreign protein in the bacterial cell (29). Particular sequences of DNA are now known to regulate transcription either by direct interaction with RNA polymerase (30) or with repressor proteins (31). Similarly regions of mRNA sequence have been shown to interact specifically with ribosomal RNA in the regulation of translation (32). The formation of stable DNA and RNA secondary structures have also been shown to be very important in the control of both transcription and translation (33,34).

1.5.2 Choice of expression system in E.coli

One of the best characterised regulatory regions of the E.coli genome is that of the tryptophan (trp) biosynthetic operon. The sequence of the entire operon has now been determined and the composition and function of the

transcription promoter, operator and attenuator have been extensively studied and understood (35). This has enabled their use to express foreign proteins in E.coli (3,36). This has also been the case with the regulatory regions of the E.coli lactose (lac) operon (37). In addition, a series of plasmids were available during the course of this work which contained derivatives of the trp promoter region and of the translation intiation point of the first gene in the operon, trpE. Genes may be easily cloned into these vectors and their expression induced in E.coli (38,39). Gene expression from the trp promoter can be simply induced by the presence of 3 β-indole acrylic acid which operates by a three-fold effect: (i) In inhibiting the binding of the repressor protein to the trp operater, (ii) in inhibiting the charging of $tRNA^{trp}$ thereby overcoming attenuation and (iii) by inhibition of the action of tryptophan synthetase, thus further reducing cellular levels of tryptophan. Using these plasmids proteins may be expressed in the mature form or fused to the initial residues of the trpE gene product, anthranilate synthetase. This seemed therefore a suitable means by which to express the synthetic gene in E.coli and the sequence was therefore designed to be inserted into these plasmids.

1.5.3 Peptide fusions

Protein fusions, in which DNA sequences of two host peptides or of host and foreign peptides are linked have been widely applied in the study of gene expression in prokaryotes (37,40,47). They have also proved useful in the examination of protein secretion and in the stabilisation of foreign peptides (20,41). Prior to rapid gene synthesis and site-directed mutagenesis the expression of fused gene products has often avoided the necessity for extremely difficult and time-consuming plasmid manipulations and has permitted more simple detection and assay of the expressed

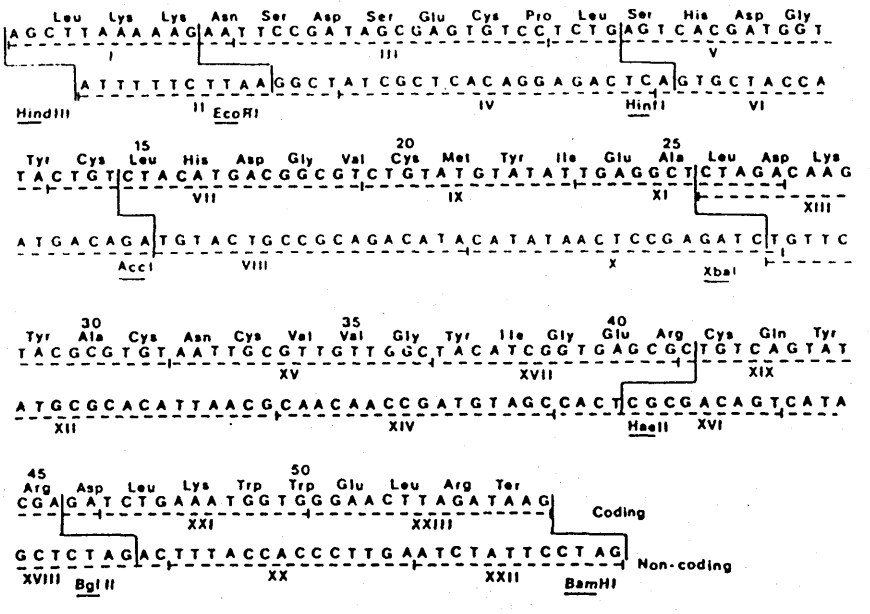


Figure 2. The synthetic β-urogastrone gene.

The synthetic β-urogastrone gene, showing the nucleotide and amino-acid sequence, the individual oligonucleotides, and the six base-pair restriction enzyme sites.

proteins. To facilitate studies upon the efficiency of gene expression and upon the stability of the β-urogastrone peptide, the synthetic gene was therefore designed to be capable of expression either as a "mature" peptide or as a peptide fusion. The design enabled fusion to the initial residues of the trpE gene product, anthranilate synthetase in the expression vector. Since β-urogastrone is known to be resistant to internal cleavage by trypsin, a linker region was designed which incorporated a trypsin labile site immediately preceding the first residue of the peptide. Subsequent cleavage of the fusion at this point yields the mature peptide. Similarly, restriction sites were included close to the C-terminal coding region of the gene to enable potential fusions to peptides such as β-Galactosidase. This provides a more simple assay (42) and easier screening of recombinants using the chromogenic substrate X-gal which is cleaved by β-galactosidase releasing a blue indolyl derivative.

The resultant gene was a 174 base-pair duplex flanked by unique Hind III and Bam HI restriction endonuclease sites. It encoded the 53 amino acids of β-urogastrone immediately followed by a TAA translation termination codon and preceded by the trypsin labile dipeptide LYS-LYS. This is depicted in Figure 2. The Hind III site allowed ligation to the Hind III site within the modified trpE gene, the unique Bam HI site facilitated the addition, removal or replacement of sequences of DNA or RNA downstream of the gene and a unique Eco RI site permitted the removal of DNA prior to the first nucleotide of the coding sequence. This enabled the expression of the mature peptide using specific linkers incorporating sequences known to function as ribosome binding sites. The synthetic gene thereby not only avoided many technical obstacles normally encountered during the cloning of a human gene but provided the potential opportunity for considerable studies upon the regulation of

the expression of a human peptide in heterologous hosts.

1.6 Design of a gene expression system in B.subtilis

1.6.1 Bacilli as hosts for foreign gene expression.

Since E.coli is the best genetically and biochemically characterised prokaryotic organism, it is not surprising that most experiments to date involving gene cloning have been carried out using specially constructed strains of E.coli as the host and derivatives of its natural plasmids and phages as vectors for cloned genes. The most widely used systems employ E.coli K12 with ColE1 derived plasmids such as pBR322 (43a and b) and derivatives of phage λ and M13 (44,45,46). In such systems well characterised regulatory regions of the E.coli chromosome such as those of the tryptophan and lactose operons as well as λ Pl and Pr (45) have been employed in the expression of plasmid borne genes. Nevertheless, there are a number of good reasons for the pursuit of efficient and versatile plasmids and gene expression systems in other prokaryote hosts, especially the bacilli and of Bacillus subtilis in particular.

Primarily, being the best biochemically and genetically studied gram-positive prokaryote Bacillus subtilis provides an obvious opportunity for comparisons of genome organisation and regulation with that of E.coli as well as the study of processes such as sporulation which cannot be examined using E.coli.

In addition, bacilli have for many years been very important organisms in industrial fermentation, in the production of amino acids, vitamins and enzymes. Since many biosynthetic genes of E.coli have already been isolated using molecular cloning it is quite possible that the cloning and manipulation of tightly regulated genes may also

be of use in the study of these processes in bacilli and could lead to improvements in the overall yields of fermentations. Indeed considering the increasing industrial interest in the use of micro-organisms for the production of enzymes and hormones in general, B.subtilis may emerge due to its non-pathogenicity and ability to secrete large quantities of particular proteins (48), as an altogether more preferable organism to E.coli for the large scale production of genetically engineered proteins. Such industrial applications of molecular cloning in B.subtilis depend on a rapid increase in the understanding of the molecular genetics of the organism to the level of that currently determined for E.coli.

However, no such ubiquitous systems as those based upon the E.coli trp or lac promoters, exist as yet to study gene expression in B.subtilis, due to the fact that far fewer groups have applied molecular cloning to the study of this organism than to that of E.coli. This in due in part to the better biochemical and genetical understanding of E.coli than of B.subtilis, and also to the fact that E.coli has proven to be a convenient and versatile host for many gene manipulation experiments whilst B.subtilis is a more difficult organism to manipulate using recombinant DNA technology. In addition, intracellular proteolysis of foreign peptides has often been a problem in B.subtilis (49). This can be at least partially overcome however by the use of protease deficient strains although these were not available during the course of this work. Development of efficient protein secretion systems for foreign peptides might also overcome this problem and reduce proteolysis of peptides to a degree below that observed in E.coli. Very recently however the development of molecular cloning and gene expression systems for use in B.subtilis and other bacilli has become a subject of increasing interest.

A number of groups have now developed plasmids for the cloning of genes in B.subtilis (50,51,52) and have used phage (53) and chromosomal (54) promoters for gene expression. This interest in the cloning and sequencing of a number of genes of the B.subtilis chromosome and those of its phage being greatly stimulated not by potential industrial applications but by the regulatory comparisons so far drawn with E.coli which have highlighted intriguing differences in the methods by which the organisms regulate gene expression.

1.6.2 Shuttle vectors

Many of the plasmids which have been constructed for the purpose of expressing cloned genes in B.subtilis, including those used in this work, have proven to be useful in demonstrating these regulatory differences and in providing ease of technical manipulation. This has been achieved by constructing these plasmids as hybrids containing two replicons, one capable of replication in E.coli and the other in B.subtilis, each carrying selectable antibiotic resistance markers (50,55,56). Such shuttle vectors are subsequently capable of propagation in either host, enabling studies on the expression of genes cloned into the plasmid in either organism. The principal advantage of these plasmids is that in addition to providing controlled comparisons of gene expression they permit much of the difficult manipulation often involved in plasmid construction to be carried out in the more easily handled E.coli host. Subsequent transformation of competent B.subtilis can then be carried out using high concentrations of purified and characterised plasmid.

These vectors however are not entirely free of problems. Often they have displayed structural instability in B.subtilis (57,58) and sometimes in E.coli (59). As yet none have been engineered extensively enough to be suitable for widespread application. There are indications nevertheless that the use of site-directed mutagenesis and rapid oligonucleotide synthesis may in future yield more versatile vectors, perhaps possessing multiple unique cloning sites, in the manner of the pUC (46,60) vectors of E.coli. The problem of plasmid structural instability in B.subtilis however, remains relatively unresolved. Although recombinational deficient (recE) hosts can be used to reduce these occurences (61), some recE independent rearrangements still occur (58). The mechanism of this event is still unclear.

In this work, careful consideration has been given to these aspects of the use of shuttle vectors and a shuttle system has been designed primarily to study the expression of β-urogastrone in both hosts, but which should prove versatile for more widespread application and which minimises the problems of vector instability. The regulatory elements of gene expression have been incorporated as easily replacable units and unlike any other reported shuttle vectors, the entire E.coli replicon, in this case pAT153 is removed prior to transformation of B.subtilis. This reduces the likelyhood of plasmid rearrangement and facilitates oligomerisation of the plasmid, necessary for uptake by competent B.subtilis cells (62). The series of plasmids derived from the parent vector has been used to examine the expression of the synthetic gene in B.subtilis and in E.coli.

The recent nucleotide sequencing of a number of cloned genes of the B.subtilis chromosome has begun to explain a perplexing contrast in the expression of genes in B.subtilis and in E.coli. That is: the observation that although B.subtilis genes are often expressed with reasonable efficiency in E.coli, normally E.coli genes are not expressed in B.subtilis. This has been shown to be due to regulatory differences between the two organisms occuring both at the level of transcription and translation. Although the most directly significant block to the expression of E.coli genes occurs at the level of translation in B.subtilis, it is the differences in transcriptional regulation which have been most intensely investigated for their role in the development of the process of sporulation. There are some indications from these studies why many E.coli genes are poorly or not at all transcribed in B.subtilis.

1.6.3 Contrasts of transcriptional regulation in E.coli and B.subtilis

As in E.coli, the RNA polymerase of B.subtilis consists of a "core" molecule comprising three proteins, α,β and β', in the ratio 2:1:1. Associated with this core molecule is another peptide known as sigma (σ) which is responsible for conferring specificity upon the polymerase to recognise and to initiate transcription at particular promoter sequences in the genome. In E.coli only one form of the σ peptide has been detected, a 70 kDal m.wt. peptide (σ^{70}) (63), although very recently there have been reports of a sigma like protein being induced during the heat shock response (64). However in B.subtilis the core polymerase can interact with a number of sigma-like proteins. The principal form of σ factor detected in vegetatively growing B.subtilis cells is a 43 kDal m.wt. peptide designated σ^{43} (65). This is very homologous to the sigma factor of E.coli. However in

the response of B.subtilis to nutrient depletion in which the bacterium enters a developmental pathway culminating in the development of the endospore at least five temporally defined classes of genes are expressed. The expression of these genes can be correlated with the appearance of sigma-like proteins which bind to and modify the promoter specificity of core RNA polymerase (66). RNA polymerase modified by the presence of a σ factor other than σ^{43} recognises and selectively transcribes different classes of promoter sequences (67). A similar form of transcriptional regulation occurs during the development of certain phage infections of B.subtilis in which the phage genome encodes novel σ factors (68) which temporally regulate its expression. The core polymerase of B.subtilis, differs however in a further respect from that of E.coli, in that it associates with an additional peptide known as the delta (69) factor. This 21 kDal m.wt. peptide appears to enhance the interaction of polymerase with promoter regions without affecting the specificity of promoter sub-class recognition.

As yet insufficient genes of the B.subtilis chromosome have been sequenced to allow extensive comparisons of the primary sequence homologies within and between these promoter sub-classes. Nevertheless some homologies have been detected and consensus sequences assigned, particularly for promoters recognised by the major form of the RNA polymerase, that containing σ^{43}. Comparisons between these and E.coli promoter sequences show that both display the same consensus hexamers of DNA sequence located at -10 bases and -35 bases from the transcription initiation point (70). However, the individual sequences of B.subtilis σ^{43} promoters display far less variation from these conserved hexamers than do their E.coli counterparts (70). This apparent difference in stringency may simply reflect the far fewer numbers of B.subtilis promoter sequences yet determined but seems to indicate a greater conservation of

DNA structure which may be required for recognition and initiation of transcription by the RNA polymerase. Certainly it is consistent with the observation that σ^{43} promoter sequences from B.subtilis are more frequently transcribed efficiently in E.coli than are E.coli promoters in B.subtilis. The use of random and site-directed mutagenesis has already clearly demonstrated that in E.coli certain nucleotides within these conserved regions are essential for efficient transcription from particular promoters (30). Similar studies in B.subtilis are likely to show whether or not B.subtilis is more sensitive to particular changes in promoter DNA sequence and may demonstrate mutations which can be made to promoter regions of E.coli in order to increase the efficiency by which they are transcribed in B.subtilis.

1.6.4 Choice of promoter

This work describes the cloning, sequence analysis and characterisation of a promoter region derived from the B.subtilis bacteriophage SPP1. Although a virulent phage, SPP1 infection does permit host DNA, RNA and protein synthesis for some time after infection, a feature more characteristic of infection by temperate phage. The linear phage genome is approximately 43 kb in length with many phage genes and proteins identified (71,72). The promoter isolated in this work has previously been shown to be one of five major early promoters on the SPP1 genome (73). It is transcribed immediately after infection and continues to be active throughout infection, either in the presence or absence of chloramphenicol (74). This is indicative of transcription by unmodified host polymerase but does not rule out the possibility of any additional transcription by modified forms of polymerase, as has been shown to be the case for phage SPO1 (75). The existence of SPP1 specified RNA polymerase modifying factors has not been demonstrated

although late phage genes do require de novo protein synthesis for transcription (74). This promoter has also been shown to initiate transcription *in vitro* more strongly than any of the other early SPP1 phage promoters, using RNA polymerase from both *B.subtilis* and *E.coli* (73). The genome of phage SPP1 has been restriction mapped in detail and the promoter shown to reside on a 1.4 kB *EcoRI* restriction endonuclease fragment. In addition the approximate position of transcription initiation within this fragment and the direction of transcription have previously been shown (73) and are verified in this work. Recent publication (76) of nucleotide sequence from this fragment differs substantially from that determined in this work as discussed in Chapter 5.

This promoter region therefore seemed appropriate for DNA sequence analysis and determination of the RNA polymerase binding sites and transcription start point in both *B.subtilis* and *E.coli*. It also provided an efficient transcription regulatory region to combine with specific regions of translational regulation for the study and comparison of heterologous gene expression in both hosts.

1.6.5 *Contrasts of translational regulation in E.coli and B.subtilis*

The role of specific nucleic acid sequences in the regulation of translation has similarly been examined and compared in *E.coli* and *B.subtilis*. As before, the vast majority of sequence data which has been determined to date is of the *E.coli* chromosome or of coliphage. Nevertheless, the DNA sequences encoding ribosome binding regions of *B.subtilis* mRNA species have indicated possible explanations for the non-translation of *E.coli* derived messenger RNA's in *B.subtilis*. This is based on the nature of the various components of the classically defined ribosome binding site. In both *E.coli* and *B.subtilis*, ribosome binding sites are

regions of an mRNA sequence at which ribosomes will specifically bind and initiate protein synthesis. They comprise a region of base complementarity to the 16S subunit of ribosomal RNA, the Shine-Dalgarno (S.D.) box (77), followed at an appropriate distance by a methionine initiator codon, usually ATG where protein synthesis initiates. The degree of complementarity to the ribosomal RNA and the number and content of the intervening bases between this region and the initiation codon are crucial for efficient initiation of translation in either organism (32). However, these paramaters appear to be more influential on efficient translation initiation in B.subtilis than they are in E.coli since the ribosome binding sites of B.subtilis and other gram-positive derived mRNA's display consistently higher complementarity to the 16S rRNA subunit than do mRNA's derived from E.coli (70). The calculated free energies of base-pairing interacton of these regions give an average value of -17 kcal for B.subtilis and -11 kcal for E.coli. This has led to the hypothesis that B.subtilis requires a stronger interaction of these regions to form an efficient translation initiation complex (78). This is not generally reflected in the other components of the translational machinery since most of these can be successfully interchanged between these organisms and also between E.coli and other gram-positive organisms such as Bacillus stearothermophilus. E.coli elongation factors EF-G and EF-Tu can complement temperature sensitive lesions in B.subtilis protein synthesis systems and vice versa (79). B.stearothermophilus 16S rRNA can be used to reconstitute 30S ribosomes with E.coli 30S ribosomal proteins and vice versa and E.coli 5S and 23S rRNA can likewise be used to reconstitute 50S ribosomes in the presence of B.stearothermophilus 50S ribosomal proteins (80). A good illustration of the molecular basis of interchangeability of some components is observed when the N-terminal amino-acid sequences of all the 30S ribosomal proteins and several of

the 50S proteins from the three organisms are compared (81). In these there is 84% correlation between the two bacilli and 50% between B.stearothermophilus and E.coli. Nevertheless, certain ribosomal proteins have been implicated in the translational differences observed between E.coli and B.subtilis in ways consistent with the differences now apparent in their mRNA ribosome binding. The E.coli ribosomal protein S1 has been shown to confer upon B.stearothermophilus the ability to translate the coliphage R17 hitherto untranslated replicase and coat proteins (82). Protein S1 is involved in the binding of mRNA to the ribosome possibly by stabilising the interaction of the S.D. region and the 16S rRNA. The implication is that this or other salt dissociable factors are required for translation of mRNA's from gram negative but not gram positive organisms. This is consistent with the stronger complementarity observed between the B.subtilis mRNA's and the 3' end of the ribosomal 16S RNA suggesting that this stronger interaction precludes the need for such factors required by E.coli. If as hypothesised, B.subtilis does require a more stringent base pairing for efficient translation then most E.coli mRNA's would not base-pair sufficiently to be translated in B.subtilis and this could explain the translational discrimination seen against most E.coli genes.

Considering this fact it is possible to design putative DNA sequences which when transcribed could act as ribosome binding sites both in E.coli and in B.subtilis. It should also be possible to evaluate in both hosts the effect of variations in the complementarity of synthetic ribosome binding sites to the 3' end of the 16S RNA subunit. The secondary structure of messenger RNA has also been shown both in E.coli (33) and in B.subtilis (60) to be crucially involved in translational efficiency. This must also be considered in conjunction with RNA primary strucure. However

using sophisticated computer assistance to predict the effects on secondary strucure of changes in primary structure of RNA as with DNA in gene design, it is possible to avoid the introduction of stable duplex structures which have been shown to exert an effect on the regulation of translation.

Taking this into account, expression systems were designed, based on the shuttle plasmid system derived from vectors pAT153 and pUB110. These incorporated the promoter from phage SPP1 with synthetic ribosome binding sites designed to effect efficient translation of β-urogastrone in both B.subtilis and E.coli. The vectors were intended to facilitate technical manipulation, to enable the expression of different genes and to allow the simple replacement of the synthetic ribosome binding region. Subsequent oligonucleotide directed mutagenesis was used to examine the effect of translation initiation from the ribosome binding region immediately downstream of the SPP1 promoter.

This then provided a versatile system to study the expression of β-urogastrone and subsequently other peptides in Bacillus subtilis. It also provided a means to examine in E.coli the initiation and efficiency of transcription from a promoter derived from a B.subtilis phage and to evaluate in both hosts the use of synthetic ribosome binding sites in gene expression.

CHAPTER 2

MATERIALS AND METHODS

2.1 Chemicals

General reagents used in this work were supplied by BDH chemicals. Amino acids, carbon sources and specialist reagents were from Sigma Biochemicals. Ultra pure acrylamide and phenol were supplied by Bethesda Research Laboratories. Sephacryl S300 was supplied by Sigma.

2.2 Radiochemicals

All radiochemicals were supplied by Amersham International.

2.3 Enzymes

All restriction endonucleases were supplied by New England Biolabs or Bethesda Research Laboratories (B.R.L.). E.coli DNA Polymerase (holoenzyme) used in blunt end generation, was from Biolabs. E.coli DNA Polymerase (Klenow fragment) used in DNA sequencing and site directed mutagenesis, was obtained from Boehringer Mannheim. Polynucleotide kinase was obtained from P.L. Laboratories and T4 DNA ligase, AMV reverse transcriptase and S1 nuclease from Biolabs. Ribonucleases A and T1 were obtained from Sigma.

2.4 Bacterial Strains

The following bacterial strains were used in this work.

E.coli K-12

HB101 - leu, pro, thi, lacY, recA, $r_k^- m_k^-$, rpsL, endA (83).

HW87 - araD139(ara-leu)del7697, (lacIPOZY)del74, galU, galK, rpsL, r_k^- ~~recA56~~, Tn10::Tc^r (84).

HW27 - thr1, leuB6, thi1, argE3, his4, proA2, lon1, lacY1, galK2, mtl1, xyl5, ara14, rpsL, tsx33, supE44 (84).

JM101 - (lac-pro)del, supE, thi1; F' $proA^+B^+$, $lacI^q$, lacZdelM15, tra del36 (85).

B.subtilis

3G18 - ade, met, trpC2 (86).

2.5 Plasmids and Phages

All plasmids used in this work are listed below.

pAT153		(87).
pUB110		(52).
pFF2	dual-replicon pAT153+pUB110.	(fig 21).
pFF4	pFF2 + SPP1 fragment 10.	(fig 21).
pFF401	pFF4 deleted in one Eco RI site.	(fig 21).
pFF501	pAT153 + β-urogastrone and R.B.S. I.	(fig 21).
pLF1	pAT153 + Eco RI to Bam HI fragment of pUB110.	(fig 11).
pUR1	C-terminal fragment of β-urogastrone cloned into pLF1.	(fig 12).
pUR2	pUR1 + Hind III site.	(fig 14).
pUR3	pUR2 + N-terminus of β-urogastrone.	(fig 15).
pWT221	pAT153 based version of pWT121	(38)
221-URO	β-urogastrone gene cloned into pWT221	(Chap.4)
pWT571		(39)
571-URO	β-urogastrone gene cloned into pWT571	(Chap.4)
pMC1403		(88)
pFF710	shuttle plasmid + SPP1 promoter R.B.S. I and β-urogastrone.	(fig 21).
pFF730	as pFF710 with R.B.S. II replacing R.B.S. I.	(Chap.5)
pFF700	as pFF710 with R.B.S. I deleted.	(Chap.5)
pFF740	as pFF700 with the SPP1 Pvu II to Eco RI fragment deleted.	(Chap.5)
pFF710R	as pFF710 with pAT153 in the reverse orientation.	(Chap.5)
pFF710L	as pFF710R with the Bam HI to Aha III fragment of pMC1403 (containing most of the lacZ gene of E.coli) inserted between the Bgl II site of β-urogastrone and the Nru I site of pAT153.	(Chap.5)

pFF730L as pFF710L with R.B.S. II replacing R.B.S. I. (Chap.5)
pFF700L as pFF710L with R.B.S. I deleted. (Chap.5)
pFF740L as pFF700L with the SPP1 Pvu II to Eco RI fragment deleted. (Chap.5)

Phage M13 mp8 and mp9 (45).

2.6 Media

Hallewell modified minimal media

1 x M9 salts, 2.5% glucose, 2.5% casamino-acids, 0.1mM $CaCl_2$, 1mM $MgCl_2$, B1 0.1mg/ml.

Luria broth (per litre): Difco bacto tryptone (10g), Bacto yeast extract (5g), NaCl (5g);adjusted to pH7.2.

2.7 Growth of bacteria

L-Agar plates were used for growth of E.coli and Tryptone Blood Agar Base plates for growth of B.subtilis.

For biological assay, the growth of cells expressing β-urogastrone from the trp promoter, was in Hallewell modified minimalmedia, supplemented where appropriate with tryptophan (40ug/ml).

In all other cases bacteria were grown in Luria broth supplemented where appropriate with antibiotics (E.coli - 100ug/ml ampicillin, B.subtilis - 5ug/ml kanamycin) both for plasmid preparation and for assay of β-urogastrone and β-galactosidase.

2.8 Synthesis of the oligonucleotides

All oligonucleotides used in this work were synthesised by the oligonucleotide synthesis group of the organic chemistry department at G.D.Searle.

Analytical grade reagents were used for the chemical synthesis of the oligodeoxynucleotide blocks. 4,4'-dimethoxytrityl,N-acyl protected nucleosides were obtained from Lancaster synthesis and further purified by silica gel chromatography. HPLC purification was performed on a Dupont 830 liquid chromatograph.

Each of the oligonucleotides of the synthetic β-urogastrone gene was synthesised by the solution phase phosphotriester procedure, coupled to rapid HPLC separation of final products from intermediates.

Starting materials for the syntheses were the 5'-O-(di-p-methoxytrityl)-3'-p-chlorophenol β-cyanoethyl phosphate derivatives of : N^6-benzoyladenosine, N^4-benzoylcytidine, N^2-isobutyryl guanosine and thymidine. 5'-dimethoxytrityl protecting groups were removed with 2% benzene sulphonic acid in chloroform/methanol (7:3 v/v) for 5 minutes at 0^oC and cyanoethyl protecting groups were removed by treatment with pyridine/triethylamine (1:1 v/v) for 2 hours at 40^oC.

Typically 80mg of the 5'-hydroxyl component was condensed with a 1.5 X molar excess of the 3'-phosphodiester component in anhydrous pyridine in the presence of a 3 X molar excess of the coupling agent 2,4,6-tri-isopropylbenzene sulphonyltetrazolide. After 1 hour or when the 5'-hydroxyl component was used up (as judged by silica TLC eluted with 10% methanol in chloroform) the reactions were quenched by the addition of 5% (w/v)

aqueous sodium bicarbonate and extracted with chloroform. The chloroform extracts were dried with sodium sulphate and chromatographed on 15-25um ODS silica using a solvent gradient from chloroform/methanol/water (2:6:3 v/v) to chloroform/methanol/water 2:6:0.5 v/v). Complete separation was obtained in under two hours and the fractions containing product extracted into chloroform, dried and evaporated under vacuum. Yields varied between 50-80%.

This process was repeated as necessary until at the completion of the synthesis, all the protecting groups were removed by sequential treatment with 0.5M N^1,N^1,N^2,N^2,-tetramethyl guanidinium 4-nitrobenzaldoximate in dioxan/water (1:1 v/v) at room temperature for 60 hours, concentrated ammonia at 70°C for 4 hours and 80% aqueous acetic acid at room temperature for 30 minutes. After evaporation, an aqueous solution of the oligonucleotide was extracted with diethyl ether and filtered. The product was purified by chromatography on a Partisil 10 SAX HPLC column, using a solvent gradient from 1M to 4M triethylammonium acetate pH 4.9 at 50°C. Sequence analysis was carried out using the mobility shift method of Wu et al (89).

Synthesis of the synthetic ribosome binding sites was carried out at a later date using solid phase phosphotriester chemistry (90). Oligonucleotides were then synthesised on a 2% cross-linked polystyrene support and syntheses carried out in a continuous flow column apparatus. Completed oligonucleotides were 5' ^{32}P phosphorylated, characterised and purified if necessary from 20% denaturing polyacrylamide gels. Primers used in the dideoxy chain termination sequencing of the SPP1 fragment were synthesised using an Applied Biosystems 380B DNA synthesiser.

2.9 Assembly of the β-urogastrone gene

2.9.1 Phosphorylation of oligonucleotides

Molar equivalents (approximately 2ug) of each of the 12 oligomers comprising the C-terminal half of the gene were phosphorylated using 4 units of T4 polynucleotide kinase in 20ul of a solution containing 50mM Tris-HCL (pH 7.8), 10mM $MgCl_2$, 0.25mM ATP (including 40uCi dried down ^{32}P labelled ATP) and 10mM BSH for 15 mins at 37°C. Due to problems of incorrect fragment annealing, larger quantities of some oligomers of the N-terminal half of the gene (up to 7ug) were phosphorylated and combined in molar ratios of 0.5:1. Following reaction the enzyme was inactivated by incubation at 90°C for 5 mins. Each oligonucleotide was then visualised by autoradiography following electrophoresis of 1ul of reaction mix in a total volume of 12ul, including 6% ficoll loading buffer, on a 20% polyacrylamide/7M urea gel.

2.9.2 Ligation of oligonucleotides

The remaining volumes of each phosphorylated oligomer were then combined as described in the results section and conditions adjusted for ligation. Ligation of oligomers was carried out using 6 units of T4 DNA ligase in a solution containing 50mM Tris pH 7.8, 10mM $MgCl_2$, 1mM ATP, 20mM DTT for 8 hours at 25°C. Ligation products were characterised by electrophoresis of 3ul aliquots on 20% native and denaturing TBE poly-acrylamide gels. Fragment sizes were accurately sized using a ladder of ligated 10bp long *Hind* III linkers run in parallel.

2.9.3 Gel Electrophoresis

All poly-acrylamide gels used other than in DNA sequencing were 30cm x 12cm x 1.5mm. They were prepared from 30% stock (29%-Acrylamide, 1% Bis-acrylamide) and run in buffer TBE (90mM Tris HCL pH 8.3, 90mM boric acid, 2.5mM

EDTA). Gels were polymerised by addition of 400ul of 10% Ammonium Persulphate and 30ul TMED to 100mls of acrylamide/TBE solution. Electrophoresis was typically carried out for 16 hours at 300V. Denaturing gels were as native gels but included 7M urea. Gels used in DNA sequencing were 30cm x 20cm x 0.8mm. They were prepared as 6% (5.8% acrylamide - 0.2%-Bis-acrylamide) stock solutions containing 1 x TBE and 7M urea. Gels were polymerised by addition of 300ul of 10% ammonium persulphate and 40ul TMED to 50mls of acrylamide/TBE/urea. Electrophoresis was carried out for 2-6 hours at a constant power of 40W, depending on the extent of sequencing required.

Agarose gels were run on flatbed systems of 20cm x 15cm or 10cm x 8cm submerged in buffer TBE. For visualisation or extraction of specific fragments less than 200 bp in length, 2% gels were used. All other gels were of 1% agarose. For extraction of fragments, low gelling temperature agarose was used. Marker fragments of phage λ DNA digested with Hind III and phage ϕX174 digested with Hae III were used on all gels. For fragment sizes see Appendix 2.

2.9.4 Extraction of assembly intermediates

Gel slices containing correctly ligated assembly intermediates were excised following superimposition of exposed autoradiographs onto native polyacrylamide gels. Extraction of DNA from gel slices was carried out by electroelution onto a 5mm column of DE52 DEAE cellulose placed between two GFC discs within a 2ml syringe. The gel slice was contained within a 1ml Gilson micropipette tip placed within the syringe and the syringe fixed within a tube gel electrophoresis apparatus. Maximum transfer of DNA to the DEAE cellulose (>90%) occurred after 1 hr of electrophoresis at 200V. The DE52 was then washed in buffer containing 2mM magnesium acetate, 0.02% w/v SDS and 20mM

EDTA and the DNA eluted in 2ml of the same buffer containing 1.1M NaCl. The solution was diluted three-fold and the DNA recovered by ethanol precipitation. Recovery from the DEAE cellulose varied between 75-90%.

2.10 Cloning

2.10.1 Restriction fragment isolation

All restriction endonucleases were used according to manufacturers' recommendations. Analytical agarose gels were 0.8%, 1%, or 1.5% and were run in 1 x TBE buffer. All restriction fragments used in the cloning of the urogastrone gene were isolated by electroelution from 1% TBE agarose gels using the same apparatus and procedure as for polyacrylamide gel elution. Restriction fragment recovery from polyacrylamide gels was by electro-elution.

Restriction fragment isolation in subsequent cloning procedures was carried out by extraction from low gelling temperature agarose. Fragments were visualised by long wave UV irradiation following staining by ethidium bromide (100ng/ml) and excised in as small a volume as possible (typically 100ul). The fragment was then melted at 65^{o}C in a total volume of 500ul in 300mM NaAc, 10mM Tris pH 8, 1mM EDTA. In all cases except isolation of the fragment for in vitro transcription, 5ug/ml carrier tRNA was also added. Agarose was then removed by three extractions with phenol (v/v 1:1), and traces of phenol removed by 2 extractions with ether (v/v 1:1). Fragments were then precipitated by addition of 2.5 volumes of ethanol and storage for 1hr at -20^{o}C and recovered by centrifugation for 10 minutes in an Eppendorf microfuge. Fragments were washed once in 70% ethanol and taken up in a volume of 20ul of water.

2.10.2 Fragment ligations

Ligation of fragments was carried out typically in a volume of 20ul containing 100mM Tris pH7.5, 10mM $MgCl_2$, 10mM β-mercaptoethanol, 2mM ATP. Ligations of fragments possessing complementary ends were carried out for 4 hrs at 25°C. Blunt end ligations were carried out overnight at 15°C.

2.10.2 Preparation of plasmid DNA

Plasmid DNA was initially prepared from E.coli by the cleared lysate method (39). This was later replaced by the alkali denaturation method of Birnbaum and Doly (91). For large scale preparations, (50 mls of cells), the use of ultracentrifugation was replaced by gel filtration. Final precipitates of DNA prepared by the alkali denaturation method were taken up in a total volume of 200ul containing 10mM Tris pH8, 1mM EDTA, 100ug/ml RNase A, 20ug/ml Rnase T1, and incubated for 30 mins at 37°C. To this was then added 20ul of 100% glycerol and the total volume loaded onto a 30cm x 15mm column of Sephacryl S300. Column buffer of 10mM Tris pH8, 1mM EDTA, 150mM NaCl was pumped at a flow rate of 2mls/min. Detection of nucleic acid was by an in-line LKB Uvicord detector linked to a Kipp and Zonen BD40 chart recorder. PLasmid DNA yields of 50-200 ug could typically be collected over a period of 30 mins in a volume of 5-7mls, entirely free of contaminating cellular RNA, and precipitated in 2.5 volumes of ethanol at -20°C.

Plasmid DNA from B.subtilis was also isolated by the cleared lysate method but no clearing spin was done. DNA was instead purified by two consecutive ultracentrifugations for 16 hours at 36K rpm in a Beckman Ti60 vertical rotor.

Small scale preparations of DNA (1ml), were carried out in E.coli by the alkali denaturation method employing only one phenol extraction step (v/v 1:1) and two ether extraction steps (v/v 2:1) prior to precipitation. Small scale preparations of B.subtilis DNA was done by the alkali denaturation method with the solution 1 incubation step carried out for 30 mins at 37°C.

2.10.4 Transformation

Transformation of all strains of E.coli was carried out according to the Ca^{2+} treatment method (92) and transformation of B.subtilis was by the method of Arwert and Venema (86). Preparation of all shuttle plasmids excepting those incorporating the lacZ fragment, for transformation of B.subtilis involved cleavage with Bam HI followed by 90°C incubaton for 10 minutes to destroy the enzyme, addition of 1/10th volume of 10x ligation buffer, 1ul ligase and ligation for 4 hours at 25°C. The entire mix was then used in the transformation. Shuttle plasmids incorporating the fragment of lacZ were linearised by cleavage with Bgl II and religated as before.

2.10.5 S1 nuclease and DNA polymerase

Treatment of DNA fragments with S1 nuclease and DNA polymerase was carried out using standard methodology.

2.11 DNA Sequencing

All DNA sequencing was carried out by the dideoxy chain termination method (93). Sequencing reactions were carried out in micro-titre wells. All primers were annealed to template DNA for 30mins at 55°C.

2.12 Site Directed Mutagenesis

Site directed mutagenesis of the SPP1 promoter region was carried out using the method of Zoller and Smith (94).

2.13 Oligonucleotide primed reverse transcription

2.13.1 RNA preparation

Total cell RNA was prepared from E.coli and B.subtilis as follows. A fresh colony was used to inoculate an overnight culture of cells, grown in L-broth under antibiotic selection. The following morning 50mls of cells in a 1 litre baffled shake flask were inoculated at $O.D._{600}$ of 0.01. Cells were grown to an $O.D._{600}$ of 0.5 wherupon 10 mls of cells were spun at 10,000 rpm for 1 minute. Cells were washed in 10 mls buffer A (10mM Tris pH 7.5, 5mM $MgCl_2$, 10mM KCl,), and re-pelleted for 1 minute. The supernatant fluid was removed and cells taken up in 2mls of buffer A plus 300ug/ml T4 lysozyme. Cells were immediately frozen in liquid nitrogen and then placed in a 65^{o}C bath for 3 minutes. A volume of 220ul of 10% SDS was then added at which point the solution became totally clear. To this solution was added 80ul of 3M NaAc pH 5.2 and an equal volume of phenol pre-heated to 65^{o}C. The mix was then vortexed for 20 seconds and spun for 5 minutes at 10,000 rpm. The aqueous layer was removed along with any interface and the process repeated until little or no interface was evident (typically 4 or 5 times). The aqueous layer was then extracted twice with an equal volume of ether. NaCl was then added to 2M and the nucleic acid precipitated by addition of 1.5 volumes of EtOH and storage at -20^{o}C for 30 minutes. The nucleic acid was then spun at 10,000 rpm for 10 minutes, dried and resuspended in 1ml H_2O. When dissolved NaCl was added to 3M and the solution incubated overnight at 4^{o}C to selectively precipitate the RNA. This was then spun as

before, washed in 70% EtOH 3x, dried and taken up in 200ul H_2O. This was then stored at -70°C. Yields were between 100-400ug from 10mls of cells.

2.13.2 Reverse transcription to identify transcription initiation point

Reverse transcription of messenger RNA prepared as above was carried out in a micro-titre well. Approximately 2ug OF RNA was annealed to approximately 2.5ng of oligonucleotide primer in a 10ul solution containing 10mM Tris pH7, 5mM $MgCl_2$, and allowed to anneal at 55°C for 30 minutes. To this was then added 1ul of dNTP mix (1mm each of GTP, CTP and TTP), 1ul of ^{35}S ATP, and 1ul of reverse transcriptase, and the mixture incubated for 30 minutes at 37°C. To this was then added 5ul of formamide dye mix (100% deionised formamide, 20mM EDTA, 0.03% XCFF and BPB) and the mixture heated to 80°C for 15 minutes. This was then electrophoresed on a normal sequencing gel along with appropriate sequence reactions as indicators.

2.14 Transcription and translation in vitro

In vitro transcription and translation of plasmids and purified fragment was carried out using a prokaryotic DNA directed translation kit (Amersham).

2.15 Plasmid copy number determination

Plasmid DNA was isolated from a cell pellet equivalent to 1ml of cells of $O.D._{600}$ 2.0, by extraction with unequilibrated phenol/chloroform. Plasmid DNA was then electrophoresed on a 1% TBE agarose gel, along with standard amounts of pAT153 DNA. Following ethidium bromide staining, a photograph of the gel was scanned using a Chromoscan 3 and

copy number determined from peak integral values.

2.16 Preparation of cells for biological assay

2.16.1 Cells containing trp promoter regulated β-urogastrone gene

Cultures of 10mls of Hallewell minimal media supplemented with L-tryptophan (40ug/ml) and ampicillin 100ug/ml, were inoculated from a glycerol stock of cells and grown overnight at 37°C. These were then used to inoculate 300ml cultures of Hallewell minimal media supplemented with ampicillin (100ug/ml) at an initial $O.D._{600}$ of 0.1 and grown at 37°C in 1l shake flasks. Growth was monitored by absorbance at $O.D._{600}$ until at approximately $O.D._{600}$-0.4 a pre-induction 1ml sample was taken for assay. Expression from the trp promoter was then induced by addition of β-IAA (final concentration 20ug/ml). Further 1ml samples for biological assay were removed at time intervals of 15, 60, 90, 120, 150, 180 and 240 minutes following induction. Cell samples were then prepared as folows.

1ml cell sample + 60ul (25% sucrose/50mM Tris pH7.4)
Kept on ice 5 minutes.
Addition of 20ul of 10mg/ml lysozyme in PBS.
Kept on ice 5 minutes.
Addition of 1ul PMSF (35mg/ml in EtOH) and 20ul 500mM EDTA pH 8.0.
Kept on ice 5 minutes.
Addition of 100ul of 0.6% Triton TX-100
Cells were then shaken vigorously at room temperature for 5 minutes and spun in an Eppendorf microfuge for 10 minutes. The supernatant volume of 200ul was removed and added to 1400ul of PBS. This was then dialysed over 24hrs at 4°C against PBS and then stored at -20°C prior to receptor binding or radio-immuno assay.

2.16.2 Cells containing SPP1 promoter regulated β-urogastrone gene

Cultures of 10mls of L.broth supplemented with appropriate antibiotics were inoculated from a fresh streak of E.coli or B.subtilis and grown overnight at 37°C. These were then used to inoculate 100 ml cultures of L.broth supplemented with antibiotics and grown at 37°C in 1l shake flasks. Growth was monitored at $O.D._{600}$ until at $O.D._{600}$ 0.3 the first 5ml sample was taken for assay. Further samples were taken at intervals of 15, 60, 90, 120, and 150 minutes. Samples were concentrated 10 fold in buffer containing 50mM Tris pH7.5, 1mM PMSF, and 5M urea and sonicated for 2 x 30 seconds on ice in a sonicator. Extracts were spun for 10 mins in a microfuge and the supernatant volumes of 500ul removed and dialysed overnight. Samples were then assayed by radio-immuno assay.

2.16.3 Cells containing urogastrone/β-galactosidase peptide fusions

Cultures of 10mls of L.broth supplemented with appropriate antibiotics were inoculated from a single colony from a fresh streak of cells and grown overnight at 37°C. These were then used to inoculate 50mls of antibiotic supplemented L.broth culture grown in 1l shake flasks. At $O.D._{600}$ the first 1ml sample was taken. Further samples were taken at intervals of 15, 45, 75, 135 and 195 minutes. All samples were concentrated twofold in buffer BB (42) and lysed by sonication for 2 x 30 seconds on ice. Samples were spun 10 mins in a microfuge and supernatant fractions removed for assay.

2.17 Assays

2.17.1 β-urogastrone receptor binding assay

Reagents

Cell Growth Medium:

RPMI 1640 (Bicarbonate buffered) + 10% foetal calf serum.

Diluent for Standards and Samples:

RPMI 1640 (Hepes buffered) + 0.1% Human serum albumin.

Cells:

Vero (African green monkey kidney)

Standards: 200, 100, 50 and 25 ng/ml.

^{125}I β-urogastrone:

After iodination and chromatography, aliquots of the peak fractions are diluted for testing in order to give about 5000cpm/10ul. At these dilutions the individual fractions are tested under assay conditons for specific and non-specific binding. Those giving maximum specific with minimum non-specific binding are pooled, re-checked in assay and stored at $-30^{o}C$.

Dilution of test samples

An estimate of the β-urogastrone concentration provides the dilution factor required to ensure that the result will fall within the limits of the standard curve. At least three replicates of each dilution are assayed.

Growth of monolayers

Vero cells are seeded into 24 well plates, 1.0ml/well at 400,000 cells/ml in RPMI 1640 + 10% feotal calf serum. Cells are incubated at 37°C in 5% CO_2 overnight.

Assay

Growth medium is removed and cells washed x3 with PBS. To each well is added 150ul diluent, 40ul of sample and 10ul of ^{125}I β-urogastrone. A complete set of standards is included in each plate. After 1 hr at room temperature the well contents are aspirated and washed x4 with PBS. 500ul of 0.1N NaOH is added to each well to lyse the cells and incubated at 37°C for at least 15 minutes. After mixing, 450ul of well contents are removed for counting. Concentrations of β-Urogastrone in the samples are determined using a standard curve prepared from the set of standards.

2.17.2 Radio-Immuno Assay

Reagents

Phophate Buffered Albumin (PBA) 0.05M:

40mM Na_2HPO_4, 11mM NaH_2PO_4, 0.1% NaN_3, 0.5% BSA.

Stored at 4°C.

Immunoprecipitin (Fixed Staphylococcus aureus cells)

Staphylococcus aureus (BRL) cells are spun at 3000 G for 10 mins and supernatant decanted. Cells are resuspended in an equal volume of PBA containing 10% w/v BSH and 3% w/v SDS and are mixed well. Cells are then heated to 95°C for 30 minutes and spun at 3000 G for 30 minutes. The supernatant is decanted and cells resuspended to 0.75 original volume in PBA and stored at 4°C. Cells are used at one hundredth dilution in the assay.

Antibody (Rabbit anti-β-urogastrone)

The dilution of antibody to be used in the assay is determined for each new batch using a series of antibody dilutions set up for assay against PBA to establish a value for non-specific antibody binding. The resulting percentage bindings are plotted against the dilutions and an appropriate antibody dilution selected.

^{125}I β-urogastrone

This is prepared as in the receptor binding assay.

Standards

Standards are prepared as batches from aliquots of a stock solution of purified β-urogastrone stored at -70°C. Serial dilutions are prepared to give a range of standard concentrations from 4ng/ml to 128 ng/ml. Each fresh batch of standards is checked by assay against current standards. Standards are stored in aliquots of 250ul at -30°C.

Assay

Samples are diluted in PBA, to ensure reults falling on the standard curve. 100ul of each sample is mixed with 100ul of antibody and 100ul of ^{125}I β-urogastrone. All samples are duplicated and a range of standards including a non-specific binding standard (PBA only) are included. Tubes are mixed thoroughly and incubated at room temperature for 18 hours. To this is then added 1.0ml of Staphylococcus aureus. Samples are then allowed to stand for at least 15 minutes and then spun at 2000 G for 20 minutes. The supernatant is then decanted and the cells drained thoroughly. Tubes are then counted. Averages are determined for duplicate samples and the value of the non-specific binding subtracted. Concentrations of β-Urogastrone in samples are then determined using the standard curve derived from the standard values.

2.17.3 β-galactosidase assay

β-galactosidase activity was measured as described by Miller (42). Plates for screening of β-galactosidase producing E.coli or B.subtilis colonies contained X-gal (40ug/ml) in L-broth or minimal media.

CHAPTER 3

ASSEMBLY AND CLONING OF THE GENE FOR HUMAN β-UROGASTRONE

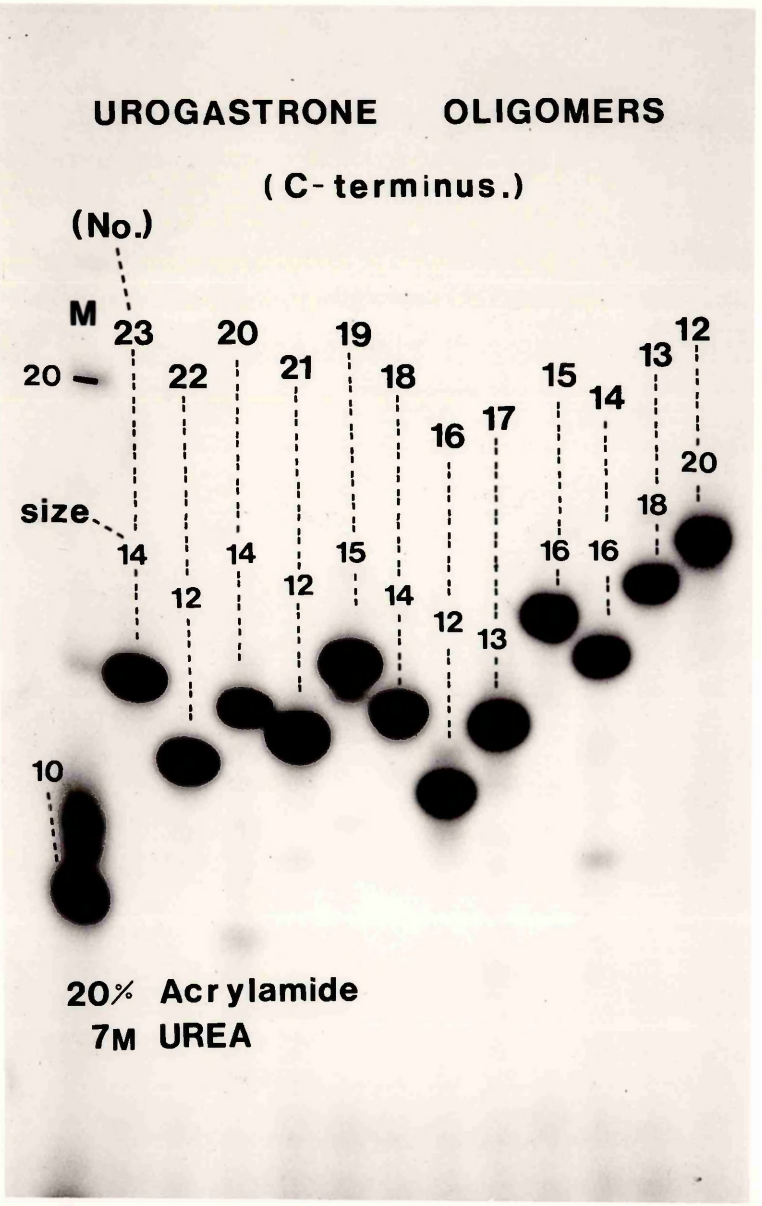


Figure 3. Oligomer identification.

The 12 oligomers of the C-terminal half of the synthetic β-urogastrone gene, visualised by auto-radiography following labelling with ^{32}P and electrophoresis on a 20% denaturing poly-acrylamide gel.

3.1 Assembly of the gene

3.1.1 Identification of the oligomers

Following deblocking, each of the oligomers to be used in the assembly of the β-Urogastrone gene possessed an hydroxyl group at both the 5′ and 3′ termini. This facilitated labelling of each fragment with ^{32}P ATP using T4 polynucleotide kinase and subsequent visualisation by autoradiography on a denaturing poly-acrylamide/7Murea gel.

Each oligomer could thus be identified as well as checked for integrity, purity and correct postitioning at each stage in the assembly of the gene.

For simplicity the gene was assembled and cloned in two almost equal parts encoding the N and C termini, the latter being carried out first and establishing a strategy for the former.

Figure 3 illustrates each of the oligomers of the C-terminal half, with for reference markers a ladder of ligated Hind III linkers of 10 bp in length, giving useful increments up to approximately 100 base pairs. Oligomers are referred to by their identification number in Figure 1.

Oligomers can be seen to electrophorese as would be expected from their relative lengths, but obviously with such small fragments the purine/pyrimidine ratio has a relevant influence on the mobility of individual oligomers. There is little evidence of contaminating fragments.

3.1.2 Pre-assembly

Prior to the full-scale assembly of the gene, two of the oligomers were tested to determine both their ability to correctly hybridise and ligate and under what conditions this was most favoured. These oligomers, numbers 12 and 13

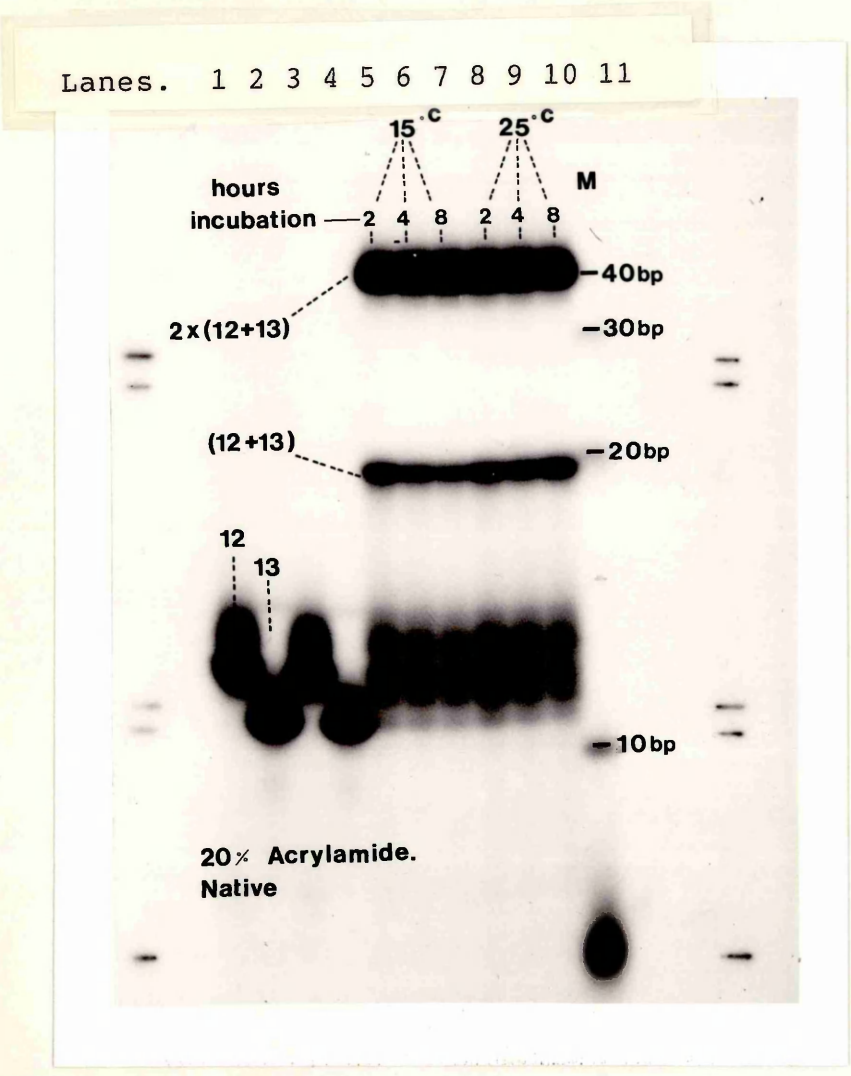


Figure 4. Dimerisation about the Xba I site.

Poly-acrylamide gel showing dimerisation occuring about the palindromic Xba I restriction site, following annealing and ligation of oligomers 12 and 13. Lanes 1 and 3 show oligomer 12 alone and Lanes 2 and 4 show oligomer 13 alone. The effect of duration and temperature of ligation are shown in lanes 5 to 10. Lane 11 shows ligated, ^{32}P labelled 10 bp Hind III linkers as markers.

form the XbaI site in the centre of the gene and upon hybridisation to one another would normally be expected to dimerise about the palindromic restriction site.

Figure 4 depicts this event, clearly illustrating the formation of the double-stranded DNA fragment 12+13 and the dimeric form 2x(12+13) whilst there is no evidence of incorrectly ligated material.

Variations in either, duration or temperature of ligation showed little effect on the products of assembly and it was decided that future ligations would be undertaken at 15^{o} C for 4 hours.

In order to demonstrate that the large band in **Figure 4** was indeed the dimerised product of oligomers 12 and 13, the products of this ligation were digested with Xba I and visualised on a native polyacrylamide gel as shown in **Figure 5**. The major band of 38 base pairs is cleaved to give the 12+13 double-stranded oligomer.

These experiments demonstrate that the oligomers can be simply built into larger molecular weight structures. By capitalising on the fact that structures will dimerise, can be subsequently cleaved around terminal restriction enzyme sites and the correctly assembled products thereby, easily identified and isolated from high resolution acrylamide gels.

3.1.3 First Stage Ligations

Of the twelve oligomers shown in **Figure 3**, eleven were now used in the first stage of the gene assembly. This was carried out in four separate reactions in which molar equivalents of overlapping oligomers were combined such that two particular fragments were ligated whilst held in correct alignment by hybridisation to the third, as illustrated in

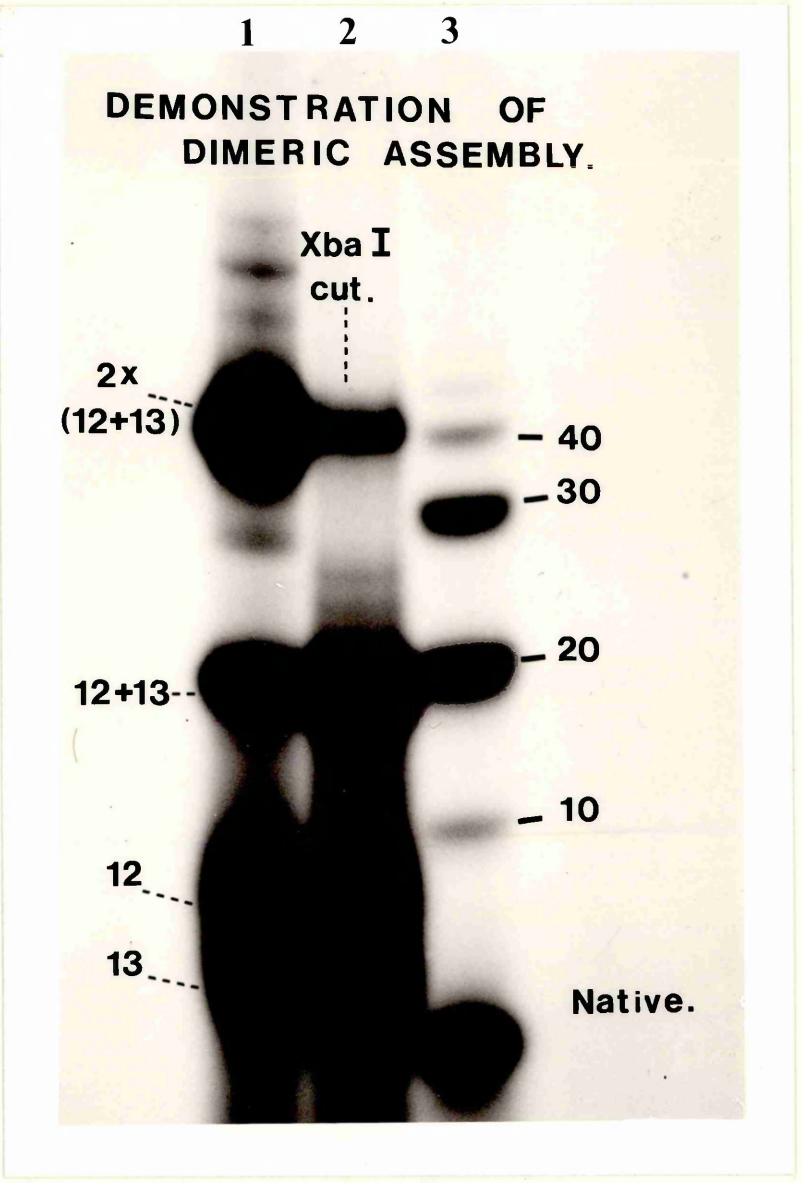


Figure 5. Cleavage of the dimeric assembly.

Lane 1 shows the dimerisation of oligomers 12 and 13, as in Figure 4. Lane 2 shows the effect of cleavage of this by Xba I. Lane 3 shows Hind III linker m.wt. markers.

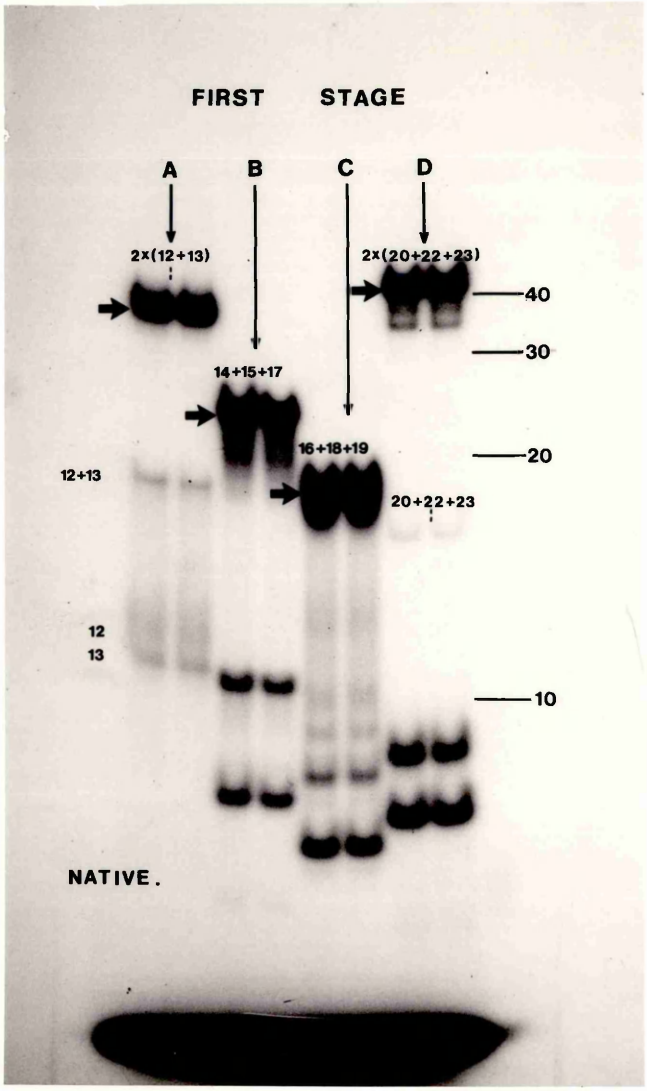


Figure 6. First stage ligations.

The first stage ligations showing the 4 initial sub-assemblies, A,B,C and D. Assembly A comprising oligomers 12 and 13 (Lanes 1 and 2); B comprising oligomers 14,15 and 17 (Lanes 3 and 4);C comprising oligomers 16,18 and 19 (Lanes 5 and 6);D comprising oligomers 20,22 and 23 (Lanes 7 and 8). Lane 9 Hind III m.wt. markers (not clearly visible).

Figure 1. The exception to this is in the ligation of oligomers 12 and 13 which was carried out as in the pre-assembly stage. The assembly of oligomers 22 and 23 will generate a terminal Bam HI site analagous to the Xba I site formed by 12 and 13 which likewise facilitates dimerisation.

Figure 6 shows each of these reactions run on a native polyacrylamide gel. In each case the major band is positioned where the correctly assembled tri-mer would be expected with unligated oligomers also clearly visible.

Each of these assembly intermediates was now removed in as small a gel slice as possible and extracted from the acrylamide by means of electro-elution. Each fragment thus isolated can be further verified by electrophoresis on an acrylamide/urea denaturing gel as demonstrated in **Figure 7**.

The importance of the isolation of correctly assembled intermediates from the native gel is highlighted by the appearance of faint bands running slightly below some of the products in **Figure 7**, especially evident in lane D. Whether or not this material would play a significant interfering role in the later stages of assembly is impossible to say. Clearly its presence along with unligated oligomers would be likely to reduce the yield of correctly assembled fragments and cloud interpretation of the fragments obtained. The exclusion of these fragments is therefore of high priority.

3.1.4 Second Stage Ligations

The products of the first stage ligations (A,B,C,D) were now subjected to further assembly following the logic outlined in **Figure 1**. Intermediates A and B were combined as were C and D along with the linking oligomer 21. Ligations were carried out as before and assembled products again separated on a 20% poly-acrylamide gel as shown in **Figure 8**.

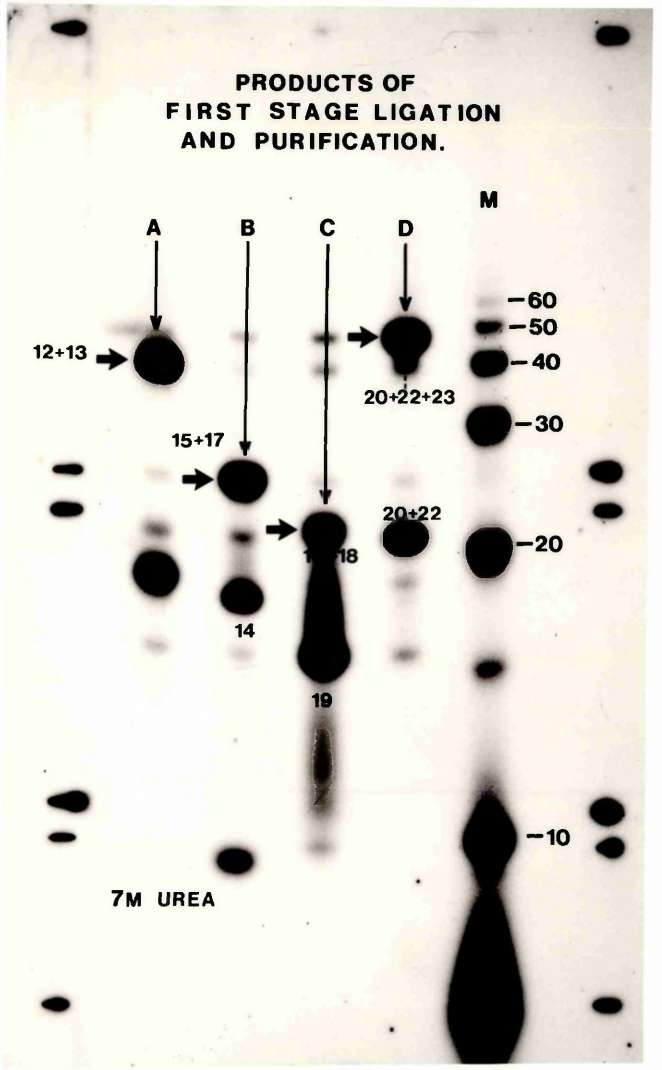


Figure 7. Extracted products of first stage.

The four purified sub-assemblies A(Lane 1); B(Lane 2); C(Lane 3) and D(Lane 4), run on a 7M urea, denaturing poly-acrylamide gel. Lane 5 Hind III m.wt. markers. Single stranded products of first stage ligations are indicated.

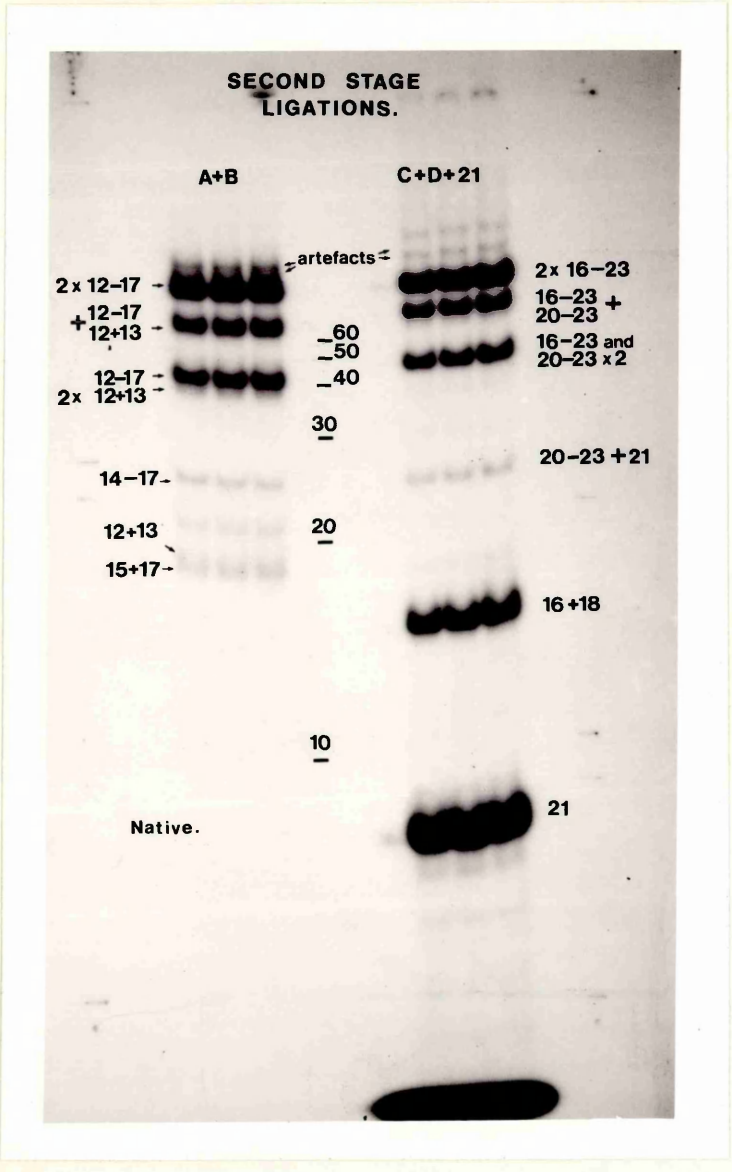


Figure 8. Second stage ligations.

Electrophoresis of second stage sub-assemblies on a 20% poly-acrylamide gel. Lanes 1-3,assembly A+B. Lanes 5-7,assembly C+D + oligomer 21. Lane 4, Hind III linkers as markers. Oligomers present in each assembly are indicated.

Bands of the expected size were obtained with in each case the major assembled band appearing to be the two desired fragments A+B and C+21+D. These bands cannot be accurately sized using the ladder of molecular weight markers but are determined as being the correct fragments based on their relative mobility to the other known fragments and upon the observation that they are cleaved as expected by the respective restriction enzymes Xba I and Bam HI as illustrated in **Figure 9**.

In each case the bands representing fragments dimerised about the terminal restriction sites are cleaved to give the expected lower molecular weight products such that in lane A+B the 2x(12-17) and (12-17+12-13) bands are cleaved giving fragments 12-17 and 12-13. Similarly in lane C+D+21 the 2x(16-23) and (16-23+20-23) bands are cleaved to give the fragments 16-23 and 20-23. Unfortunately in the latter case the material loaded on to the gel following cleavage was at lower concentration to that prior to cleavage making the bands more difficult to see. Nevertheless this represents good evidence that the two major products in **Figure 8** are indeed the correctly assembled fragments of the gene. These fragments were now recovered taking care to avoid minor contaminants which are evident running slightly above the major bands.

3.1.5 Final Stage Assembly

The Xba I to Bam HI fragment representing the right half or C-terminus of the gene was now arrived at by ligation of the two fragments thus derived from the second stage of assembly.

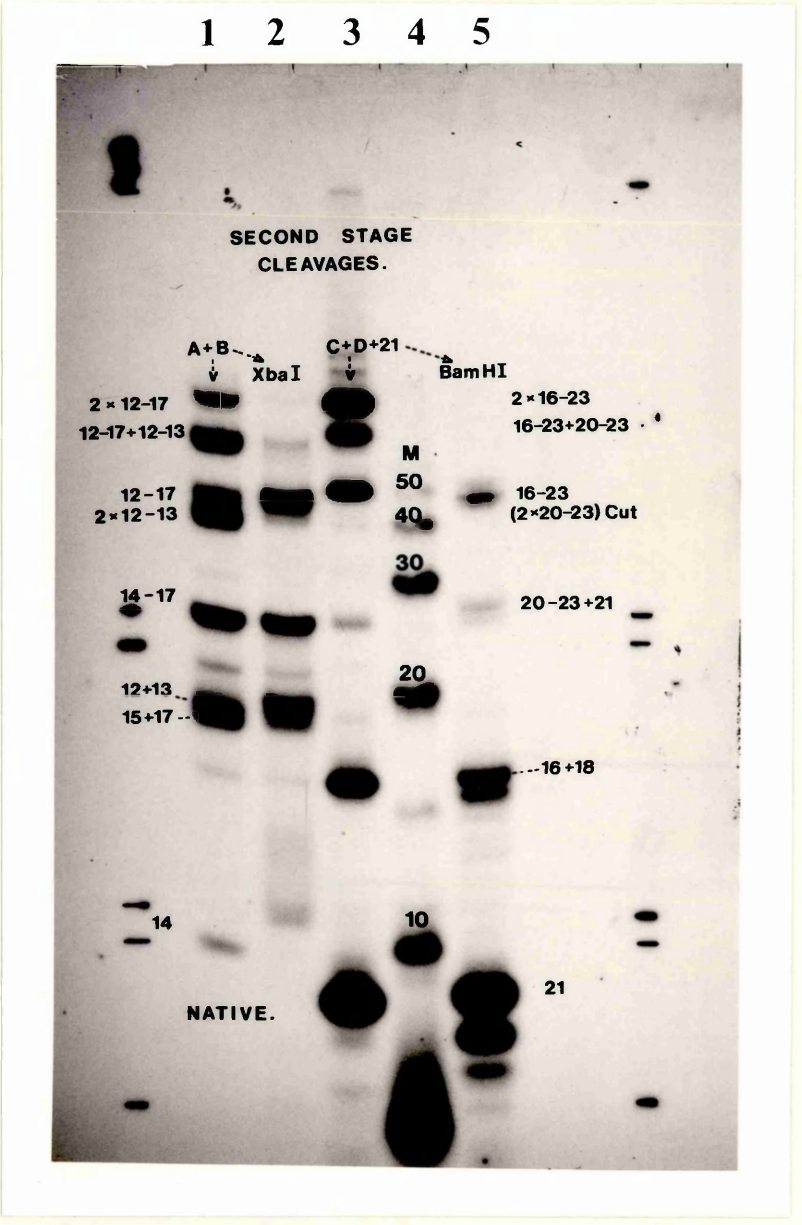


Figure 9. Cleavage of second stage ligations.

Electrophoresis on a 20% poly-acrylamide gel of fragments generated upon cleavage of second stage sub-assemblies. Lane 1, sub-assembly A+B; Lane 2, cleavage of sub-assembly A+B by Xba I; Lane 3, sub-assembly C+D + oligomer 21; Lane 5, cleavage of sub-assembly C+D + oligomer 21 by Bam HI. Lane 4, Hind III linkers as markers.

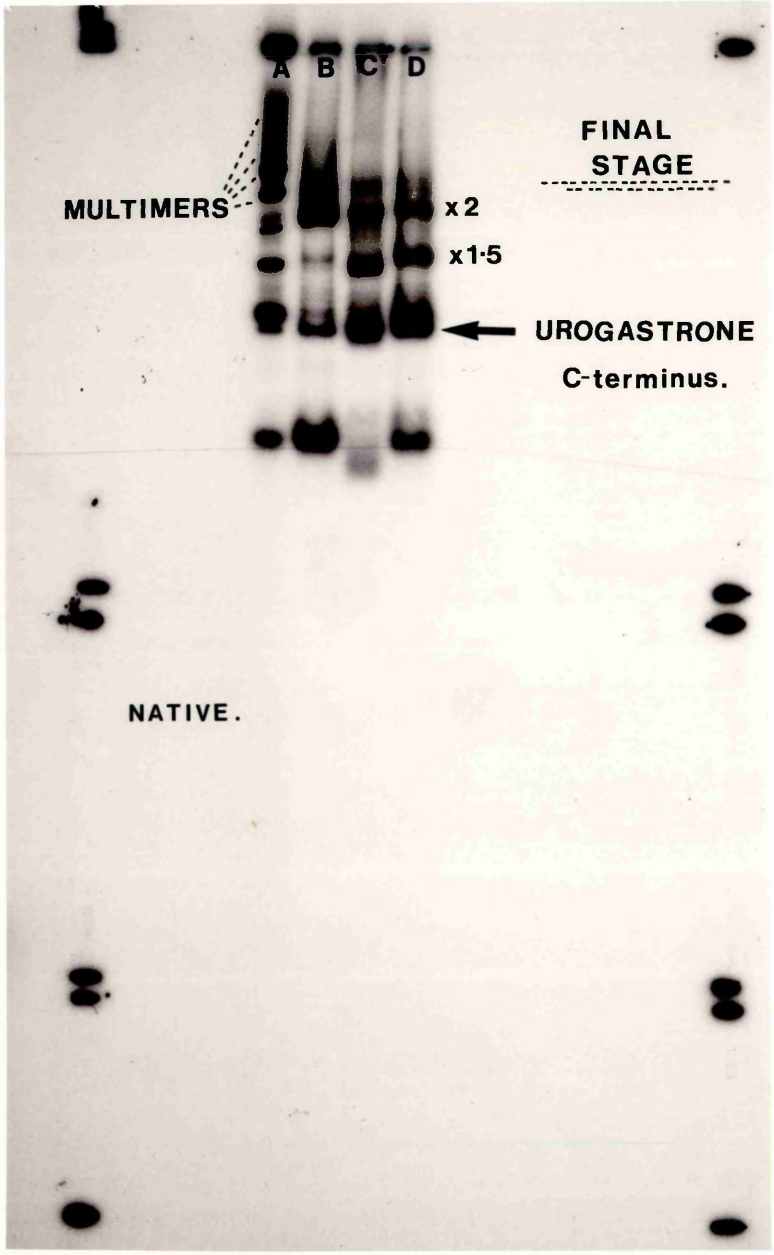


Figure 10. Final stage assembly.

Fragments generated in the final assembly stage of the C-terminal half of the gene. Lane A, shows the result of ligation of sub-assemblies (A+B) + (C+D +oligomer 21). Concatameric multimers ligated about the terminal Xba I and Bam HI restriction sites are indicated. Lane B, shows Xba I cleavage, and Lane C, Bam HI cleavage of these fragments. Lane D shows double cleavage with Xba I and Bam HI. The fragment most likely to represent the C-terminal fragment of the gene, and a possible dimer and intermediate structure, are also indicated.

Following ligation of these two fragments, a series of fragments is expected with incremental size representing the concatenation of the dimers about the terminal restriction enzyme sites. Cleavage with either of these enzymes should yield molecules dimerised about the opposite site as well as some monomeric molecules. Simultaneous cleavage with both enzymes would be expected to yield primarily the Xba I to Bam HI fragment. This can be seen to be the case in **Figure 10**.

Lane A represents the concatemers formed by ligation of the two fragments isolated in **Figure 8**. In lane B these fragments are cleaved with Xba I and in lane C with Bam HI. Lane D shows the result of cleavage with both enzymes to give what appears to be the desired fragment of the Urogastrone gene. This fragment was isolated as before and cloned into the vector pLF1.

The importance of gel purification at each stage of the process and particularly at this final stage in the series of oligonucleotide assembly reactions is now very apparent, since there is only a small percentage of the original oligomers present in what is now believed to be the correctly constructed fragment of the gene. The presence of excess quantities of monomeric or partially assembled oligomers would severely interfere with the cloning step, into the plasmid vector. This is especially evident with the terminal fragments incorporating the restriction sites which would be very likely to saturate the ends of the vector preventing the incorporation of the C-terminus of the gene and possibly necessitating considerable screening of clones.

As it was, of 8 clones screened by restriction cleavage with Xba I and Bam HI each proved to contain a fragment of approximately 88 base pairs, indicative of the C-terminus of the β-urogastrone gene.

3.2 Cloning of the synthetic gene

3.2.1 Construction of the cloning vector

Subsequent to the isolation of the desired fragment of DNA, a plasmid vector was chosen for the cloning step. Most work to date involving the cloning of fragments of DNA in E.coli has made use of the ColE1 derived vector pBR322 since it is a stable and versatile multi-copy plasmid with few occurences of structural instability. The plasmid however used in all the cloning stages of the β-urogastrone gene was pAT153, a copy number mutant of pBR322 which offered several advantages over the parent plasmid.

Primarily, pAT153 is deleted in the region of the DNA containing the origin of transfer, present in pBR322 and unlike the former is consequently non-mobilisable. This was very relevant at the time of cloning since this work was subject to the guidelines of G.M.A.G. and carried out under appropriate conditions subject to biological safety categorisation. Using a vector which was incapable of conjugal transfer contributed to the assignment of category 1 containment to this work, which greatly assisted in the technical ease of manipulations. Furthermore the elevated copy number of pAT153 provided additional advantages. Larger quantities of plasmid DNA could be more readily isolated and genes cloned in the vector, being present in higher numbers of copies may be more highly expressed and therefore more easily detected, characterised and subsequently purified.

However, pAT153 did not possess a unique restriction cleavage site for XbaI, necessary in the cloning of the C-terminus of the gene. This site was therefore introduced by means of subcloning onto pAT153 a fragment of approximately 840 base pairs from a different plasmid,

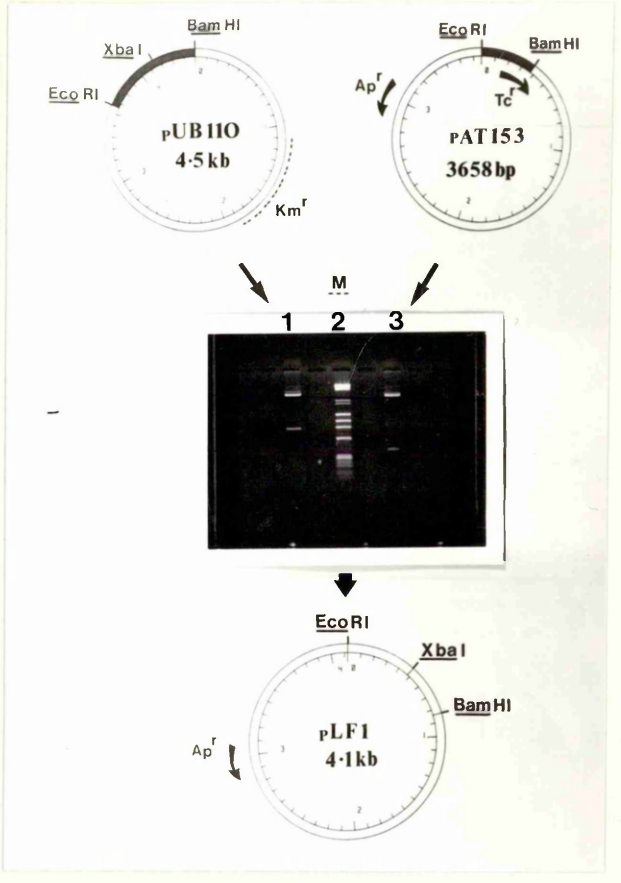


Figure 11. Generation of plasmid pLF1.

Following cleavage by Eco RI and Bam HI and agarose gel electrophoresis, the small fragment of pUB110 (Lane 1) was isolated along with the large fragment of pAT153 (Lane 2). These were ligated to generate plasmid pLF1. Lane 3 contains m.wt. markers consisting of a Hind III digest of phage lambda and a Hae III digest of phage ϕX174. (See appendix 2 for fragment sizes).

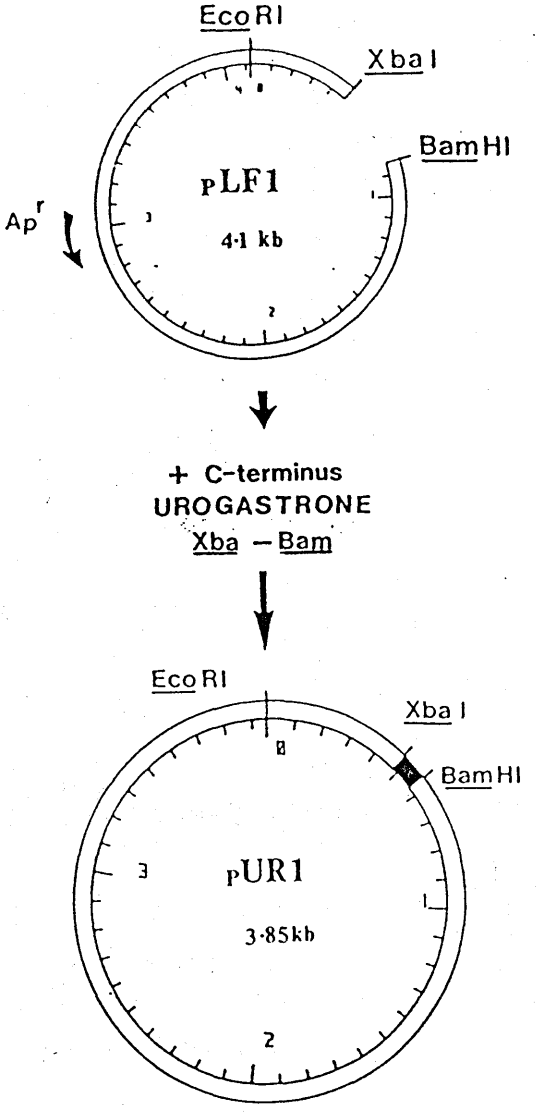


Figure 12. Cloning of the C-terminal fragment of the β-urogastrone gene.

Following cleavage with Xba I and Bam HI the large fragment of pLF1 was isolated as shown. This was then ligated to the isolated fragment of the β-urogastrone gene, giving plasmid pUR1.

pUB110 as shown in **Figure 11**. Plasmid pUB110 was originally isolated from Staphylococcus aureus and is as yet relatively poorly characterised. A partial restriction map is however available and the plasmid is known to carry resistance to kanamycin. This is not located on or near the fragment subcloned into pAT153, which did not confer any obvious novel phenotype nor show any structural instability when propagated in E.coli.

Each plasmid was cleaved with EcoRI and BamHI and the appropriate fragments (as arrowed) extracted by electro-elution from a 1% agarose gel. The resultant plasmid pLF1 when cleaved with Xba I and Bam HI could then be used for the cloning of the β-urogastrone fragment, as illustrated in **Figure 12**.

3.2.2 Analysis of C-terminal clones of β-urogastrone

Following transformation of competent HB101 with pLF1 cleaved as in **Figure 12** and ligated to the extracted fragment of the β-urogastrone gene, 24 colonies were isolated of which eight were subjected to initial restriction analysis.

The insert was entirely excised as well as being cleaved with each of the enzymes designed to be present in the correctly assembled fragment. Part of this analysis is presented in **Figure 13**. Each of four clones can be seen to liberate an identical fragment of around the expected size of 88 base pairs (A) upon cleavage with Bam HI and Xba I. Each also contains an additional Hae II site not present in the vector pLF1 but which maps at the predicted position in the gene, cleaving the largest vector fragment to yield an additional 77 base pair fragment (B). Further cleavage with enzymes Fnud II and Hph I yielded the fragments expected from cleavage of the sites in the β-urogastrone gene. Since these cleavages strongly suggested that the

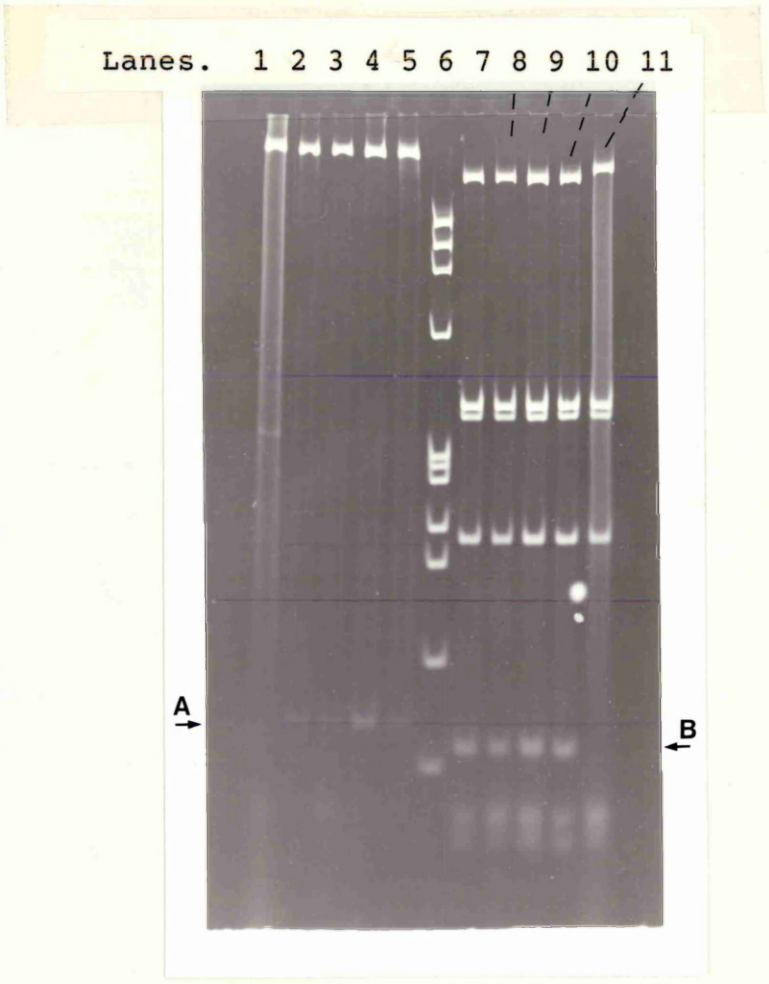


Figure 13. Identification of pUR1.

Restriction analysis of four putative pUR1 isolates. Lane 1; pLF1 cleaved by Xba I and Bam HI. Lanes 2-5; isolates 1-4 cleaved by Xba I and Bam HI. Lane 6; m.wt. markers ϕX174 rf DNA cleaved by Hae III. Lanes 7-10; isolates 1-4 cleaved by Hae II Lane 11; pLF1 cleaved by Hae II. The 88bp fragment of the β-urogastrone gene is arrowed (A), in lanes 2-5. The 77bp additional Hae II fragment, due to the extra site present within the β-urogastrone gene is also arrowed (B).

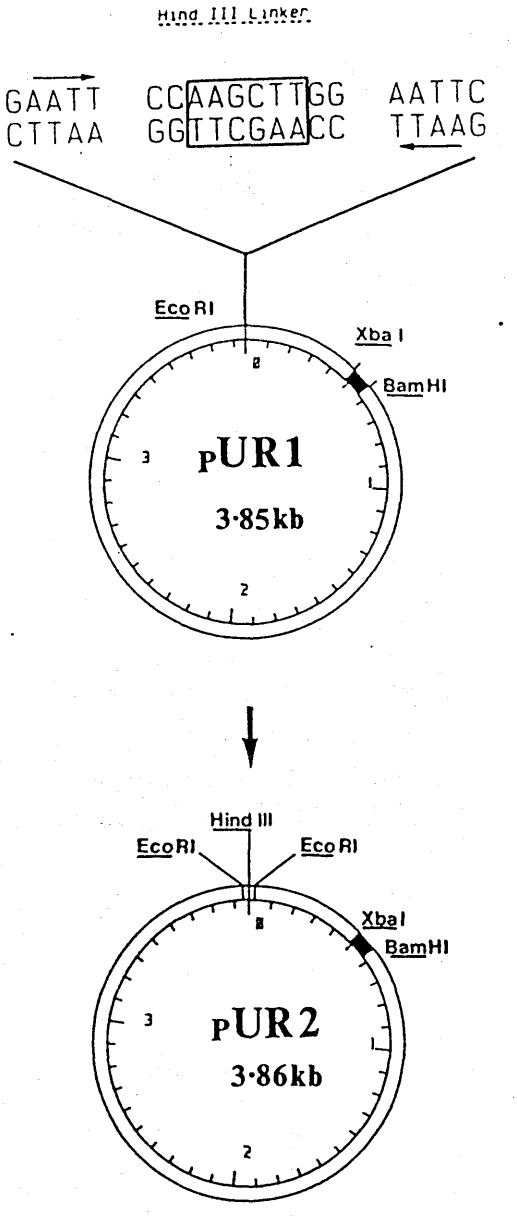


Figure 14. Generation of plasmid pUR2.

The sequence of the Hind III linker inserted at the unique Eco RI site of plasmid pUR1 is shown at the top. The restriction sites present in the resulting plasmid pUR2 are shown at the bottom.

correct fragment had been cloned, one of these isolates designated pUR1 was prepared for the cloning of the N-terminus of the gene.

3.2.3 Construction of pUR2

The fully assembled N-terminus of the gene was flanked by restriction sites for Hind III and Xba I necessitating the presence of unique targets for these enzymes in the cloning vehicle. A Hind III site was therefore introduced at the unique Eco RI site of pUR1 by means of linker insertion. The sequence of the linker was such that it reconstituted the Eco RI site on either side as diagrammed in **Figure 14**. Subsequent cleavage with Hind III and Xba I yielded the necessary vector for cloning of the fragment encoding the N-terminus of β-urogastrone. This was designated pUR2.

3.2.4 Cloning of the N-terminal fragment and analysis of the full gene

The N-terminal fragment of the gene was assembled using essentially the same methodology as outlined for the C-terminus. It was then ligated to the purified large plasmid fragment obtained from Xba I and Hind III cleavage of pUR2, as illustrated in **Figure 15**.

Five isolates were prepared and cleaved with enzymes Hind III and Bam HI. As shown in **Figure 16**, one released a fragment of the approximate size to be the β-urogastrone gene. This fragment was extracted and cloned into M13 vectors mp8 and mp9 in which the total nucleotide sequence of the gene was determined on each strand. The sequence obtained, as shown in **Figure 17** was exactly that of the correctly assembled oligomers of the synthetic gene.

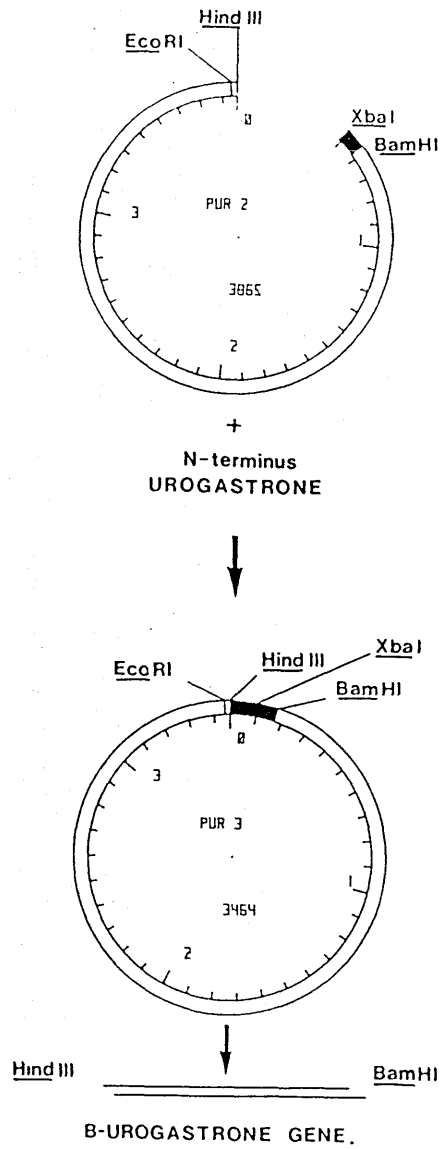


Figure 15. Cloning strategy for the N-terminal fragment of the β-urogastrone gene.

The large fragment of plasmid pUR2, as illustrated at the top, was isolated following cleavage by Hind III and Xba I. This was ligated to the assembled N-terminus of the gene yielding the full length insert.

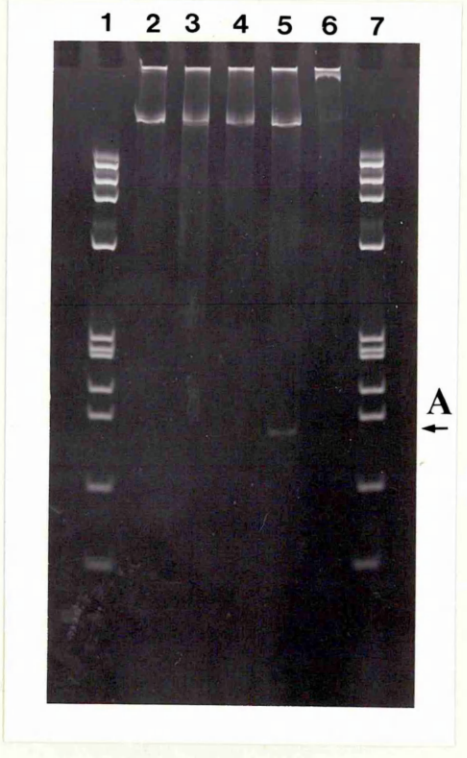


Figure 16. Identification of pUR3.

Restriction of five putative pUR 3 isolates. Lanes 1 and 7; m.wt. markers ϕX174 rf cleaved by Hae III. Lanes 2-6; five isolates cleaved by Hind III and Bam HI. Lane 5 (isolate 4), contains a fragment of the approximate size of the β-urogastrone gene (174bp), as arrowed (A). This migrates slightly below the 194bp band of the m.wt. markers.

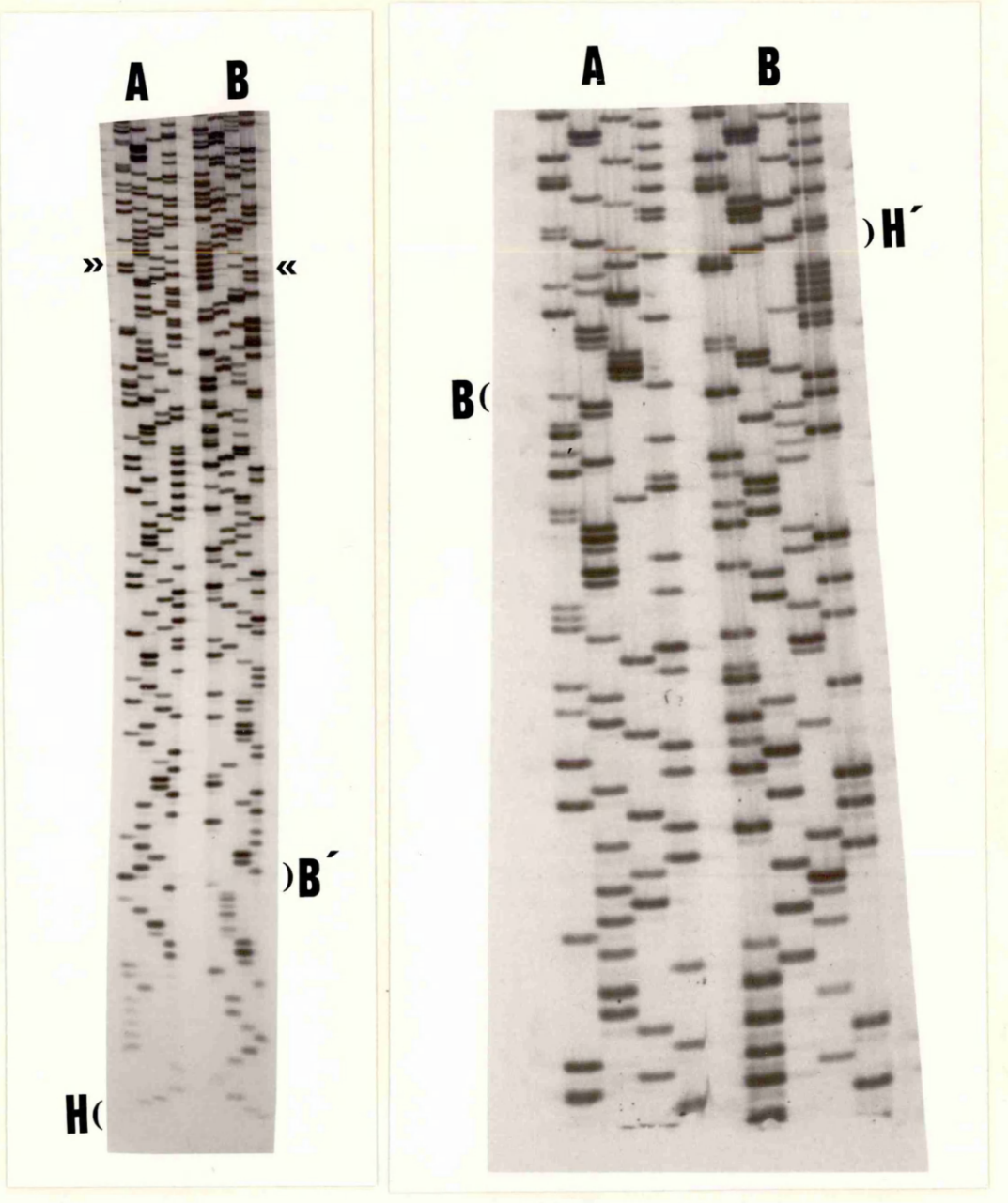


Figure 17. Sequence of the cloned β-urogastrone gene.

Sequence of both strands of the β-urogastrone gene. The gene was cloned between the Hind III and Bam HI sites of M13 vectors mp8 and mp9 and sequenced in both directions using universal sequencing primer (Biolabs). Track A reads from the Hind III (H) site at the bottom towards the Bam HI (B) site at the top. Track B reads from the Bam HI (B′) site at the bottom towards the Hind III (H′) site at the top. Sequence lanes in both cases are in the order A.G.C.T. Sequence at the top of the gel was discerned by extended electrophoresis. Chevrons indicate the point of overlap between the gels.

3.2.5 Discussion

These results demonstrate the feasibility of total gene synthesis as a means to the cloning of a DNA sequence encoding a given peptide. At the time of assembly the urogastrone gene was the largest synthetic gene to have been cloned although subsequently several larger genes have been cloned, both in these laboratories and in others (95). The methodology employed however, has in many cases been closely based upon strategies used or developed in this work. Much of this strategy is self-evident although it is important to stress the relevance of several factors carefully considered in this work. Primarily, it is important to consider in detail the logic of the gene assembly and the eventual destination of the gene for expression, in the specific incorporation of restriction sites. The identification of sequence sub-assemblies and the subsequent versatility of the gene for manipulation are greatly facilitated by incorporation of useful sites. The systematic purification of sub-assemblies significantly reduces the possibility of eventual cloning of wrongly ligated material or the saturation of cloning vector restriction ends with high concentrations of unligated monomer. Finally the usefulness of computer assistance cannot be over-stressed. The programs used in the elimination of repeating structures in this work avoided time-consuming comparisons by eye and have now been superseded by more sophisticated and more rapid programs. This underlines the dynamic nature of this sort of approach, since the improvement of computer assistance and indeed of gene synthesis now allow faster and more thorough design of much larger oligonucleotides. Synthetic genes can now be assembled from fragments of 30 to 40 nucleotides in length of purity equal or greater to those used in this work. Considering the similar progression which is occuring in the efficiency of automated protein sequence determination, the prospect of gene synthesis becoming the major source of cloned DNA is becoming increasingly likely.

CHAPTER 4

EXPRESSION OF β-UROGASTRONE IN E.coli USING trp PROMOTER REGULATED EXPRESSION VECTORS

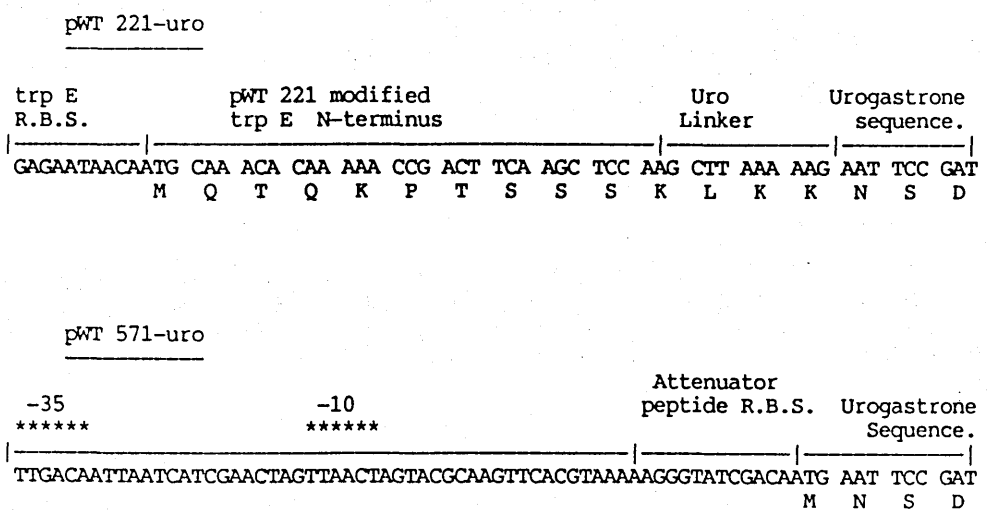


Figure 18. Plasmids pWT221-URO and pWT571-URO.

The DNA and amino acid sequences at the N-terminus of β-urogastrone in pWT221-URO and pWT571-URO.

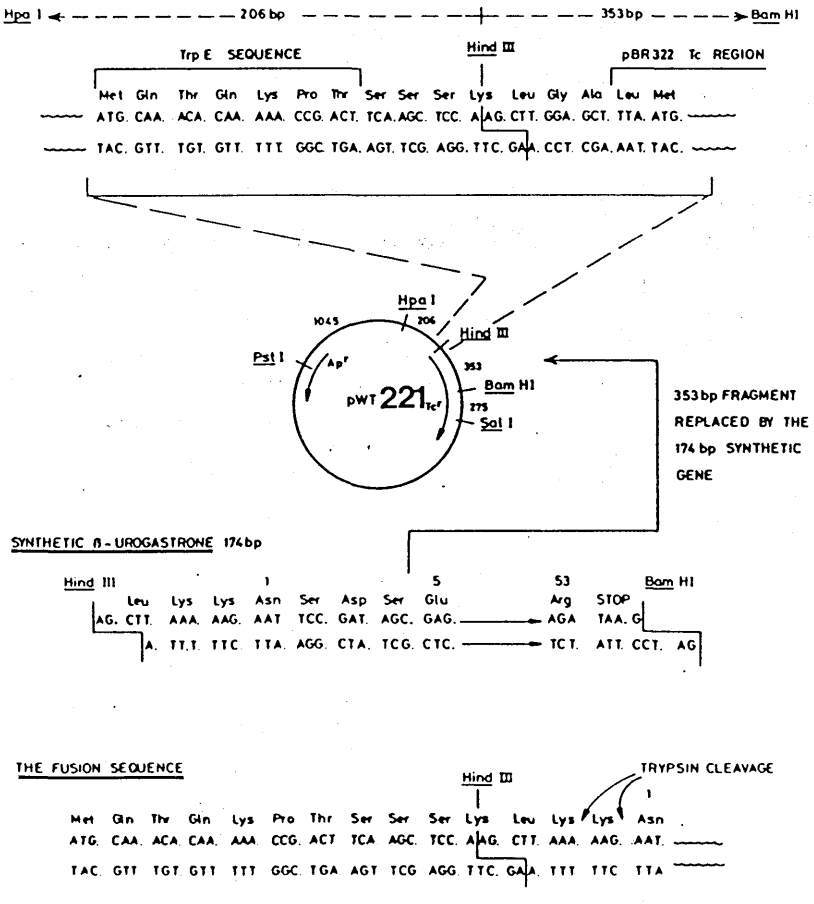


Figure 19. Derivation of plasmid pWT 221-URO.

4.1 Plasmid Construction

Following total DNA sequencing the β-Urogastrone gene was cloned into expression plasmids containing the promoter and operator regions of the tryptophan (trp) biosynthetic operon of E.coli K12. These were obtained from Dr W.Tacon at G.D.Searle. Two plasmids were constructed in which β-Urogastrone could be expressed either as a mature peptide or incorporating a 14 residue N-terminal fusion as shown in **Figure 18**. This fusion comprised the 4 residue linker region preceded by the first 10 residues of the Trp E protein, anthranilate synthase.

Plasmid 221-URO was formed by ligating the 174 bp Hind III to Bam HI fragment encoding β-urogastrone form pUR3 into the similarly cleaved expression plasmid pWT221 as shown in **Figure 19**. Plasmid 571-URO was constructed by ligating the 163 bp Eco RI to Bam HI fragment of the β-urogastrone gene also from pUR3 into similarly cleaved pWT571. This expression vector lacks the attenuator region of the trp promoter, originally being designed to potentiate greater transcription rates than the wild type promoter in the presence of low levels of tryptophan. The plasmid was constructed using a specific linker region incorporating an Eco RI site at the beginning of the attenuator polypeptide coding sequence. This allows expression of mature β-Urogastrone inserted at this position using the translation initiation point of the trp attenuator polypeptide.

Plasmid ligations were transformed into host HB101 and correctly assembled plasmids identified in each case by restriction cleavage. Copy numbers of both plasmids were determined at approximately 100 per cell.

Graph 1. Expression of Trp E fused urogastrone in HB101.

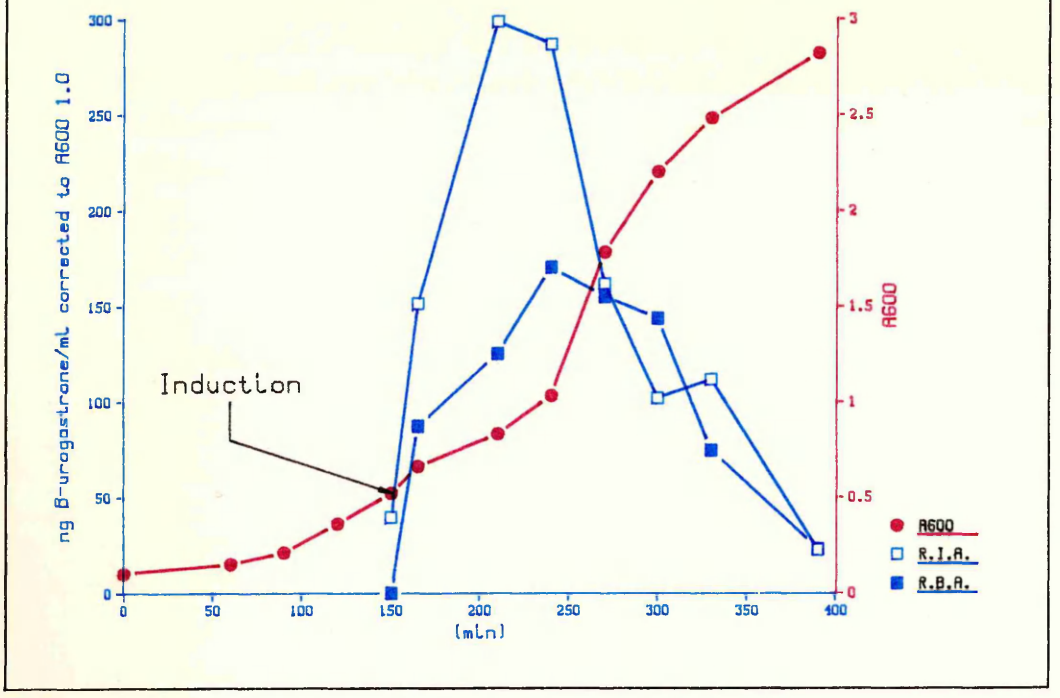


Table 1

Sample	ND	ND	ND	ND	1	2	3	4	5	6	7	8
Time	0	60	90	120	150	165	210	240	270	300	330	390
A_{600}	0·1	0·15	0·21	0·36	0·52	0·66	0·83	1·03	1·78	2·2	2·48	2·82
ng.uro/ml to A_{600} 1·0 R.I.A.	-	-	-	-	39·6	151·5	298·7	286·8	161·7	101·8	111·6	22·6
ng uro/ml to A 1·0 R.B.A.	-	-	-	-	0	87·2	125·3	170·5	155	143·6	74·5	23

Graph 1. and Table 1.

Expression of Trp E fused urogastrone in HB101, as detected by receptor binding and radio-immuno assay.

4.2 Expression of β-urogastrone

4.2.1 β-urogastrone expression in HB101

The expression of β-Urogastrone in HB101 using these vectors was now examined by two forms of biological assay. Firstly in a receptor binding assay, β-urogastrone in bacterial extracts was used to compete against ^{125}I labelled urine derived β-urogastrone for EGF cell surface receptors on cultured fibroblasts. It was also measured using a radio-immuno assay employing antiserum isolated from rabbits immunised with conventionally isolated human β-Urogastrone. Expression of Trp E fused urogastrone was measured using both forms of assay but mature β-Urogastrone was assayed initially by radio-immuno assay only, due to technical problems with the receptor binding assay. Expression of both mature and Trp E fused urogastrone was induced by the addition of 3 β-indole acrylic acid and the expression levels measured before and after induction. Following induction, β-urogastrone expression was measured throughout the exponential and early stationary growth phases of HB101. Cell growth and the presence of Trp E fused urogastrone during this time are shown in **Table 1** and plotted on **Graph 1**. All samples of control cultures of HB101 alone showed no detectable activity within the range of both forms of assay.

4.2.2 β-urogastrone expression in a protease deficient host strain

To examine the possible effect of proteolytic degradation of β-urogastrone, the expression of plasmid pWT 571-URO was examined in host HW27 which is an E.coli K12 derivative defective in the lon gene. The product of the lon gene is an ATP dependent protease and genetic studies have shown that mutants in this gene have a reduced capacity to degrade abnormal proteins (96). Cell growth and the presence of native β-urogastrone as detected by radio-immuno assay

Graph 2. Expression of native B-urogastrone in HB101 and HW27.

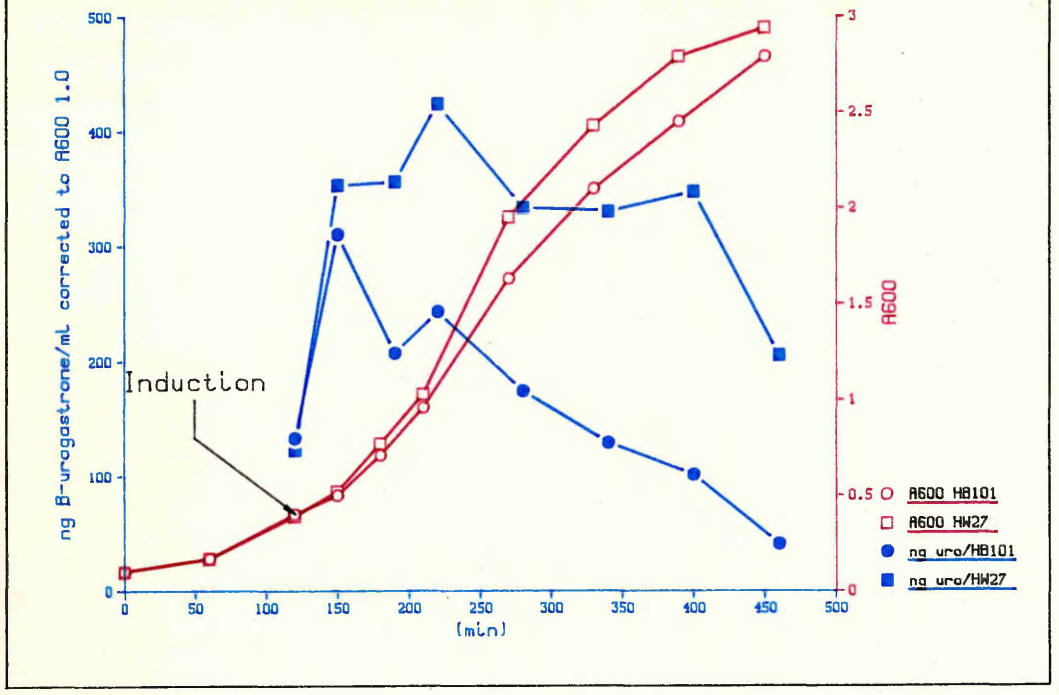


Table 2

Sample	ND	ND	1	2	3	4	5	6	7	8
Time	0	60	120	150	180	210	270	330	390	450
HB 101 A_{600}	0·1	0·17	0·405	0·5	0·71	0·96	1·63	2·1	2·45	2·79
HB 101 ng.uro/ml to A_{600} 1·0	–	–	133	310	207	243	174	129	101	40·8
HW 27 A_{600}	0·1	0·17	0·39	0·52	0·77	1·03	1·95	2·43	2·79	2·94
HW 27 ng.uro/ml to A_{600} 1·0	–	–	123	353	356	424	333	330	347	205

Graph 2. and Table 2.

Expression of native β-urogastrone in HB101 and HW27 as detected by radio-immuno assay.

were monitored throughout exponential and early stationary phases and are shown in **Table and Graph 2**. A culture of HW27 alone showed no detectable activity within the range of either assay.

4.2.3 Discussion

Clearly the synthetic gene is expressed from the *trp* promoter in both plasmid constructions, and both forms of the peptide molecule are synthesised in HB101. Following induction of both Trp E fused and mature β-urogastrone, there is a rapid increase in intracellular levels. These are maintained between 100-300 ng/ml as detected by radio-immuno assay throughout the exponential growth phase. These initial experiments have since been extensively repeated (E.Cook personal communication) indicating similar expression levels of Trp E fused and mature β-urogastrone to those observed here. Levels drop considerably as the cells enter stationary phase, most likely due to an increase in proteolytic degradation as protein turnover is markedly increased at this stage in the growth of *E.coli*. In addition, other heterologous peptides have been shown to induce proteolysis during their expression in *E.coli* (102). The susceptibility of β-urogastrone to proteolytic degradation is further indicated by the fact that higher levels of the peptide are detected throughout growth in the *lon* strain HW27. It is also indicated by the fact that higher levels of Trp E fused urogastrone are detected by the radio-immuno assay which may detect partially degraded or unfolded protein than by the receptor binding assay which most likely requires the intact and correctly folded molecule. It should however be stressed that although β-urogastrone activity is assumed, the biological assay used here, only measures the EGF activity of the peptide. It is conceivable although unlikely that some partially degraded material retains EGF activity whilst having lost β-urogastrone activity.

Following these experiments a considerable programme of work was initiated both to increase the expression levels of β-urogastrone in E.coli and to study the turnover of the peptide. Plasmids of copy number elevated above that of pAT153 and incorporating tandem trp promoter regions have been used to give levels of 2-3ug of mature and Trp E fused urogastrone per ml of cells (A_{600} 1.0) (T.Smith personal communication). Pulse labelling has also shown the half-life of mature β-urogastrone in E.coli to be about 2 minutes and of Trp E fused urogastrone to be 5 minutes (M.Carrier personal communication). The activity of Trp E fused urogastrone has also been shown to be identical to the native molecule and the trypsin labile site within the linker region can be effectively cleaved without affecting the other trypsin cleavage sites of the peptide. These are most likely to be folded within the molecule and therefore inaccessible to the enzyme.

The preliminary results presented in this work however do illustrate the potential, assuming a recovery of 50% to isolate from a 20l fermentation the equivalent active protein formerly requiring 1000l of urine. Given the improvements discussed above that potential may be considerably greater.

However if sufficient material was to be isolated to allow investigation of its possible therapeutic potential and to allow detailed study of the molecule, large quantities would be required. Should the rapid proteolysis remain a significant factor or if problems were encountered with purification then alternative hosts to E.coli might have to be considered. In particularly an organism which unlike E.coli might alleviate these problems by secretion of the peptide. The feasibility of an alternative host to E.coli for the expression of β-urogastrone was therefore

examined. Since bacilli have been extensively used in industrial fermentation (as discussed in Chapter 1) and have the capacity to secrete large quantities of protein, a system was designed for the expression initially of β-urogastrone and potentially of subsequent peptides in bacilli. In addition, since B.subtilis is the best characterised of the bacilli and has been a subject of very recent interest the system was designed with B.subtilis as the initial host but with the potential for the transfer to alternative bacilli. Following a study of the comparative regulation of gene expression in B.subtilis and E.coli a series of vectors were then designed.

CHAPTER 5

DESIGN OF A SYSTEM FOR GENE EXPRESSION IN B.subtilis

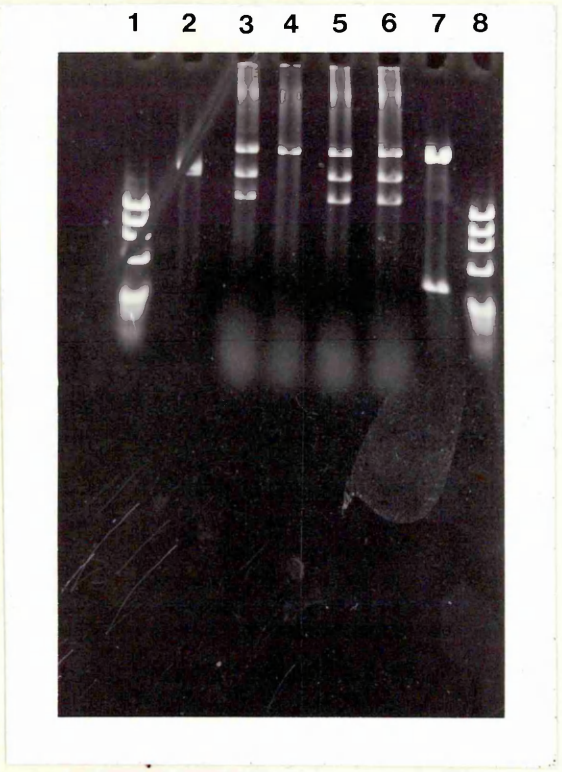


Figure 20. Identification of pFF2.

Restriction analysis of putative pFF2 isolates. Lanes 1 and 8; m.wt. markers ϕX174 rf DNA cleaved by Hae III. Lanes 2-6; five isolates cleaved by Eco RI, Bam HI and Bgl II. Lane 7; pAT153 cleaved by Eco RI and Bam HI as an additional m.wt. marker. Lanes 3 5 and 6 (isolates 2,4 and 5), contain the expected fragments of the correctly ligated plasmid. Lanes 2 and 4 contain incorrectly assembled plasmids.

5.1 Design of the expression vectors

5.1.1 Design of the shuttle vector

Primarily a dual replicon or shuttle vector was designed as a precursor to potential expression vectors. This plasmid was intended to be capable of selectable propagation in both E.coli and B.subtilis in order to facilitate DNA manipulation steps in the more easily handled E.coli with subsequent transformation of B.subtilis for expression studies. Plasmid pAT153 was chosen as the E.coli replicon since it had been successfully used in all previous cloning experiments of this work. Plasmid pUB110 was chosen as the B.subtilis replicon as it had been previously used quite successfully in dual replicon plasmids (55), and had covenient situation of restriction sites for ligation to pAT153. Furthermore it had already been handled in this work, proving useful in the generation of pLF1 for the cloning of the β-urogastrone gene.

5.1.2 Construction of the shuttle vector

Both plasmids were cleaved with enzymes Eco RI and Bam HI and the larger fragments electroeluted from 1% agarose gels, following staining with ethidium bromide and visualisation under long wave UV irradiation. These fragments are represented by the larger fragments in lanes 1 and 3 of **Figure 11**. They were then ligated, transformed into HB101 and ampicillin resistant colonies selected. Of 5 colonies streaked onto L-agar plates containing 5ug/ml kanamycin, only 3 proved to be resistant and grew. Plasmid DNA was prepared from all five and restricted with Eco RI, Bam HI and Bgl II. The correctly ligated vector should show cleavage fragments of 3.28kb, 1.5kb and 2.2kb since the pAT153 fragment is excised intact and the pUB110 fragment is cleaved at the single Bgl II site approximately 1.5kb from the Bam HI site. Figure 20 shows that clones 2,4 and 5 (all kan^r) show these fragments and that the others are

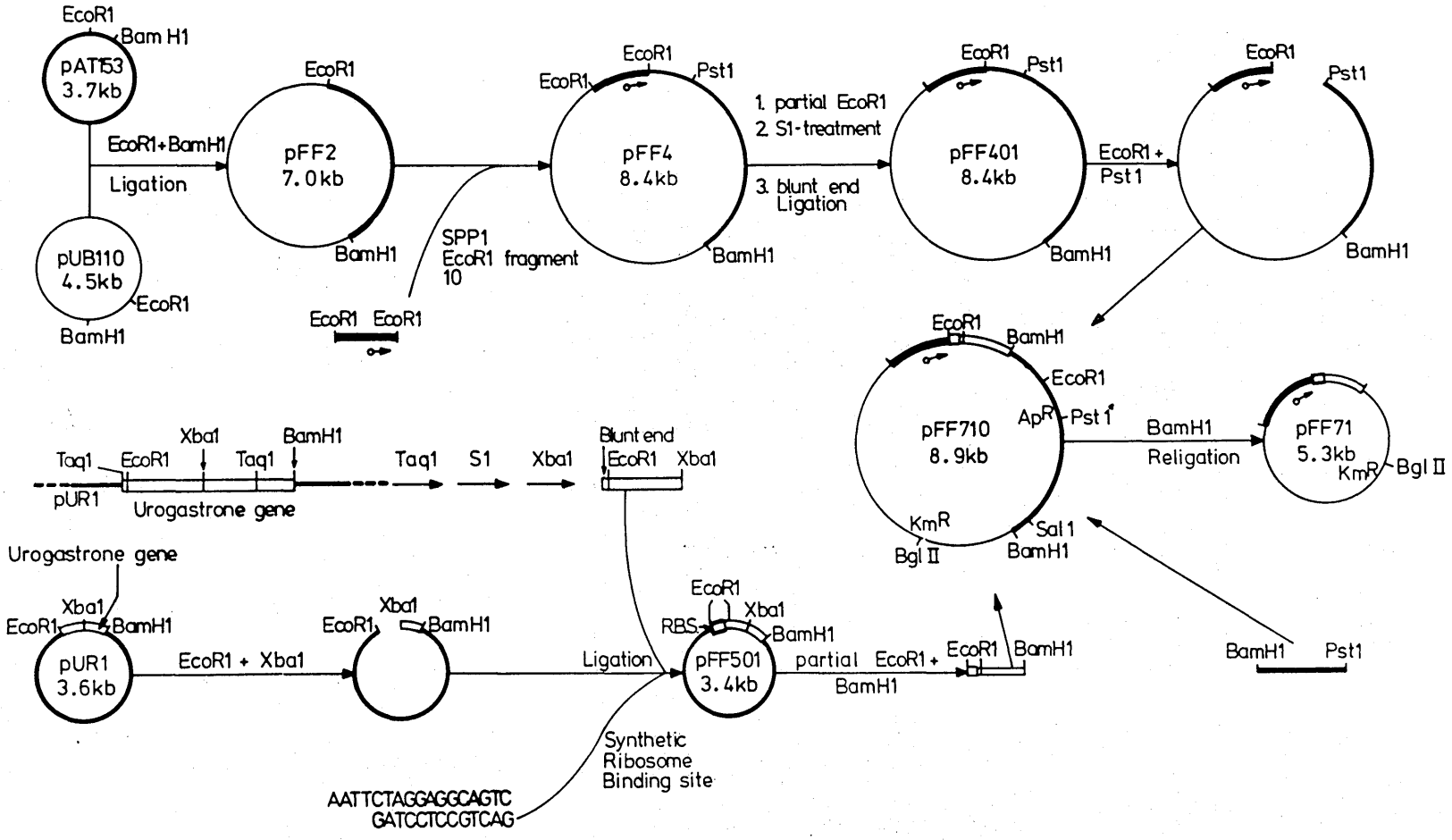


Figure 21. Derivation of plasmids pFF710 and pFF71.

incorrectly ligated. Clone 2 was designated pFF2 and stored as a glycerol stock. Plasmid DNA was then prepared for cloning of the SPP1 fragment.

5.1.3 Cloning and sequencing of Eco RI fragment 10 of SPP1

Total digestion of B.subtilis phage SPPI genome with the restriction endonuclease EcoRI produces 15 fragments, ranging approximately from 350 bp to 7.6 Kb in length (97). The 1.4 kb fragment 10 carries one of the five major early promoters of the phage genome (73). As mentioned this promoter is efficiently transcribed both in vivo and in vitro. Fragment 10 of the SPP1 genome was kindly supplied by Dr.T.Trautner at the Max-Planck Institute in Berlin. This fragment was cloned into the unique Eco RI site of shuttle vector pFF2 and further plasmid manipulations carried out as shown in **Figure 21**. Clones were isolated containing the fragment in either orientation as determined by cleavage both at the unique Xba I site asymmetrically placed within the fragment and at the single Pst I site in pAT153. The desired orientation was that in which transcription proceeded towards the pAT153 replicon as depicted. This plasmid was named pFF4. In order to facilitate further plasmid constructions the upstream Eco RI cleavage site was destroyed by partial digestion with Eco RI followed by nuclease S1 treatment and blunt end ligation. This resulted in the vector pFF401.

This vector provided both the precursor for the plasmids which were subsequently used to study urogastrone expression plus a source of the SPP1 promoter for DNA sequence analysis and transcription mapping. Several restriction enzymes were then tested for the presence of cleavage sites within the SPP1 fragment. Cleavage with the enzyme Pvu II revealed a single site within the fragment approximately 500 bp upstream of the Eco RI site of pFF401. Since the promoter had been roughly approximated to be close

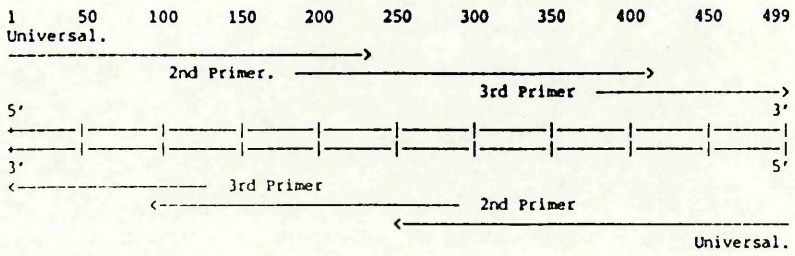


Figure 22. Sequencing strategy for the 499bp SPP1 fragment.

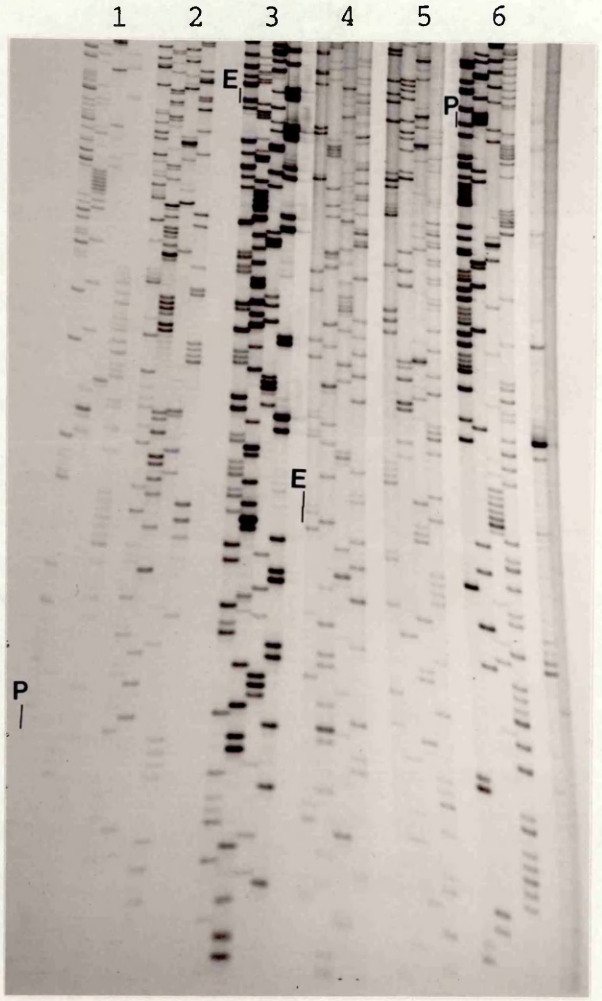


Figure 23. Sequence analysis of the 499bp SPP1 fragment.

Tracks 1-3; top strand universal primer, 2nd primer, 3rd primer. Tracks 4-6; bottom strand universal primer, 2nd primer, 3rd primer. All tracks contain sequence lanes in the order A.G.C.T. The position on either strand of the Eco RI site (E) and Pvu II site (P) are indicated.

to this Eco RI site it seemed likely that it would be incorporated within the 500bp Pvu II to Eco RI fragment. Double cleavage with these enzymes released this fragment, which was then cloned into the M13 derived, phage vectors mp8 and mp9, each cleaved with Eco RI and with Sma I. Since DNA fragments cleaved with Pvu II and Sma I possess blunt ends they may be ligated together. The sequence of this fragment was then determined on both strands by means of the dideoxy chain termination method of DNA sequencing. After sequencing of the fragment termini, four additional specific sequencing primers were designed to facilitate the completion of sequencing of the entire fragment, as shown in **Figure 22**. Part of this analysis is shown in **Figure 23** and the entire sequence of the fragment along with the position of known restriction endonuclease sites shown in **Figure 24**.

5.1.4 Analysis of the SPP1 fragment

Only two open reading frames capable of encoding a peptide larger than 40 residues are present. The larger of these initiates with an ATG codon at base 266 and continues to the end of the fragment, potentially encoding a peptide of at least 78 residues. This is preceded by a region strongly characteristic of a B.subtilis ribosome binding site, incorporating the highly conserved sequence 5′ GGAGG 3′ positioned seven bases prior to the potential ATG start. The other open frame initiates with an ATG codon at position 96 and terminates with a TAG stop codon overlapping the S.D. box of the second open frame, potentially encoding a peptide of 52 residues. This is preceded by a region less characteristic of a B.subtilis ribosome binding site but which nevertheless possesses homology to the 3′ end of the 16S subunit of rRNA. It is not known which gene products are expressed from the promoter at this region of the SPP1 genome although two peptides of around 50 and 75 amino acids in length have been detected in E.coli minicells carrying the 1.4 kB SPP1 fragment cloned onto pBR322 (76). The

A

CAGCTGAATGAAAAATATTCATCTTTTTTTAAAAATAGTGGTTGCCTTTCTATGTTTTCTATGTTTTAATAG
AATCATAGAGAGGGGGGGACAACATGAACGAGCAAAAGAGAGTTAAGAAAAACCTTGAAGTAAGCCCGACAG
TGTGGAAGAAAGCTCACATCATGAAGGCTGACACAGGCAAGAAACTCTATGAAATTTTAGAAGAAGCAATCA
ACGAAATGTATGAGCGGCACGAGCCGAAACACTTTTAGGAGGGGTTAAAATGGCGAGGGTAGTGGCAAACAC
GAAGGACATGCCACATGAGGAATGGTTGGCACTTCGTAAGCACGGCATTGGCGGCAGTGATGCGGCAAAGGT
GTTAGGAGTCAGCAAATACGGTAGCCCGTTAACAGCTTACATGGAGAAGAAAGGAATGTACACCCCGAAAGT
TAGCGAAGCGACGGAAGAAGCCGCTAGGTGGGGGACGATCATGGAGCCAGTTCTTCGGGACGAATTC

B

SPP1. Pvu II - Eco RI 499 bp fragment.

Enzyme	No. Cuts	1	83	166	250	333	416	499
Aha III	1.	1						
Alu I	3.1			1			1	
Dpn I	1.							1
Dra I	1.	1						
EcoRI	1.							1
Fnu4H I	4.				1	11	1	
HinC II	1.						1	
HinF I	2.		1			1		
Hpa I	1.						1	
Mbo I	1.							1
Mbo II	5.			1	1		1	1 1
Mnl I	4.		1		1 1	1		
Nla III	6.		1	1		11	1	1
Nla IV	1.							1
Pvu II	1.1							
Rsa I	1.						1	
Sau3A I	1.							1
SfaN I	1.					1		
Tthl11 II	1.					1		

19 enzymes found

Figure 24. DNA sequence and restricion map of the 499bp SPP1 Pvu II-Eco RI fragment.

The 499bp sequence of the Pvu II to Eco RI fragment of the SPP1 genome is shown at the top (A). The approximate position of known restriction endonuclease sites within the fragment is shown below (B).

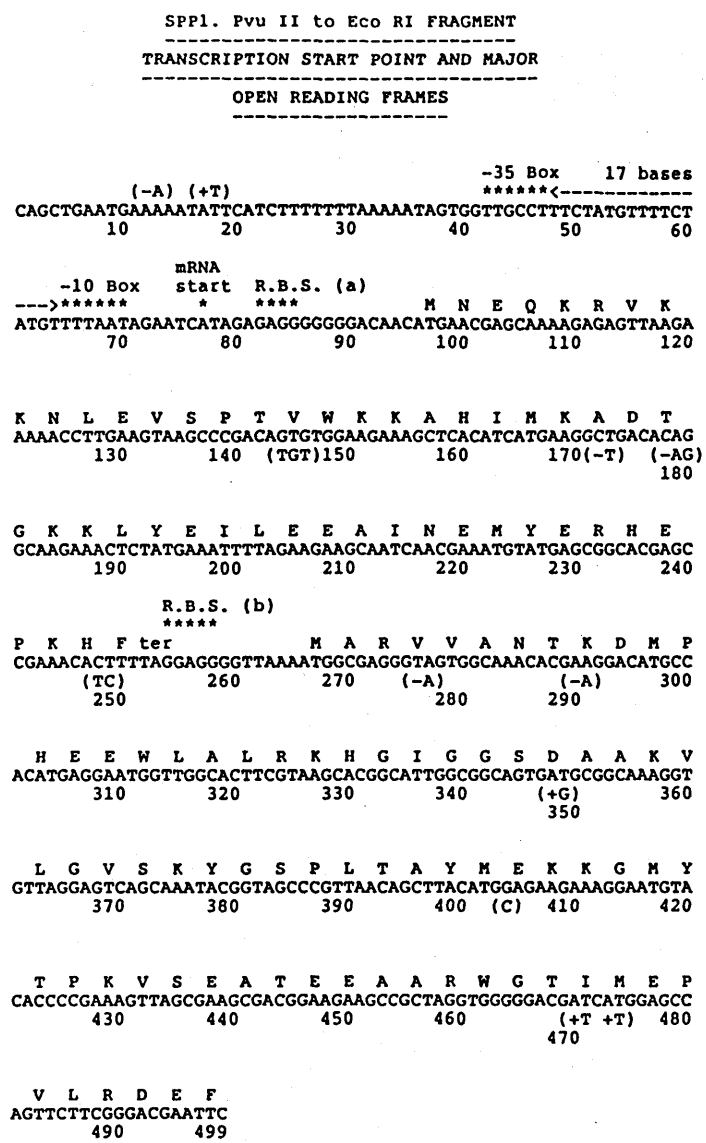


Figure 25. Characterisation of the 499bp SPP1 fragment.

The DNA sequence of the Pvu II to Eco RI fragment showing the promoter region and potential peptides arising from open reading frames greater than 40 amino acid residues in length. Sequence differences of Tailor et al (see text) are shown in brackets above or below the DNA sequence. Deletions are indicated by (-) and insertions by (+) below the preceding base. The -10 and -35 regions and the transcription initiation point are indicated by asterisks. The two putative ribosome binding sites (R.B.S. (a) and (b)) are also shown by asterisks.

position and sequence of the potential peptides which would be encoded by these open reading frames are shown in **Figure 25**.

The 499 bp DNA sequence was then subjected to statistical computer analysis to examine the likelyhood that these open frames do encode peptides (98). This involved the determination of di and tri-nucleotide frequencies for specific codons in each possible frame. Due to the frequency of occurence of particular amino acids, coding regions of DNA unlike non-coding regions display a preferential use of particular nucleotide combinations independent of any effect of codon usage. It is possible within a given sequence to measure these frequencies in every frame and to search for the presence of this bias. A value or score can then be assigned at every position in the sequence for each frame. A sequence which consistently scores highest along an open reading frame indicates strong similarity in nucleotide content to known coding regions and suggests that the open frame does encode a peptide.

As shown in **Figure 26**, the first of the two major open frames within the SPP1 fragment scores highly by this analysis along its entirety. The second open frame although not consistently scoring as highly as the first, nevertheless also indicates strong similarity as measured by this method to known coding regions of DNA.

The amino acid composition and codon usage of the potential peptides were also examined and compared to the typical usage in proteins of *E.coli* and *B.subtilis*. These showed no obvious major differences, except for the fact that both of these peptides would be very rich in charged amino-acids. The amino acids and codons present are shown in

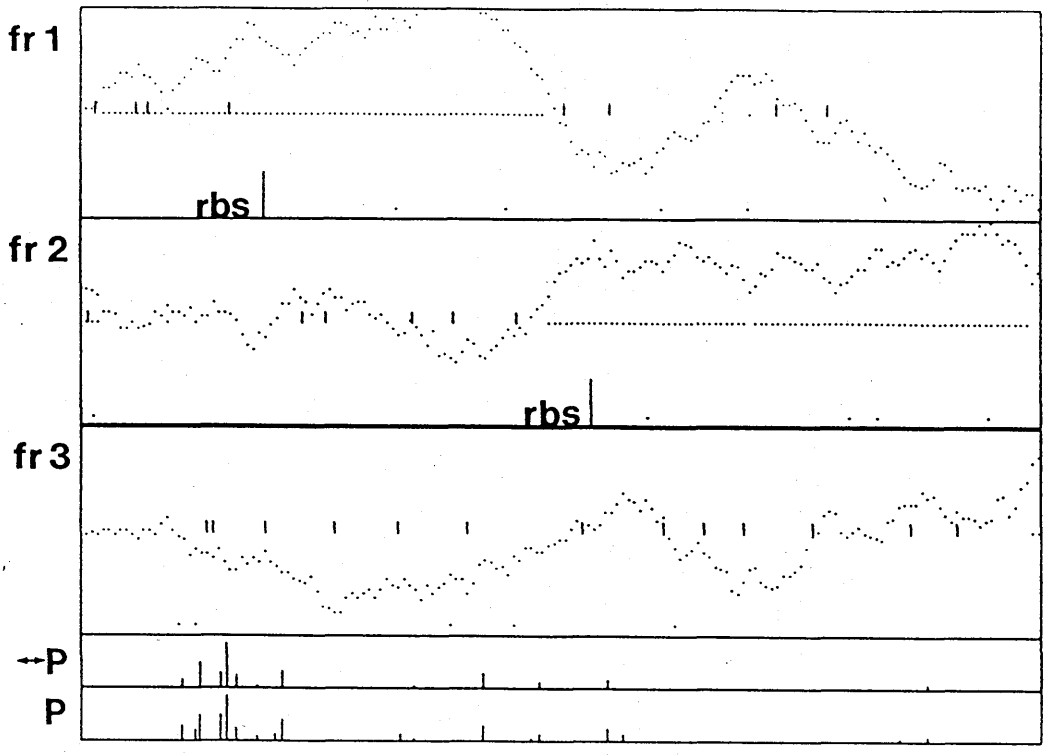


Figure 26. Statistical computer analysis of the SPP1 Pvu II to Eco RI fragment.

For each of the three reading frames (fr1, fr2, fr3), the scores attained for positional base preference are indicated by a point, above or below the centre line of each of the top three boxes. The frame scoring highest is indicated by an additional point on the centre line of the box. Scores are indicated at 3bp intervals. Stop codons are shown by a small vertical line on the centre line. Potential initiation codons (ATG only), are shown by a point along the bottom of each box. Regions showing strong homology to E.coli ribosome binding sites are indicated by a vertical line at the bottom of each box, and are labelled (rbs). The bottom box (P), indicates regions displaying homology to the conserved -10 and -35 hexamers of E.coli promoters. The box above (<->P) shows this analysis corrected for appropriate hexamer spacing.

AMINO ACID COMPOSITION AND CODON USAGE

OF FIRST OPEN READING FRAME

1 UUU Phe	0 UCU Ser	2 UAU Tyr	0 UGU Cys
0 UUC Phe	0 UCC Ser	0 UAC Tyr	0 UGC Cys
1 UUA Leu	0 UCA Ser	0 UAA ***	0 UGA ***
0 UUG Leu	0 UCG Ser	0 UAG ***	1 UGG Trp
1 CUU Leu	0 CCU Pro	0 CAU His	0 CGU Arg
1 CUC Leu	0 CCC Pro	3 CAC His	0 CGC Arg
0 CUA Leu	0 CCA Pro	1 CAA Gln	0 CGA Arg
0 CUG Leu	2 CCG Pro	0 CAG Gln	1 CGG Arg
1 AUU Ile	0 ACU Thr	0 AAU Asn	0 AGU Ser
2 AUC Ile	0 ACC Thr	3 AAC Asn	1 AGC Ser
0 AUA Ile	2 ACA Thr	4 AAA Lys	1 AGA Arg
3 AUG Met	0 ACG Thr	5 AAG Lys	0 AGG Arg
1 GUU Val	2 GCU Ala	0 GAU Asp	0 GGU Gly
0 GUC Val	0 GCC Ala	1 GAC Asp	1 GGC Gly
1 GUA Val	1 GCA Ala	5 GAA Glu	0 GGA Gly
1 GUG Val	0 GCG Ala	3 GAG Glu	0 GGG Gly

AMINO ACID COMPOSITION AND CODON USAGE

OF SECOND OPEN READING FRAME.

0 UUU Phe	0 UCU Ser	0 UAU Tyr	0 UGU Cys
1 UUC Phe	0 UCC Ser	3 UAC Tyr	0 UGC Cys
2 UUA Leu	0 UCA Ser	0 UAA ***	0 UGA ***
1 UUG Leu	0 UCG Ser	0 UAG ***	2 UGG Trp
2 CUU Leu	0 CCU Pro	1 CAU His	1 CGU Arg
0 CUC Leu	0 CCC Pro	1 CAC His	0 CGC Arg
0 CUA Leu	2 CCA Pro	0 CAA Gln	0 CGA Arg
0 CUG Leu	2 CCG Pro	0 CAG Gln	1 CGG Arg
1 AUU Ile	0 ACU Thr	0 AAU Asn	1 AGU Ser
1 AUC Ile	1 ACC Thr	1 AAC Asn	3 AGC Ser
0 AUA Ile	1 ACA Thr	3 AAA Lys	0 AGA Arg
5 AUG Met	3 ACG Thr	4 AAG Lys	2 AGG Arg
2 GUU Val	2 GCU Ala	1 GAU Asp	1 GGU Gly
1 GUC Val	1 GCC Ala	2 GAC Asp	3 GGC Gly
1 GUA Val	3 GCA Ala	5 GAA Glu	2 GGA Gly
2 GUG Val	3 GCG Ala	3 GAG Glu	1 GGG Gly

Figure 27. Amino acid content and codon usage of the potential peptides from both major open reading frames of the 499bp SPP1 fragment.

Figure 27. A search of the NBRF database revealed no significant homology of either of these potential peptides to any known sequences.

5.1.5 Position of transcription initiation

To determine whether either or both of these putative peptides were encoded within transcripts originating in this region of the genome it was now necessary to determine the position of the SPP1 promoter and the exact position of transcripton initiation. This was determined by means of reverse run-off transcripts initiated on mRNA using specific oligonucleotide primers. This was carried out on mRNA isolated from both B.subtilis and E.coli containing the cloned 1.4 kB fragment on plasmids pFF71 and pFF710 respectively. Prior to this however the 499 bp DNA sequence was subjected to a further computer analysis which searched for sequences homologous to the consensus hexamers found at -10 (TATAAT) and -35 bp (TTGACA) from E.coli transcription start sites. This is useful in allowing predictions of the position at which to prime DNA synthesis for run-off experiments in order to obtain conveniently sized reverse transcripts. The result of this analysis is shown in **Figure 26**.

It is very likely that all SPP1 early promoters are transcribed by the σ43 major form of the B.subtilis RNA polymerase. Since all σ43 promoter sites display strong homology to these consensus sequences found in E.coli it was likely that this form of analysis would detect the location of the promoter if it was present on this fragment. The content of any homologous hexamers as well as correct spacing are measured in this analysis. Using both of these criteria a single site was determined within this fragment which displayed strong homology to known promoter regions of E.coli and B.subtilis. This is illustrated in Figure 25. The -35 and -10 hexamers of this potential promoter site were

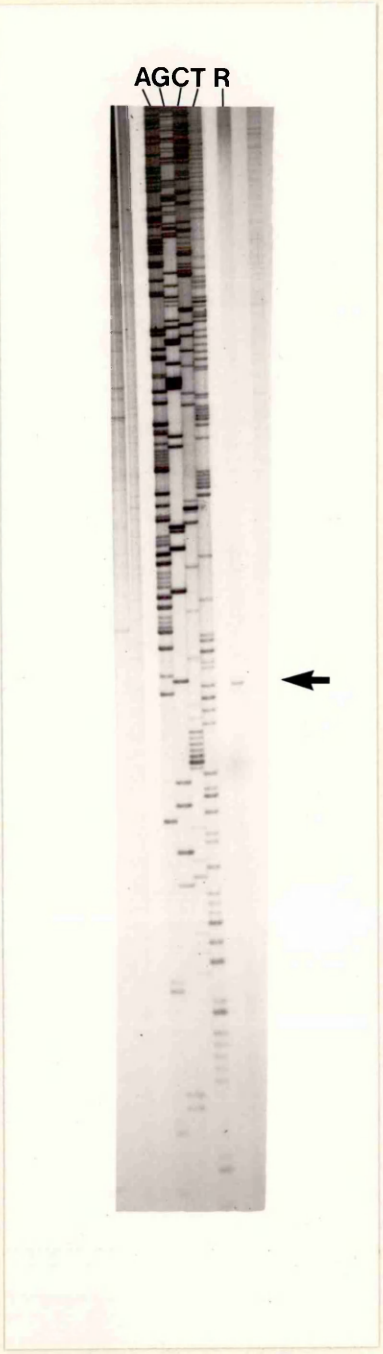


Figure 28. Initiation point of transcription by E.coli RNA polymerase from the SPP1 promoter.

Sequence lanes as indicated are in the order A.G.C.T. A single reverse transcript, initiated on RNA isolated from E.coli cells containing plasmid pFF710, is visible in Lane R. This band migrates to the A residue at position 77 of the SPP1 sequence.

strongly homologous to the conserved regions of E.coli and B.subtilis promoters. The -35 region contained the highly conserved bases TGG as well as the C at position 5, and the -10 region possessed 5 out of 6 bases homologous to the consensus sequence. The 17bp spacing of these regions and the presence of an A-T rich region (25 out of 28 bases are A or T) preceding the -35 region is also highly typical of promoter regions. The figure also shows the position of sequences homologous to ribosome binding sites of E.coli. Both potential ribosome binding sites on the fragment are highlighted by this analysis as would be expected of ribosome binding sites functional in B.subtilis, the first site closely following the potential promoter site.

Transcription reverse run-off experiments were conducted using one of the primers previously used in the sequencing of the SPP1 fragment and were co-electrophoresed with a sequence ladder also generated by this primer. This was first carried out on mRNA isolated from E.coli. The result of this analysis (shown in **Figure 28**) is that a single major band is generated which migrates to position 77 of the sequence, the predicted initiation point of transcription. This analysis was then extended to mRNA isolated from B.subtilis. As shown in **Figure 29** the major product of reverse transcription also migrates to this position. Additional bands can be seen higher in the sequence. If these are genuine reverse transcripts they must originate from the unsequenced region of the SPP1 fragment and their significance cannot be fully determined at present.

5.1.6 Translation from the SPP1 fragment

To examine the nature of any peptides synthesised from the transcript, the 499 bp fragment was subjected to transcription and translation in vitro using an S30 extract of E.coli. As shown in **Figure 30** one peptide at least is

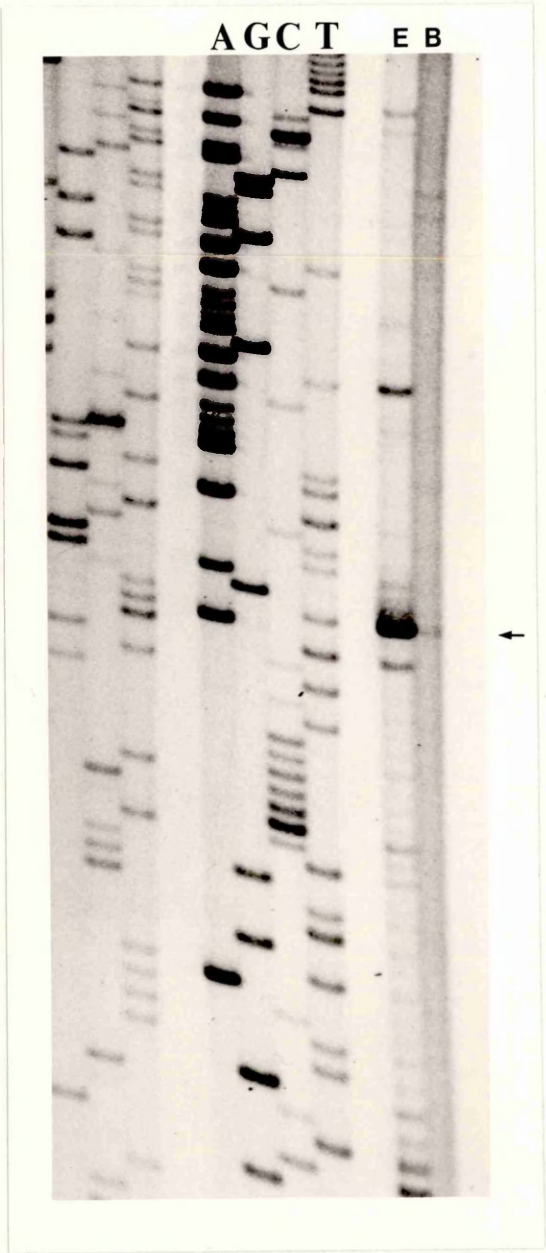


Figure 29. Initiation point of transcription by B.subtilis RNA polymerase from the SPP1 promoter.

Sequence lanes as indicated are in the order A.G.C.T. Lane E; reverse transcripts initiated on RNA isolated from E.coli cells containing plasmid pFF710. Lane B; reverse transcripts initiated on RNA isolated from B.subtilis containing plasmid pFF71. Reverse transcription in both cases and DNA sequence, primed using sequencing primer 3 on the bottom strand (shown in Figure 22). The major band in lane B can be seen migrating at position 77 of the sequence, along with the major band in lane E.

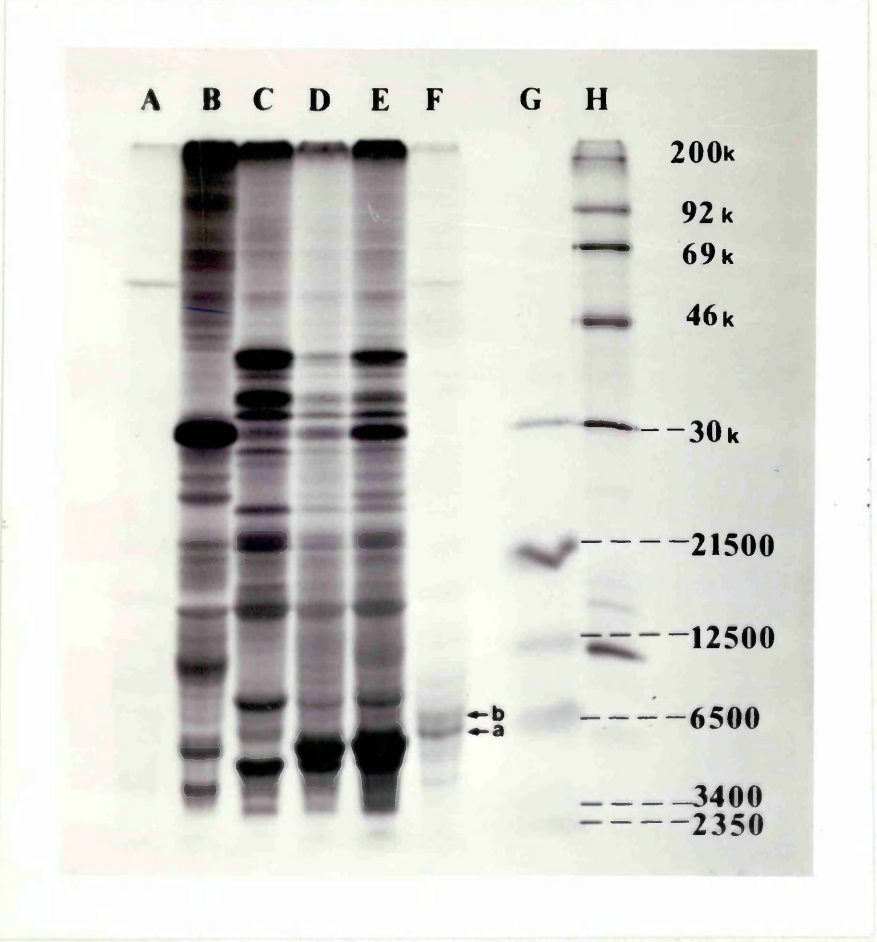


Figure 30. Transcription and translation in vitro of plasmids pFF740, pFF700, pFF730 and the 499bp fragment of SPP1.

Electrophoresis on a 12.5% poly-acrylamide gel, of ^{35}S-methionine labelled proteins, synthesised in vitro from plasmid DNA, using an S30 extract of E.coli (Amersham). In each case, 1ug of plasmid DNA was used, except for lane F which contained 100ng of purified fragment. Lane A; control lane containing no DNA. Lane B; control lane containing pAT153 DNA. Lane C; plasmid pFF740. Lane D; plasmid pFF700. Lane E; plasmid pFF730. Lane F; PvuII to Eco RI fragment of SPP1. Arrows (a and b), indicate two proteins identified in lane F, which are not present in the control lane A. Lanes G and H; ^{14}C labelled m.wt. marker proteins. Sizes as indicated at the right of the figure.

synthesised from this fragment. Plasmids pFF740 (lane C) and pFF700 (lane D) are identical except for the presence in the latter of the 499bp Pvu II to Eco RI fragment (data shown below). Lane D clearly shows a strong additional band indicative of a peptide of approximate m.wt. 6000 Dal. The first open frame of the fragment would generate a peptide of m.wt. 6300 Dal. The peptide is also synthesised from pFF730 (lane E) which is identical to pFF700 except for the presence of the synthetic ribosome binding site R.B.S. II. The increased intensity of the band in lane E may indicate synthesis of β-urogastrone (m.wt. 6.2kDal) co-migrating with the phage peptide. The purified fragment (lane F) also directs the synthesis of the additional peptide seen in Lane 3. This appears to be migrating slightly more slowly, possibly due to the widely differing levels of protein present. Although some contaminating bands are present in lane F comparison with the blank (lane A) shows that a further slightly larger band is also present in lane F, possibly corresponding to the 78 codon second open reading frame. The potential encoded peptide would have a m.wt. of 8.4 kD which is in agreement with the second band observed. This peptide appears to be expressed at lower levels in vitro than the smaller peptide. Since the stop codon of the first open frame overlaps the Shine-Dalgarno region of the second, this may be due to inefficient coupling of the translation of the two peptides by the cell extract.

5.1.7 Discussion

It is clear that transcription initiates both in E.coli and in B.subtilis at the adenosine at position 77 of the SPP1 sequence. This precedes the first putative ribosome binding site within the fragment thereby further indicating that the first open frame does encode a peptide. Although B.subtilis promoter regions are known to function in E.coli, this is the first illustration that the precise position of initiation is identical in both organisms. Additional

reverse transcripts appear in lane E of Figure 29 which are absent in Figure 28. Since the same mRNA was used in both cases this probably represents RNA degradation during storage.

The sequence of the transcript running towards the Eco RI site was searched by computer for the presence of any inverted repeats. These sequences are often capable of forming stable stem-loop structures characteristic of transcription terminators in E.coli and B.subtilis. No obviously stable structure was detected and it is presumed that this transcript does not terminate before the Eco RI site.

A minimum of one peptide is synthesised from the phage fragment and possibly the N-terminus of a second. This is in agreement with recently published data (76) in which minicells of E.coli were shown to express two peptides from the 1.4 kb SPP1 fragment. The smaller of these is a peptide of approximate m.wt. 6.3 kDal and the larger is a peptide of about 9.4 kDal m.wt. The larger peptide if encoded by the second open frame shown in this work would extend an additional 8 codons into the sequence of pBR322 generating a peptide of approximate m.wt. 9.3 kDal. This is therefore also in agreement.

However the data of the other authors (Tailor et al) is in considerable disagreement with these results concerning the DNA sequence of the 499bp fragment, over which 16 differences exist of which 10 are frameshifts and 6 are substitutions. All but one of the substitutions and 3 single base frameshifts occur in the sequence of the small peptide in which there are three amino acid substitutions and one amino acid deletion. There is in addition one silent mutation which appears to be part of a two base transversion. These variations are highlighted in Figure 25.

It is highly unlikely that these changes in sequence represent allelic variation since seven of the eight DNA sequence differences within the region encoding the small peptide contribute to amino-acid variations. Allelic variants would be much more likely to include a greater number of silent mutations. Furthermore the amino-acid replacements are not at all conservative since in one instance a glycine is replaced by an arginine two residues distant from a deleted aspartate and in another instance a cysteine replaces a valine. It is equally unlikely that the sequence determination in this work is in error since the entire sequence of the region has been determined on both DNA strands with only two stretches of 6 bases over which sequence could not be determined unambiguously on both strands. In both of these regions the sequence could however be determined completely unambiguously on one strand and neither region is in disagreement with the other authors. The second open reading frame detected in this work is not found by the other authors since frameshift mutations cause it to be terminated after only 13 codons. They attribute the 9.4 kDal protein to an unknown open frame upstream of the Pvu II site. If the sequence of Tailor et al is correct it appears highly coincidental that a protein of the appropriate molecular weight to arise from the open frame detected in this work is also found by the other authors. It is also highly coincidental that the open frame is preceded by a ribosome binding site typical of those recognised by B.subtilis. The DNA sequence determination carried out by these authors was obtained by the Maxam and Gilbert method (99) which I have found to be less reliable and reproducible than the dideoxy chain termination method used in this work. The data presentation on the paper is also incorrect, having assigned a valine residue to a TGT cysteine codon. For these reasons I believe their to be incorrect. Tailor et al also

data

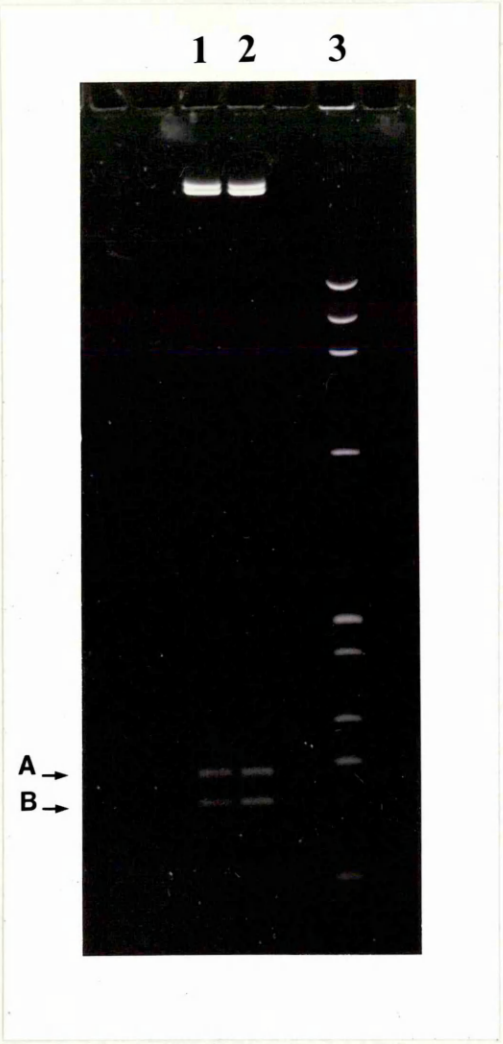


Figure 31. Isolation of the β-urogastrone gene and the first synthetic ribosome binding site from pFF501.

Electrophoresis on a 5% poly-acrylamide gel of the β-urogastrone gene plus ribosome binding site I, cleaved from plasmid pFF501. Lanes 1 and 2; pFF501 cleaved with Bam HI and partially with Eco RI. Lane 3; m.wt. markers ϕX174 rf cleaved by Hae III. Both lanes 1 and 2 show the 186bp fragment (A), of the gene plus the ribosome binding site, and the 163bp fragment of the gene alone (B). Fragment A migrates slightly below the 194bp marker band as expected. Lane 1 represents 1/3 of complete cleavage by Eco RI and lane 2 represents 1/2 of complete cleavage. The two high m.wt. plasmid fragments can be seen at the top of lanes 1 and 2.

propose the transcription initiation point to be that identified in this work although no data is shown to accurately determine this fact.

5.2 Construction of the expression plasmids

All the vectors involved in the expression of the β-urogastrone gene in B.subtilis were derived initially from pFF2 allowing all plasmid manipulations to be carried out in E.coli. The plasmids although complex in initial construction were designed where possible to simplify technical manipulation and facilitate replacement of the promoter and ribosome binding sites. **Figure 21** summarises the steps in the derivation of the first expression vector pFF710.

As shown, the ribosome binding site for translation of mature β-Urogastrone was inserted using a synthetic linker ligated to the blunted Taq I site of 221-URO. Since the linker possessed an upstream Eco RI site, the synthetic ribosome binding site and initiator ATG could be subsequently removed on a 23 bp Eco RI fragment and replaced by an alternative sequence. The β-urogastrone gene preceded by the ribosome binding site was ligated with pFF401 to form pFF710. This involved isolation of the Bam HI to partial Eco RI fragment of pFF501. This was extracted from a 5% polyacrylamide gel as illustrated in **Figure 31**. A linker region incorporating the remaining sequence of pAT153 on the Bam HI to Pst I fragment was also included to generate a full length copy of the plasmid. Prior to transformation of B.subtilis this expression vector was cleaved with Bam HI and the larger fragment isolated and ligated. This removed all pAT153 sequences and generated a large number of oligomeric molecules enabling efficient transformation of competent Bacillus subtilis. Following transformation of

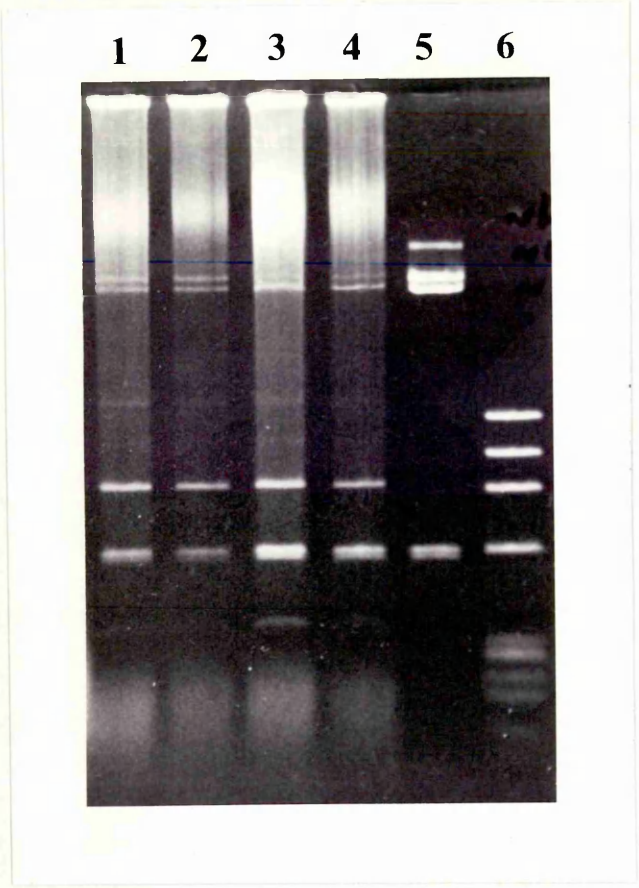


Figure 32. Structural stability of plasmid pFF71.

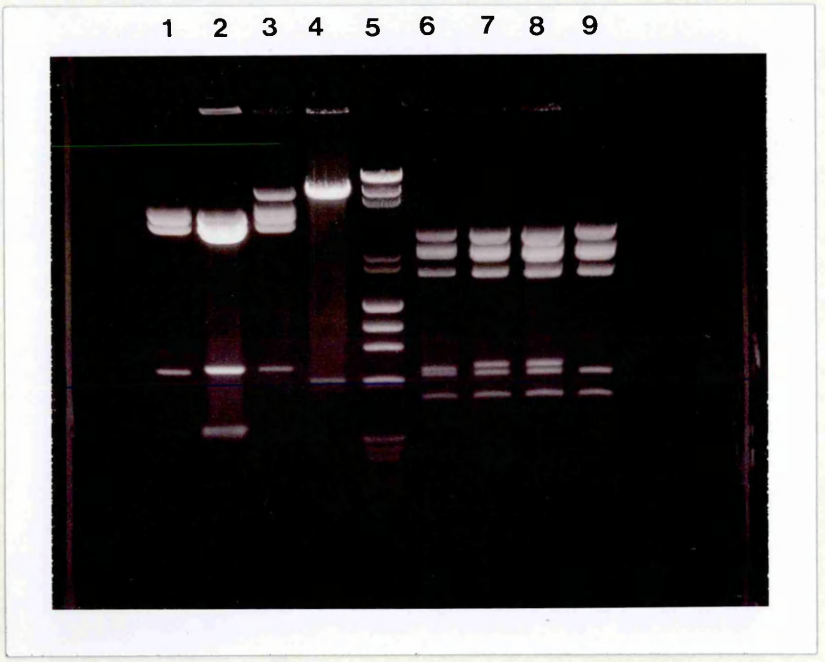
Electrophoresis on a 1% agarose gel of pFF71 plasmid DNA isolated from 3G18. Lanes 1-4; plasmid DNA isolated from cells in shaking culture, having undergone 10 (Lane 1), 20 (lane 2), 50 (lane 3) and 100 (lane 4) cell divisions following inoculation by a single kanamycin (5ug/ml) resistant colony. Plasmid in each case is shown following restriction by Pvu II and Xba I. Lane 5; pFF710 plasmid DNA isolated from HB101 and restricted by Pvu II and Xba I as a control. Lane 6; m.wt. markers ϕX174 rf cleaved by Hae III.

3G18 with plasmid pFF71 and re-streaking, plasmid integrity was verified by restriction analysis. Although exhaustive stability studies were not possible in the course of this work plasmid samples were prepared following growth of the transformed cells selected on kanamycin (5ug/ml). After approximately 10, 20, 50 and 100 generations including successive dilution into fresh medium no structural rearrangement of plasmid DNA was evident. This is shown in **Figure 32**. Plasmid DNA was cleaved with Xba I and Pvu II. Two fragments of approximately 600bp are generated from the sites within the SPP1 fragment and within the β-urogastrone gene. These can also be seen in the digest of pFF710. A small band of roughly 340bp is also generated in pFF71 from the Pvu II site of pUB110 to the Xba I site of β-urogastrone. This does not arise in pFF710 due to the presence of pAT153. The large fragment of pFF71 is approximately 3.65kb. The bands of approximately 900bp and 4.25kb in lanes 1-4 are due to partial Xba I cleavage. Plasmid pFF710 also shows bands of roughly 3.65kb (as pFF71) and 4.0kb (due to pAT153 plus the small fragment of pUB110 and part of β-urogastrone).

The other vectors used in this work were then generated from pFF710, transformed into E.coli and B.subtilis and plasmid integrity verified by restriction analysis. The steps involved in the construction of these vectors is shown below and restriction analysis shown in **Figure 33**.

pFF 730:

The Eco RI site within the pAT153 replicon of pFF710 was subsequently removed by replacement of the small Pst I to Bam HI fragment with that of a pAT153 derivative pFF153D, in which the Eco RI site had been destroyed by treatment with S1 nuclease. The 23 bp Eco RI fragment of pFF710 was



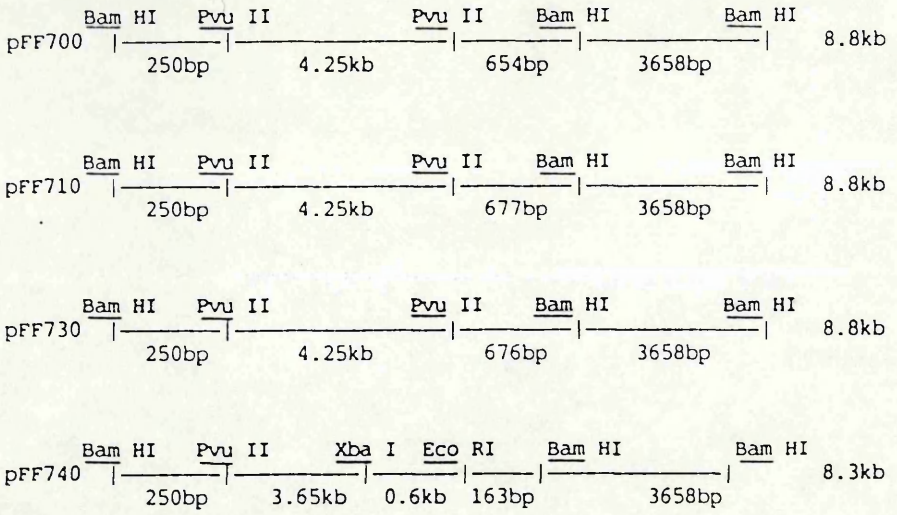


Figure 33. Restriction analysis of the expression vectors used in B.subtilis.

Electrophoresis on 1% agarose, following excision of the SPP1 and β-urogastrone region from plasmids, pFF700 (Lane 1), pFF710 (Lane 2), pFF730 (Lane 3). Cleavage in each case was by Pvu II and Bam HI. Lane 4, cleavage of pFF740 by Eco RI and Xba I. Expected fragments shown below. Lane 5, m.wt. markers (see appendix 2). Lanes 6-9; cleavage by Hinc II of plasmids, pFF700L (Lane 6), pFF710L (Lane 7), pFF730L (Lane 8), pFF740L (Lane 9). Plasmid maps shown in Figure 36.

then replaced by a 22 bp fragment encoding a different ribosome binding site giving plasmid pFF730. Plasmids carrying the correctly orientated fragment were detected by double cleavage with enzymes Acc I and Cla I and subsequent electrophoresis on a 20% poly-acrylamide gel.

pFF 700:

The 23 bp Eco RI fragment of pFF710 was deleted by Eco RI cleavage and plasmid religation in a volume of 100 ul to give plasmid pFF 700. This plasmid is therefore identical in structure to pFF710 and pFF730 except for the deletion of the ribosome binding site. This provides a vector for subsequent insertion of ribosome binding sites and a control by which to measure the gene expression of the previous two plasmids.

pFF 740:

The SPP1 promoter was deleted by cleavage with Eco RI and partial cleavage with Pvu II. The Eco RI site was then filled using the Klenow fragment of Pol I and the plasmid religated. Since cleavage with Pvu II leaves a blunt end with a 3′ guanine nucleotide this then regenerated the Eco RI site at which position other promoters could in future be inserted. This plasmid was also used to demonstrate the absence of the SPP1 peptide in **Figure 30** (lane C).

5.3 Design of the synthetic ribosome binding sites

The two synthetic ribosome binding sites, designed and used in this work were based upon consensus sequences derived from a number of chromosomally and phage encoded ribosome binding sites from Bacillus subtilis and related

R.B.S. I

```
                Frame 1          Frame 2       Frame 3
                 ***              ** *          ***
GAATTCTAGGAGGCAGTCACAATG AAT TCC GAT AGC
                         M   N   S   D   S
```

R.B.S. II

```
                Frame 2          Frame 1                Frame 3
                 ***              ** *                   ***
GAATTCATAGGAGGCAATCGATG AAT TCC GAT AGC-- TAA
                        M   N   S   D   S
```

Figure 34. Termination codons within the synthetic ribosome binding sites.

Termination points for peptides initiating within the SPP1 fragment are shown for each possible reading frame. A peptide reading through R.B.S. II in frame 3 would terminate at the end of the β-urogastrone gene. All other frames terminate within or close to the synthetic ribosome binding sites.

gram-positive organisms. Primarily, they included the highly conserved Shine-Dalgarno (S.D.) region (5′ GGAGG 3′) situated between 5 and 12 nucleotides upstream of the ATG initiation codon on mRNA species of B.subtilis. Since these expression vectors were designed prior to the DNA sequencing of the SPP1 fragment they had to take account of a likely open reading frame continuing from the Eco RI site. Termination codons were therefore introduced as shown in **Figure 34**, to minimise the possibility that the readthrough of a peptide would reduce the initiation efficiency. The sequence of the SPP1 fragment later showed that frame 1 encodes a fragment of phage peptide which will terminate at the S.D. box of R.B.S.1 in pFF 710 and at a stop codon overlapping the ATG initiator codon of pFF 730. Computer search of the resulting sequence in both expression vectors showed no obviously stable stem-loop secondary structure in this region. The synthetic oligomers were flanked by the appropriate restriction sites necessary in the construction and the second site alone included a Cla I cleavage site to ease identification of correctly cloned and orientated molecules prior to DNA sequencing. The synthetic fragments and resulting ribosome binding sites within the expression vectors are illustrated in **Figure 35**.

In order to fully evaluate the effect on translational efficiency of variations in factors such as the degree of S.D. box complementarity and the number and nature of the bases between this region and the ATG codon, a large number of different ribosome binding sites would need to be tested. To begin this sort of analysis, the second site used in this work was designed to show a greater degree of complementarity to the 16S rRNA subunit and the length of the spacer region used was more typical of those found on native mRNA molecules of B.subtilis. It might therefore be expected to initiate translation more efficiently. The effect of these differences were evaluated both in

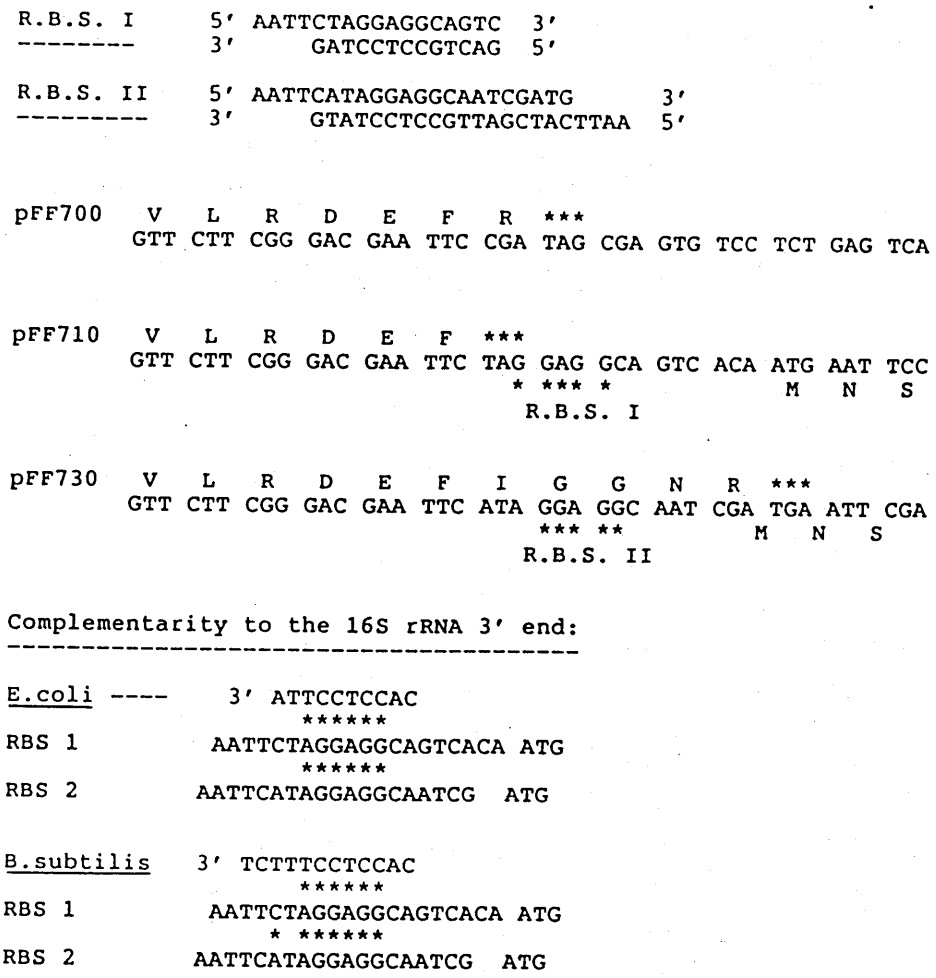


Figure 35. The synthetic ribosome binding sites.

The DNA sequence of the oligomers comprising the two synthetic ribosome binding sites is given at the top of the figure. The resulting ribosome binding sites within the expression plasmids are shown below. The highly conserved Shine-Dalgarno regions are indicated by asterisks below the DNA sequence in the case of plasmids pFF710 and pFF730. The N-terminal amino acid sequence of β-urogastrone is also shown below the DNA sequence of these plasmids. The C-terminal residues of the putative SPP1 peptide in each case is shown above the DNA sequence. Termination codons are indicated by asterisks above the DNA sequence. Complementarity of the Shine-Dalgarno regions to the 3′ end of the 16S rRNA in each host is shown at the bottom.

B.subtilis and also in E.coli which displays a far lesser stringency of sequences which can act as effective ribosome binding sites. Both of the synthetic sites could therefore be reasonably expected to function effectively in E.coli in which it seemed likely that differences in translational efficiency would be less pronounced than in Bacillus subtilis.

5.4 Expression of β-urogastrone in E.coli and B.subtilis

Cell extracts were prepared from cultures of HB101 alone and HB101 containing plasmids pFF710, pFF730, pFF700 and pFF740. Each was then examined for the presence of β-urogastrone by radio-immuno assay as described in Chapter 2. Similarly, extracts from cultures of 3G18 alone and 3G18 carrying plasmids pFF71, pFF73, pFF70 and pFF74, were prepared and assayed for β-urogastrone activity. The results (summarised in **Table** 3) were extremely inconclusive. No β-urogastrone activity was detected in any samples of either host in the absence of plasmid or in the presence of plasmids pFF700, pFF70, pFF740 or pFF74. However the levels of β-urogastrone detected in the other strains was extremely variable throughout the growth range of the cells over which samples were taken (A_{600} 0.3-1.0). The levels of β-urogastrone varied between 30-200 ng/ml of cells (corrected to an A_{600} of 1.0) in E.coli and between 30-150 ng/ml of cells (corrected) in B.subtilis. No pattern was evident and no consistent correlation possible between levels detected in cells expressing β-urogastrone from R.B.S. I or R.B.S. II. From these results the only conclusion possible is that β-urogastrone is synthesised using the SPP1 promoter and either synthetic ribosome binding site in both E.coli and B.subtilis. The levels are not significantly different between hosts, nor are they significantly different from the levels of β-urogastrone

Table 3

Cell Sample	ng/uro ml to A_{600} 1·0
HB 101	B.R.
HB 101+pFF 74·0	B.R.
HB 101+pFF 700	B.R.
HB 101+pFF 710	30 - 200
HB 101+pFF 730	30 - 200
3G18	B.R.
3G18 +pFF 70	B.R.
3G18+pFF 74	B.R.
3G18+pFF 71	30-150
3G18+pFF 73	30-150

Table 3. Expression of β-urogastrone in HB101 and 3G18 using the SPP1 promoter and the synthetic ribosome binding sites.

detected in cells expressing the native peptide under control of the trp promoter (described in Chapter 3).

Clearly the system designed for gene expression in B.subtilis was functional but accurate evaluation of the translational efficiency of the synthetic ribosome binding sites required a more consistent level of a detectable peptide to be present.

In an attempt to provide a faster and more simple means to assay gene expression the β-urogastrone gene was then fused to a fragment of the lac Z gene of E.coli generating a peptide fusion. This allows rapid and accurate detection of β-galactosidase using a simple assay or X-gal indicator medium and has been widely applied in the study of gene expression in heterologous systems. In addition, it had emerged from the further studies on the expression of the β-urogastrone gene using the trp expression vectors that the mature peptide had a much shorter half-life (2mins) than the Trp-E fused molecule (5 mins). Since the fragment of the Trp E peptide appeared to somehow stabilise the molecule in E.coli it seemed that fusion to β-galactosidase might have a similar effect and may therefore be a more sensitive and accurate measure of the expression efficiency of the shuttle vectors. Since much larger protein fusions to β-galactosidase have been previously used (100) it was likely that the enzyme would retain activity with an N-terminal fragment of β-urogastrone.

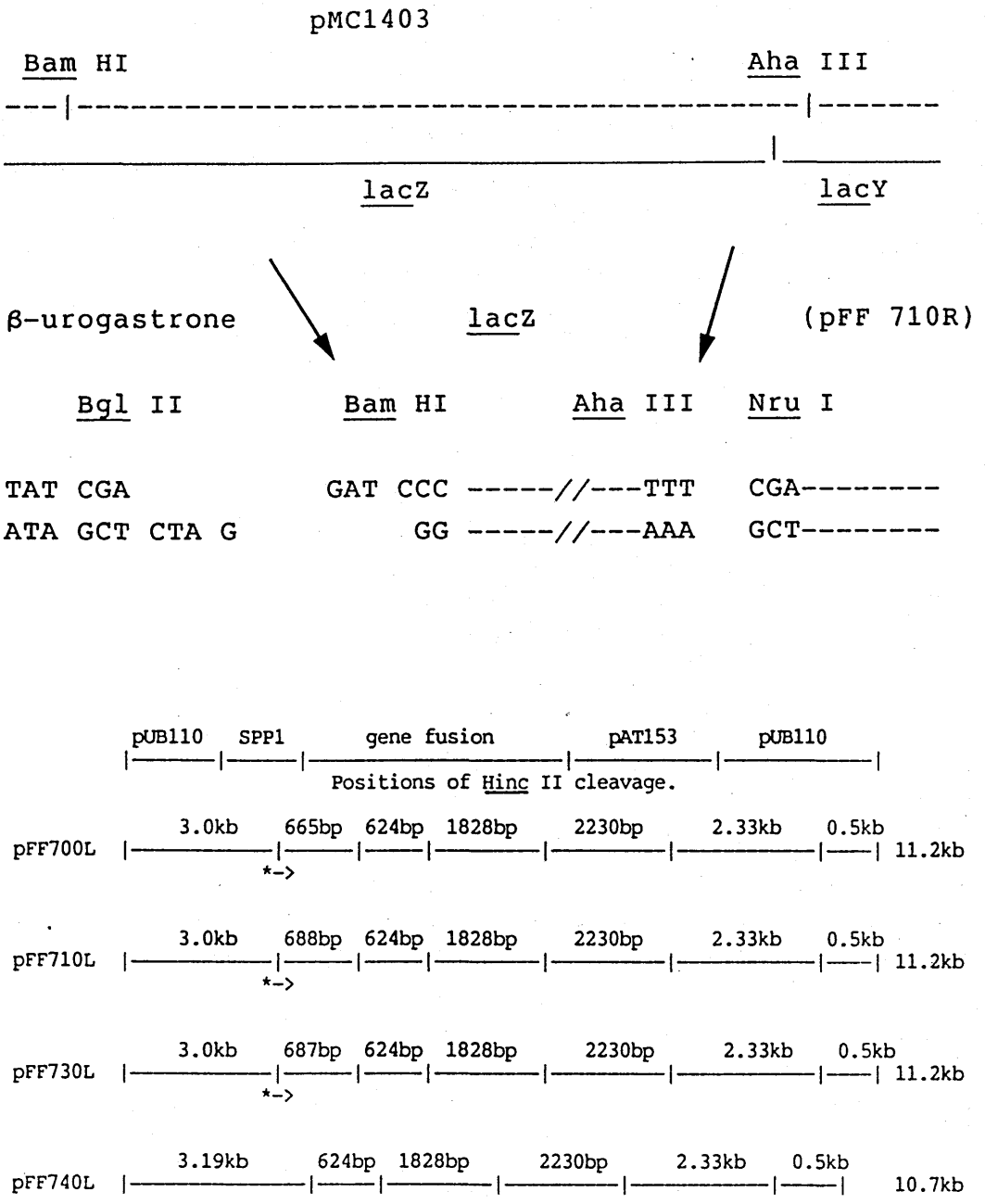


Figure 36. Generation of plasmid pFF710L.

The 3122bp, Bam HI to Aha III fragment of the lacZ gene from pMC1403, was inserted into plasmid pFF710R as shown at the top of the figure. The resulting plasmid pFF710L and derivatives, pFF730L, pFF700L and pFF740L are illustrated below with expected Hinc II cleavage fragments indicated. The position of the SPP1 promoter is also indicated (*->). The result of Hinc II cleavage of each plasmid is shown in Figure 33.

5.5 lacZ fusion peptides

A fragment encompassing almost all of the lacZ gene was excised from plasmid pMC1403 on a 3kb fragment as shown in **Figure 36**. The Bam HI site was ligated to the Bgl II site within the β-urogastrone gene fusing the two peptides in the correct reading frame. For the purpose of this construction a derivative of pFF710 was used in which the pAT153 sequence is in the reverse orientation. The Aha III site within the first few residues of the lac Y gene was ligated to the Nru I site within pAT153 since both enzymes generate blunt ends upon cleavage. The resultant fusion named pFF710L thereby included very little extraneous sequence. The small Xba I fragment from each of vectors pFF730, pFF 700, and pFF 740 was cloned into Xba I cleaved pFF710L to generate lacZ fusion equivalents of each of the vectors used to study native β-urogastrone expression. Since these vectors do not carry an easily excisable pAT 153 replicon this was not removed prior to transformation of B.subtilis. Oligomers were instead generated by cleavage at the unique Bgl II site within the kanamycin resistance gene of pUB110.

Since the host E.coli strain used has a deletion of the entire chromosomal lacZ gene and B.subtilis strain 3G18 possesses no β-galactosidase activity, all enzyme activity is derived from the plasmid borne fusions. Colonies of HB101 carrying plasmids pFF710 and pFF730 turned blue on X-gal indicator medium, as did 3G18 carrying plasmids pFF71 and pFF73. Relative blueness could not accurately be determined although in each host the presence of R.B.S. II led to a greater degree of blue colour. In HB101 the presence of plasmid pFF700 caused the development of a pale blue colour over 2-3 days indicating very low levels of β-galactosidase present. This must be due to fortuitous translational initiation on transcripts originating from the SPP1 promoter.

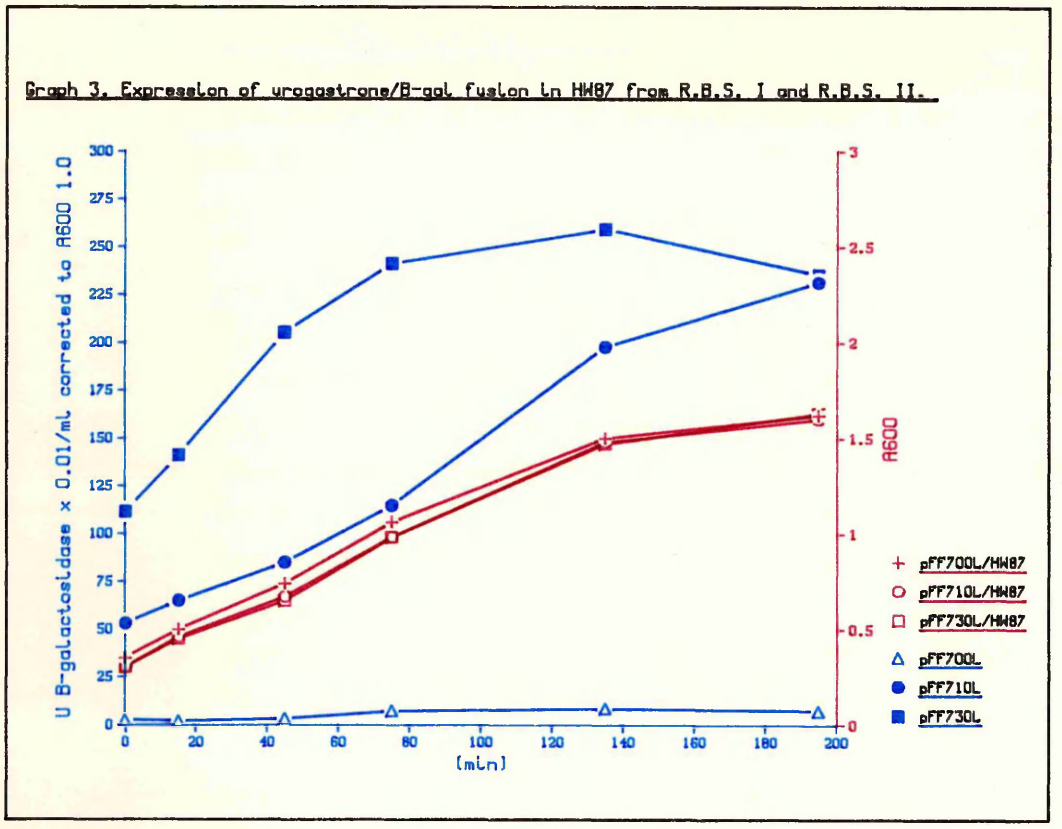


Table 4

	Sample	1	2	3	4	5	6
	Time	0	15	45	75	135	195
A_{600}	pFF700L	0·35	0·5	0·74	1·06	1·5	1·62
	pFF710L	0·3	0·46	0·67	0·98	1·48	1·6
	pFF730L	0·3	0·45	0·65	0·98	1·47	1·63
U.B-gal to A_{600} 1·0	pFF700L	236	185	330	716	848	713
	pFF710L	5287	6488	8485	11460	19760	23130
	pFF730L	11138	14097	20520	24090	25910	23524

Graph 3. and Table 4. Expression in HW87 of urogastrone/β-galactosidase fusion peptide.

5.6 Expression of urogastrone-Galactosidase peptide

The β-Galactosidase activity in both hosts carrying each of the four plasmids was now accurately determined by spectrophotometric assay and is shown in **Tables 4 and 5**. No β-galactosidase activity was detected as expected in either HW87 or 3G18 when no plasmid was present. Similarly in HW87 carrying plasmid pFF740L no activity was detected throughout the growth of the cells (A_{600} 0.3-1.4). However very low levels of activity were detected in HW87 carrying plasmid pFF700L and quite substantial levels were expressed from pFF710L and pFF730L. In 3G18 no β-galactosidase activity was detected throughout the growth of the cells carrying pFF70 or pFF74 (A_{600} 0.3-1.5) but low levels were detected in the presence of pFF71 and pFF73. The relative levels of enzyme were corrected for protein concentration in cell extracts and normalised to A_{600} 1.0 to allow comparison. Time courses of expression for each host/plasmid combination remained consistent over repeat experiments showing both in *E.coli* and *B.subtilis* a roughly twofold higher level of β-galactosidase activity in cells expressing *lacZ* fused urogastrone from R.B.S. II than R.B.S. I. This is illustrated in **Graphs 3 and 4**.

Copy numbers of plasmids from E.coli were estimated by densitometric scanning of plasmid extracts on 1% agarose gels, showing all four plasmids to have copy numbers of approximately 50. Although accurate copy number determinations could not be carried out in *B.subtilis* due to frequently observed inconsistency of plasmid DNA isolation it is unlikely that the structural differences of these plasmids, especially the minor differences between pFF71 and pFF73 would bring about significant differences in copy number. Extensive plasmid stability studies were not carried

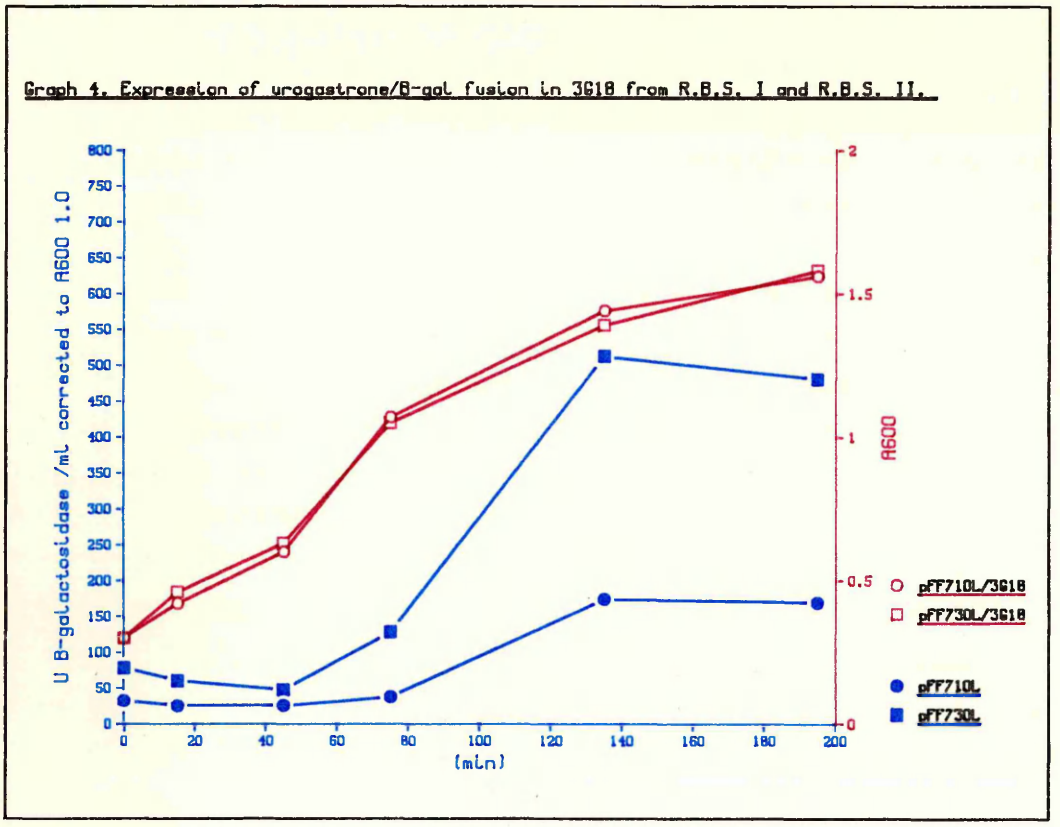


Table 5

	Sample	1	2	3	4	5	6
	Time	0	15	30	60	90	120
A_{600}	pFF 710L	0·3	0·42	0·6	1·07	1·44	1·56
	pFF 730L	0·3	0·46	0·63	1·05	1·39	1·58
UB-gal to A_{600} 1.0	pFF710L	320	245	250	376	1735	1685
	pFF 730L	778	594	469	1281	5124	4800

Graph 4. and Table 5. Expression in 3G18 of urogastrone/β-galactosidase fusion peptide.

out although plating on X-gal indicator medium following growth of all initially blue host/plasmid combinations showed no evidence of segregants reduced in blue colour.

5.7 Discussion

These results demonstrate that both synthetic ribosome binding sites are capable of initiating translation in B.subtilis and in E.coli. This supports the data on the expression of β-urogastrone. However, these results also show that higher β-galactosidase activities are observed when using R.B.S. II rather than R.B.S. I. Although additional data is required to exclude the possibility that the difference is due to variations in levels of synthesis or breakdown of mRNA, the more likely explanation is that the two ribosome binding sites have differing efficiencies of translational initiation. This could be due to subtle alterations in mRNA secondary structure or to the effect of the changes in the S.D. region affecting the interaction with 16S rRNA. No significant secondary structure was observed by computer search of the region around either ribosome binding site but it is interesting to note that R.B.S. II does display increased homology to the 3′ end of the 16S rRNA than R.B.S. I. In addition, translation from the open frame preceding the synthetic ribosome binding sites would be likely to influence the efficiency of translation. Although constraints of time and technical difficulties precluded accurate determinations of mRNA levels of the strains used in this work, the results do show that small sequence variations in synthetic ribosome binding sites can lead to large differences in observed protein levels in both organisms. The much lower enzyme activities obtained in B.subtilis may reflect the differing stringency of gene expression of the organisms or higher levels of proteolysis. It must be emphasised that these studies are

preliminary and represent only the beginning of the development of a suitable system for gene expression in B.subtilis. Both proteins used are to differing degrees foreign to both organisms. Ideally a native bacillus protein, or an E.coli protein showing high conservation of structure and function with B.subtilis would be more appropriate in studies of gene expression. Furthermore, the study by specific mutation of a known region of translational initiation in B.subtilis would allow a more controlled evaluation of the influence of minor sequence variation. The extension of the concepts examined in this work and the further development of this system to examine these possibilities are now considered.

CHAPTER 6

GENERAL DISCUSSION

AND

CONCLUSIONS

It is quite clear that the cloning in E.coli and expression of β-urogastrone demonstrated in these results, illustrates the value of total gene synthesis in achieving bacterial expression of a eukaryotic peptide. Very large quantities of human urine were previously required to obtain similar quantities of β-urogastrone to that detected in these preliminary expression studies. In addition the subsequent increases in bacterial expression levels have yielded much higher quantities of the peptide than could have otherwise been realistically obtained. Similar observations have of course been made upon the expression in bacteria of other human genes derived from natural sources. In this respect it is important to stress that many of the significant advantages evident in the application of sophisticated technology allowing peptide expression from a synthetic gene are largely practical. This is very important since technical difficulties can often be severely limiting in instances where theoretical knowledge is not , a fact exemplified by the problems discussed in the isolation of human mRNA and that of β-urogastrone in particular. It is also illustrated in the necessity for careful consideration of many aspects in the design of the gene. The value of the logic applied to the gene assembly, permiting relatively trouble free isolation of the full sequence and of restricion cleavage and gel purification were clear. These factors can largely influence the value of this approach as a faster and more feasible alternative to cDNA or chromosomal gene cloning. The benefit of many of the other factors included in the gene design can only be evaluated after extensive studies on the expression of the gene. For example the inclusion of the facility for peptide fusions did allow more accurate assessment of gene expression but also showed interesting differences in the protease sensitivity of the native and Trp E fused β-urogastrone molecules. This illustrates the point that perhaps the greatest value of this work lies not neccesarily in the demonstration of

expression of the synthetic gene but in the facility to easily modify the peptide itself or the expression system. This has already potentiated many studies arising from observations made upon β-urogastrone expression.

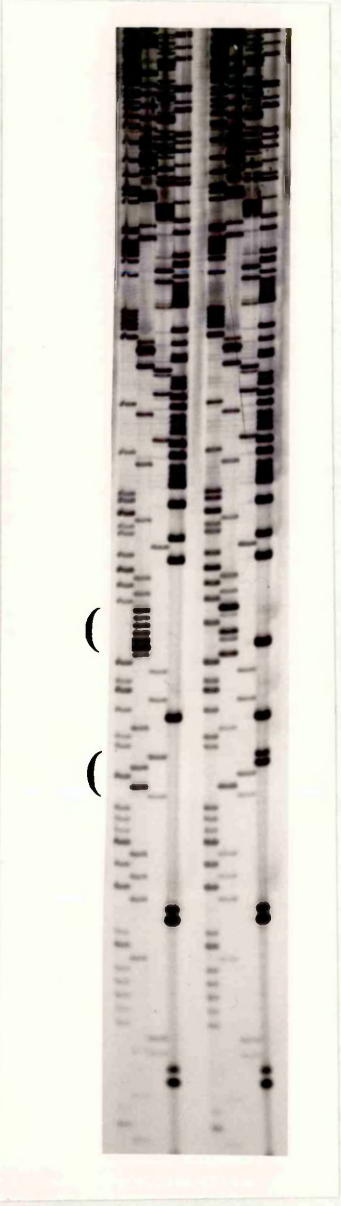
Programmes of work are currently investigating the different susceptibilities of the native and Trp E fused molecules to proteolytic degradation in E.coli. Expression of β-urogastrone has been used to evaluate expression from plasmids containing tandemly repeated trp promoters and from plasmids carrying mutations in copy number regulation. In addition modifications have been made to the N-terminal linker region to leave only a single lysine residue. This follows observations of incompletely cleaved protein following treatment with trypsin. Modifications have also been made to the C-terminal region to incorporate a stretch of arginine residues. This has allowed the development of a novel technique of peptide purification (101). Since the C-terminal residue of native β-urogastrone is an arginine residue both the N-terminal and C-terminal peptides can be removed by trypsin digestion. The combination of these developments has enabled the isolation from E.coli, and purification to homogeneity of a several hundred milli-grammes of β-urogastrone. This has enabled current investigations into the crystallisation and the determination of the 3-D structure of the molecule and the commencement of extensive clinical trials to examine its medical potential.

The examination of the expression of β-urogastrone in B.subtilis did not show significant differences in maximal levels of the peptide detected by radio-immuno assay from that detected in E.coli. Although the apparent levels of peptide present were lower, the inconsistency of the results allow no realistic evaluation of gene expression. It is highly likely that rapid proteolysis of the peptide occurs

in B.subtilis as has now been shown in E.coli. Furthermore, the high levels of β-urogastrone which have now been purified from E.coli have removed the immediate necessity for the use of an alternative host organism as a source of the peptide. It is therefore important to distinguish at this point the two aspects of the study of β-urogastrone expression in B.subtilis. Although an initial evaluation of the levels of the peptide were desired in the consideration of B.subtilis as a suitable alternative host, the synthetic gene was used also in the study of regulatory aspects of gene expression. The use of β-galactosidase fusions to permit a more accurate and reproducible evaluation of expression, illustrates that a gene of prokaryote origin would have been more appropriate for this latter purpose. A peptide native to B.subtilis or possibly an E.coli peptide would have been more likely to be resistant to intracellular proteolysis therby introducing fewer variables to the examination. Had an appropriate gene been available at the time, a more controlled assessment of the function of synthetic ribosome binding sites could have been undertaken. In addition constraints upon time did not permit the full characteristion of the promoter region prior to the design of these expression systems. This neccessitated the inclusion of termination codons within the synthetic ribosome binding sites which led to variations in the termination points of the second open reading frame of the phage fragment. Nevertheless, several aspects of this study do indicate the worth of the development of this system. The SPP1 promoter used has previously been shown to be a very efficient promoter and is likely to be very effective in the expression of heterologous genes. It is likely that the 52 residue peptide is efficiently expressed in B.subtilis. Since the exact point of transcription initiation is now known this promoter could form the basis of a more useful expression system. It is also clear that synthetic ribosome binding sites can be used in the initiation of translation

in B.subtilis. Although it is not possible to determine the reason for the differences in observed expression the design of these regions can clearly influence the efficiency of translation. Therefore a more controlled system based upon these results may be very effective in B.subtilis. To pursue this possibility two separate alterations have been made to the SPP1 fragment by site directed mutagenesis which should allow the expression of alternative genes. Transcription would again be under control of this promoter and translation initiated from the ribosome binding site of the small phage peptide. The structure of these variants are illustrated in appendix 1. They have been designed to introduce specific restriction enzyme sites at the initiation region of the small phage peptide allowing alternative genes to be inserted and expression examined using modified versions of the native regulatory region of the phage. The restriction sites introduced (Eco RI and Bam HI) have been chosen for the particular inclusion of two genes now available which should allow more effective analysis of the effect of minor changes in promoter and ribosome binding site. These are: i) the aspC gene of E.coli K12, encoding the aspartate aminotrasferase (AAT). This has been recently cloned in this laboratory (103), and ii) exonuclease II of Staphylococcus aureus (J-I. Flock personal communication). These genes contain the appropriate restriction sites for combination with the SPP1 variants. The aspartate aminotransferase can be over-expressed one-hundred fold in E.coli without significant proteolytic degradation and the exonuclease can similarly be expressed to very high levels in S.aureus. The vectors used for gene expression in B.subtilis in this work were clearly functional to a degree. The study of more appropriate subject peptides should be more useful in the development of an expression system. In this event the eventual

introduction of the β-urogastrone gene at this Eco RI site would remain a possibility and the expression of the peptide in B.subtilis could yet represent a viable option.



Wild type:
AGAGAGGGGGGGACAAC ATG AAC GAG CAA AAG AGA
        * *   * **
AGAGAGGAGGTGACAAC ATG AAT TCG CAA AAG AGA
Mutant:

Appendix 1. Figure 1. Modification to the SPP1 fragment by site-directed mutagenesis.

Track 1; Wild type sequence of SPP1 DNA. Track 2; Sequence of the mutation introduced to the SPP1 fragment. Both tracks show sequence lanes as read, in the order A.G.C.T. Position of the sequence changes is indicated by brackets. The wild type and mutant sequences are illustrated below, with differences indicated by asterisks. The resulting Eco RI site of the mutant is underlined.

APPENDIX 1

Two mutations were introduced by site-directed mutagenesis, within the ribosome binding site of the small phage peptide. These introduced cleavage sites for the restriction endonucleases Eco RI and Bam HI. The wild type sequence and that of the two mutants are shown below. The restriction endonuclease cleavage sites are underlined. The sequence begins with the first base of the mRNA. The DNA sequence of the wild type and mutant 1 are shown in **Figure 1**.

wild type:
ATAGAGAGGGGGGGACAAC ATG AAC GAG CAA AAG AGA

mutant 1:
ATAGAGAGGAGGTGACAAC ATG AAT TCG CAA AAG AGA

mutant 2:
ATAGAGAGGAGGATCCAAC ATG AAC GAG CAA AAG AGA

These mutations permit the insertion of the S.aureus exonuclease gene (mutant 1) and the E.coli aspC gene (mutant 2). The exonuclease gene possesses an Eco RI site at the N-terminus of the peptide and the aspC gene possesses a Bam HI site introduced by site-directed mutagenesis within the ribosome binding site. The sequence of the N-terminal region of the mutated aspC gene and the resulting SPP1 derived expression vector are shown overleaf. The Bam HI site is underlined.

wild type aspC gene:

```
ATAATGGAACCTCGTC ATG TTT GAG
```

mutated aspC gene:

```
ATAATGGATCCTCGTC ATG TTT GAG
                 M   F   E
```

SPP1-aspC expression plasmid:

```
ATAGAGAGGAGGATCCTCGTC ATG TTT GAG
                      M   F   E
```

APPENDIX 2

Molecular weight markers comprising restriction digests of the DNA of phages λ and ϕX174 generated fragments of sizes shown below:

λ *Hind* III digest.

1.	23,130 bp
2.	9,419 bp
3.	6,557 bp
4.	4,371 bp
5.	2,322 bp
6.	2,028 bp
7.	564 bp
8.	125 bp

ϕX174 *Hae* III digest.

1.	1,353 bp
2.	1,078 bp
3.	872 bp
4.	603 bp
5.	310 bp
6.	281 bp
7.	271 bp
8.	234 bp
9.	194 bp
10.	118 bp
11.	72 bp

REFERENCES

1. Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A.D. (1979) Expression in Escherichia coli of chemically synthesised genes for human insulin. Proc.Natl.Acad.Sci, 76:106-110.

2. Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Miozzari, G., Crea, R., and Seeburg, P.H. (1979) Direct expression in Escherichia coli of a DNA sequence coding for human growth hormone. Nature, 281:544-548.

3. Itoh, S., Mizukami, T., Matsumoto, T., Nishi, T., Saito, A., Oka, T., Furuya, A., Takaoka, C., and Taniguchi, T. (1984) Efficient expression in Escherichia coli of a mature and a modified human Interferon-β_1. DNA, Vol 3.(2) 157-165.

4. Manjusri, D. (1982) Epidermal Growth Factor:Mechanisms of action. International review of Cytology. Vol 78:233-256.

5. Gregory, H., Holmes, J.E., Willshire, I.R. (1977) Urogastrone levels in the urine of normal adult humans. J. Clin. Endocrinol. Metabolism, 45:668-672.

6. Heitz, P.U., Kasper, M., Van Noorden, S., Polak, J.M., Gregory, H., and Pearse, A.G.E. (1978) Immunohistochemical localisation of urogastrone to human duodenal and submandibular glands. Gut, 19:408-413.

7. Elder, J.B., Williams, G., Lacey, E., Gregory, H. (1978) Cellular localisation of human urogastrone/epidermal growth factor. Nature, 271:466-467.

8. Gregory, H. (1975) Isolation and structure of urogastrone and its relation to epidermal growth factor. Nature, 257:325-327.

9. Savage, C.R., Inagami, T. and Cohen, S. (1972) The primary structure of epidermal growth factor. J.Biol.Chem, 247:7612-7621.

10. Koffman, C.G., Elder, J.B., Ganguli, P.C., Gregory, H., and Geary, C.G. (1982) Effect of urogastrone on gastric secretion and serum gastrin concentration in patients with duodenal ulceration. Gut, 23:951-956.

11. Bower, J.M., Camble, R., Gregory, H., Gerring, E.L., and Willshire, I.R. (1975) The inhibition of gastric acid secretion by epidermal growth factor. Experientia, 31:825-6.

12. Hollenberg, M.D., and Gregory, H. (1977) Human urogastrone and mouse epidermal growth factor share a common receptor in cultured human fibroblasts. Life. Sciences, 20:267-274.

13. Hollenberg, M.D., and Gregory, H. (1980) Epidermal growth factor-Urogastrone: Biological activity and receptor binding of derivatives. Mol. Pharmacol, 17:314-320.

14. Elder, J.B., Ganguli, P.C., Gillespie, I.E., Delamore, W.I., and Gregory, H. (1975) Effect of urogastrone in the Zollinger-Ellison syndrome. Lancet, 2:424-427.

15. Davis, R.J., Czech, M.P. (1984) Tumour-promoting phorbol diesters mediate phosphorylation of the epidermal growth factor receptor. J.Biol.Chem, 259:8545-8549.

16. Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J., and Waterfield, M.D. (1984) Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. Nature, 307:521-527.

17. Massague, J. (1983) Epidermal growth factor-like transforming growth factor. J.Biol.Chem, 258:13606-13613.

18. Abelson, J. (1979) RNA processing and the intervening sequence problem. Ann.Rev.Biochem, 48:1035-69.

19. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) in Molecular Cloning, A laboratory manual, pp270-274.

20. Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H., Bolivar, F., Boyer, H. (1977) Expression in Escherichia coli of a chemically synthesised gene for the hormone somatostatin. Science, 198:1056-1063.

21. Hsuing, H.M., Brousseau, R., Michniewicz, J., and Narang, S. (1979) Synthesis of human insulin gene. Part 1. Development of reversed phase chromatography in the modified triester method. Its application in the rapid and efficient synthesis of eight deoxyribonucleotide fragments constituting human insulin A DNA. Nucleic acids Res, 6:1371-1385.

22. Grantham, R., Gautier, C., and Gouy, M. (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res, 8:1893-1912.

23. Ikemura, T. (1981) Correlation between the abundance of Escherichia coli transfer RNA's and the occurence of the respective codons in its protein genes. J.Mol.Biol, 146:1-21.

24. Gitt, M.A., Wang, L., Doi, R. (1985) A strong sequence homology exists between the major RNA polymerase σ factors of Bacillus subtilis and Escherichia coli. J.Biol.Chem. Supplemental material, 260:7184-7185

25. Ogasawara, N. (1985) Markedly unbiased codon usage in Bacillus subtilis. Gene, 40:145-150.

26. Borer, P.N., Dengler, B., Tinoco, I. (1974) Stability of ribonucleic acid double-stranded helices. J.Mol.Biol, 86:843-853.

27. Sreedhara Swamy, K.H., and Goldberg, A.L. (1981) E.coli contains eight soluble proteolytic activities, one being ATP dependent. Nature, 292:652-654.

28. Geggenheimer, P., and Apirion, D. Processing of procaryotic ribonucleic acid. (1981) Microbiol.Rev, 45:502-541

29. Amann, E., Brosius, J., and Ptashne, M. (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli. Gene, 25:167-78.

30. Hawley, D.K., and McClure, W.R. (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res, 11:2237-2255.

31. Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M., and Sauer, R.T. (1980) How the λ repressor and Cro work. Cell, 19:1-11.

32. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S., and Stormo, G. (1981) Translational initiation in prokaryotes. Ann.Rev.Microbiol, 35:365-403.

33. Iserentant, D., and Fiers, W. (1980) Secondary structure of mRNA and efficiency of translation initiation. Gene, 9:1-12.

34. Rosenberg, M., and Court, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann.Rev.Genetics, 13:319-353.

35. Yanofsky, C., Platt, T., Crawford, I.P., Nichols, B.P., Christie, G.E., Horowitz, H., VanCleemput, M., and Wu, A.M. (1981) The complete nucleotide sequence of the tryptophan operon of E.coli. Nucleic Acids Res, 9:6647-68.

36. Velan, B., Cohen, S., Grosfeld, H., Leitner, M., and Shafferman, A. (1985) Bovine interferon genes. Structure and expression. J.Biol.Chem, 260:5498-5504.

37. Charnay, P., Perricaudet, M., Galibert, F., Tiollais, P. (1978) Bacteriophage lambda and plasmid vectors allowing fusion of cloned genes in each of the three translational phases. Nucleic Acids Res, 5:4479-4494.

38. Tacon, W., Carey, N., and Emtage, S. (1980) The construction and characterisation of plasmid vectors suitable for the expression of all DNA phases under the control of the E.coli tryptophan promoter. M.G.G, 177:427-438.

39. Tacon, W.C.A., Bonass, W.A., Jenkins, B., and Emtage, J.S. (1983) Expression plasmid vectors containing Escherichia coli tryptophan promoter transcriptional units lacking the attenuator. Gene, 23:255-265.

40. Aksoy, S., Squires, C.L., and Squires, C. (1984) Translational coupling of the trpB and trpA genes in the Escherichia coli tryptophan operon. J.Bact, 157:363-367.

41. Talmadge, K., and Gilbert, W. (1980) Construction of plasmid vectors with unique Pst I cloning sites in a signal sequence coding region. Gene, 21:235-241.

42. Miller, J.H. (1972) in Experiments in Molecular Genetics, Cold Spring Harbour Laboratory. pp 402-404.

43a. Bolivar, F., Rodriguez, R.L., Betlach, M.C., and Boyer, H.W. (1977a) Construction and characterisation of new cloning vehicles. I. Ampicillin resistant derivatives of the plasmid pMB9. Gene 2:75-93.

43b. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977b) Construction and characterisation of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95-113.

44. Blattner, F.R., Williams, B.G., Blechl, A.E., Deniston-thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., and Smithies, O. (1977) Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science, 196:161-9.

45. Messing, J., and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene, 19:269-276.

46. Vieira, J., and Messing, J. (1982) The pUC plasmids, an M13mp-7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene, 19:259-268.

47. Linn, T., and Ralling, G. (1985) A versatile multiple and single copy vector system for the in vitro construction of transcriptional fusions to lacZ. Plasmid, 14:134-142.

48. Keay, L. (1972) Proteases of the genus bacillus. in Proceedings of the IV I.F.F. "Fermentation technology today" pp289-298.

49. Ohmura, K., Shiroza, T., Nakamura, K., Nakayama, A., Yamane, K., Yoda, K., Yamasaki, M., and Tamura, G. (1984) A Bacillus subtilis secretion vector system derived from the Bacillus subtilis alpha amylase promoter and signal sequence region, and secretion of Escherichia coli β-lactamase by the vector system. J.Biochem, 95:87-93

50. Ehrlich, S.D. (1978) DNA cloning in Bacillus subtilis. Proc.Natl.Acad.Sci, 75:1433-1436.

51. Williams, D.M., Duvall, E.J., and Lovett, P.S. (1981) Cloning restriction fragments that promote expression of a gene in Bacillus subtilis. J.Bacteriol, 146:1162-1165.

52. Gryczan, T.J., Contente, S., and Dubnau, D. (1978) Characterisation of Staphylococcus aureus plasmids introduced by transformation into Bacillus subtilis. J.Bact, 134:318-329.

53. Williams, D.M., Schoner, E.J., Duvall, E.J., Preis, L.H., and Lovett, P.S. (1981) Expression of Escherichia coli trp genes and the mouse dihydrofolate reductase gene cloned in Bacillus subtilis. Gene, 16:199-206.

54. Yansura, D.G., and Henner, D. (1984) Use of the Escherichia coli lac repressor and operator to control gene expression in Bacillus subtilis. Proc.Natl.Acad.Sci, 81:439-443.

55. Band, L., Yansura, D.G., and Henner, D.J. (1983) Construction of a vector for cloning promoters in Bacillus subtilis. Gene, 26:313-315.

56. Goze, A., and Ehrlich, S.D. (1980) Replication of plasmids from Staphylococcus aureus in Escherichia coli. Proc.Natl.Acad.Sci, 77:7333-7337.

57. Gryczan, T.J., and Dubnau, D. (1978) Construction and properties of chimeric plasmids in Bacillus subtilis. Proc.Natl.Acad.Sci, 75:1428-1432.

58. Uhlen, M., Flock, J-I., and Philipson, L. (1981) RecE independent deletions of recombinant plasmids in Bacillus subtilis. Plasmid, 5:161-169.

59. Primrose, S.B., and Ehrlich, S.D. (1981) Isolation of plasmid deletion mutants and a study of their instability. Plasmid, 6:193-201.

60. Duvall, F.J., Williams, D.M., Lovett, P.S., Rudolph, C., Vasantha, N., and Guyer, M. (1983) Chloramphenicol inducible gene expression in Bacillus subtilis. Gene, 24:171-177.

61. Dubnau, D., and Cirigliano, C. Genetic characterisation of recombination deficient mutants of Bacillus subtilis. J.Bact, 117:488-493.

62. Canosi, U., Morelli, G., and Trautner, T.A. (1978) The relationship between molecular structure and transformation efficiency of some Staphylococcus aureus plasmids isolated from Bacillus subtilis. M.G.G., 166:259-267.

63. Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S., and Gross, C.A. (1981) The nucleotide sequence of the cloned rpoD gene for the RNA polymerase sigma subunit from E.coli K12. N.A.R, 9:2889-2904.

64. Landick, R., Vaughn, V., Lau, E.T., VanBogelen, R.A., Erickson, J.W., and Neidhardt, F.C. (1984) Nucleotide sequence of the heat shock regulatory gene of E.coli suggests its protein product may be a transcription factor. Cell, 38:175-182.

65. Gitt, M.A., Wang, L., Doi, R. (1985) A strong sequence homology exists between the major RNA polymerase σ factors of Bacillus subtilis and Escherichia coli. J.Biol.Chem. 260:7178-7185.

66. Losick, R., and Pero, J. (1981) Cascades of sigma factors. Cell, 25:582-584.

67. Doi, R.H. (1982) Multiple RNA polymerase holoenzymes exert transcriptional specificity in Bacillus subtilis. Arch.Biochem.Biophys, 214:772-781.

68. Talkington, C., and Pero, J. (1979) Distinctive nucleotide sequences of promoters recognised by RNA polymerase containing a phage coded "σ-like" protein. Proc.Natl.Acad.Sci, 76:5465-5469.

69. Achberger, E.C., and Whiteley, H.R. (1981) The role of the δ peptideof the Bacillus subtilis RNA polymerase in promoter selection. J.Biol.Chem, 256:7424-7432.

70. Moran, C.P., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Soneneshein, A.L., Pero, J., and Losick, R. (1982) Nucleotide sequences that signal the initiation of transcription and translation in Bacillus subtilis. M.G.G, 186:339-346.

71. Behrens, B., Luder, G., Behncke, M., Trautner, T.A., and Ganesan, A.T. (1979) The genome of B.subtilis phage SPP1. Physical arrangement of phage genes. M.G.G, 175:351-357.

72. Esche, H., Schweiger, M., and Trautner, T. (1975) Gene expression of bacteriophage SPP1. Phage directed protein synthesis. M.G.G, 142:45-55.

73. Stuber, D., Morelli, G., Bujard, H., Montenegro, M.A., and Trautner, T.A. (1981) Promoter sites in the genome of B.subtilis phage SPP1. M.G.G, 181:518-521.

74. Montenegro, M.A., and Trautner, T.A. (1981) In vivo transcription of Bacillus subtilis bacteriophage SPP1. M.G.G, 181:512-517.

75. Costanzo, M., and Pero, J. (1983) Structure of a Bacillus subtilis bacteriophage SPO1 gene encoding a RNA polymerase σ factor. Proc.Natl.Acad.Sci, 80:1236-1240.

76. Tailor, R., Bensi, G., Morelli, G., Canosi, U., and Trautner, T.A. (1985) The genome of Bacillus subtilis phage SPP1: Structure of an early promoter. J.Gen.Microbiol, 131:1259-1262.

77. Shine, J., and Dalgarno, L. (1975) Determinant of cistron specificity in bacterial ribosomes. Nature, 254:34-38

78. Band, L., and Henner, D.J. (1984) Bacillus subtilis requires a "stringent" Shine-dalgarno region for gene expression. DNA, 3:17-21.

79. Dubnau, E., Pifko, S., Sloma, A., Cabane, K., and Smith, I. (1976) Conditional mutants in the translational apparatus of Bacillus subtilis. M.G.G, 147:1-12.

80. Nomura, M. (1973) Assembly of bacterial ribosomes. Science, 179:864-873.

81. Yaguchi, M., Visentin, L.P., Nazar, R.N., and Matheson, A.T. (1978) in "Biochemistry of thermophily" (S.M.Friedman ed) pp169-191 Academic Press New York and Higo, K.I., Itoh, T., Kumazaki, T., and Osawa, S. (1980) in "RNA Polymerase tRNA and ribosomes: Their genetics and evolution" (S.Osawa, H.Oseki, H.Uchida and T.Yura eds.), pp655-666. Univ. Tokyo Press, Tokyo.

82. Isono, S., and Isono, K. (1975) Role of ribosomal protein S1 in protein synthesis: Effects of its addition to Bacillus stearothermophilus cell-free system. Eur.J.Biochem, 56:15-22.

83. Boyer, H.W., and Roulland-Dussoix, D. (1969) A complementation analysis of the restriction and modification of DNA in Escherichia coli. J.Mol.Biol, 41:459-472.

84. Nugent, M.E., Primrose, S.B., and Tacon, W.C.A. (1983) in "Developments in Industrial Microbiology" a publication of the Society for Industrial Microbiology-1983. Vol.24 Chap.21 pp271-285.

85. Messing, J., Crea, R., and Seeburg, P.H. (1981) A system for shotgun DNA sequencing. Nucleic Acids Res, 9:309-321.

86. Arwert, F., and Venema, G. (1975) Transformation in Bacillus subtilis. Fate of newly introduced transforming DNA. M.G.G, 123:185-198.

87. Twigg, A.J., and Sherratt, D. (1980) Trans-complementable copy-number mutants of plasmid ColE1. Nature, 283:216-218.

88. Casadaban, M.J., Chou, J., and Cohen, S.N. (1980) In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: Escherichia coli plasmid vectors for the detection and cloning of translational initiation signals. J.Bact, 143:971-980.

89. Chen-Pei, D., Jay, E., Bahl, C.P., and Wu, R. (1976) A reliable mapping method for sequence determination of oligodeoxyribonucleotides by mobility shift analysis. Anal.Biochem, 74:73-93.

90. Ito, H., Ike, Y., Ikuta, S., and Itakura, K. (1982) Solid phase synthesis of polynucleotides. VI. Further studies on polystyrene co-polymers for the solid support. Nucleic Acids Res, 10:1755-1769.

91. Birnboim, H.C., and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res, 7:1513-1523.

92. Cohen, S.N., Chang, A.C.Y., and Hsu, L. (1972) Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of Escherichia coli by R-factor DNA. Proc.Natl.Acad.Sci, 69:2110-2114.

93. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc.Natl.Acad.Sci, 74:5463-5467.

94. Zoller, M.J., and Smith, M. (1982) Oligonucleotide directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. Nucleic Acids Res, 10:6487-6500.

95. Edge, M.D., Greene, A.R., Heathcliffe, G.R., Meacock, P.A., Schuch, W., Scanlon, D.B., Atkinson, T.C., Newton, C.R., and Markham, A.F. (1981) Total synthesis of a human leukocyte interferon gene. Nature, 292:756-762.

96. Gottesman, S., and Zipser, D. (1978) Deg phenotype of Escherichia coli lon mutants. J.Bact, 133:844-51.

97. Ratcliff, S.W., Luh, J., Ganesan, A.T., Behrens, B., Thompson, R., Montenegro, M.A., Morelli, G., and Trautner, T.A. (1979) The genome of Bacillus subtilis phage SPP1. The arrangement of restriction endonuclease generated fragments. M.G.G, 168:165-172.

98. Staden, R. (1984 a,b,c) a)Computer methods to locate signals in nucleic acid sequences. b) Graphic methods to determine the function of nucleic acid sequences. c) Measurements of the effects that coding for a protein has on a DNA sequence and their use for finding genes. Nucleic Acids Res, (a) 12:505-519, (b) 12:521-538 (c) 12:551-67.

99. Maxam, A.M., and Gilbert, W. (1977) A new method for sequencing DNA. Proc.Natl.Acad.Sci, 74:560-564.

100. Silhavy, T.J., Bassford, P.J., and Beckwith, J.R. (1979) in "Bacterial outer membranes: Biogenesis and functions." ed. Inouye, M. New York. Wiley Chap.4.pp 218-232.

101. Smith, J.C., Derbyshire, R.B., Cook, E., Dunthorne, L., Viney, J., Brewer, S.J., Sassenfeld, H.M., and Bell, L.D. (1984) Chemical synthesis and cloning of a poly(arginine)-coding fragment designed to aid polypeptide purification. Gene, 32:321-327.

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

103. Fotheringham, I.G., Dacey, S.A., Taylor, P.P., Smith, T.J., Hunter, M.G., Finlay, M.E., Primrose, S.B., Parker, D.M., and Edwards, M.E. (1986) The cloning and sequence analysis of the aspC and tyrB genes from Escherichia coli K12. Biochem.J. (February 1986 in press).