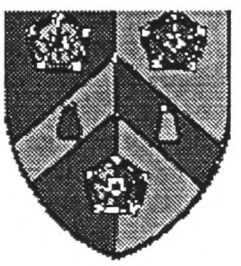
Expression and Mutagenesis of Bacteriorhodopsin, an Integral Membrane Protein

A thesis submitted to the Board of the Faculty of Physical Sciences of the University of Oxford in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Inderjit Kaur Sidhu



Wolfson College, Oxford

Trinity Term 1998

Expression and Mutagenesis of Bacteriorhodopsin; an Integral Membrane Protein

Inderjit K. Sidhu, Wolfson College, Oxford

A thesis submitted to the University of Oxford for the degree of D.Phil

Trinity term, 1998

Although integral membrane proteins represent nearly a quarter of the genes present in both prokaryotes and eukaryotes, progress in this area of research is often hindered due to the nature of their hydrophobic environment. Elucidating the folding pathway of these proteins is essential to understand many membrane mediated biological processes such as signal transduction, ion transport and chemotaxis.

The wealth of structural and genetic information on bacteriorhodpsin renders it an ideal model system for the study of membrane proteins. Detailed studies however, necessitate efficient methods for its overexpression and purification. Previous expression systems have reported difficulty in obtaining good yields and simple purification procedures.

This thesis investigates a variety of alternative expression and purification systems for the bacterio-opsin gene in *Escherichia coli*. With sufficient protein, site directed mutagenesis is performed to mutate three proline residues present in the membranous region of bacteriorhodopsin to alanine. The folding kinetics of these mutants is investigated using stopped flow fluorimetry to determine whether proline isomerisation is responsible for a slow step in the folding pathway of bacteriorhodopsin. Comparison of the results with those of the folding kinetics of wild type showed proline isomerisation not to be responsible for the slow step in the pathway. More recent studies have suggested that the slow step may be due to refolding conditions and lateral pressure the lipids impose upon the protein as well as pH.

Separate structural studies using mass spectrometry aimed to study the rates of isotopic exchange of amide and side chain protons in bacteriorhodopsin. Low resolution results obtained using matrix assisted laser desorption ionisation mass spectrometry (MALDI-MS) prompted the investigation of electrospray ionisation mass spectrometry (ESI-MS). Techniques for sample preparation were optimised by investigating a variety of solvent systems and initial deuteration experiments performed.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof. Sabine L. Flitsch for all her help and support throughout this project as well as my co-supervisor Dr. Paula J. Booth. Many thanks go to Dr. Karen Sage and especially Dr. Dominic Campopiano for their advice and guidance on molecular biology procedures. Thanks go to Evonne W. Chung and Dr. Carol V. Robinson for the mass spectrometry studies and to Dr. Hui Lu for help with the kinetic studies using stopped-flow fluorimetry. I would also like to thank Prof. H. G. Khorana for supplying the *bO* gene and to Dr. Koichi Koyama for supplying the *bR* antibody. Thanks go to Dr. Gareth Roberts for proof-reading this thesis. Finally, thanks go to EPSRC for their financial support throughout the course of this project.

Special thanks to my family - Mum Ji, Dad Ji, Harjit (Haj), Rajinder (Raj), Harpreet (Bits 'n' pieces) and Gurjit (Stevie Boy). I would especially like to thank Lynnette for always being there as well as flatties Hercules, Frenchie, Bella, Scotlass, Mar, Jezza and Corin. Thanks also go to 'the Lardy Boys', 'the amorphous love blob', the 'Jesus lads' and my 'Boating Buddies'.

Last but not least, thanks to the Wright brothers for inventing the aeroplane and thereby enabling me to see the wonders of the world during my research years.

For mum
- simply the best

Contents

Contents.....v
Abbreviations.....viii

Chapter 1: Introduction

1.1 Introduction.....2
1.1.1 Membrane protein folding.....2
1.1.2 Bacteriorhodopsin as a protein folding model.....4
1.1.3 Bacteriorhodopsin as a protein folding pump.....9
1.1.4 Folding kinetics of bacteriorhodopsin.....12
1.1.5 Protein production for structure-function studies.....14

1.2 Heterologous expression system for bacteriorhodopsin.....15
1.2.1 *Halobium salinarium*.....15
1.2.2 Yeast.....19
1.2.3 *Escherichia coli*.....21

1.3 Aims of this project.....26

Chapter 2: Materials and methods

2.1 Materials.....29
2.1.1 General reagents.....29
2.1.2 Enzymes.....29
2.1.3 Commercially available kits.....29
2.1.4 Monoclonal antibodies for bR.....29
2.1.5 DNA vectors.....30
2.1.6 Bacterial cell stocks.....30
2.1.7 Laboratory stocks.....31
2.1.8 General solutions and buffers.....31
2.1.9 Media.....31

2.2 Equipment.....32

2.3 Methods.....33
2.3.1 Microbiology of *Escherichia coli*.....33
2.3.2 Manipulation and analysis of DNA.....35
2.3.3 Analysis of protein.....41
2.3.4 Protein purification.....42
2.3.5 Growth of *Halobacterium salinarium*.....46
2.3.6 Analysis of membrane proteins by mass spectrometry.....49

Chapter 3: Develpoment of a prokaryotic expression system for the *bO* gene

3.1 Introduction.....52

3.2 Results and discussion.....53
3.2.1 *bO* gene manipulation to incorporate an *NdeI* site.....53
3.2.2 General methodology employed to generate expression vectors containing *bO* gene.....55
3.2.2.1 Studies using the pPL1 expression vector.....55
3.2.2.2 Studies using the pRH1090 expression vector.....60
3.2.2.3 Studies using the pT7-7 expression vector.....64
3.2.2.4 Studies using the pET16b expression vector.....66
3.2.2.5 Studies using the pCYB2 expression vector.....72
3.2.2.6 Studies using the pT7IMPACT expression vector.....76
3.2.2.7 Studies using the pLEX expression vector.....90

3.3 Conclusion.....97

Chapter 4: Purification of expressed *bO*

4.1 Introduction to the purification of *bO*.....101

4.2 Results and discussion.....102
4.2.1 pET16b/*bO* purification using a His column.....102
4.2.1.1 Preparation of His column resin.....103
4.2.1.2 Preparation of the cell extract.....104
4.2.2 pET16b/*bO* purification using a solvent extraction procedure.....106
4.2.3 Purification of T7IMPACT/*bO*.....106
4.2.4 Purification of pLEX/*bO*.....109
4.2.4.1 Membrane extraction to purify *bO*.....109
4.2.4.2 Ion exchange chromatography to purify *bO*.....109

4.3 Conclusion.....112

Chapter 5: Construction of proline mutants and subsequent kinetic investigations

5.1 Introduction.....115
5.1.1 Proline residues in bacteriorhodopsin.....116

5.2 Results and discussion.....116
5.2.1 Construction of proline mutants.....116
5.2.2 Sequence analysis of proline mutants.....121
5.2.3 Expression studies on the proline mutants using the pLEX expression system.....122
5.2.4 Refolding yields of proline mutants.....122
5.2.5 Kinetic studies on proline mutants using stopped-flow fluorimetry.....124

5.3 Conclusion.....128

Chapter 6: Mass spectrometry in membrane protein analysis

6.1 Introduction.....131
6.1.1 Matrix assisted laser desorption ionisation mass spectrometry (MALDI).....132
6.1.2 Electrospray ionisation mass spectrometry (ESI-MS).....134
6.1.3 Hydrogen/deuterium exchange (HDX) studies on proteins using MS.....137

6.2 Results and discussion.....139
6.2.1 HDX studies using MALDI.....139
6.2.2 HDX studies using ESI-MS.....140

6.3 Conclusion.....150

Chapter 7: General conclusions

7.1 A summary of the principal finding of this project.....152

References.....155
Appendices.....170

Abbreviations

2D	two-dimensional
3D	three-dimensional
aa	amino acid
bR	bacteriorhodopsin
bO	bacterioopsin
bp	base pairs
BSA	bovine serum albumin
CBD	chitin binding domain
CD	circular dichroism
Da	dalton
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	1,4-dithiothreitol
EDTA	ethylenediamine tetracetate
ESI-MS	electrospray ionisation mass spectrometry
HDX	hydrogen/deuterium exchange
i. d	internal diameter
IMP	integral membrane protein
IPTG	isopropyl-β-thio-D-galactoside
KDa	kilo dalton
LB	Luria Bertaini
MALDI-MS	matrix assisted laser desorption ionisation mass spectrometry
OD	optical density
PEG	polyethylene glycol
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNAse	ribonuclease
NMR	nuclear magnetic resonance
SD	Shine Dalgarno
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl polyacrylamide gel electrophoresis
TEMED	N,N,N'N'-[tetramethyl]ethylenediamine]
UV	ultraviolet

Amino Acid residues of Proteins

Ala	alanine	Arg	arginine
Asn	aspargine	Asp	aspartic acid
Cys	cysteine	Gln	glutamate
Glu	glutamine	Gly	glutamic acid
His	histidine	Ile	isoleucine
Leu	leucine	Lys	lysine
Met	methionine	Phe	phenylalanine
Pro	proline	Ser	serine
Thr	threonine	Trp	tryptophan
Tyr	tyrosine	Val	valine

This thesis was prepared using Microsoft® word 6.0.1 on an Apple Power Macintosh LC475. References were prepared using Endnote Plus 2.1; graphs were prepared using Cricket graph, Chemdraw Pro. Plasmid maps were prepared using MacPlasmap. Photographic images were prepared using an Apple Scanner. DNA sequence manipulation and analysis was prepared using DNA Strider. The text is double spaced and set in Palatino 12 point.

Chapter 1:

Introduction

Chapter 1 : Introduction

1.1 Introduction

Integral membrane proteins (IMPs) are encoded by a quarter of the genes present in both prokaryotes and eukaryotes (Saraste and Walker, 1998) yet so little is known about them. Although many structure-function studies have been carried out on soluble proteins (Matthews, 1993), the study of membrane proteins has progressed much more slowly due to the hydrophobic environment in which they are found. Elucidating the folding pathway of these proteins is essential to understand many membrane mediated biochemical processes such as signal transduction, ion transport and chemotaxis, as misfolding can lead to malfunction and disease.

1.1.1 Membrane protein folding

The transmembrane domain of IMPs is composed largely of amino acid residues with nonpolar side chains. In the absence of water, the hydrophilic peptide bonds are driven to form hydrogen bonds with each other. An α-helical structure accommodates the intramolecular hydrogen bonding requirements and is thus the favoured structural motif for proteins in the bilayer. The folding of helical membrane proteins has been proposed to be a two-stage process in thermodynamical terms (Engelman and Steitz, 1981; Popot and Engelman, 1990). In the first stage, transmembrane α-helices insert into the phospholipid bilayer where they may represent autonomous folding domains. The second stage

involves the side by side packing between these already established transmembrane α-helices to form the α-helical bundle found in the native fold of many membrane proteins. Generally speaking, experimental evidence favours this hypothesis (Bormann and Engelman, 1992) although membrane protein folding may be a lot more complex in detail (Riley *et al.*, 1997). The two stage model can serve as a useful paradigm for membrane protein structure prediction. General features of helix bundle membrane protein architecture have also been noted (Bowie, 1997). .

The folding of IMPs has been proposed to be affected by the pressures they incur from the surrounding membrane lipids. This hypothesis was suggested by Gruner (1985) and has been supported by experimental observations (Lindblom *et al.*, 1986; Keller *et al.*, 1993). The lateral pressures affecting the rigidity of a monolayer are illustrated in Figure 1.0. The hydrophobic polypeptide chains experience a positive outward pressure due to chain collisions during bond rotational motion. A positive and repulsive pressure can also arise in the headgroup region from steric hindrance, hydration or electrostatic change but the hydrogen bonding between the head groups gives rise to a negative pressure. The interface experiences a negative pressure as the hydrophobic chains cluster together to exclude water.

Figure 1. 0 Differential lateral pressure exerted on a monolayer.
Adapted from Booth *et al.*, 1997.

Any pressure exerted on a bilayer may subsequently affect the behaviour and folding of any protein embedded within it. With this in mind, Booth (*et al.*, 1997) investigated the refolding kinetics of an integral membrane protein by varying the lipid bilayer rigidity. Since the external pressure and temperature also affect the lateral profile of a monolayer, the micromechanic environment of the bilayer was altered by changing the composition of the lipid membrane at a constant temperature and pressure. They succeeded in their studies by showing that the rate limiting process of bacteriorhodpsin is affected by both bilayer bending rigidity and pH.

1.1.2 Bacteriorhodopsin as a membrane protein folding model

We have chosen the integral membrane protein, bacteriorhodopsin (bR) as a model system for our structure-function studies. bR shares its heptahelical structure with the mammalian vision receptor, rhodopsin and other G-protein coupled receptors (Hargrave, 1991; Schertler *et al.*, 1993; Donnelly and Findlay, 1994). bR, a light-driven proton pump, was first discovered in 1971 by Oesterhelt and Stoeckenius as the sole protein constituent of the purple membrane of *Halobacterium salinarium* (previously known as *H. halobium*). This high-salt loving archaebacterium can be found in abundance in the enclosed shallows of San Francisco Bay. It was initially characterised using electron microscopy (Blaurock and Stoeckenius 1971) and christened bacteriorhodopsin after the previously discovered rhodopsin as they both utilise retinal as their chromophore. Pioneering structural studies were carried out shortly after its initial characterisation by Henderson and Unwin in 1975 to elucidate its structure. The presence of the protein in the membrane as well ordered two-dimensional crystals

rendered it suitable for structural analysis by electron microscopy. Low dose electron microscopy combined with electron diffraction techniques were used to obtain a density contour map of bacteriorhodopsin at 7Å resolution. From this density map, a three-dimensional model consisting of seven transmembrane α-helices was constructed (figure 1.1).

Figure 1.1 3-D model of bR constructed from an electron density contour map at 7Å
(figure taken from Henderson and Unwin, 1975)

It was not known at this point what the connections between the helices were as the resolution was too low but in 1990, Henderson went on to resolve the three-dimensional structure of bacteriorhodopsin to 3.5Å resolution.

The entire course of its 248 amino acid polypeptide chain as well as the position of retinal was revealed. The α-helices form a bundle to provide a binding pocket for the retinal which is linked as a protonated schiff base (PSB) to the ε-amino group of Lys-216. Further advancements in the molecular structure were refined by high resolution electron crystallography (Grigoreiff *et al.*, 1996; Kimura *et al.*, 1997).

Data was limited up until this point due to difficulty in obtaining images from the highly tilted specimens. The gap in the results meant there was no interpretable electron-potential density observed for the loop regions outside the membrane. Grigorieff and his coworkers (1996) collected 30 new images to obtain an improved density map using electron diffraction data corrected for diffuse scattering together with additional phase information. This allowed a new atomic model to be constructed which included surface loops. Although a better view of the transmembrane portion of the molecule was available, the surface structure was still ambiguous. Kimura and his colleagues (1997) went on to collect the remaining data to fill in the 'missing gaps' resulting in a model showing almost all of the amino acid side chains from residues 2 to 231 including all loop regions on the surface (figure 1.2).

Figure 1.2 A ribbon representation of bR
(figure taken from Grigorieff *et al.*, 1996)

More recently, a novel strategy has been used to obtain microcrystals of bR that diffract to 2Å resolution (Pebay-Peyroula *et al.*, 1997). Since the late 1970s, protein crystallographers have been attempting to get well ordered 3D crystals of bR. Although several attempts have been reported (Henderson and Shotton, 1980), the crystals obtained were not suitable for analysis. The conventional strategy employed to crystallise membrane proteins up until a year ago was using protein-detergent micellar complexes (Rosenbusch, 1990). Since then, a novel method has been developed by Landau and Rosenbusch who crystallised bR using lipidic cubic phases rather than protein detergent complexes (Landau and Rosenbusch, 1996; Landau *et al.*, 1997). These lipidic cubic phases are a mixture of lipid and water molecules that posses cubic symmetry (Lindblom and Rilfors, 1989; Luzzati, 1997). The majority are classed as 'micellar' (micelles separated by polar regions) or 'bicontinous' (figure 1.3). The latter is shown as a complex interconnecting bilayer and was used to crystallise bR.

Figure 1.3 Schematic model of a bicontinuous cubic phase.
Composed of lipid, water and a membrane protein, the continuous bilayer surface is connected by a series of aqueous channels shown in black. The enlarged inset shows how a membrane protein (oval) may interact with the bilayer and channel (figure taken from Landau and Rosenbusch, 1996).

These lipidic cubic phases provide a continuous 3D bilayer matrix which aids the nucleation and growth of the desired crystals (Landau and Rosenbusch, 1996). By stacking sheets of 2D crystals, one can form a 3D lattice which would result in some type 1 crystals (figure 1.4). These allow a high degree of contact between the polar and nonpolar regions of the protein. Type II crystals on the other hand (figure 1.4), are the ones that are usually formed by membrane proteins and allow very little protein-protein interaction to take place. The use of lipidic cubic phases can help crystallographers grow the long sought after type 1 crystals.

Figure 1.4 Schematic representation of two types of membrane protein crystals
(figure taken from Gouaux, 1998)

The above strategy was used to obtain microcrystals of bR and as a result, the x-ray structure of bR was solved to 2.5Å by using previous electron crystallographic data as a model (Pebay-Peyroula *et al.*, 1997). They found that most of their data overlapped although a few differences were noted in the loop and side chain regions. Studies identified key residues and solvent molecules implicated in the proton transfer process and defined them with greater certainty. Earlier this year, Gouaux (1998) reviewed the use of this novel strategy to obtain crystals that diffract to 2Å. The strategy may also prove to be an invaluable tool in the structure determination of other membrane proteins.

1.1.3 Bacteriorhodopsin as a light driven proton pump

The x-ray structure of bR *via* lipidic cubic phases (Pebay-Peyroula, 1997; Gouaux, 1998) together with high resolution electron crystallographic data (Grigoreiff, 1996; Kimura, 1997) have contributed to a greater understanding of the proton translocation mechanism in the photocycle (Leuke *et al.*, 1998).

The retinal chromophore is linked *via* a schiff base linkage to lys^{216}. The Schiff base linkage divides the interhelical cavity into extracellular and cytoplasmic 'half-channels'. This is shown in Figure 1.5 along with the possible mechanism of proton translocation by identifying some of the key amino acids residues currently thought to be involved in the photocycle.

Upon absorption of a photon of light, the all-*trans* retinal chromophore isomerises to the 13-*cis* form (Braiman and Mathies, 1982) as shown in figure 1.6.

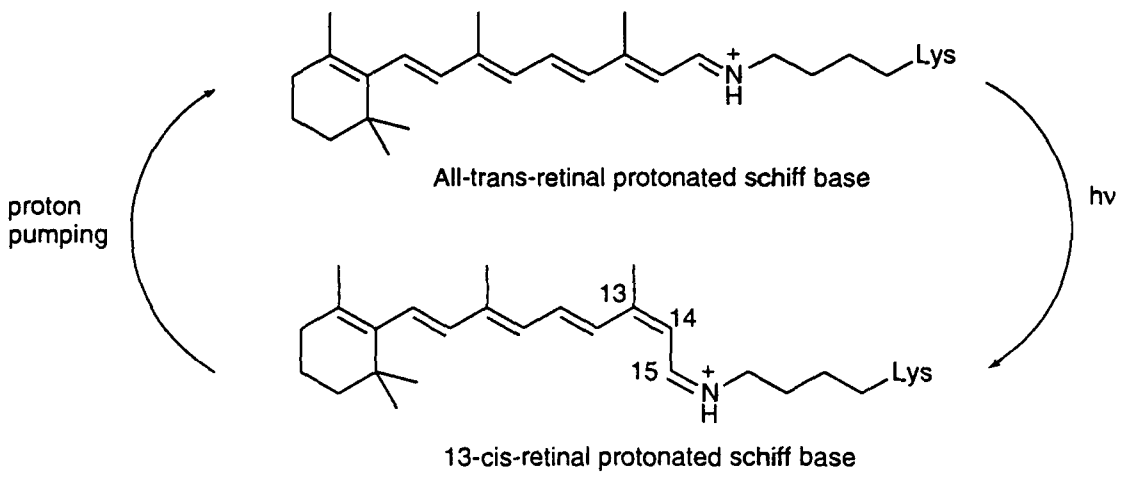


Figure 1.6 The isomerisation of retinal upon absorption of light

Photoisomerisation initiates the translocation of a proton across the membrane *via* a series of conformational changes (Krebs and Khorana, 1993; Lanyi, 1995; Khorana, 1993). The change in geometry along the retinal backbone allows Asp^{85} to move closer to the Schiff base and accept a proton. The proton then enters the external media *via* a multiple pathway involving Arg^{82}, Thr^{205} and Glu^{9}, the latter of which is situated 5Å away from Thr^{205} near the opening of the channel. It is thought that water molecules form a hydrogen bond network that facilitates the translocation of the proton *via* these amino acid residues.

The alternative pathway would be for the proton to be transferred from Arg^{82} to Glu^{204} *via* an intermediate water molecule to Glu^{9}, Glu^{74} or Glu^{194} before being released into the external medium. The Schiff base is reprotonated by Asp^{96} perhaps by a conformational change of the protein. Thermal re-isomerisation of the retinal to the ground state completes the photocycle (figure 1.5).

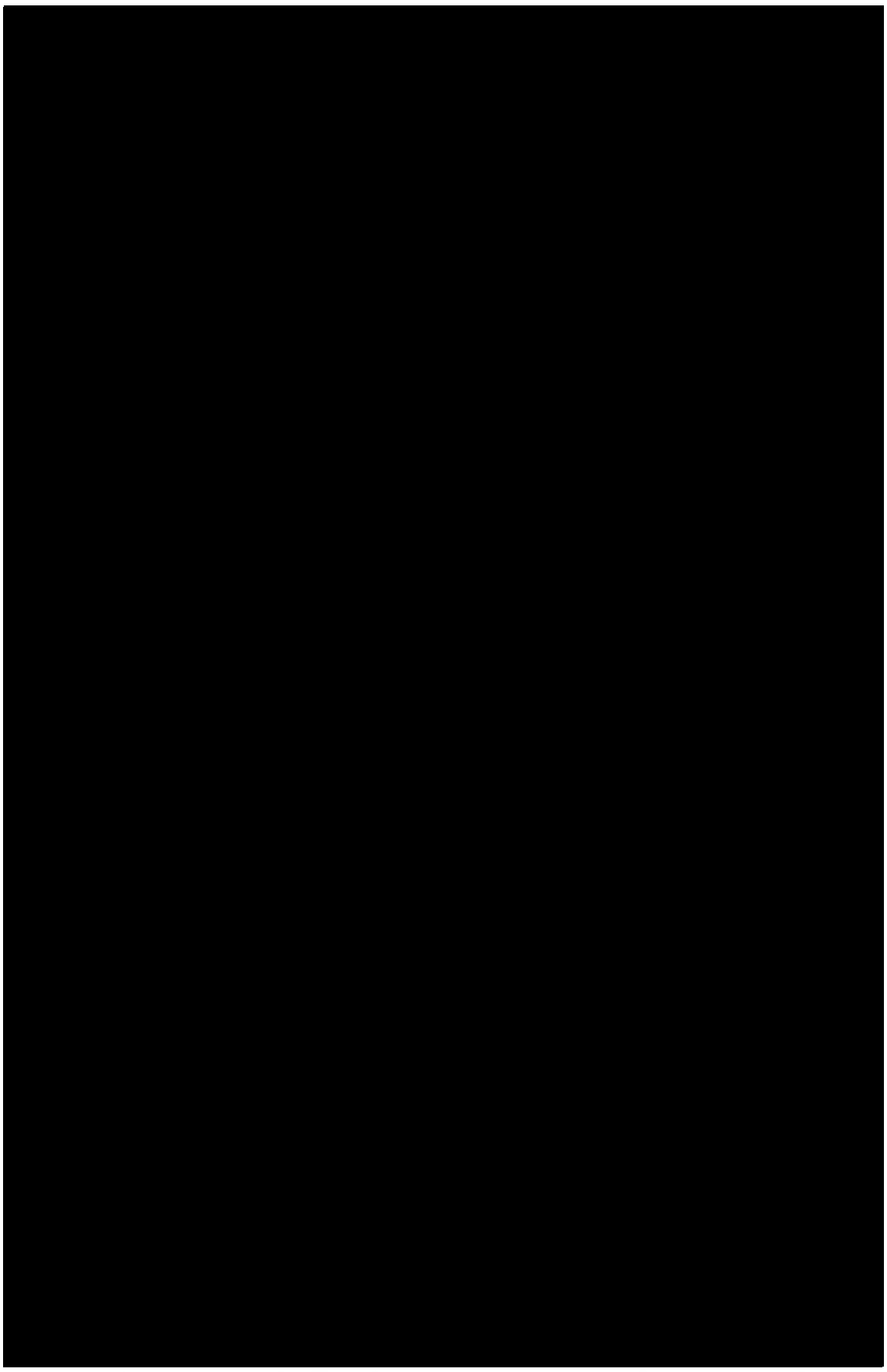


Figure 1.5 Schematic representation of the proton pathway in bR with selected structural details. Scheme has been drawn to correlate with the electron density maps and is consequently distorted. Solid and open blue circles indicate water molecules with the solid circles being well-defined in the model (adapted from Gouaux, 1998).

Folding Kinetics of Bacteriorhodopsin

One of the most attractive features of bR as a model system for membrane protein folding is its ability to spontaneously refold in vitro from a fully denatured state with out accessory proteins (London and Khorana, 1982; Huang *et al.*, 1981). Since this first example, such refolding has been shown for other membrane proteins including porins (Eisele and Rosenbusch, 1990) and photosynthetic antennae complexes (Paulsen *et al.*, 1990). For bR, regeneration efficiencies are easily determined by the extent of recovery of the characteristic purple absorption band of bound retinal. Fluorescence spectroscopy has been used to follow the time course of folding events in bR mixed detergent/lipid micelles (Booth *et al.*, 1995). A substantial amount of secondary structure is formed even in the absence of retinal (London and Khorana, 1982) which allowed various kinetic stages and probable intermediates to be identified in the pathway. Their results suggest the following sequential folding scheme (figure 1.7):

Figure 1.7 Folding pathway of bR
Adapted form Booth *et al.*, 1995.

Although denatured *bO* in SDS possesses a 60% helical content, it cannot bind retinal (London and Khorana, 1982). Folding in DMPC/CHAPS micelles results in the formation of I_o, via the I_1 intermediate, which is able to bind retinal and form functional bR. Circular dichroism studies have shown that I_o has the same amount of secondary structure as bR (London and Khorana, 1982). Studies have sought to elucidate the factors involved in each of these conversion states. Initial folding events are known to be independent of retinal binding. The mixing of SDS and DMPC/CHAPS micelles result in a rise in fluorescence which may be due to the formation of α helices or incorporation of the protein into the micelles. The formation of I_o from I_1 is a very slow phase which is influenced by retinal. In the presence of retinal, a decay in fluorescence is observed and a rise in its absence. It is thought that this slow kinetic phase may be a result of cis-trans isomerisation (Schmid, 1993) of the three membrane embedded proline residues (chapter 5; Henderson, 1990). The other possibility is the rearrangement of some of the helices to form a more native structure for the binding of retinal. Since I_o cannot directly proceed to bR, the non-covalent I_R intermediate was proposed.

Further studies (Booth, 1997; Booth *et al.*, 1997) suggest that the I_o to I_1 step is affected by both refolding conditions and lipid composition. Thus, proline isomerisaton may only be a contributing factor in this step. Whatever the role, our studies aim to further investigate the proline isomerisation by generating some proline mutants and investigating their folding.

1.1.5 Protein production for structure-function studies

Structure and function studies rely heavily on access to large amounts of purified protein which can sometimes be difficult to come by. Material from the natural hosts can often be unavailable or present in only very low concentrations making its isolation difficult. Protein engineering and gene cloning have provided alternative sources of protein for detailed studies. In 1974, Chang and Cohen demonstrated that an amp^R gene originally from *S. aureus* could function in an unrelated bacterium *E. coli*. Since then, it has been shown that a gene from one organism can be expressed in any other. Major advances in recombinant DNA technology have lead to the cloning and sequencing of genes corresponding to particular membrane proteins. Subsequent genetic engineering has rendered them suitable for expression in a variety of host cells. Heterologous expression systems are widely utilised in protein production due to their ease of manipulation and their potential to yield higher levels of protein when compared to the native source. In the ideal situation, this would allow access to large quantities of purified protein for structural studies. Unfortunately, this is not always the case and many problems can arise when the expression system is exploited beyond its capacity resulting in the production of a protein of lower quality or authenticity in terms of its sequence and structure. If the expressed protein is insoluble, inactive, forms inclusion bodies or is not folded to its correct functional structure, its biological activity can be impaired or even completely eliminated. If one tries to express a gene in a host where the codon bias differs immensely (for example, the expression of a human gene in *E. coli*), the chances of translational errors are increased.

Certain codons will be overridden and amino acids will be incorporated at the host's convenience which may result in frameshifts (Kurland and Gallant, 1996). The hosts normal proofreading mechanisms may also be pushed to the limit during overexpression.

To date, there is no way in which an accurate prediction can be made as to whether a particular protein will be expressed and processed correctly in a particular host. The expression level of a given gene varies unpredictably between different systems (Das, 1990).

1.2 Heterologous expression systems for bacteriorhodopsin

A variety of expression hosts and purification systems for *bO* have been reported in the literature. An historical overview of these systems is presented below:

1.2.1 *Halobium salinarium*

Until recently, the homologous expression of *bO* mutant genes was not possible due to a lack of practical transformation procedures. Cline *et al.* (1989) developed an efficient and reliable transformation procedure for halobacterium using polyethylene glycol (PEG) producing spheroplasts (partially degraded cell walls but intact cytoplasmic membranes). The cells are grown to OD_{550} 1.0-1.5 as younger cells can resist spheroplast formation (Cline and Doolittle, 1987). As the cultures approach stationary phase, they are susceptible to lysis and thus should be used within 24 hours. The culture is pelleted and resuspended in spheroplast solution which contains salt, buffer and sucrose. Within 5 min of adding EDTA,

spheroplasts are formed by chelation of divalent cations. Mg^{2+} is essential for both the chelation and transformation. Cells can be frozen at this stage for at least a month if required. The transfection of Halobium halobacteria is linearly dependent on the amount of DNA added with 1 μg of DNA per transfection being the upper limit. With Halobium halobacteria, the transfection step takes only 5 min. An equal volume of PEG solution is added to the spheroplasts. Its final concentration in the solution is critical (Cline and Doolittle, 1987). Transformation levels are the highest during the 10-20 mins after incubation. The cells are diluted with spheroplast dilution solution before plating to dilute the PEG which reverses cell aggregation and brings down the Mg^{2+} levels which terminates transformation and stabilises the spheroplasts.

The transformation procedures as described above (Cline and Doolittle, 1987; Cline *et al.*, 1989; Lam and Doolittle, 1989) were developed by Ni *et al.* (1990) to produce an efficient system for the expression of bR in its natural host. The transformation procedure was carried out as previously described (Cline *et al.*, 1989). The *bop* gene was excised from pBOP (from D. Oesterhelt) along with a fragment of plasmid 455 (Lam and Doolittle, 1989) containing the Mev^R gene [confers resistance to the HMG-CoA reductase inhibitor mevinolin (6 α–methyl compactin)] to construct vector p319 (Ni *et al.*, 1990). This pUC derived vector also contains the *bla* gene as well as the *brp* and *bop* genes of *H. halobium*. The *bat* gene is essential for the expression of *bop* although it is not known why (Betlach *et al.*, 1989). A single transformant was selected and pelleted after growth. It was shaken overnight in basal salts in an attempt to increase the production of bR. The purple cells were the transformed ones while the orange-yellow ones were formed as a result of the

presence of carotenoids. Spectroscopic data revealed the synthesis of functional bR. It was calculated that 125 mg of bR could be produced from a 75 L fermentation. The state of the transformed plasmid has not yet been investigated but at least the system allows production of mutant bR.

Krebs *et al.* (1991) constructed a multicopy plasmid for the expression of the *bO* gene in *Halobacterium halobium* (figure 1.9). The plasmid was transformed into *H. halobium* as previously described (Cline and Doolittle, 1987) and produced 25-40% of bR compared to the wild type. The structure and function of the purified purple membrane was identical to that of the wild type (Krebs *et al.*, 1991). By expressing *bop* mutants in *H. halobium*, the protein is purified as purple membrane and thus allows structure-function studies in the native lattice.

Figure 1. 9 Construction of the *bop* expression vector

The ISH*11* sequence is indicated by the solid black area, the bop sequence by the dotted region and the GRB1 sequence by the unshaded area. pMPK29 is formed by the spontaneous insertion of the ISH*11* sequence in the non-coding region of pGRB1. pMPF35 contains only 70 bp of the ISH*11* sequences the rest was replaced with pUC18 and a synthetic duplex. The *EcoR*I fragment containing pUC was then removed to produce pMPK52 as it lowered the plasmid copy number of several pGRB1 derivatives in *H. halobium* (adapted from Krebs *et al.*, 1991).

Ferrando *et al.* (1993) constructed an *E. coli-H. halobium* shuttle expression vector called pEF191 (figure 1.10). pWL102ΔBA is a derivative of pWL102 which is an *Hf. volcanii E. coli* shuttle vector (Lam and Doolittle, 1989). pEF191 is a phasmid containing both an *E. coli* and halobacterial *ori* as well as *bla*, Mev^R and the *bop* gene. Due to the presence of the phage f1 intergenic region, pEF191 can be isolated as single stranded DNA allowing site directed mutagenesis and sequencing to be performed directly without the need for further cloning steps. pEF191 was transformed into *H. halobium* strain L33 (Cline *et al.*, 1989; Ferrando *et al.*, 1993).

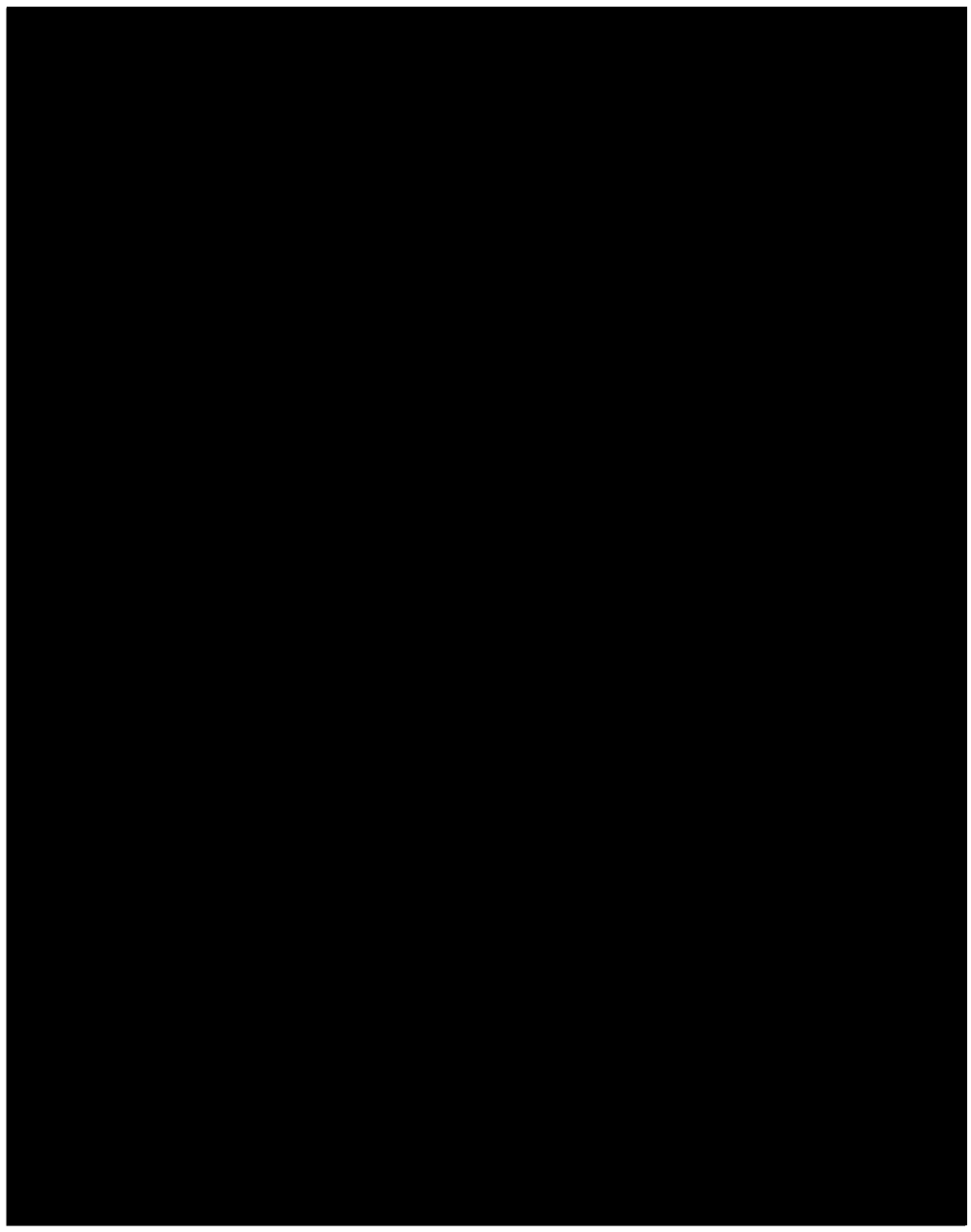


Figure 1.10 Construction of the *E. coli-H. halobium* shuttle vector, pEF191
The bop containing fragment is excised from pEF1100 and inserted into pEF1290 which also contains the f1 *ori*. The final construct was obtained by excising the *EcoR*I-*Hind*III fragment from pEF1290 (continuing bop-f1) and ligating it into pWL102ΔBA which had been digested with the same enzymes (adapted from Ferrando *et al.*, 1993).

A model was proposed for the homologous recombination between the plasmid and chromosomal *bop* (Ferrando *et al.*, 1993) and confirmed by Southern blotting. The expression system produced functional bR and allowed the construction of mutant bR. To use halobacteria as a suitable expression system, the pEF191 expression vector requires further development.

Despite the efforts of all of the above groups, the production of bR using homologous expression is still difficult.

1.2.2 Yeast

Yeast has also been investigated as an expression host for the natural bacterioopsin gene. Hildebrandt *et al.* (1989) isolated a 5.1 kb fragment containing the *bO* gene from *H. halobium* and cloned it into BR322. From this resulting construct, a 1.6 kb *BamHI-Hind*III fragment containing the *bO* as well as some non-coding regions on either end of the gene was excised and replaced with synthetic duplex DNA. The duplex introduced *Not*I and *Hin*dIII sites as well as a stop codon. The *bop* gene was then introduced into the yeast expression vector, pEVP11 (figure 1.11).

Figure 1.11 Plasmid map of pEVBOp-Yeast expression system for the *bO* gene

The *bop* gene is located downstream of the alcohol dehydrogenase (adh)promoter of *S. pombe*. The selection marker for *S. pombe* is leu^{2+}(adapted from Hildebrandt *et al.*, 1989).

Transformation of pEVBOp into yeast cells, strain leu 1-32 h⁻ was carried out as described by Ito *et al.* (1983). The transformants selected on minimal plates without leucine and grown for 2 days at 30°C to stationary phase. Cells were pelleted by centrifugation and vortexed with glass beads to break them. The production of mature protein was detected by immonoblotting without any degradation. Spectroscopic data showed the regeneration of *bO in vivo* by the addition of retinal. In contrast, the regeneration of *bO* in *E. coli* has to be performed in lipid vesicles (Dunn *et al.*, 1987; Karnik *et al.*, 1987). To study the kinetics of bR in the photocycle, *S. pombe* protoplasts were used (Oesterhelt, 1982) which gave similar results to native purple membrane. Overall, the system was reported to have a three-fold advantage. Firstly, it allowed the expression of *bO* as a mature protein without the need to modify the coding region. Secondly, proteolysis was eliminated and thirdly regeneration studies could be carried out *in vivo*. Hilderbrandt *et al.* (1991) later showed that a short leader sequence precursor could influence and increase the expression of protein using the yeast expression system. They expressed both the original *bop* gene and a shorter one where the presequence region was deleted. Both accumulated in the plasma membrane of the yeast cells. Growing cells on minimal media with 2% glucose showed a weaker colour for the mature protein compared to the precursor protein. Increasing the concentration of glucose increased the expression levels for both. The overexpression of precursor protein changed the morphology of the yeast cells to form a coryneform shape while the mature protein retained the rod-cell shape.

Although functional bR is produced in the heterologous host as discussed above (Hildebrandt *et al.*, 1989: Hildebrandt *et al.*, 1991), the yield and purification are still substantial problems.

1.2.3 *Escherichia coli*

Some groups have directed their efforts towards *E. coli*. This bacterium has been exploited by many researchers as an heterologous expression system for the last 50 years. The tremendous wealth of information regarding its genetics and physiology render it a suitable host for gene expression. A wide range of expression vectors available for use with *E.coli* are currently available from commercial suppliers. One may select a vector to combine ease of purification with other desirable properties such as high transformation efficiencies, convenient selectable markers for transformants and recombinants as well as the ability to clone reasonably large pieces of DNA. The rapid generation of biomass is also very beneficial. *E.coli* is capable of directing 60-65% of its total protein synthesising capacity to a single protein (Pouwels, 1991). The major drawbacks of using *E. coli* in gene expression is the lack of post-translational modification and the possibility that the expression of certain genes may be toxic and even lethal to the cell.

Karnik *et al.* (1987) wished to carry out structure-function studies on *bO* for which it was desired that the protein produced should resemble the native molecule. Previous expression systems produced the protein as a fusion which was adequate for initial studies (Dunn *et al.*, 1987; Hackett *et al.*, 1987; Nassal *et al.*, 1987; Braiman *et al.*, 1987) but not for structure-function studies. They also acknowledged the problem of low expression levels and with both problems in mind, constructed a variety of expression vectors to an attempt to solve these problems. In order to produce *bO* that resembles the native molecule, a few modifications were carried out to the *bO* gene. Mature *bO* has a serine residue at the C-terminus but the gene encodes for an additional aspartic acid residue. Although this residue would be

post-translationally removed in *H. halobium, E. coli* is not able to do this and thus the codon was removed from the gene sequence. In addition to this, the gene contains a 419 bp non-coding sequence after the stop codon which was also removed. The N-terminus of the gene was modified by the addition of a methionine residue and a ribosome binding site. Both modification details are described in Karnik *et al.* (1987). The vectors were constructed using strong promoters such as phage λ and T5 promoter (Karnik *et al.*, 1987). The ribosome binding sequence was also optimised in an attempt to increase expression levels of *bO* in *E. coli* (Jay *et al.*, 1984). Translation of *bO* messenger RNA appeared to be a limiting factor. The Shine-Dalgarno and initiator codon sequence as well as the distance between them was important for efficient translation (Gold *et al.*, 1981) but were not the cause of low expression levels of *bO*. The sequence following the initiation codon was found to be very critical in translation. One of the systems they selected involved the thermoinducible phage promoter system. The *bO* produced is associated with the cytoplasmic membrane in a denatured form, which exerts a toxic effect on the host. The half-life of *bO* in the cytoplasm of *E. coli* is 8-10 min (Karnik *et al.*, 1987). The protein produced under these conditions is unstable rendering its purification complex (Braiman *et al.*, 1987; Miercke *et al.*, 1991) and thus the overall yield of homogeneous *bO* is quite low although enough was obtained for their studies. The purification of *bO* is discussed in further detail in chapter 4. N-terminal fusions of the *bO* gene were also constructed but were not very useful due to cleavage problems.

Nassal *et al.* (1987) went on to synthesise the *bO* gene with 30 unique restriction sites to facilitate site directed mutagenesis. Oligonucleotides were synthesised and

the entire gene constructed from four synthetic fragments. The gene was expressed under the control of the pL promoter. This *bO* gene and the expression system was used as a starting point in our studies and is thus discussed in further detail in chapter 3.

Further attempts to increase expression yields of *bO* in *E. coli* came in 1990 by Karnik *et al.* At the time, the state and location of *bO* expressed in *E. coli* was not known. However, studies carried out by Karnik *et al.*, (1990) found *bO* and other *E. coli* proteins to be distributed between the inner and outer membrane. Processed protein was located in the inner membrane but the addition of *E. coli* signal sequences produced protein that was largely unprocessed. The presequence found in *bO* in *H. halobium* was found to produce a substantially degraded protein where the presequence was not cleaved and thus the protein was unable to fold and bind retinal.

1.2.3.1 Production of *bO* as a fusion protein in *E. coli* and associated problems

Shand *et al.* (1991) expressed *bO* in *E. coli* as a fusion protein with 13 heterologous amino acid residues at the N-terminus. They constructed two expression vectors which are described in detail in Shand *et al.* (1991). 82% of expressed *bO* was localised to the *E. coli* cytoplasmic membrane. The half-life of *bO* was 26 min which compared to 8-10 min reported by Karnik *et al.* (1987) where the *bO* has only one single heterologous N-terminal methionine. This allows more time for the protein to be processed for purification as less degradation occurs. Immunological studies reported 10-20 fold higher yields than had been reported by any other system.

1.2.3.1.1 Cleavage from fusion protein using IgA protease

Pompejus *et al.* (1993) aimed to replace the loop regions and chain terminal segments of bR by exogenous polypeptide modules to investigate their functions. This would perhaps lead to the construction of biological membrane with predefined functions. The expression levels of *bO* and the purification procedure were still poor at the time, thus they developed an alternative expression system. The *bO* gene was redesigned and synthesised with an increased G/C content for optimum expression in *E. coli* (Holm, 1986). A synthetic oligonucleotide containing eight restriction sites was constructed. The frame was then cloned into a phasmid[#] and in between these sites, DNA encoding for each of the seven helices was inserted successively by several rounds of insertion mutagenesis to form the *bos* gene for expression. For expression, this gene was fused to a modified lacZ gene to construct vector pLZ1bos (figure 1. 12).

Figure 1.12 Vector for *bos* gene expression in *E. coli*

This plasmid contains colE1 origin of replication (*ori*), a β-lactamase gene (*bla*), the wild type *lac* prmoter/operator region, *bos* gene fused to a mofified *lac*Z gene (*lac*Z454). The IgA protease cleavage site is shown by the arrow (adapted form Pompejus *et el.*, 1993).

[#] A phasmid is a plasmid vector carrying a lambda attatchment site which enables it to insert into the lambda phage genome. This can be then propagated as a plasmid or phage in appropriate *E. coli* strains

Using this phasmid, the *bO* fusion protein was formed as insoluble inclusion bodies which eliminated any toxic effects that may have been exerted on the host or attack by cellular proteases. 150-200 mg/L fusion protein was produced using this procedure. The junction between the two proteins contains the IgA protease cleavage site. After cleavage in aqueous suspension, the *bO* gene was selectively extracted into organic solvent (chapter 4). After cleavage, 30-50 mg pure *bO* were obtained per litre of *E. coli* culture. Cleavage adds an additional Thr-Pro to the N-terminus and an Asp residue to the C-terminus. Although the system allows a high expression level of the *bO* gene, it relies of IgA protease which is relatively expensive and not very accessible

1.2.3.1.2 Cleavage from fusion protein using Factor Xa

Chen *et al.*, (1996) aimed to produce and characterise *bO* mutants with altered hydrophobicity. Their studies required access to large quantities of protein that was not inserted into the host membrane. This eliminated some of the existing expression system including the system developed by Pompejus *et al.*, (1993) where although the fusion protein was formed as insoluble inclusion bodies, it was expensive to cleave *bO* from the fusion using IgA protease. Thus, they began by designing and synthesising a gene for *bO* (Chen *et al.*, 1994) containing 29 flanking unique restriction sites which would facilitate mutagenic studies at the loop and helix regions. The gene also contained a C-terminus epitope, 1D4, which allows detection of expression by immunoblotting, facilitates purification and perhaps allow crystallisation as a complex. The gene was cloned into a MBP expression vector (pMALc2; Maina *et al.*, 1988) and expressed as a 68 kDa fusion under the control of a tac promoter which is IPTG induced. Approximately 170 mg/L of

fusion protein was formed as insoluble inclusion bodies. The inclusion bodies were washed to remove the water soluble proteins and some membrane proteins. They were then solubilised in 8M urea and then dialysed. Cleavage with factor Xa was costly and gave poor results and so they investigated the use of trypsin. By optimising the concentration and length of reaction time, the fusion could be selectively cleaved at one of the 14 possible trypsin cleavage sites. The purified protein was regenerated with all-trans retinal and showed similar properties to bR from *H. salinarium*.

1.3 Aims of this project

Bacteriorhodopsin as described above provides an attractive model system to study the folding pathway of membrane embedded proteins. Detailed studies however, necessitate efficient methods for its overexpression and purification. Rapid and reliable access to large amounts of purified bR will enable us to study further the structure and function of this protein in greater detail. Previous expression systems have reported difficulty in obtaining good yields and simple purification procedures. Since many *E. coli* studies were carried out in the mid 1980s, the objectives of this project are :

To investigate alternative expression and purification systems for the *bO* gene in *E. coli*. With sufficient protein, site directed mutagenesis will be carried out to engineer three proline mutants in the membranous region of bR. Results of the folding of these proline mutants would complement previous work carried out on the folding pathway of bR (Booth *et al.*, 1995) where a slow step in the pathway, *bO* to I_1, may be due to proline isomerisation.

Separate structural studies on bacteriorhodopsin (chapter 6) are investigated using mass spectrometry. Two recent techniques: matrix assisted laser desorption ionisation mass spectrometry and electrospray ionisation mass spectrometry have allowed the analysis of a wide range of biomolecules which were previously not possible due to the nature of the sample. These structural studies are performed by measuring the rate of isotopic exchange of amide and side chain protons in bacteriorhodopsin which can easily be measured by increase in molecular masses. Kinetic and thermodynamic properties can be studied using this technique. Combining these *H. Halobium* with other techniques can allow *in vitro* folding pathways to be studied.

Chapter 2:

Materials and methods

2.1 Materials

2.1.1 General reagents

General reagents and solvents were AnalaR grade and were obtained from Boehringer Manheim Ltd., Sigma Chemical Company Ltd., Aldrich Chemical Company Ltd., Oxoid Ltd., Difco Laboratories and Gibco-BRL unless otherwise stated. Solutions were sterilised where required by autoclaving at 121°C at 20 pounds per square inch (p. s. i.) for 20 min. Thermolabile solutions were sterilised by passing through a 0.22 μm Ministart® filter.

2.1.2 Enzymes

Restriction endonucleases and other enzymes were purchased from Promega Corporation or New England Biolabs Inc. unless otherwise stated, and were used according to the supplier's recommendations.

2.1.3 Commercially available kits

Wizard® Plus SV Miniprep Kit	Promega Corporation
Wizard® Plus SV Midiprep Kit	Promega Corporation
Prep-A-Gene DNA Purification System	BIO-RAD
Thermosequenase	Amersham

2.1.4 Monoclonal antibodies for bR

C3-2 (IgG1)	Kind gift from Koichi Koyama (Tsukuba Research Consortium, Japan)	(Koyama *et al.*, 1994)
N7-3 (IgG2a)	Kind gift from Koichi Koyama (Tsukuba Research Consortium, Japan)	(Koyama *et al.*, 1994)

2.1.5 DNA vectors

pUC19	Stratagene Cloning Systems
pBluescript® II KS	Strategene Cloning Systems
pJM109	Strategene Cloning Systems
pPL1	Kind Gift from Prof. H. Gobind Khorana (M.I.T, USA)
pRH1090	Kind Gift from Dr. John Sutherland (Oxford, UK)
pET16b	Strategene Cloning Systems
pT7-7	Kind Gift from Dr. Dominic Campopiano (Edinburgh, UK)
pCYB2	New England Biolabs Inc.
pT7-IMPACT	Constructed using the above two vectors
pLEX	Invitrogen Corp.

2.1.6 Bacterial cell strains

Strain	Genotypes and remarks
G1724	F^- λ^-,lacIq, ampC::PtrpcI, mcrA, mcrB, INV(rnnE). This strain is included for growth of pLEX and optimal expression from the P_L promoter. This strain contains cI repressor under control of *trp* promoter.
HMS174	recA1, *hsd*R, rif^r A recombination-deficient strain used for high-level expression of genes cloned into expression vectors containing bacteriophage T7 promoter. Bacteriophage T7 RNA polymerase is provided by infection with a bacteriophage λ that carries bacteriophage T7 gene*1* (Campbell et al. 1978: Studier Moffat 1986).
BL21 (DE3)	F-,ompT, $hsdS_B$ ($r_B^-m_B^-$), *gal, dcm*(DE3), (λcI*ts* 857 *ind*1 Sam7 nin5 lacUV5-T7 gene1). A strain employed for high level expression of genes cloned into expression vectors containing bacteriophage T7 promoter. Bacteriophage T7 RNA polymerase is carried on the bacteriophage λ DE3, which is integrated into the chromosome of BL21 (Studier and Moffat 1986).
B834 (DE3)	F-, ompT, $hsdS_B$ ($r_B^-m_B^-$), *gal, dcm, met(DE3).*
C600	$e14^-$ ($McrA^-$) *supE44 thi-1 thr 1 levB6 lacYI tonA21.*

2.1.7 Laboratory stocks

Most stocks were prepared and stored at room temperature except the following:

0.1M DTT, 10mg/ml BSA, 1mM IPTG, 100 mg/ml ampicillin which were all stored at -20°C.

2.1.8 General solutions and buffers

10 X M9 Salts	M9 Salts were purchased from Sigma, prepared as instructed, autoclaved and stored at RT.
1M $MgCl_2$	Dissolve 20.33g in 100 ml dH_2O. Autoclave and store at RT.
100mg/ml ampicillin	Dissolve 1g ampicillin in 10 ml dH_2O. Filter sterilise and store at -20°C.
50% glucose	Dissolve 50g glucose (dextrose) in 100 ml dH_2O. Filter-sterilise and store at RT.
TE Buffer	10 mM Tris-HCl, 0.1M EDTA pH 8.0.

2.1.9 Media

Luria Bertaini (LB)	1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, pH 7.4.
LB agar	LB medium plus 1.5% bacto-agar.
RM medium	1 x M9 salts, 2% casamino acids, 1% glycerol, 1 mM, 100 µg/ml ampicillin.
RM agar	RM medium plus 0.5% glucose and 1.5% bacto-agar.
Induction medium	1 x M9 salts, 0.2% casamino acids, 0.5% glucose, 1 mM $MgCl_2$, 100 µg/ml ampicillin.
Basal salt	25% NaCl, 2% $MgSO_4$, 0.3% NaCitrate, 0.2% KCl.
Basal medium	Basal salt plus 1% peptone.
Basal agar	Basal medium plus 2% bacto-agar.

As the antibiotics used in this work were heat sensitive, autoclaved media and molten agar were cooled to below 55°C before addition of sterile antibiotic. Ampicillin was added to a final concentration of 100 μg/ml and chloramphenicol to 34 μg/ml.

2.2 Equipment

Polymerase chain reactions (PCR) were carried out using a Perkin Elmer Thermal cycler. Analytical and preparative electrophoresis of DNA samples was performed using a Northumbria Biologicals Ltd. "mini" electrophoresis tank. Ultraviolet/visible absorbance measurements were made using a PYE UNICAM PU 8800 UV/VIS spectrophotometer. DNA fragments were photographed using a Polaroid Gelcam and Electrophoresis hood EPH7.

Shake flask fermentations were carried out in New Brunswick Scientific G25R orbital shakers.

Centrifugation of small samples in Eppendorf tubes were carried out in an MSE Micro Centaur centrifuge. Larger samples were centrifuged in a Beckman J2-21 centrifuge, Sorvall RC 5C plus or a Beckman L8-70M ultracentrifuge. Sonication was performed using an MSE SONIPREP 150 sonicator. For organic extractions, an IKA Labortechnik, Janke and Kunel Ultra-Turrax T25 homogeniser was used.

Protein electrophoresis was carried out using the BIO-RAD Mini-PROTEAN vertical gel electrophoresis apparatus. For immunoblotting, resolved protein bands were transferred onto Problott™ (ABI) membrane in a BIO-RAD Trans-Blot cell. Proteins were eluted from polyacylamide gels using a BIO-RAD Model 422 Electro-Eluter. Polyacylamide gels were dried using membrane film from BIO-RAD.

DNA sequencing in Oxford was performed as a service by Dr. Annette Prescott at the Sir William Dunn School of Pathology on an Applied Biosystems ABI Prism™ Model 373A, version 1.2.1 automated DNA sequencer. In Edinburgh, DNA sequence analysis was carried out by Ms. Nicola Preston at the Darwin Building on an Applied Biosystems ABI Prism™ Model 377, version 2.1.1 automated DNA sequencer.

2.3 Methods

Standard microbiological techniques were performed throughout this work as described in *Molecular Cloning - A Laboratory Manual* (Sambrook J., Fritsch E. F., Maniatis T. (1989) Cold Spring Harbor, Cold Spring Laboratory Press) and *Current Protocols in Molecular Biology* (Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A., Struhl K., Eds. (1989), New York, Wiley Interscience).

2.3.1 Microbiology -*Escherichia coli*

2.3.1.1 Growth of bacterial cultures

E.coli strains were cultured in LB (see section 2.1.9) containing the appropriate antibiotics at either 30°C or 37°C at 250 rpm. Bacterial growth was monitored spectrophotometrically at 600 nm.

2.3.1.2 Preparation of *E. coli* competent cells

The *E. coli* strain was streaked out onto LB plate using sterile techniques and incubated for 16 hours at 37°C. A single colony was used to inoculate a starter culture of LB (10 ml) and grown to saturation. 200 µl of this overnight starter culture was used to inoculate 10 ml of LB containing 200 µl of sterile 1 M $MgCl_2$

and grown with shaking to an OD_{600} = 0.2 at 37°C. Transformation buffer (TB) (500 µl of 500 mM $CaCl_2$, 500 µl of 1 M Tris pH 8.0 and 4 ml of dH_2O) was prepared and chilled on ice. The cultures were centrifuged (2000 rpm, 4°C for 10 min), resuspended in 4 ml TB and left on ice for 30 min. Cells were centrifuged again and the pellet resuspended in 400 µl of TB and kept on ice until required.

2.3.1.3 Transformation of *E. coli* competent cells with plasmid DNA

50-100 ng of plasmid DNA was added to 100 µl of competent cells in an Eppendorf tube and mixed gently before being left on ice for half an hour. The cells were heat shocked for 5 min at 37 °C to allow the uptake of plasmid DNA. 1 ml of LB was added to the eppendorf and the cells incubated shaking at 37 °C for one hour. They were then centrifuged in a microcentrifuge and the pellet resuspended in 150 µl LB. The cells were plated out onto LB plates containing the appropriate antibiotic and incubated at 37 °C O/N to produce single colonies.

2.3.1.4 Harvesting and lysis of *E.coli*

Liquid cultures were centrifuged in a pre-cooled centrifuge (10 000 rpm, 10 min) and the supernatant removed by decanting. Lysis procedures used were both chemical (lysozyme/EDTA/SDS) and mechanical (homogeniser). After lysis, insoluble cell debris was pelleted by centrifugation (conditions as above) leaving a reasonably clear supernatant. In the case of membrane proteins, the supernatant was discarded and the pellet kept aside.

2.3.1.5 Purification of plasmid DNA

Small scale DNA purification was carried out using Wizard® Plus SV kits by Promega as described below:

- cells were pelleted from a 1-10 ml culture and resuspended in 250 µl of cell resuspension solution. The solution was transferred to an Eppendorf tube.
- Cells were lysed by adding 250 µl cell lysis solution and mixing gently.

  10 µl of alkaline protease solution was added and the tube gently inverted 4 times and incubated at room temperature. This step was omitted if using an *end*A+ strain. 350 µl neutralisation solution was added to the Eppendorf tube and gently inverted to mix. Centrifugation was carried out for 10 min to obtain a cleared lysate which was transferred to a Wizard® *Plus* SV miniprep column avoiding the carry over of any white precipitate and spun for 1 min.
- 750 µl of column wash solution was added and spun for 1 min.
- 250 µl of column wash solution was added and spun for 2 min.

The plasmid DNA was eluted by adding 110 µl nuclease free water and microcentrifuged for 1 min.

Large scale purification was carried out in exactly the same way but using a maxi-prep column also supplied by Promega.

2.3.2 Manipulation and analysis of DNA

2.3.2.1 Ethanol precipitation of DNA

Sodium acetate at pH 6.0 was added to a final concentration of 0.3 M, followed by 2.5 to 3 volumes of absolute ethanol. After incubation at -20°C, the DNA was

pelleted by centrifugation (13 000 rpm, 30 min). The pellet was washed with chilled 70% ethanol to remove salt contamination, and then was dried briefly on the bench and resuspended in the appropriate volume of TE buffer (2.1.7) at pH 8.0 or water.

2.3.2.2 Quantification of DNA

Concentration of DNA was determined by its absorbance at 260 nm using the following values: OD_{260} = 1 for 50 µg/ml double-stranded DNA or 33 µg/ml single-stranded DNA. The OD_{280} value was also read and the ratio of OD_{260}/OD_{280} calculated to estimate the purity of DNA solutions. An OD_{260}/OD_{280} ratio between 1.6-2.0 was considered satisfactory. The concentration of double-stranded DNA was also estimated by its fluorescence intensity in an ethidium bromide-stained mini-agarose gel. A known concentration of a DNA molecular weight standard (Promega) was run for comparison.

2.3.2.3 Restriction endonuclease digest of both vector and insert DNA for cloning

Digestion of DNA was carried out using the buffers supplied by the manufacturers for each enzyme. Incubations were carried out at 37°C with the exception of *Sma1* and *Bsm1* which were incubated at 25°C and 55°C respectively. Depending on the amount of DNA to be digested and the number of units of enzyme used, the period of incubation varied between 2 and 16 hours. Digestions with more than one enzyme were carried out simultaneously in a suitable common buffer whenever possible. When the buffer requirements were incompatible, digestions were performed sequentially purifying the DNA (see sections 2.3.2.5 and 2.3.2.6) between digests.

2.3.2.4 DNA ligations

DNA ligation was performed in sterile Eppendorf tubes in the following manner: DNA restriction fragments (10:1 (mol/mol) insert/vector) 10 - 20 µg/ml; 10 x T4 DNA ligase buffer, 2 µl; T4 DNA ligase (400 U/µl); 10 mM ATP and left at room temperature overnight.

2.3.2.5 Agarose gel electrophoresis

To prepare a 1% gel, 1 g of agarose was melted in 100 ml of TAE (40 mM Tris-HCl, 2 mM EDTA, 24 mM acetic acid, pH 7.7) in a microwave. Ethidium bromide (0.5 µg/ml) was added and the gel poured into the cast. Gel was run at 100 V for 1 hr.

2.3.2.6 Recovery of DNA from agarose gels

This was carried out using the Prep-A-gene kit from BIO-RAD.

The DNA required to be purified was excised from the agarose gel, chopped up finely on a glass plate and then placed into a microcentrifuge tube. The volume of the gel slice was estimated by centrifugation by comparison with an empty microcentrifuge tube (e.g ⅓). Based on the gel volume, 3 times the amount of DNA purification kit binding buffer (1 ml) was added to the gel slice and agitatated gently to dissolve. The tube was heated for several minutes at 37-55 °C to further assist dissolving the agarose after which Prep-A-Gene matrix (20 µl) was added (5 µl matrix was required for every microgram of DNA to be absorbed). The tube was flicked to mix and incubated for 5-10 min at room temperature. Frequent agitation was employed during this binding step to assist DNA binding to the Prep-A-Gene matrix. Centrifugation was carried out in a microcentrifuge for 30 secs to

pellet the DNA-containing Prep-A-Gene matrix. The supernatant was removed using a pipette and discarded. The pellet was resupended in DNA purification kit binding buffer equivalent to 25 times the amount of matrix added (375 µl) using brief vortexing. The pellet was again centrifuged for 30 secs and the supernatant disposed of. A further 2 washes were carried out using the wash buffer as described previously. After the last wash, a pipette was used to remove all traces of liquid left in the tube. The tube was then centrifuged once more and any remaining liquid again removed with a pipette. It was important to remove all traces of wash buffer since ethanol and high salt concentrations can inhibit enzyme activity. The DNA was eluted from the matrix by adding at least 1 pellet volume of elution buffer (15 µl) and incubating for 5 min at 37°C. After centrifugation, the supernatant containing the purified DNA was transferred to a clean eppendorf recovering at least 85% of DNA. An additional 10-15% was recovered by repeating this step. The combined supernatants were centrifuged to remove any last traces of Prep-A-Gene matrix leaving pure DNA in 30 µl of elution buffer. This method exhibits DNA recovery in excess of 85% ready for further experimental procedures.

2.3.2.7 Polymerase chain reaction (PCR)

The polymerase chain reaction was used to amplify specific fragments between two oligonucleotide primers *via* a succession of "cycles" where the DNA template is denatured, allowed to anneal with the primers and a new strand synthesised using a thermostable DNA polymerase. Each PCR reaction was carried out in a sterile 0.5 ml Eppendorf tube in a volume of 50 µl:

10 x Vent polymerase buffer	5 µl
100 x acetylated BSA	0.5 µl
100 mM $MgSO_4$	1-20 µl
25 mM dNTP	2.5 µl
forward primer	100 ng
reverse primer	100 ng

The reaction mixtures were overlayed with mineral oil to prevent evaporation and were subjected to a varying number of cycles of amplification in a DNA thermal cycler. Generally, each cycle encompassed three stages: the DNA was (depending on the T_m[1] value of the primers) and for the final extension stage, the temperature was set at 72°C for 1-2 min and held at 28°C. After amplification, the PCR products were analysed by agarose gel electrophoresis (see section 2.3.2.5) and stored at -20°C.

2.3.2.8 PCR Screen

A grid was marked onto an LB/amp plate dividing it up into 9 colonies. Using containing a plug of cotton wool (to prevent cross contamination), a colony was replicated from the ligation plate to the grided plate. Using the same tip, the rest of the colony was pipetted into 50 µl of dH_2O in an Eppendorf tube. This was repeated for the rest of the colonies. The Eppendorf tubes were heated at 95°C for 5 min. Samples were spun in a microcentrifuge and 5 µl of the supernatant transferred into a PCR tube and used as a template for PCR analysis.

---

[1] T_m value is a rough estimate of the temperature at which half of the molecules will be dissociated from a complementary DNA strand and is calculated according to the following formula: $T_m = 4(G+C) + 2(A+T)$ °C.

2.3.2.9 DNA sequencing

Wizard miniprep kits from Promega were used to produce good quality DNA templates for sequencing reactions. Reactions were carried out under sterile conditions in microcentrifuge tubes as described below using the Thermosequenase kit by Amersham. The following were combined in an Eppendorf tube: Sequencing reagent premix (8 µl); ds DNA template (1.0 µg); primer (5.0 pmol); dH_2O (to a final volume of 20 µl). The tubes were vortexed and spun briefly and overlaid with 20 µl of paraffin oil.

Each sample was vortexed, placed in a thermal cycler and heated as follows:

- *96°C for 30 seconds, 45°C for 15 seconds, 60°C for 4 minutes* (30 cycles)
- *4°C (hold)* (1 cycle)

20 µl of the reaction mixture was carefully removed with a pipette and transferred to a fresh Eppendorf tube containing 7 µl of 7.5 M ammonium acetate. 2.5 volumes (68 µl) of 100% ethanol was added and the tube placed in the freezer overnight. Each sample was spun for 30 minutes and the supernatant carefully removed. The pellets were washed with 70% ethanol and the supernatant removed before drying. 4 µl of loading dye was added to each sample, heated and loaded onto an automatic sequencer.

2.3.3 Analysis of protein

2.3.3.1 SDS-PAGE gel electrophoresis

Proteins were analysed using a BIO-RAD Mini-PROTEAN® Π vertical gel at 50 mA per gel under denaturing conditions using the Laemmli discontinuous buffer system[2]. Separating and stacking gels were prepared using acrylamide (30% w/v) and N,N-bis-methyleneacrylamide (0.8% w/v) to which APS (0.05% w/v) and TEMED (0.1% w/v) were added to initiate polymerisation.

Separating gel	12 %	15 %
dH_2O	3.35 ml	2.35 ml
1.5 M Tris pH 8.8	2.5 ml	2.5 ml
10% SDS	100 µl	100 µl
Acylamide/bis	4 ml	5 ml
10% APS	50 µl	50 µl
TEMED	10 µl	10 µl

Stacking gel	12%	15%
dH_2O	3.05 ml	3.05 ml
1.5 M Tris pH 8.8	1.25 ml	1.25 ml
10% SDS	50 µl	50 µl
Acylamide/bis	650 µl	650 µl
10% APS	50 µl	50 µl
TEMED	20 µl	20 µl

Protein bands were visualised by staining with Coomassie Brilliant Blue [EtOH (50% (w/v) , AcOH (10%), Coomassie Blue R250 (0.2% w/v)].

---

[2] Laemmeli, U.K (1970). Cleavage of the structural proteins during the assembly of the head of bacteriophage T7. Nature, **227**, 680-685.

2.3.3.2 Western blotting

Only 1/50 - 1/100th of the protein loaded for coomassie blue R250 is required for detection by western blotting. Following SDS-PAGE, the resolved protein bands were transferred onto Problott™ membrane which was pre-soaked in 100% methanol for 5 minutes. Electroblotting was carried out at 50 V for 2.5 hours at 4°C in transfer buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 15% (v/v) methanol). The membrane was then incubated shaking for 5 hours at room temperature in Blocking buffer (20 mM Tris-HCl pH 7.6, 37 mM NaCl, 0.1% (v/v) Tween-20, 5% (w/v) non-fat milk powder (e.g. Marvel). Fresh blocking buffer was used to incubate the membrane with the bR C-terminus primary antibody at a dilution of 1:1000 (v/v) for 5 hours. The membrane was then washed with blocking buffer (3 x 15 minutes) and incubated with the secondary antibody (peroxidase-conjugated rabbit IgG to mouse IgG at a dilution of 1:500 (v/v) for 1 hour. The membrane was again washed with blocking buffer (3 x 15 minutes) and once in dH_2O. For chemical detection, the membrane was incubated in 20 ml dH_2O containing 0.01 g *o*-dianisidine, 0.1 mM imidazole and 200 µl hydrogen peroxide until bands were visible.

2.3.4 Protein purification

2.3.4.1 Elution of proteins from SDS-PAGE gels

This was carried out using the electro-elution system by BIO-RAD resulting in 70-100% recovery. The membrane caps were stored at 60°C in elution buffer (tris-base 25 mM, glycine 192 mM and SDS 0.1%) for at least an hour before use. The gel slice was loaded into the pre-assembled apparatus as described in the manufacturer's instruction manual and eluted whilst stirring (8-10 mA, 3-5 hrs, 4°C).

2.3.4.2 Purification of *bO* using a His column

The following solutions were prepared:

8x Binding buffer	4x Elute buffer	8x Charge buffer
40 mM imidazole	4 M imidazole	400 mM $NiSO_4$
4 M NaCl	2 M NaCl	
160 mM Tris-HCl pH 7.9	80 mM Tris-HCl pH 7.9	
6 M guanidine-HCl pH 7.9	6 M guanidine-HCl pH 7.9	

8x Wash buffer	4x Strip buffer
480 mM imidazole	4 M imidazole
4 M NaCl	2 M NaCl
160 mM Tris-HCl pH 7.9	80 mM Tris-HCl pH 7.9
6 M guanidine-HCl pH 7.9	

The stock charge, binding, wash and elute buffers were diluted 1 x with sterile dH_2O before use. As a precaution against degradation of protein, the column and cell extract were prepared simultaneously and the chromatography was performed on the same day.

2.3.4.2.1 Preparation of His column resin

The polypropylene column was packed with His-bind resin under gravity (one column volume). It was then charged and equilibrated as described below:

3 volumes sterile dH_2O
5 volumes 1x charge buffer
3 volumes 1x binding buffer

2.3.4.2.2 Preparation of the cell extract

A 1.5 g pellet was resuspended in 15 ml binding buffer and was homogenised 3 x 30 sec with 30 sec intervals using a IKA Labortechnik homogeniser. The sample was

then centrifuged at 10 000 rpm for 15 min at 4°C and the supernatant discarded. The pellet was resuspended in 7.5 ml binding buffer and then 75 µl 1 M $MgCl_2$ and 75 µl DNAse (2 mg/ml stock) was added and the sample incubated at room temperature for 20 min. The sample was then homogenised and centrifuged as above and the supernatant discarded once again. The pellet was solubilised in 7.5 ml binding buffer, homogenised (3 x 30 sec) and then centrifuged at 20 000 rpm for 20 min at 4°C to remove any insoluble material so as to avoid blocking the column.

2.3.4.3 Purification of *bO* by solvent extraction

5 M NaCl (300 µl), 1 M Tris pH 8.0 (125 µl), 0.5 M EDTA (40 µl), PMSF (0.0025 g dissolved in dioxane) and lysozyme (0.01 g dissolved in 100 mM Tris pH 8.0) were added to 10 g of cell pellet on ice and the mixture stirred until it became thick. It was then frozen in an ethanol ice bath for 30 min. The pellet was thawed at room temperature before adding 1 M $MgCl_2$ (50 µl), 0.1 M $CaCl_2$ (50 µl) and a small spatula tip of both DNAse and RNAse. The mixture was stirred and left to incubate at room temperature for 20 min. It was then sonicated for 1 x 30 min with 1 min intervals. The sample was transferred into ultracentrifuge tubes and centrifuged at 21 000 rpm for 1.5 hours at 4°C. The supernatant was removed and the pellet was freeze dried overnight. 1 g of dried pellet was mixed with 4 ml of chloroform:methanol:water:triethylamine 100:100:25:1 and the sample thoroughly homogenised. It was then centrifuged at 3000 rpm for 15 min at 4°C. The supernatant was saved and the extraction step repeated a further three times on the remaining pellet. The supernatants were combined and an equal volume of

water was added to create a phase separation. Both the top and the bottom layers were discarded leaving behind the white lipidacious interface containing the *bO*.

2.3.4.4 Ion exchange chromatography to purify *bO*

The following solvent mixtures were prepared:

Solvent A	$CHCl_3$:MeOH:dH_2O:Et_3NH_2	100:100:25:1
Solvent B (anhydrous)	$CHCl_3$:MeOH:Et_3NH_2	100:100:1
Solvent C (150 mM salt)	485 ml solvent A, 4.32 ml glacial acetic acid and 10.43 ml triethylamine	
Solvent C (30 mM salt)	dilute solvent C (150 mM) 1/5 with solvent A	
Solvent D (150 mM salt)	485 ml solvent B, 4.32 ml glacial acetic acid and 10.43 ml triethylamine	
Solvent D (30 mM salt)	dilute solvent D (150 mM) 1/5 with solvent B	
Solvent E (1 M salt)	321.4 ml solvent A, 23 ml glacial acetic acid, 55.6 ml glacial acetic acid	

Since chloroform is present in all the above solvents, the column, tubing fittings and all other equipment had to be chloroform resistant. The white lipidacious pellet obtained after the solvent extraction step was dissolved in enough solvent B to get it into solution. It was then centrifuged at 3000 rpm for 10 min to pellet any insoluble debris. An equal volume of water was added and the interface retained. The above step was repeated using solvent D which contains 30 mM salt. The sample was once again centrifuged to remove any insoluble material before loading it onto the column.

A glass column (15 cm x 1 cm i.d) was packed with a DEAE Tris acryl ion-exchange resin (10 ml) under gravity. The column was pre-equilibrated using solvent C (30 mM salt). The sample was loaded onto the column and the column was

washed with solvent C (30 mM salt) until the A_{280} peak returned back to baseline (approximately 20 ml). The protein was eluted using 100 ml of a 30 mM to 150 m M salt gradient.

2.3.5 Growth of *Halobium salinarium*

2.3.5.1 Maintenance of *Halobium salinarium* on solid media

Using a sterile inoculating loop, a single colony of *H. salinarium* was isolated and streaked out onto a solid Basal medium plate (see section 2.1.9) and incubated at 37 °C for 2 weeks after which purple colonies were predominantly visible. The plates were stored for up to 3 months at 4°C after which the cultures were re-plated.

2.3.5.2 Growth of *H. salinarium* in liquid culture

A single colony of *H. salinarium* was used to inoculate 50 ml Basal media (see section 2.1.9) in a 250 ml conical flask. The mouth of the flask was plugged with loose cotton wool and the culture grown at 39°C in the dark (or flasks were covered with aluminium foil) using an incubator shaker (200 rpm). Bacterial growth was monitored spectrophotometrically at 560 nm and 660 nm. The halophile took six to seven days to reach stationary phase (A_{560} = 1.0). This starter culture was used to inoculate 1 L of Basal medium. The cells were grown for a further six days and harvested at A_{560}=1.0.

2.3.5.3 Isolation of purple membrane

The *H. salinarium* culture was harvested by centrifugation for 20 min at 10 000 rpm. The supernatant was poured off and the pellet resuspended in 30-

40 ml of Basal salt. Solid DNaseI was added to the suspension (enzyme to suspension ratio of 1:5 (w/v mg/ml) and incubated at room temperature for 1 hour with occasional stirring.

The DNaseI treated suspension was dialysed against NaCl (0.1 M) overnight. This was then transferred into ultracentrifuge tubes and centrifuged for 30 min at 30,000 rpm. The supernatant was poured off and the pellet resuspended in a small volume of 0.1 M NaCl. The suspension was homogenised using a hand held homogeniser and then centrifuged. The above procedure was carried out once again using 0.1 M NaCl and a further two times using dH_2O water. The pellet was finally resuspended in the minimum volume of dH_2O possible and sodium azide (0.025%) was added to prevent microbial growth. The purple membrane suspension was stored at 4°C.

2.3.5.4 Delipidation of bR from PM suspension

The concentration of bR in the aqueous purple membrane was calculated by recording the A_{250} and A_{560} using a quartz cuvette and dH_2O as a blank. The protein concentration in PM suspension was determined using $\varepsilon_{280} = 2.38$ cm^2/mg and the purity of refolded protein calculated using A_{280}/A_{560}. To the known volume of PM suspension (15-20 mg/ml) was added a solvent mix of $CHCl_3$:MeOH:triethylamine (100:100:1), (CMT) at a suspension to solution ratio of 1:10 (v/v) and vortexed. The aim was to try and dissolve the pellet in the minimum volume of CMT and although the solvent ratio was not critical, the addition of excess solvent was undesirable. Retinal was removed from the resulting clear yellowish suspension using NH_2OH (10 M, 10 µl was added per mg of protein) to give a clear solution. An equal volume of sodium phosphate buffer (0.1 M, pH 6.0) was added to this

suspension and vortexed. Centrifugation was carried out at 3 000 rpm for 15 min. The white bR pellet formed at the interface between the upper aqueous and lower organic phase was isolated and resuspended in dH_2O. After centrifugation, the delipidated *bO* was redissolved in SDS at a protein/detergent ratio of 1:5 (w/w) to give a final concentration of 2 mg/ml. The A_{280} of the protein solution was measured to check the protein concentration and divided into 100 µl aliqouts and stored at -20 °C.

2.3.5.5 Preparation of micelles

DMPC (0.20 g), sodium phosphate (pH 6.0, 0.1 ml, 0.1 M) and dH_2O (7.0 ml) were mixed together in a glass vial and stirred at room temperature for 2 hours. CHAPS (2.0 ml, 10%) was added to the milky suspension and stirred until it became clear. This was then sonicated for 30 min and stored at 4°C for up to 2 weeks.

2.3.5.6 Determination of degree of refolding of *bO*

The following stock solutions were required: a micellar solution as described above, a stock solution of *bO* (2 mg/ml), retinal (2 mg/ml), SDS (10% [w/v]) and sodium phosphate buffer (pH 6.0, 0.1M). In an eppendorf was added 750 µl of the micellar solution, 75 µl of sodium phosphate, *bO* (0.2 mg/ml) and SDS [0.1-y]% (w/v) where y is the % contribution made by the SDS in the *bO* stock solution. The *bO* concentration was determined by measuring the A_{280} and to this was added an appropriate amount of retinal to give an equimolar protein:ligand ratio. Retinal was added in the dark as it undergoes *trans-cis* isomerisation upon exposure to light and *bO* can only bind *trans*-retinal. The mixture was left to equilibrate

overnight at room temperature. The A_{280}/A_{560} ratio was calculated to determine the degree of refolding (α) using the equation below:

$$\alpha\,(\%) = \frac{A_{280} \,/\, A_{560} \text{ of PM suspension} \times 100}{A_{280} \,/\, A_{560} \text{ of mixture}} \qquad \text{where } 0 \leq \alpha \leq 100$$

A_{280} is the protein absorbance, A_{560} is the retinal absorbance and PM is purple membrane.

2.3.6 Analysis of membrane protein by mass spectrometry

2.3.6.1 Phase separation technique for sample preparation in ESI-MS

Samples for electrospray ionisation mass spectrometry measurement were prepared using the phase separation method established by Barnidge and co-workers (1990). The method is outlines in figure 2.1. 100 µl of purple membrane solution was combined with 100 µl of a 1% aqueous solution of β-octyl glucoside in 2% formic acid. To this was added a solvent mixture of chloroform:methanol:water (2:5:2) (v/v) and 1% acetic acid. This was vortexed for 10 seconds and then centrifuged at 10,000 g for 2 min. The final solution resulted in a phase separation with 900 µl in the upper aqueous layer and 100 µl in the lower layer. The lower chloroform rich phase was injected into the mass spectrometer using chloroform:methanol:water (2:5:2) (v/v) and 1% acetic acid if acarrier solvent was required. Deuteriated samples were prepared in the same way except they were deuteriated (2.3.5.3) before the extraction procedure.

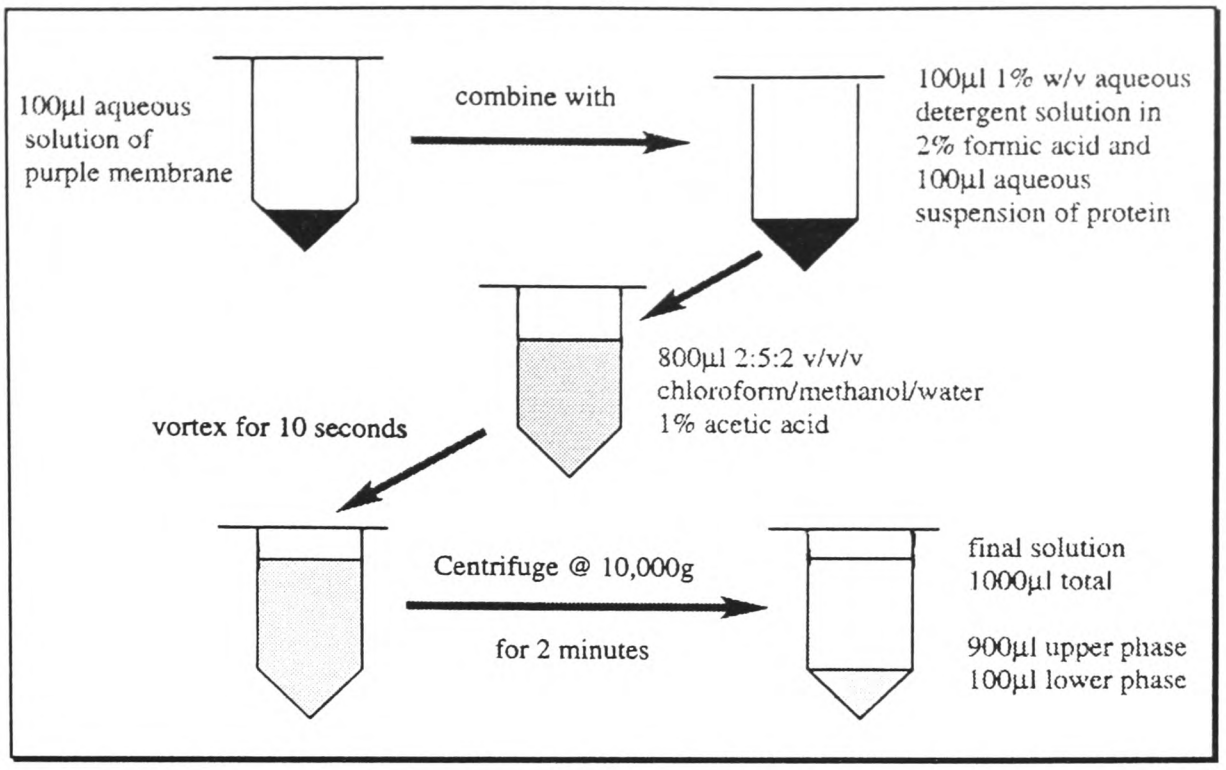


Fig 2.1 - Phase separation procedure for isolation of membrane proteins

2.3.6.2 Exchange of solvent system after preparation of samples by the phase separation method

Samples were prepared as described in section 2.3.5.1 and the lower chloroform rich layer transferred to a glass vial. The solvent was removed using a Büchi Rotavapor R-114. The sample was then resuspended in 200 µl of the solvent system under investigation.

2.3.6.3 Deuteriation of purple membrane

Purple membrane (0.2 mg) was concentrated using a nanosep 10 K column purchased from Flowogen. The sample was centrifuged for 10 min at 13000 rpm and the amount of H_2O removed was measured using a syringe, the H_2O was replaced with D_2O and the sample mixed thoroughly. The above procedure was repeated a further two times in an attempt to exchange as much of the H_2O present in the sample for D_2O. The final volume of purple membrane was made up to 200 µl by the addition of more D_2O and left overnight at room temperature.

Chapter 3:

Development of a Prokaryotic Expression System for the bacterio-opsin gene

Chapter 3: Development of a Prokaryotic Expression System for the bacterio-opsin gene

3.1 Introduction

Whether or not a recombinant protein can be produced in significant quantities in *E. coli* depends very much on the individual gene concerned. To carry out the expression of a foreign gene in *E.coli,* it must be encoded in to the expression vector under the control of the *E. coli* expression signals which direct transcription and translation of the gene. Although many features are individual to a plasmid, many features of gene expression can be applied to all expression systems and be further optimised to improve the expression levels of a given gene. Some features may be advantageous in one particular sense but not so useful in another and so all these factors must be taken into consideration in order to reach a compromising situation.

Major advances in recombinant DNA technology over the past decade have prompted the investigation of modern expression systems for the *bO* gene. This chapter describes the PCR mediated mutagenesis of the synthetic *bO* gene and its subsequent cloning into seven selected bacterial expression vectors. Transformation of the recombinant DNA molecules into six different *E. coli* cell lines has allowed expression studies to be performed and expression systems to be compared. Growth conditions including choice of media, temperature, induction times/levels were also optimised. The seven bacterial expression vectors chosen along with the six bacterial cell lines that were used to carry out expression studies are listed below:

Expression vectors: pPL1, pRH1090, pT7-7, pET16, pCYB2 pT7 IMPACT and pLEX.

E. coli strains: G1724, HMS174, BL21(DE3), BL21(DE3)pLysS, B834(DE3) and C600.

3.2 Results and discussion

3.2.1 *bO* gene manipulation to incorporate an *NdeI* site

The *bO* gene (Nassal *et al.*, 1987) was excised from the pSB02 vector (kind gift from H. Gobind Khorana) as an *HindIII/ EcoRI* fragment and cloned into pBluescript at the *EcoRV* site. Polymerase chain-reaction (PCR) mediated mutagenesis of the synthetic *bO* gene was performed with an oligonucleotide primer to incorporate the *NdeI* restriction site [5′-d(CATATG)-3′] to overlay the ATG initiation codon of the gene (figure 3.1). This would facilitate the cloning of the *bO* gene into some of the selected expression vectors allowing in-frame insertion.

NdeI

5′ G **CAT ATG** CAA GCT CAA ATT ACT GG 3′

start codon

Oligonucleotide coding for the *Nde* I start codon

Figure 3.1 Oligonucleotide primer to incorporate *NdeI* restriction site into *bO* gene

The PCR product was cloned into the pUC19 vector and the mutation was confirmed by DNA sequencing (see figure 3.3). The pUC19 sequencing primers M13 and NR are shown in figure 3.2.

pUC19 SEQUENCING PRIMERS

M13 (22 mer) 5′- CTTCGCTATTACGCCAGCTGGC- 3′

NR (21 mer) 5′- CCGCTCACAATTCCACACAAC-3′

Figure 3.2 pUC19 sequencing primers

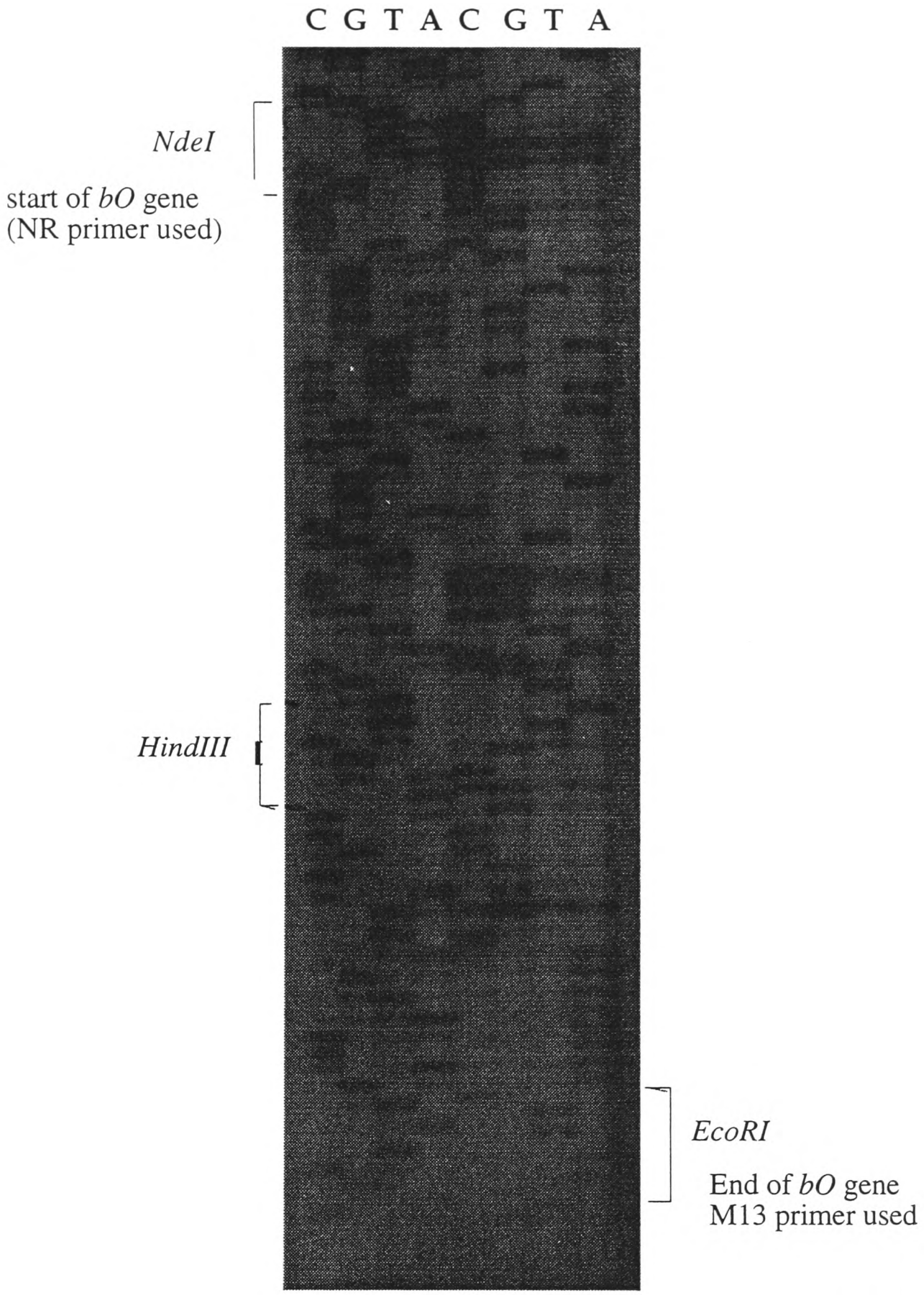


Figure 3.3 Section of *bO* gene sequence showing *NdeI* restriction site incorporated at start of gene

3.2.2 General methodology employed to generate expression vectors containing the *bO* gene.

After PCR mediated mutagenesis of the *bO* gene to incorporate the *NdeI* restriction site, it was digested with the appropriate restriction enzymes and ligated into expression vectors cut with compatible enzymes. Some vectors allowed insertion of the *bO* gene without any PCR manipulation. *E. coli* cell lines were then transformed with the different vectors. Expression was carried out by growing the cells to a known OD_{600} ~1.0 and inducing them by adding a chemical to the growth medium and/or increasing the temperature. Since this depends on the individual expression vector, it is discussed separately for each case. The cells were grown for a further period of time and harvested. Expression levels were investigated by immunoblotting. The overall results are summarised in Table 3.1.

3.2.2.1 Studies using the pPL1 expression vector

3.2.2.1.1 Constructing the pPL1/*bO* expression vector

pPL1 is a 2.8 kb expression vector that allows insertion of the *bO* gene nine nucleotides downstream of the Shine-Dalgarno (SD) sequence. Expression is under the control of the λP_L promoter (Nassal *et al.*, 1987). Construction of the pPL1/*bO* expression vector is outlined in figure 3.4 with the polylinker region shown in figure 3.5. The *bO* gene was excised from pBluescript as an *HindIII/EcoRI* fragment and cloned into the pPL1 expression vector which was pre-cut using the same restriction enzymes. The resultant construct was introduced into the JM109 cloning host. Appropriate restriction digests using *HindIII/EcoRI* were performed to check the presence of the *bO* gene. Restriction fragments were analysed by agarose gel electrophoresis as shown in figure 3.6.

Figure 3.4 Construction of pPL1/*bO* (adapted from Nassal *et al.*, 1987)

Figure 3.5 Polylinker region of pPL1 (adapted from Nassal *et al.*, 1987)

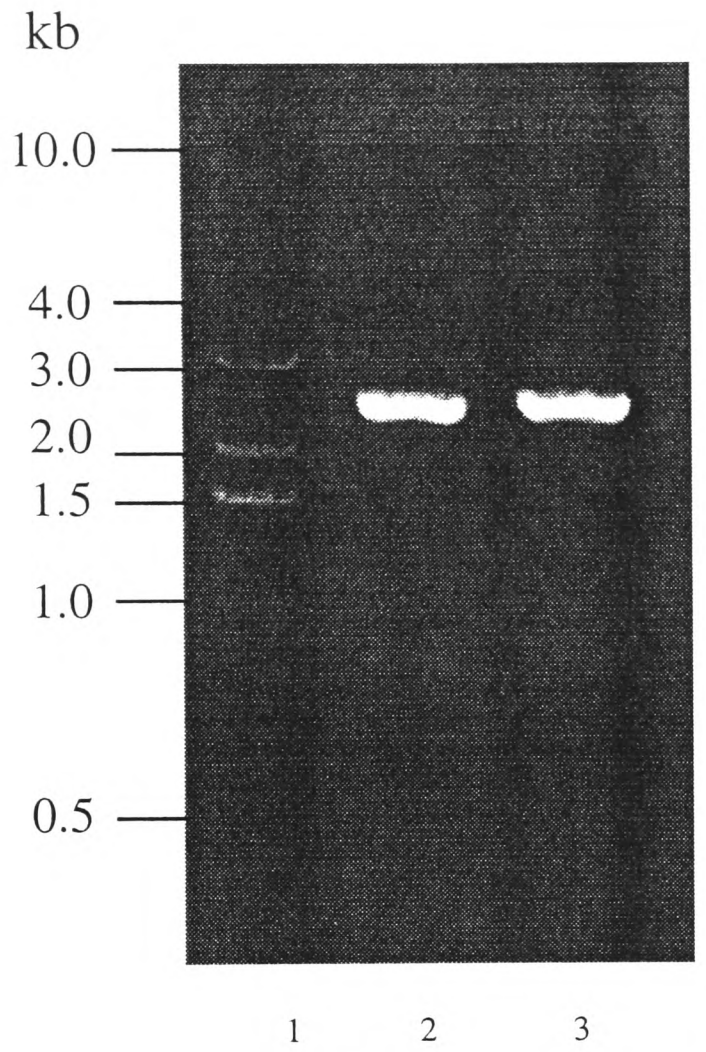


Figure 3.6 Agarose gel analysis of pPL1/*bO* restriction digests

Lane 1:1 kb DNA ladder from NEB as indicated
Lane 2: pPL1/*bO* digest with *HindIII/ EcoRI*
Lane 3: another pPL1/*bO* sample also digested with *HindIII/ EcoRI*

3.2.2.1.2 Expression of *bO* in pPL1

The *E. coli* cells were co-transformed with pPL1/*bO* and the pc1857 repressor plasmid which carries the kanamycin and ampicillin resistance genes (Karnik *et al.*, 1987). They were grown at 30°C to mid-exponential phase and then induced by shifting the temperature up to 42°C. The cells were grown for a further hour and then harvested. Expression levels were detected by SDS-PAGE analysis.

3.2.2.1.3 Discussion on expression of *bO* using pPL1 system

The pPL1 thermoinducible expression system was used by Khorana and his colleagues (Karnik *et al.*, 1987; Nassal *et al.*, 1987) back in the late 80s to express the *bO* gene in *E. coli*. The pPL1 expression vector was originally constructed from pPLc28 where the polylinker had been replaced with a synthetic duplex containing new restriction sites and destroying some existing ones.

The *bO* gene was cloned into the pPL1 vector and expressed as described in section 3.2.2.1.1 and 3.2.2.1.2 respectively. At this point, we did not have access to a bR antibody and so expression levels were detected by SDS-PAGE analysis. Since the yields of *bO* produced in *E. coli* are low, it was important to carry out the organic extraction procedure (see chapter 4) in order to detect any expression using SDS-PAGE analysis. Figure 3.7 shows the expression levels obtained using this system which were really quite poor. Barely visible by SDS-PAGE analysis, the yield was estimated to be 0.3 mg/L when compared to a known amount of authentic *bO* sample.

Since low yields have always been a problem in the expression of *bO*, it was decided to explore alternative expression systems.

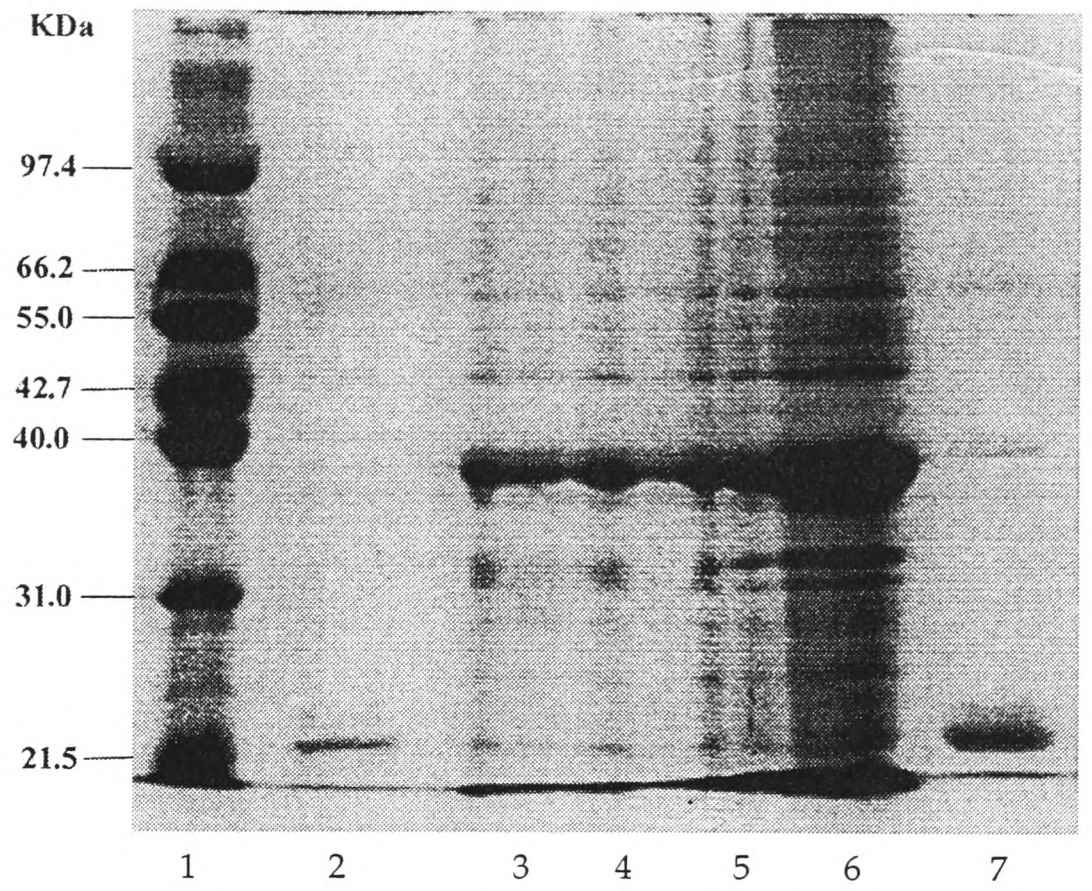


Figure 3.7 Expression and purification of *bO* from *E. coli* using the pPL1 expression system.

Lane 1: Protein Markers, mid-range promega as indicated
Lane 2: Purple membrane standard.
Lanes 3, 4, 5 and 6: increasing concentrations of the same sample of *bO* expressed and purified using the organic extraction procedure.
Lane 7: Purple membrane standard.

3.2.2.2 Studies using the pRH1090 expression vector

3.2.2.2.1 Constructing the pRH1090/*bO* expression vector

pRH1090 (kind gift from Dr. J. Sutherland, Oxford) is a 5.8 kb expression vector that contains both the chloramphenicol and kanamycin resistance gene. Expression is under the control of the Ptrc promoter. Construction of the pRH1090/*bO* expression vector is outlined in figure 3.8. The *bO* gene was excised from pBluescript as an *NdeI/HindIII* fragment. Cloning into the *NdeI/HindIII* site of pRH1090 resulted in the loss of kanamycin resistance gene and so the resultant construct carried only the chloramphenicol resistance gene. The presence of the *bO* gene was confirmed by restriction digest using the appropriate restriction enzymes and subsequent analysis using agarose gel electrophoresis as shown in figure 3.9.

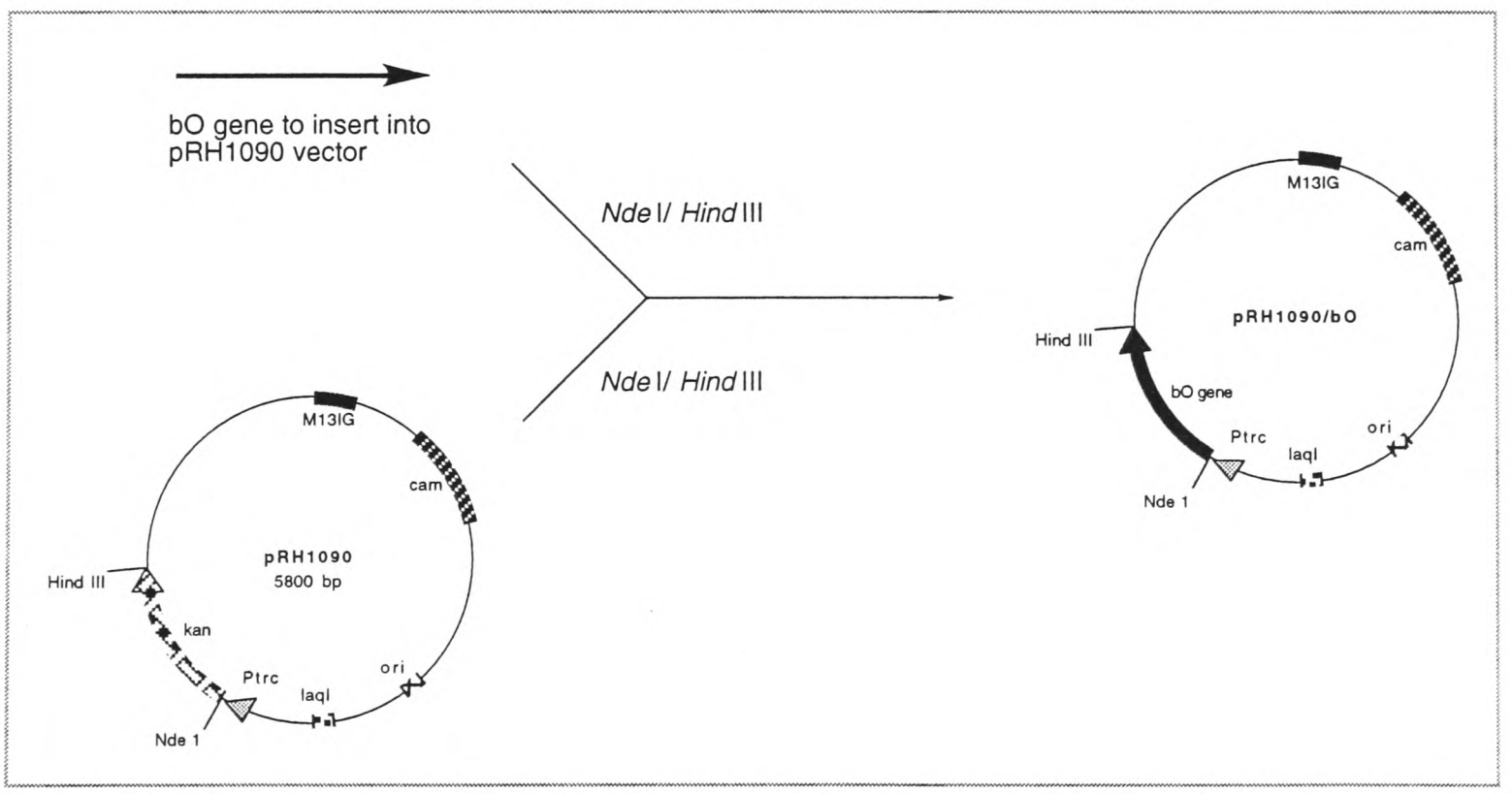


Figure 3.8 Construction of pRH1090/*bO*

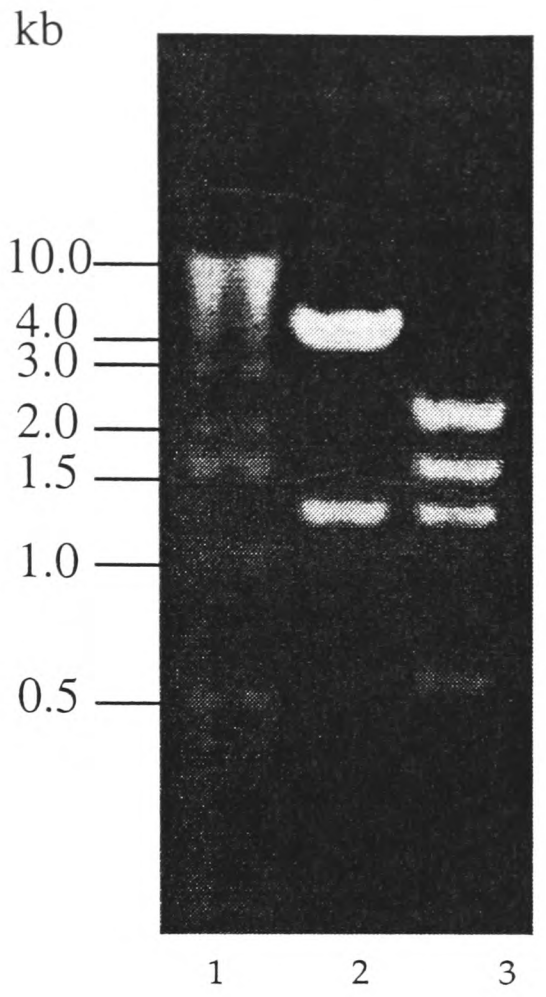


Figure 3.9 Agarose gel analysis of pRH1090/*bO* restriction digests

Lane 1: 1kb DNA ladder from NEB as indicated
Lane 2: pRH1090/*bO* digest with *NdeI/ BamHI* (10, 1274, 4526 bp)
Lane 3: same sample as above digested with *EcoRI/ BamHI* (531, 1264, 1628, 2377 bp)

Restriction fragment analysis confirms the *bO* gene was successfully cloned into the pRH1090 vector.

3.2.2.2.2 Expression of *bO* in pRH1090

The pRH1090/*bO* plasmid was transformed into *E. coli* cells and grown at 37°C to mid-exponential phase. IPTG was added at a final concentration of 1 mM to induce expression. The cells were grown for a further hour at 37°C and then harvested. Expression levels were detected by SDS-PAGE analysis.

3.2.2.2.3 Discussion on expression of *bO* using the pRH1090 system

The *bO* gene was inserted into the pRH1090 expression vector (3.2.2.2.1) and expressed (3.2.2.2.2). Since we did not have an antibody for bR at this stage of our investigations, expression levels were detected by SDS-PAGE.

A small scale expression showed promising results (figure 3.10). The majority of the protein was collected as insoluble aggregates and separated by SDS-PAGE. The correct band was then excised and the protein eluted using electro-elution. Our results estimated the expression of *bO* using this system to be approximately 400 mg/L. This was an extraordinary result (figure 3.11).

N-terminal protein sequence analysis was preformed on the protein but unfortunately, it did not match the N-terminal sequence of *bO* (figure 3.12).

N-terminal protein sequence obtained:

Ser His Ile Gln Arg Glu Thr Ser Cys Ser Arg Pro Arg Leu Asn Ser Asn Met Asp Ala

bO sequence:

Met Gln Ala Gln Ile Thr Gly Arg Pro Glu Trp Ile Trp Leu Ala Leu Thr Ala Leu Met

Figure 3.12 N-terminal protein sequence analysis of protein expressed using pRH1090/*bO* in C600

The sequence obtained from the N-terminal analysis was entered into a protein databank search on the internet and it was found to match that of aminoglycosidase 3′ phosphotransferase (kanamycin kinase type 1 from *E. coli*). Further analysis of the DNA sequencing results revealed that small sections of *bO* gene had been deleted and replaced with sequences from the kanamycin gene. The difference in the N-terminal sequencing results may therefore be the result of a frameshift. The result was rather unexpected as all previous stages of analysis showed the successful insertion of the gene and identified restriction fragments of the correct size by agarose gel electrophoresis. The cause of the gene interruption is unknown and may be due to instability of the gene in that particular system. Thus we were unable to express the *bO* gene using this expression system.

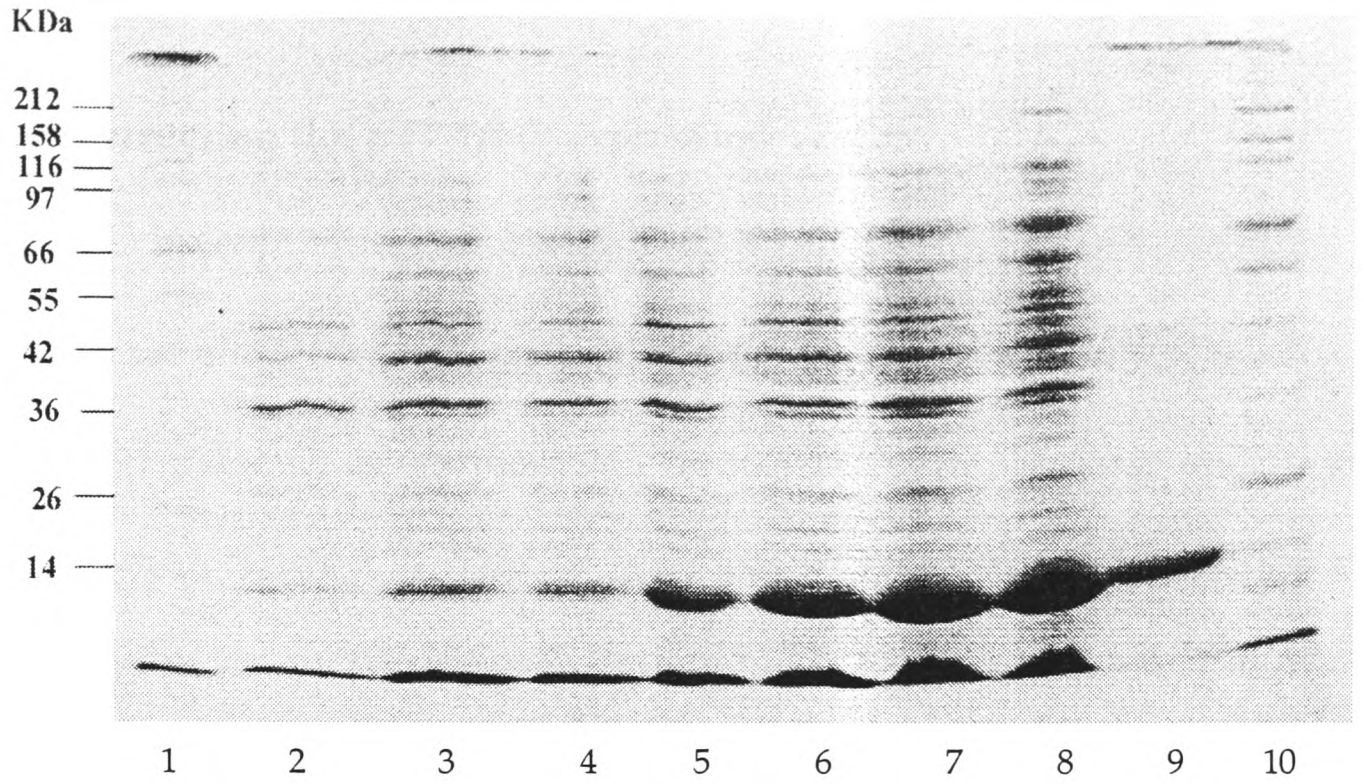


Figure 3.10 Expression of *bO* using the pRH1090 expression system

Lane 1: Protein markers as indicated. Lane 2: uniduced cells. Lane 3 to 8: induced cells 30 min, 1 hour, 2 hours, 3 hours and 4 hours after induction respectively. Results show increasing levels of expression with time for a protein whose electrophoretic mobility is identical to that of purple membrane. N-terminal sequencing results however indicated that the protein expressed was not *bO*. Lane 9: purple membrane standard. Lane 10: protein markers

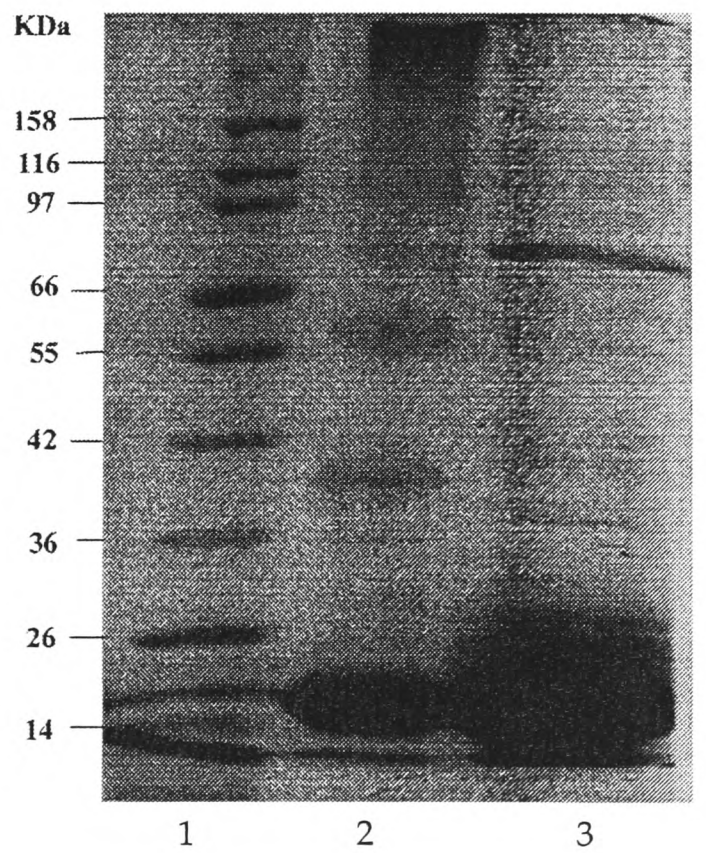


Figure 3.11 Electroelution of *bO* expressed using the pRH1090 system.

Lane 1: protein markers as indicated. Lane 2: purple membrane. Lane 3: electroelution of protein expressed using the pRH1090 expression system. Further analysis showed the protein was not *bO*.

3.2.2.3 Studies using the pT7-7 expression vector

3.2.2.3.1 Constructing the pT7-7/*bO* expression vector

pT7-7 is a 2473 bp expression vector derived from pBR322 which contains the T7 RNA polymerase promoter ϕ10 and an ampicillin resistance gene. Construction of this expression plasmid is outlined in figure 3.13. The cloning and expression region is shown in figure 3.14. The *bO* gene was inserted from pBluescript into the polylinker region as an *NdeI/EcoRI* fragment. It was introduced into the JM109 cloning host and the plasmid DNA screened for the presence of the *bO* gene using the appropriate restriction enzymes. Fragments were analysed using agarose gel electrophoresis as shown in figure 3.15.

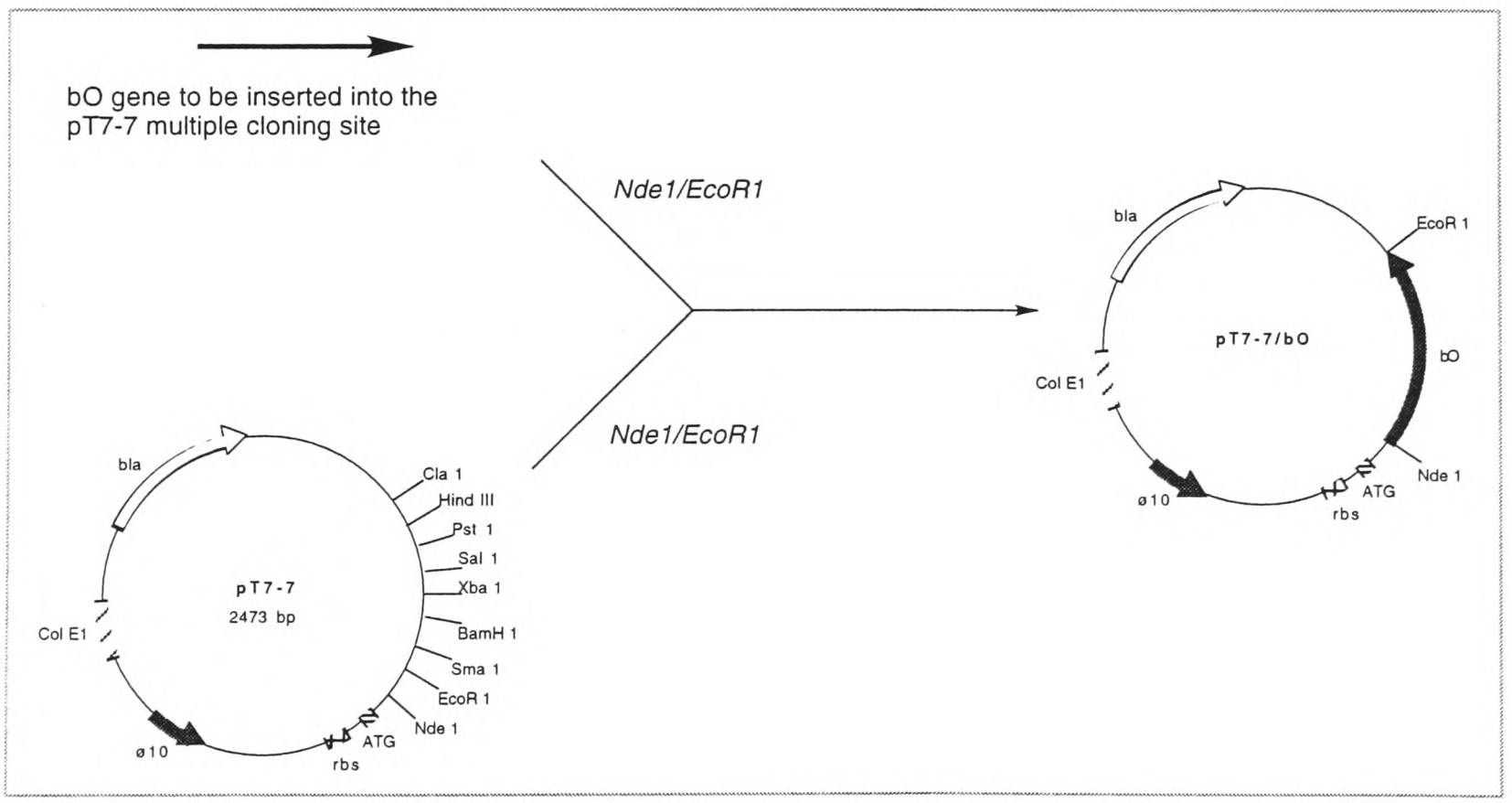


Figure 3.13 Construction of pT7-7/*bO*

```
                                                                T7promoter
CGATTCGAACTTCTCGATTCGAACTTCTGATAGACTTCGAAATTAATACGACT

 T7 promoter (cont.)
CACTATAGGGAGACCACAACGGTTTCCCCTAGAAATAATTTTGTTTAACTTTA

                                Met  Ala  Arg  Ile  Arg  Ala  Arg  Gly  Ser  Ser
AGAAGGAGATATACAT ATG GCT AGA ATT CGC GCC CGG GGA TCC TCT
      rbs              NdeI             EcoRI        SmaI        BamHI

 Arg  Val  Asp  Leu  Gln  Pro  Lys  Leu  Ile  Ile  Asp
ACA GTC GAC CTG CAG CCC AAG CTT ATC ATC GAT
XbaI      SalI PstI            HindIII      ClaI
```

Figure 3.14 pT7-7 cloning and expression region

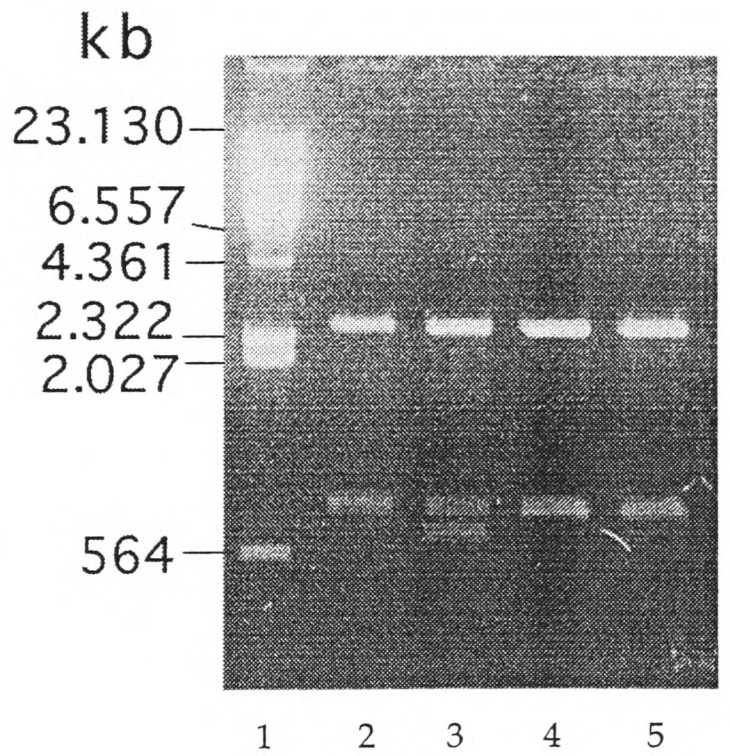


Figure 3.15 Agarose gel of pT7-7/*bO* restiction digests

Lane 1: λ DNA *HindIII* fragments form Gibco BRL™. Lanes 2, 3, 4 and 5 show different samples of pT7 7/*bO* digested with *NdeI* and *EcoRI*. Sample 3 was discarded as it did not appear to contain the correct insert.

3.2.2.3.2 Expression of *bO* in pT7-7

pT7-7/*bO* was transformed into the selected *E. coli* bacterial strains and grown at 37°C to OD_{600} value 1.0. IPTG was added to a final concentration of 1 mM and the cells grown for a further hour. They were then harvested and the expression levels were detected by immunoblotting.

3.2.2.3.3 Discussion on expression of *bO* using the pT7-7 system

The pT7-7 vector was a kind gift from Dr. Dominic Campopiano (Edinburgh, UK). It contained suitable restriction sites to clone the *bO* gene (figure 3.11). Insertion at the *NdeI* site allowed the gene to be cloned in-frame and eliminated the addition of any amino acid residues at the N-terminus. The gene was cloned and expressed in all the selected bacterial strains (3.2.2.3.1 and 3.2.2.3.2 respectively). Unfortunately, no expression was detected in any of the cell lines (Data not shown). It was thought that perhaps the *bO* gene was extremely toxic to the host strain.

3.2.2.4 Studies using the pET16b expression vector

3.2.2.4.1 Constructing the pET16b/*bO* expression vector

pET16b is a 5711 bp expression vector that allows expression of target genes using a bacteriophage T7 RNA polymerase promoter. The host cell provides the T7 RNA polymerase. It contains the ampicillin resistance gene and a hexahistidine coding sequence which is fused to the target protein and may aid in its subsequent purification. The methodology used to insert the *bO* gene into the pET16b

expression vector is outlined in figure 3.16. The cloning and expression region is shown in figure 3.17. The *bO* gene was excised from pBluescript as an *NdeI/EcoRI* fragment and digested with the exonuclease Klenow fragment to create a blunt site. The gene was then inserted into the polylinker region of pET16b which was pre-excised with *NdeI* and *BamHI* and also digested with exonuclease Klenow fragment to create a blunt end. The resultant construct was introduced into the JM109 cloning host. To check insertion of the *bO* gene into the pET16b vector, the appropriate restriction digests were performed and the resultant fragments analysed by agarose gel electrophoresis as shown in figure 3.18.

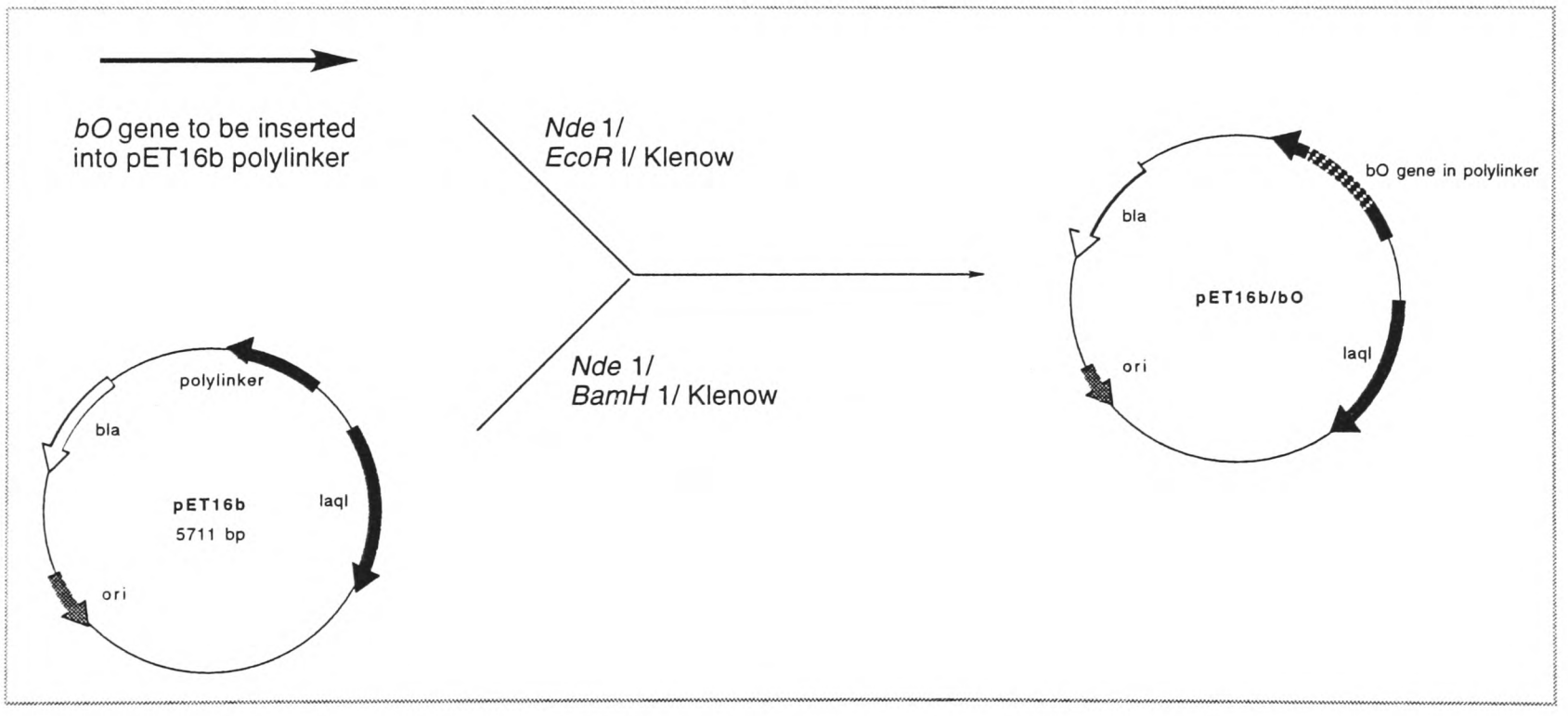


Figure 3.16 Construction of pET16b/*bO*

BglII — *T7 promoter* — *lac operator*

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCC

XbaI — *NcoI*

CCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCCATCATCATCATCAT

Met Gly **His His His His His**

NdeI *XhoI* *BamHI*

CATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGCTCGAGGATCCGGCTGCT

His His His His His Ser Ser Gly His Ile Glu Gly Arg His Met Leu Glu Asp Pro Ala Ala

Factor Xa

EspI

AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCC

Asn Lys Ala Arg Lys Glu Ala Glu Leu Ala Ala Ala Thr Ala Glu Gln End

EcoRV

CTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATC

pET16b Cloning/Expression Region
(modified from Novagen catalogue)

Figure 3.17 pET16b cloning and expression region (pET system manual, 6th edition)

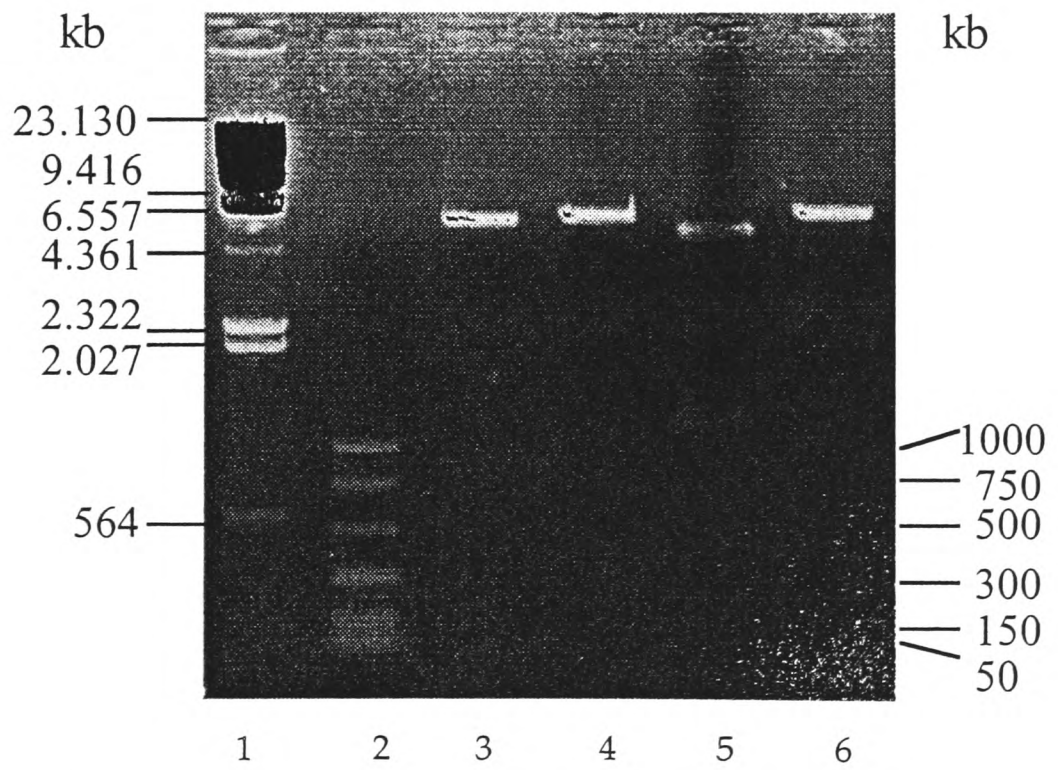


Figure 3.18 Agarose gel analysis of pET16b/*bO* digests

Lane 1: λ DNA/*HindIII* fragments from Gibco BRL™. Lane 2: PCR markers from Promega. Lane 3: Digest of pRH1090/*bO* with *Nde* I-linearises vector. Lane 4: Digest of pRH1090/*bO* with *EcoR* 1-linearises vector. Lane 5: Digest of pRH1090 with *Nde* I/ *EcoR* I- cuts out *bO* gene. Lane 6: Digest of pRH1090/*bO* with *EcoICR*

3.2.2.4.2 Expression of *bO* in pET16b

To initiate expression, the recombinant plasmid was transferred to host *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. Such hosts are lysogens of bacteriophage DE3, a λ DNA derivative that carries a DNA fragment containing the *lacI* gene and *UV5* promoter as well as the gene for T7 RNA polymerase. Once this lysogen is formed, only the lacUV5 promoter is known to direct transcription of the T7 RNA polymerase gene which is inducible by IPTG allowing transcription of the target DNA in the plasmid. Cells were grown at 37°C induced at OD_{600} 1.0 by addition of IPTG to a final concentration of 1 mM. They were grown for an hour and then harvested. Expression levels were detected by immunoblotting.

Only the HMS174 cell strain allowed the expression of *bO* although it was leaky (figure 3.19). Note that the band is slightly higher compared to the authentic *bO* sample due to the presence of a His tag. Expression using B834 is also shown as a comparison (figure 3.20) as this strain along with all the others did not allow the expression of *bO*.

3.2.2.4.3 Discussion on the Expression of *bO* using the pET system

The pET16b vector belongs to a series of vectors which were originally developed by Studier and colleagues (1990). This *E. coli* based expression system allows target genes to be cloned under the control of strong bacteriophage T7 transcription and (optionally) translation signals. Expression requires the host cell to provide a source of T7 RNA polymerase. The RNA polymerase of bacteriophage T7 is unlikely to recognise a DNA promoter unless it is related to T7 DNA (Dunn and Studier, 1983) rendering it a more selective system. It is also five times more

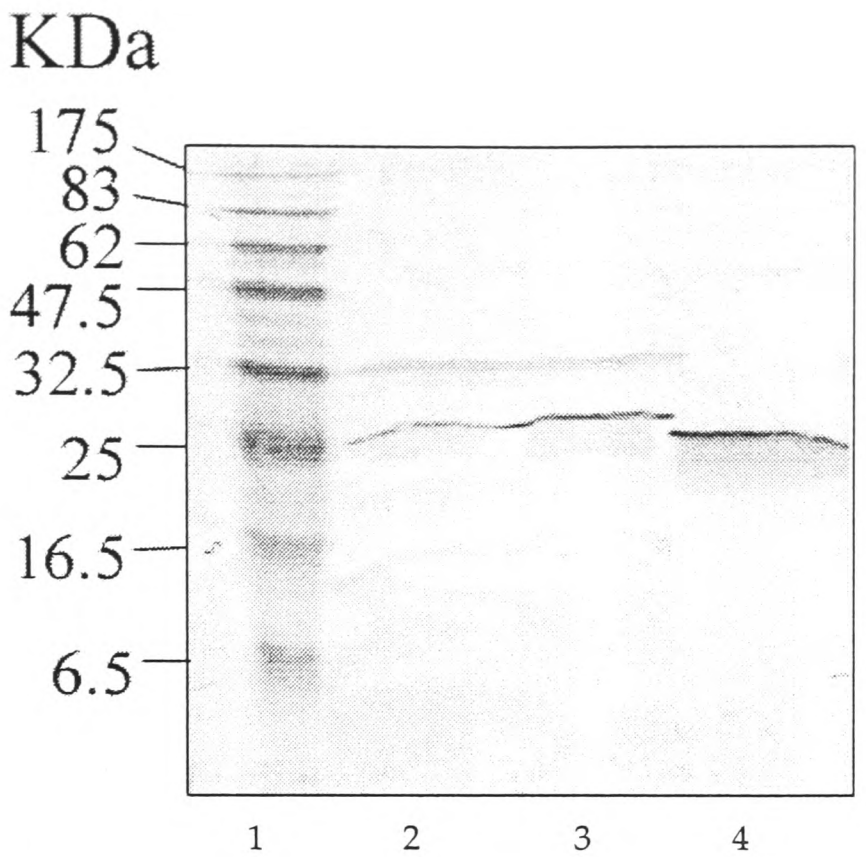


Figure 3.19 Immunoblot of *bO* expression in pET16b in HMS174

Lane 1: Prestained protein markers, broad range from NEB. Lane 2: uninduced cells show leaky expression. Lane 3: induced cells. Lane 4: purple membrane standard (0.3 μg)

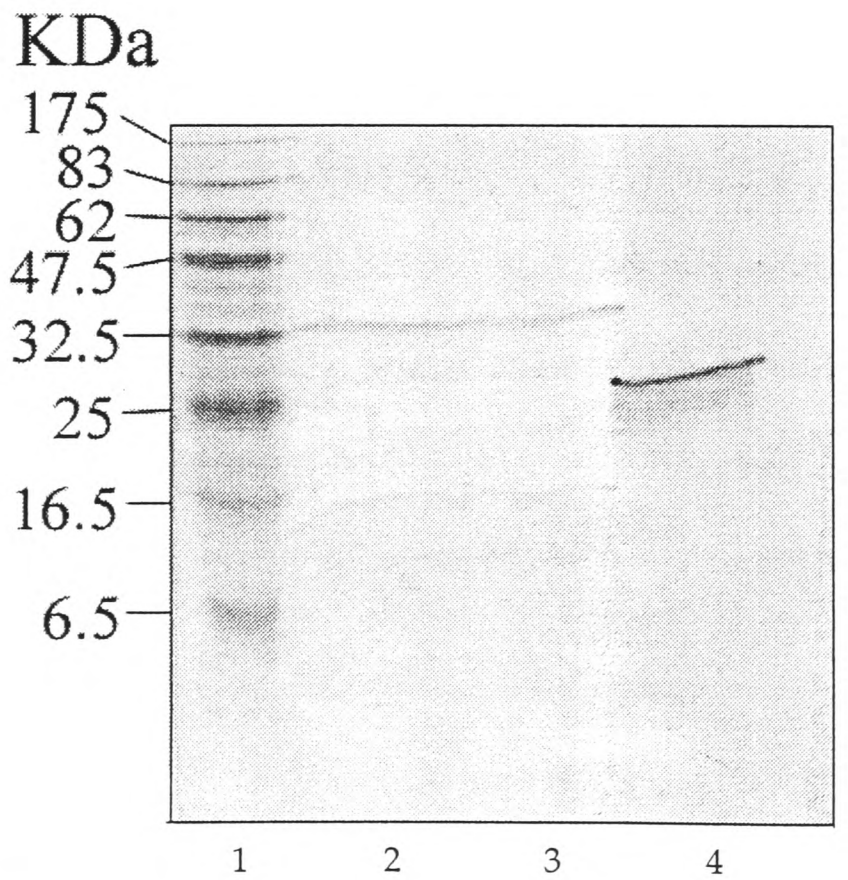


Figure 3.20 Immunoblot of *bO* expression in pET16b in B834

Lane 1: Prestained protein markers, broad range from NEB. Lane 2: uninduced cells Lane 3: induced cells (no expression). Lane 4: purple membrane standard (0.3 μg)

efficient than *E. coli* RNA polymerase (Golomb and Chamberlain, 1974). If the target gene is potentially toxic to the host cell, the unstable plasmid can be cloned using hosts that do not contain the T7 RNA polymerase gene. It can then be transformed into a host cell that does contain a chromosomal copy of the T7 RNA polymerase gene under the control of lacUV5 which requires the addition of IPTG to induce expression.

Novagen have further developed on the original idea by Studier *et al* (1990) to produce a wide variety of vectors with added features to allow easy cloning, detection of the protein and subsequent purification. One such vector we selected for our studies was the pET16b vector which also contains a polyhistidine-coding sequence which can be exploited to purify the fusion protein *via* metal-affinity chromatography (see chapter 4). The *bO* gene was cloned and expressed in the pET16b vector as described in sections 3.2.2.4.1 and 3.2.2.4.2 respectively. Results obtained *via* immunoblotting revealed that the protein was expressed only in the HMS174 strain at 0.1 mg/L and no expression was observed for any of the other cell lines. Usually, the HMS174 cell strain is regarded not as an expression host but one used for the initial cloning and maintenance of the target DNA in a pET vector (Studier *et al.*, 1990). Perhaps the *bO* gene was too unstable in the other host strains and exerted toxicity upon them which is why expression was only observed in the host it was most stable in.

3.2.2.5 Studies using the pCYB2 expression vector

3.2.2.5.1 Constructing the pCYB2 expression vector

pCYB2 is a medium copy number *E. coli* cloning and expression vector which is 6843 bp in length. It contains the gene for ampicillin resistance, the M13 origin, the ColE1 origin and the lacIq gene. The *bO* gene was cloned into this vector as illustrated in figure 3.21. The cloning and expression region is shown in figure 3.22. The pCYB2 polylinker region allows in-frame insertion of the *bO* gene upstream of the intein region and contains an *NdeI* restriction site for translation initiation. The *bO* gene was excised from pBluescript as an *NdeI/NotI* fragment. This resulted in the loss of seven base pairs from the C-terminus of the *bO* gene which were considered not to play an important role in the structure and function of the protein. The *NotI* restriction site was chosen since it is possible to create a blunt site using the Klenow enzyme which will allow the gene to be cloned into the *SmaI* site of pCYB2. Using the *SmaI* site, cleavage of the fusion protein occurs at the peptide bond between glycine and the N-terminal cysteine and thus the protein has an additional glycine residue present at the C-terminus. The resultant construct was introduced into the JM109 cloning host. To check insertion of the *bO* gene into the pCYB2 expression vector, the construct was digested with the appropriate restriction enzymes (*PstI, HindIII, BamHI, MluI, BspHI* and *XhoI*) and the corresponding fragments analysed by gel electrophoresis as shown in figure 3.23.

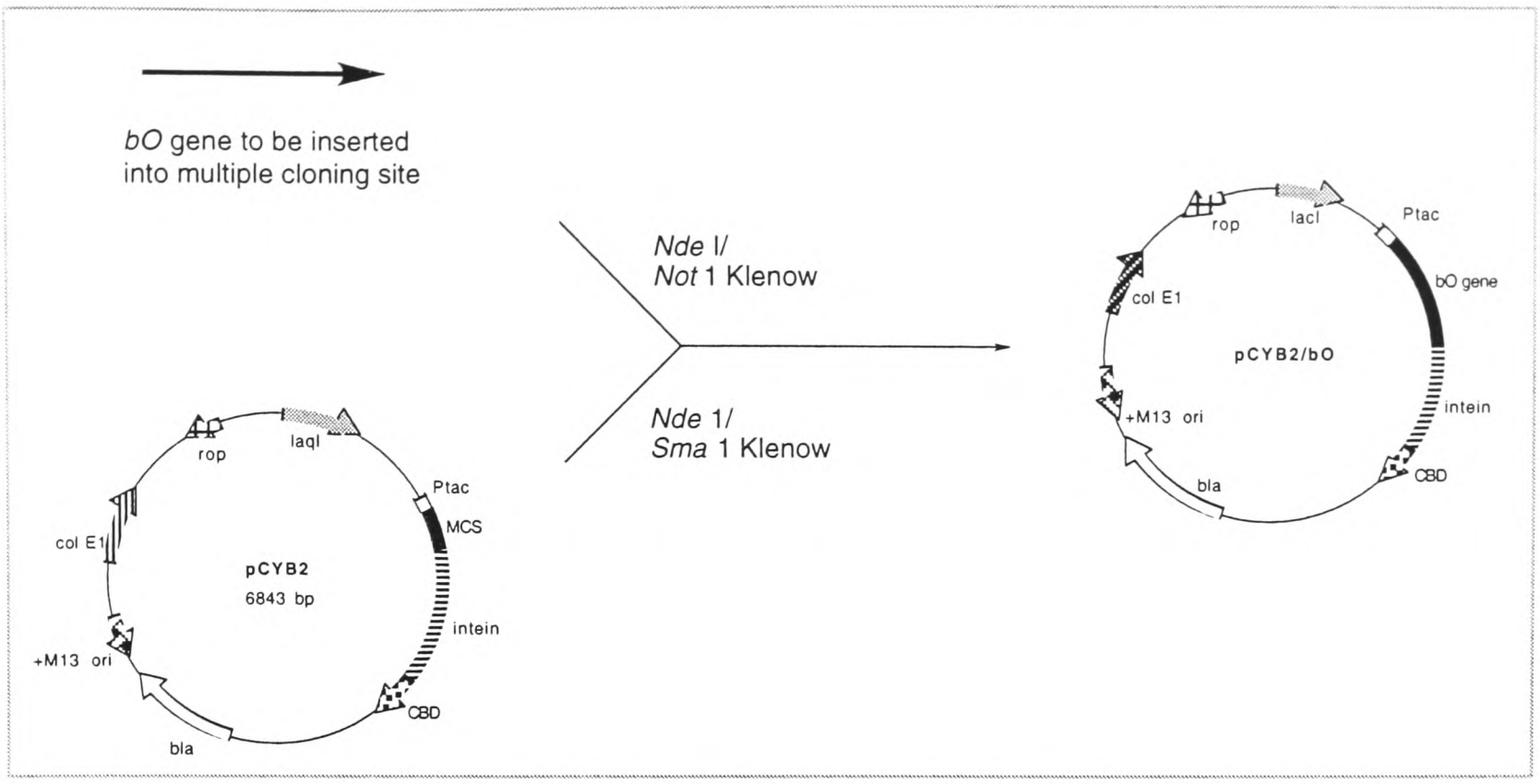


Figure 3.21 Construction of pCYB2/*bO*

P_{tac} Forward sequencing primer →

5'... G ACC GGA TAA CAA TTT CAC ACA GGA AAC AGC CAG TCC GTT TAG GTG TTT TCA CGA

Met Ala Ser Ser Arg Val Asp Ser Arg Glu

GCA CTT CAC CAA CAA GGA CCA TAG CAT ATG GCT AGC TCG CGA GTC GAC TCT AGA GAA

NdeI *NheI* *NruI* *SalI* *XbaI* *EcoRI*

Intein→

Phe Leu Glu Pro Gly CysI

TTC CTC GAG CCC GGG TGC TTT GCC AAG GGT ACC AAT CTT TTA ATG GCG GAT GGG TCT ATT

XhoI *SmaI*

GAA TCT ATT GAA AAC ATT GAG GTT GGT AAT AAG GTC ATG GGT...3'

← Intein Reverse Sequencing Primer

Figure 3.22 pCYB2 Cloning/expression region (modified from New England Biolabs manual)

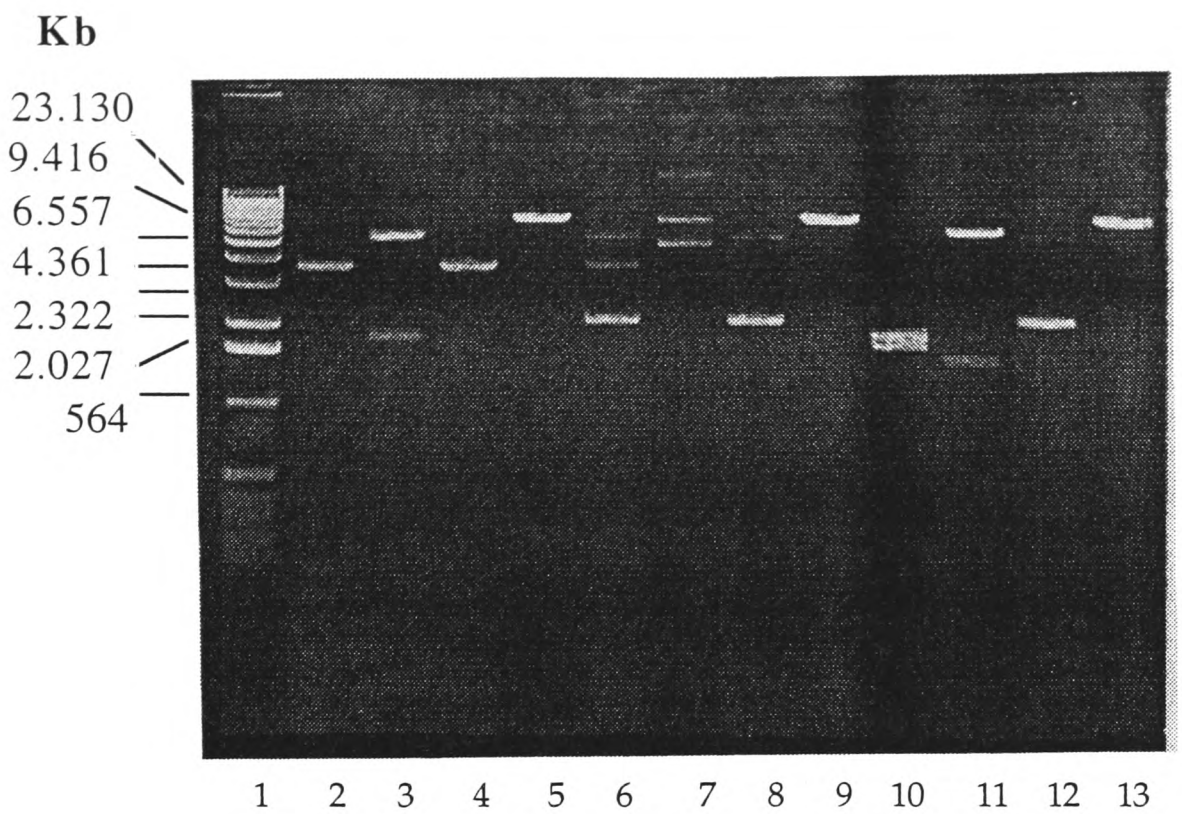


Figure 3.23 Agarose gel of pCYB2/*bO* restriction digests

Lane No.	Sample	Restriction Enzyme
1	λ DNA/*Hind*III fragments from Gibco BRL™	-
2	pCYB2/bO	*Pst* I
3	pCYB2/bO*	*Pst* I
4	pCYB2/bO	*Hind* III
5	pCYB2/bO*	*Hind*III
6	pCYB2/bO	*BamH* I
7	pCYB2/bO*	*BamH* I
8	pCYB2/bO	*Mlu* I
9	pCYB2/bO*	*Mlu* I
10	pCYB2/bO	*BspH* I
11	pCYB2/bO*	*BspH* I
12	pCYB2/bO	*Xho* I
13	pCYB2/bO*	*Xho* I

Two samples were differentiated using *. Sample pCYB2/bO was found to contain the correct insert and thus pCYB2/bO* was discarded.

3.2.2.5.2 Expression studies on pCYB2/*bO*

pCYB2/*bO* was transformed into the selected *E. coli* bacterial strains (HMS174, BL21(DE3), BL21(DE3)pLysS, B834(DE3) and C600) and grown at 37°C to OD_{600} value 1.0. IPTG was added to a final concentration of 1 mM and the cells grown for a further hour. They were then harvested and the expression levels were detected by immunoblotting. By using the bR antibody (kind gift from Koichi Koyama from Japan, Koyama *et al.*, 1994). we hoped to detect the expression of the fusion protein but no bands were observed. It was thought that perhaps the structure of the fusion prevented the bR antibody binding to the region of interest. Thus, we decided to use an antibody to the intein region. Unfortunately, it was found to be very unspecific as many bands were highlighted even in a control sample where no intein was present. Since results by immunoblotting seemed unreliable, we decided to analyse the expression by SDS-PAGE which is usually not possible due to the low expression levels of *bO* produced but the expression of a protein as a fusion would perhaps allow detection. No protein was observed at this stage either.

3.2.2.5.3 Discussion on the Expression of pCYB2/*bO*

Detectable amounts of *bO* were not produced using this system. It was thought that perhaps the promoter was not strong enough for a *bO* as the expression of this protein proves lethal to the host. Manufacters of the expression system recommend cloning the of the fusion protein under the control of a strong pT7 promoter if it is toxic. At the time, they did not have a license to produce a commercial vector for this purpose and so the pT7IMPACT expression vector was constructed.

3.2.2.6 Studies using the pT7IMPACT expression vector

3.2.2.6.1 Construction the pT7IMPACT/*bO* expression vector

The T7IMPACT expression vector was constructed using a combination of the pT7 -7 and pCYB2 expression vectors which have been previously described in figures 3.13 and 3.21 respectively. The overall strategy used to construct this vector is outlined in figure 3.24. The aim was to create an expression vector where the *bO* gene could be constructed as a fusion to the intein and CBD regions (for purification purposes) under the control of a T7 promoter which would provide a tighter control on the expression. The intein/CBD region was excised from pCYB2 using *NdeI/PstI* and cloned into the pT7-7 vector using the same sites. The resultant construct was designated pT7-7/Int/CBD. Using the expression vector pT7-7/bO (see section 3.10), the *bO* gene was excised using the *NdeI* and *NotI* restriction enzymes. The *NotI* site was digested with the exonuclease Klenow fragment to create a blunt site. The pT7-7/Int/CBD vector was excised simultaneously using the *NdeI* and *SmaI* restriction enzymes. The *SmaI* site again allowed a blunt site to be created after digesting with the exonuclease Klenow fragment. The *bO* gene was cloned into the pT7-7/Int/CBD vector to produce a new expression vector which was designated pT7IMPACT/*bO*. The construct was cloned into the JM109 cloning host and screened for the presence of the *bO* gene using restriction analysis with the appropriate enzymes on an agarose gel as shown in figure 3.25.

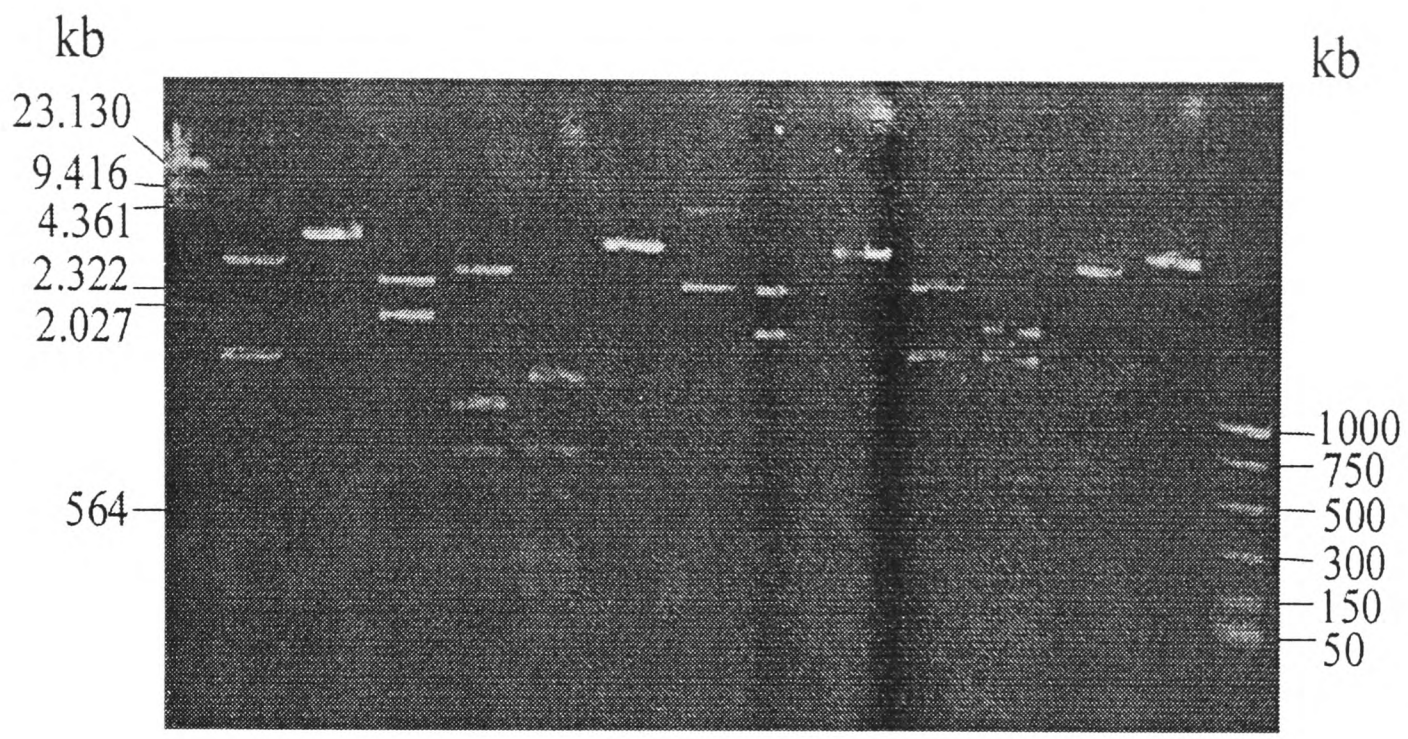


Figure 3.25 restriction digest of pT7IMPACT/*bO*

Lane No. in gel	Restriction enzyme	No. of cuts in bO	No. of cuts in pT7 7 vector	No. of cuts in Int/CBD
1	*AatII*	1	1	0
2	*AvaI*	1	0	0
3	*BamHI*	1	0	1
4	*BglI*	1	1	1
5	*HpaI*	7	0	1
6	*NcoI*	1	0	0
7	*NotI*	0	0	0
8	*PstI*	1	1	0
9	*SalI*	1	0	0
10	*SpHI*	1	0	1
11	*SspI*	1	1	1
12	*XbaI*	1	1	0
13	*XhoI*	1	0	0

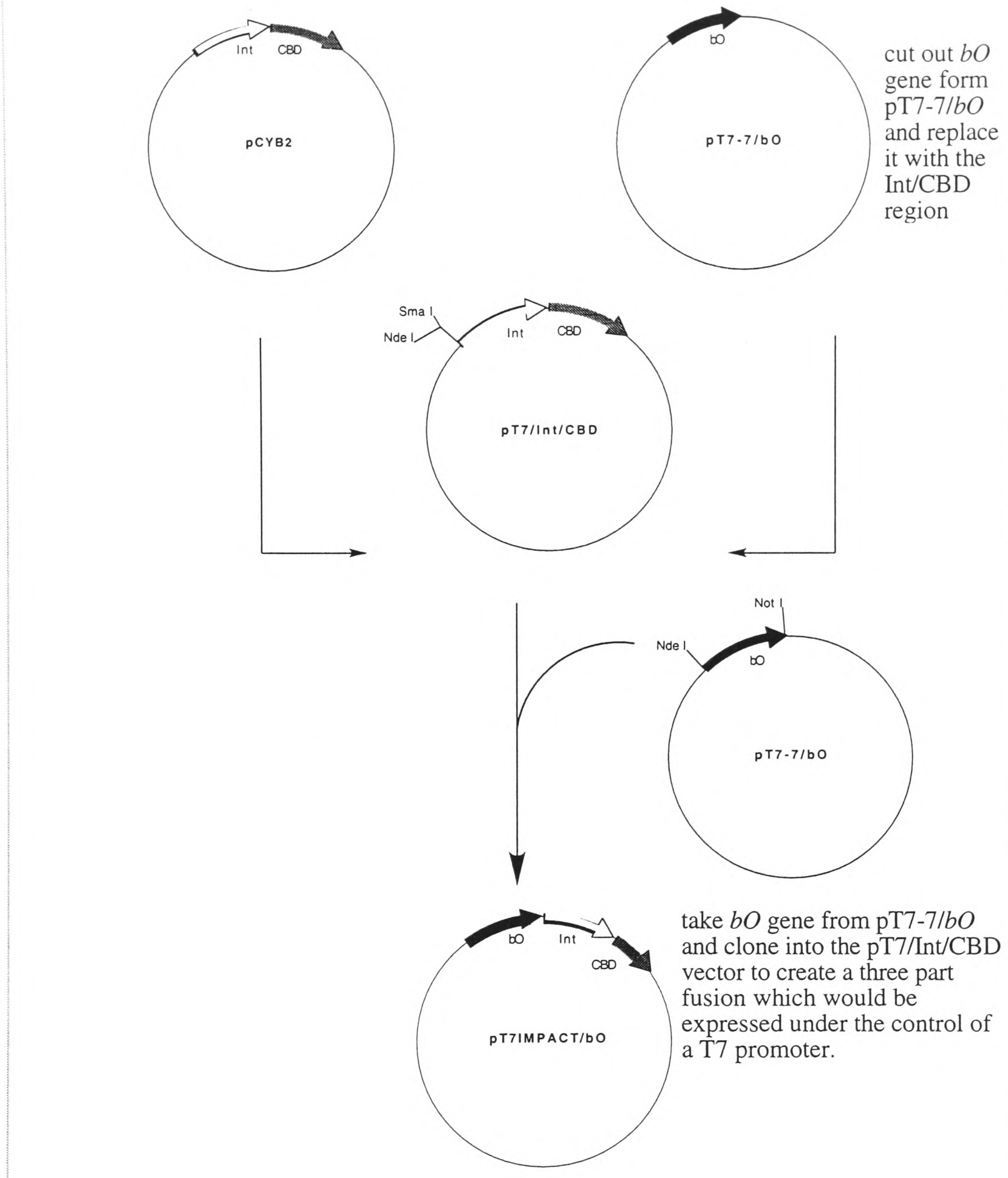


Figure 3.24 Construction of pT7IMPACT/bO

3.2.2.6.2 Expression studies on PT7IMPACT/*bO*

pT7IMPACT/*bO* was transformed into the selected *E. coli* bacterial strains and grown at 37°C to OD_{600} value 1.0. IPTG was added to a final concentration of 1 m M and the cells grown for a further hour. They were then harvested and the expression levels were detected by immunoblotting. Expression was observed in 4 of the cell lines, B834, BL21, BL21pLysS and HMS174 (figures 3.26, 3.27, 3.28, and 3.29 respectively). These showed increasingly higher levels of *bO* production and were also less leaky with the HMS172 strain showing no leaky expression at all.

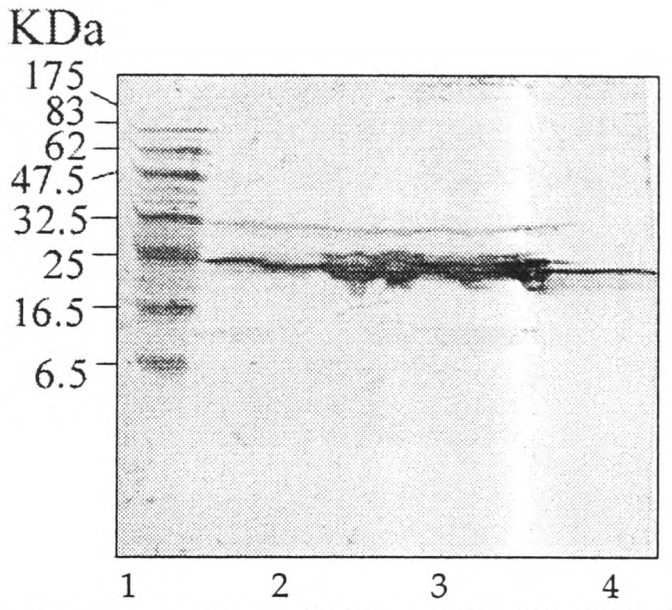


Figure 3.26 Expression of *bO* in pT7IMPACT in B834
Lane 1: Protein markers as indicated. Lane 2: uniduced cells. Lane 3: induced cells. Lane 4: purple membrane standard (0.3 µg)

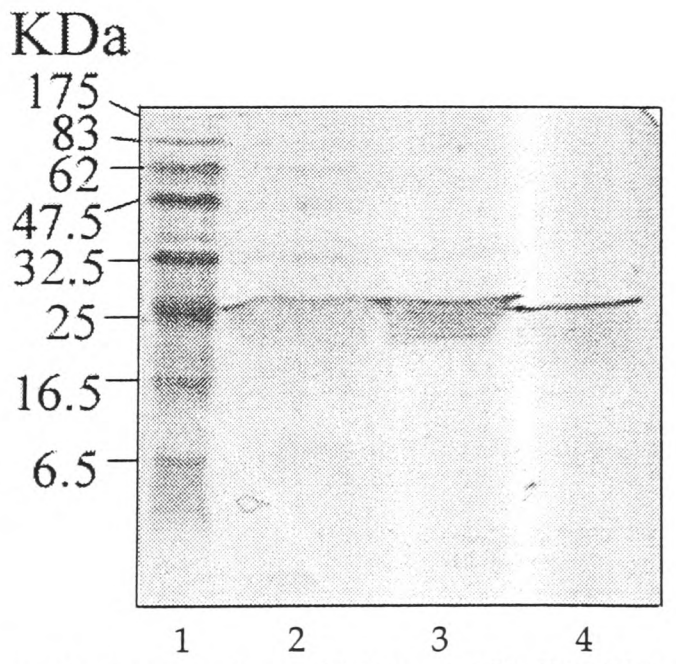


Figure 3.27 Expression of *bO* in pT7IMPACT in BL21
Lane 1: Protein markers as indicated. Lane 2: uniduced cells. Lane 3: induced cells. Lane 4: purple membrane standard (0.3 µg)

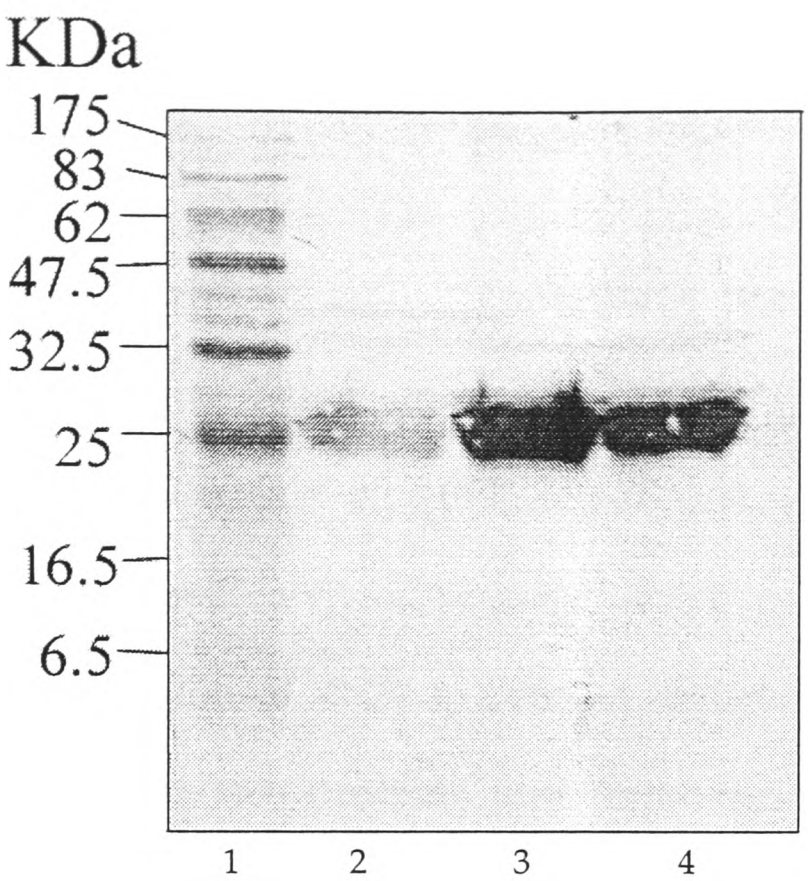


Figure 3.28 Expression of *bO* in pT7IMPACT in BL21pLysS
Lane 1: Protein markers as indicated. Lane 2: uniduced cells. Lane 3: induced cells. Lane 4: purple membrane standard (0.3 μg)

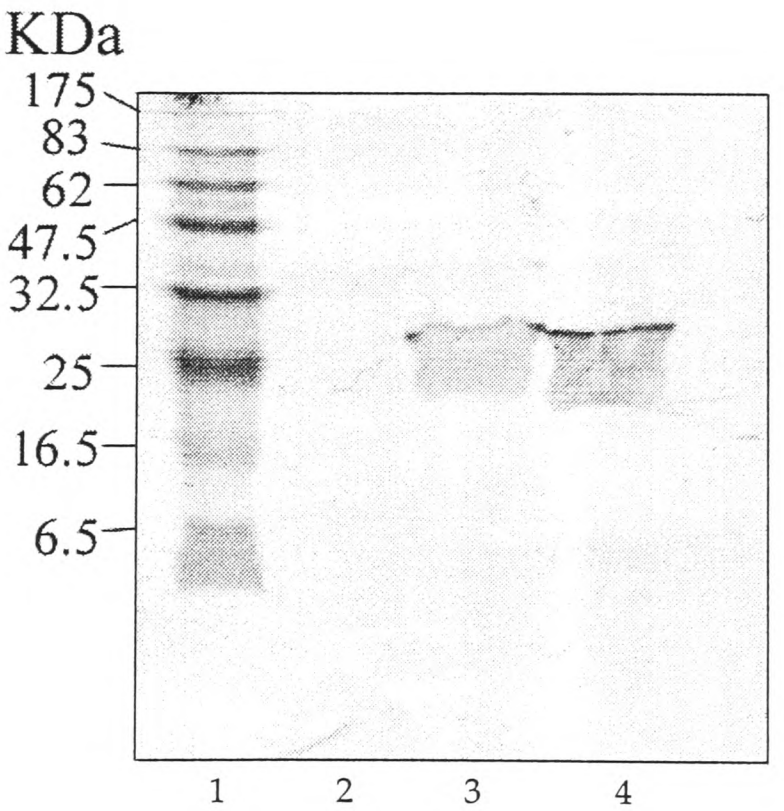


Figure 3.29 Expression of *bO* in pT7IMPACT in HMS174
Lane 1: Protein markers as indicated. Lane 2: uniduced cells. Lane 3: induced cells. Lane 4: purple membrane standard (0.3 μg)

3.2.2.6.3 Discussion on the Expression of pCYB2/*bO* and pT7IMPACT/*bO* using the IMPACT system

3.2.2.6.3.1 Protein splicing

Several archaeal, eubacterial and eukaryotic genes have been recently discovered to contain in-frame insertions. These intervening polypeptide domains called inteins [int(ernal) prot(ein)] (Perler *et al.*, 1994) are excised not at the RNA level, but at the protein level *via* a self-catalysed process known as protein splicing. The intein is excised after translation of RNA to protein (figure 3.30). The single gene sequence actually encodes for two proteins and once the intein is excised, the flanking amino- and carboxyl-terminal domains known as exteins, ligate *via* formation of a peptide bond to yield an active protein. The synthesis of this native bond distinguishes this process from simple autoproteolysis (Cooper *et al.*, 1993).

Figure 3.30 Protein Splicing (left pathway) and RNA splicing (right pathway).
Adapted from the NEB Transcript. A scientific newsletter from New England Biolabs. January 1997, Vol 18, No. 2, pg. 1-5.

In an excellent review by Yang Shao and Stephen B. H. Kent (Shao and Kent, 1997), the occurrence and mechanism of this protein splicing phenomenon is described in great detail. The first demonstrated example of protein splicing involves the protein product of the *Saccharomyces cerevisiae* TFP1 gene, which encodes the 69 kDa catalytic subunit of vacuolar ATPase (Kane *et al.*, 1990; Hirata *et al.*, 1990). The 119 kDa precursor is actually interrupted by a 50 kDa codon stretch of intein sequence which is precisely excised to produce the 69 kDa vacuolar ATPase subunit (Kane *et al.*, 1990). By using Northern blot analysis, the possibility of RNA splicing was ruled out (Kane *et al.*, 1990; Hirata *et al.*, 1990). Since both proteins were produced at equal rates, it was evident that a precursor protein was synthesised which was post-translationally processed into two proteins. Since this initial reported example of protein splicing, many further examples have been discovered in all three kingdoms of life. From the vacuolar ATPase of the yeast *Candida tropicalis* (Gu *et al.*, 1993) to the RecA proteins of the mycobacteria *Mycobacterium tuberculosis* (Davis *et al.*, 1991) and *Mycobacterium leprae* (Davis *et al.*, 1994) to DNA polymerases in thermophilic archaebacteria *Thermococcus litoralis* (Hodges *et al.*, 1992) and *Pyrococcus* species GB-D (Perler *et al.*, 1992). It is interesting to note that these archaeal DNA Polymerase genes have two inteins present in each gene (Cooper and Stevens, 1995).

3.2.2.6.3.2 Chemical Mechanism of Protein splicing

Protein splicing still occurs if the intein is transferred into heterologous proteins suggesting that all the information required for splicing to take place is contained in the intein and the first C-extein residue (Davis *et al.*, 1994). The sequence of the exteins does not make much of a difference in the splicing process. In contrast, the

inteins share a sequence homology which is highly conserved at both the splice junctions (Pietrokovski, 1994). Towards the carboxyl terminus of each splice junction, there is always a hydroxyl or thiol containing residue such as serine, threonine or cysteine. The carboxyl terminus of the intein contains an asparagine and usually a histidine in the penultimate position. This His-Asn sequence is usually preceded by several hydrophobic residues (Shao and Kent, 1997). Some inteins do have Gly-Asn instead of the usual His-Asn (Shao and Kent, 1997). Several theories were proposed for the protein splicing mechanism (Wallace, 1993; Cooper *et al.*, 1993; Clarke, 1994; Xu et al., 1994) but since it was difficult to purify the precursor proteins or splicing intermediates due to *in vivo* splicing, in vitro studies could not be performed to back-up any of the proposed mechanisms.

The real breakthrough came when researchers at New England Biolabs were able to insert the *Psp* pol intein-1 from the DNA polymerase of the extremely thermophilic *Pyrococcus sp.* GBD between the maltose binding protein and *Dirofilaria immitis* paramyosin and observe *in vitro* splicing at elevated temperatures (26-65°C) (Xu *et al.*, 1993). Thus expression of this fusion (hereby referred to as MIP) in *E. coli* at low temperatures (12-15°C) provided a source of unspliced precursor which proved that protein splicing is a self-catalysed, post-translational process. The study of protein splicing in mesophiles was more challenging because temperature regulation could not be used to study the mechanism even if the intein was inserted into a foreign protein. This problem was overcome by developing an *in vitro* splicing system using the *Sce* VMA intein from the 69 kDa vacuolar ATPase subunit of *Saccharomyces cerevisiae* (Chong *et al.*, 1996). Studies on *Psp* pol intein-1 and Sce VMA intein suggest that the protein splicing mechanism is the same for thermophiles and mesophiles.

The mechanism can be divided into 4 steps which are diagramatically outlined in figure 3.31 and discussed in detail below (Shao and Kent, 1997):

Step 1: formation of an ester intermediate by an N-O acyl rearrangement

This step results in the formation of an ester intermediate between the carboxyl terminus of the N-extein and the serine hydroxyl sidechain of the intein. Under acidic conditions, the free amine group is protonated and formation of the ester intermediate is favoured via a hydroxyoxazolidine intermediate. Neutral or alkaline pH conditions favour formation of the amide. In nature, the ester must proceed to further chemical reactions to avoid this unfavourable equilibrium. Replacing the serine with cysteine (Shao *et al.*, 1996) results in the formation of a thioester which are more likely to undergo nucleophilic attack at the nitrogen rather than oxygen esters. This results in higher rates of protein cleavage for a cysteine thiol group when compared to a serine hydroxyl group (Shao *et al.*, 1996). Using different nucleophiles and characterising the products, Shao *et al.*, were able to confirm that an N-O or N-S acyl rearrangement takes place. Their studies were backed by experimental data from Xu and Perler (1996).

Step 2: formation of a branched intermediate by transesterification

The hydroxyl side chain of serine at the downstream splice junction attacks the ester intermediate formed in step 1. The reaction is analogous to that of a serine protease. During splicing studies, Xu *et al.*(1993) identified a slow migrating species whose kinetic data resembled that of an intermediate. The species was isolated using SDS-PAGE and the N-terminus examined by Edman degradation. Two amino termini were identified indicating the peptide bond upstream of the splice junction was broken and that the migrating species was a branched molecule.

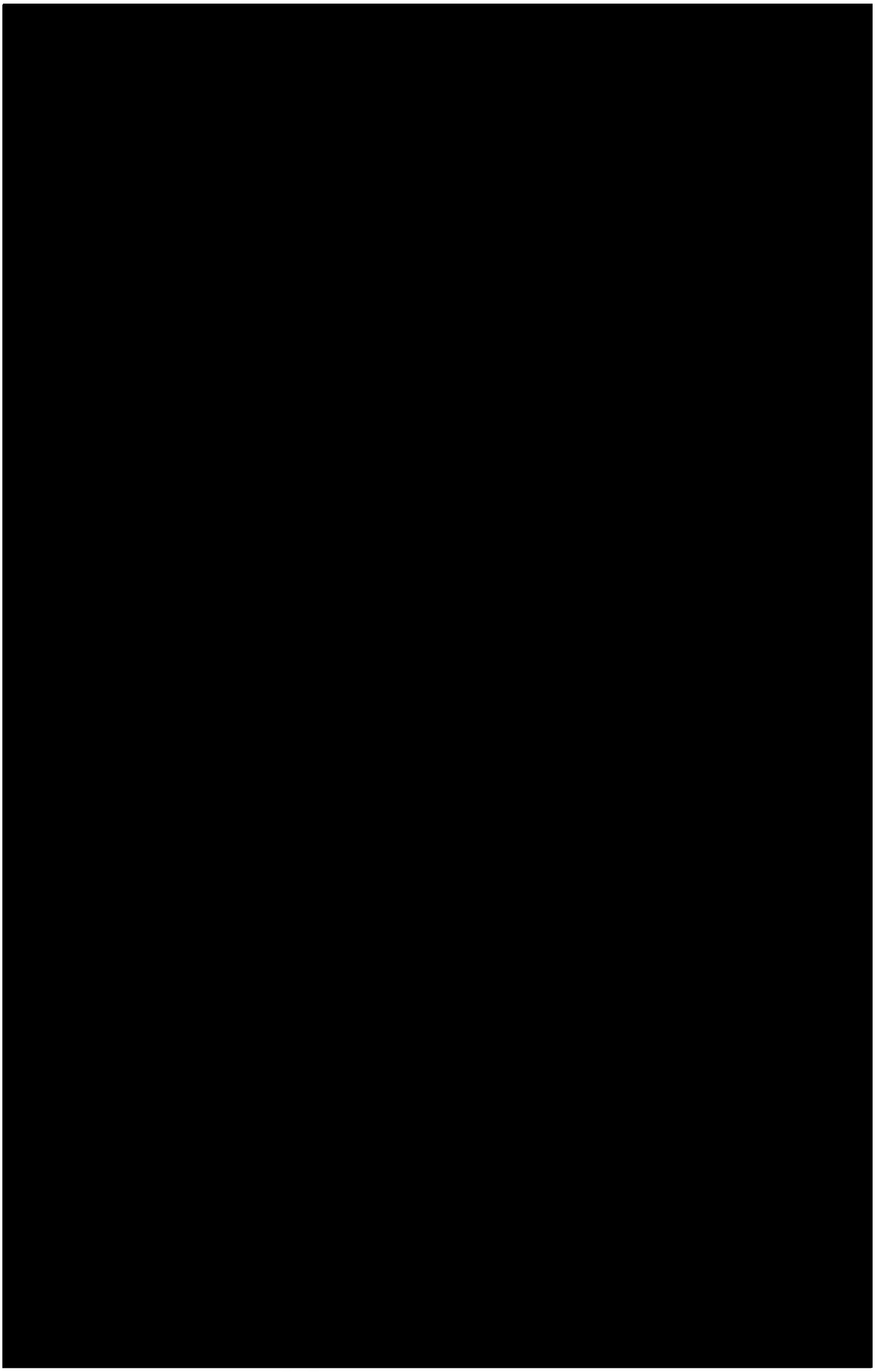


Figure 3.31 Proposed mechanism for protein splicing of the *Psp* pol intein-1
(adapted from Chong *et al.*, 1996; Shao and Kent, 1997).
Splicing of the Sce VMA intein from the 69 kDa vacuolar ATPase subunit of *Saccharomyces cerevisiae* proceeds with four analogous chemical steps except that the serine residues shown in the diagram are replaced by cysteine so that steps 1 to 4 are N-S and S-N acyl shifts respectively.

Step 3: excision of the intein

This is an irreversible step where the asparagine cyclises at the carboxyl terminus of the excised intein to form an aminosuccinimide residue.

Step 4: spontaneous O-N acyl rearrangement

The unstable transient products undergo N-O acyl rearrangements to form stable amides. Apart from in protein splicing, there are no other examples of this rearrangement for proteins at neutral or high pH apart from chemical synthesis.

Researchers at New England Biolabs cleverly developed a novel one-step purification system using the mechanism of protein splicing and called it IMPACT I™ (Intein Mediated Purification with an Affinity Chitin-binding Tag). The system uses the intein from the *Saccharomyces cerevisiae* VMA1 gene which has been modified to undergo self-cleavage at the amino terminus in the presence of thiols (such as DTT, β-mercaptoethanol or cysteine) at low temperatures.

The target gene was inserted into the multiple cloning site of the pCYB vector to create a 3-part fusion with the intein and chitin binding domain (CBD) from *Bacillus circulans*. The CBD allows affinity purification of the fusion protein. Expression in *E. coli* is under the control of an IPTG inducible Ptac promoter. When a crude cell extract is passed over a chitin column, the 3-part fusion protein binds to the column while all other contaminants are washed through. A reducing agent such as DTT or β-mercaptoethanol is added to induce an on column cleavage at 4°C overnight. The target protein is released into the eluent while the intein-CBD fusion remains bound to the column.

Reasons for choosing the IMPACT I™ system

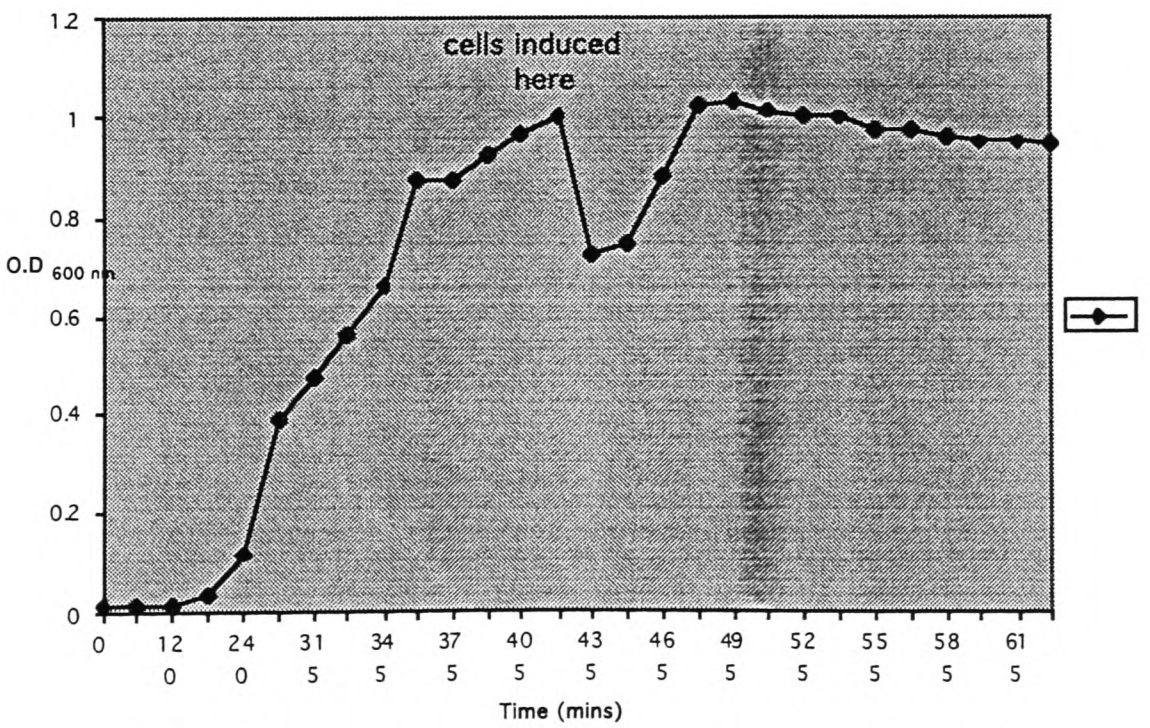
This system was chosen for its many advantages which are listed below:

- rapid and simple purification of native proteins without an affinity tag which may alter its properties
- simultaneous purification and cleavage of fusion proteins

  tight transcriptional control and high levels of expression
- release from fusion partner at C-terminus without the use of expensive proteases
- entire purification at 4°C eliminating incubation at higher temperature which may destabilise some proteins
- label C-terminus of the target protein
- chitin is an abundant organic substance that is stable, inexpensive and reusable.

Since *bO* is toxic to *E. coli*, it was thought that the addition of a bulky fusion protein such as the intein/CBD domain would prevent the protein from inserting itself into the *E.coli* membrane and thus avoid exerting a toxic effect on the host. Instead it was hoped that the fusion protein would be insoluble and form inclusion bodies from which we would then be able to extract the protein. The other attractive feature was the one-step, on column cleavage which would eliminate the normal solvent extraction procedure used to purify membrane proteins as well as the further chromatographic steps involved to obtain a pure sample.

The gene was cloned into the pCYB2 vector and expressed as described in section 3.2.7.1 and 3.2.7.2 respectively. The growth rate of the cells was noted to dramatically decrease after induction. In the process of trying of produce *bO*, it was thought the protein may have been to toxic to the host cell which was slowing down the rate of cell growth and eventually killing the cells. We were unable to detect any expression even using the bR antibody. Expression of the toxic protein under control of a more tightly controlled promoter such as T7 may perhaps eliminate the possibility of any leaky expression. The pT7IMPACT vector was designed for this purpose and the gene was cloned and expressed in this vector as described in sections 3.2.8.1 and 3.2.8.2 respectively. Even under the control of the T7 promoter, some leaky expression was detected. The rate of cell growth after induction dropped quite dramatically but then picked up again indicating that the cells were coping with the production of the protein. The growth curve for T7IMPACT/*bO* in HMS174 is shown in graph 3.1.



Graph 3.1 Growth curve of T7IMPACT/*bO* in HMS174

Immunoblot analysis identified the expressed protein to run identical with the native bR sample. However, this was not the expected result as the *bO* was expressed as a fusion to the intein and CBD binding domain proteins. Thus, we would only expect to see the expressed *bO* at the native position after the on-column cleavage step using DTT or another thiol containing compound. Even with the elimination of DTT in the SDS-PAGE sample buffer, we still observed this *in vivo* cleavage. Unfortunately, we found the intein antibody not to be extremely selective which would have allowed us to determine if we still expressed a large amount of protein as a fusion. We therefore relied on the bR antibody which was only able to detect the cleaved *bO* and not any that may have been expressed as a fusion. Thus perhaps the yield of the *bO* was split between the inclusion bodies that we were able to detect by immunoblotting and a fusion protein that we were not able to detect by immunoblotting. In order to purify this protein, we would have to use both the chitin column and the organic extraction method to ensure we were not losing any protein. So, although the detected expression levels were higher than we had previously detected in other systems, there were a few problems that would have to be dealt with in the purification. For this reason, the IMPACT system was not further pursued since the next expression system that we investigated, the pLEX system proved to be the more convenient.

3.2.2.7 Studies using the pLEX expression vector

3.2.2.7.1 Constructing the pLEX/*bO* expression vector

The 2.9 kb expression vector pLEX (from Invitrogen), is claimed to be especially useful in the expression of proteins that are toxic to *E. coli*. It allows expression of the *bO* gene using a tightly regulated tryptophan-inducible expression system under the control of the bacteriophage lambda P_L promoter. The *c*I repressor protein which is coded for on the bacterial chromosome under the *trp* promoter controls expression at the P_L promoter. The *bO* gene was excised from pBluescript as an *NdeI/EcoRI* fragment and inserted into the polylinker region of pLEX (figure 3.33) using the strategy illustrated in figure 3.32 which was pre-excised with *NdeI* and *EcoRI*. The resultant construct was introduced into the JM109 cloning host. To check insertion of the *bO* gene into the pLEX expression vector, the appropriate restriction digests were performed and the resultant fragments analysed by agarose gel electrophoresis as shown in figure 3.34.

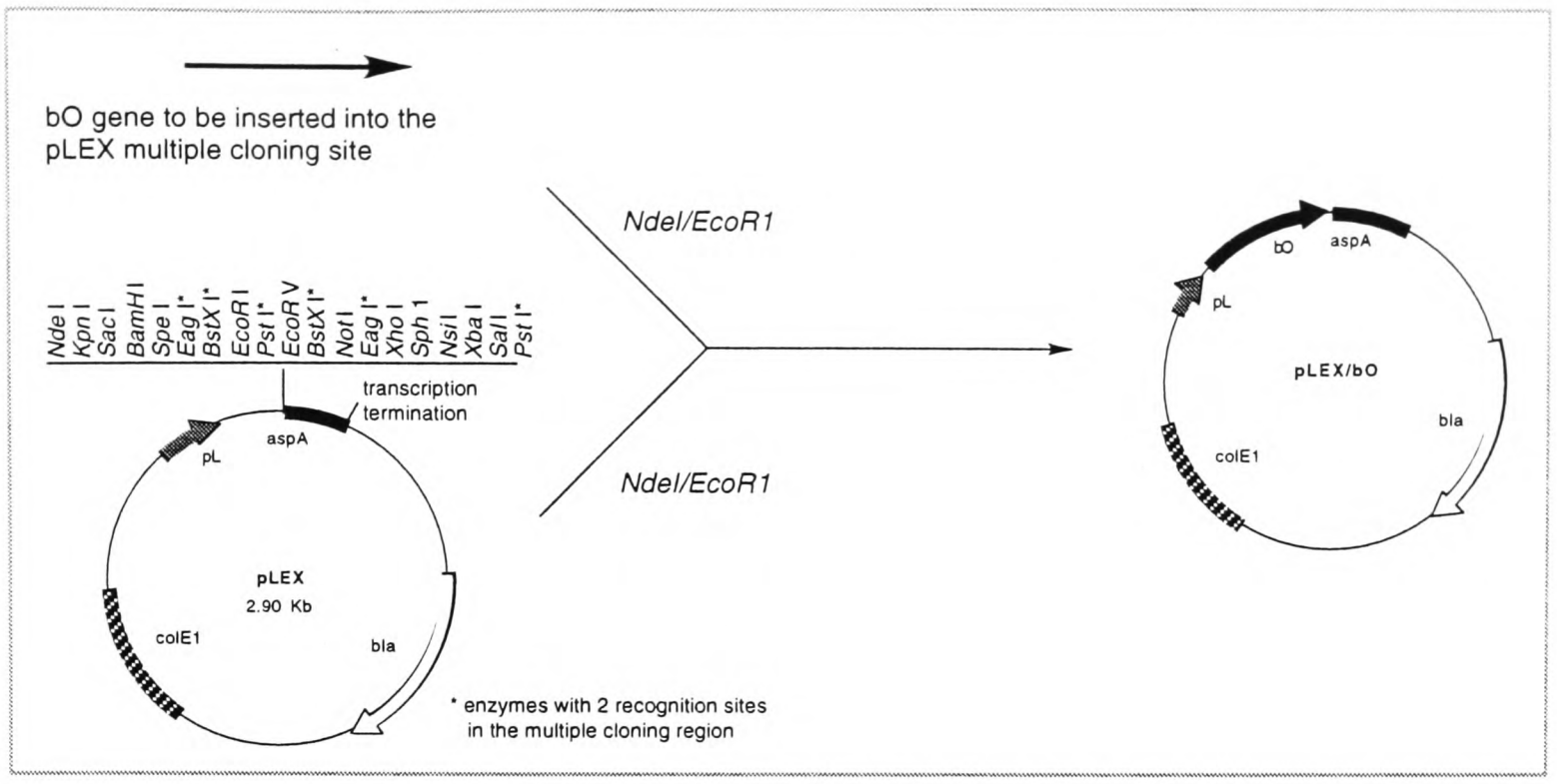


Figure 3.32 Construction of pLEX/*bO*

Figure 3.33 pLEX Cloning/expression region (modified from Invitrogen Instruction Manual)

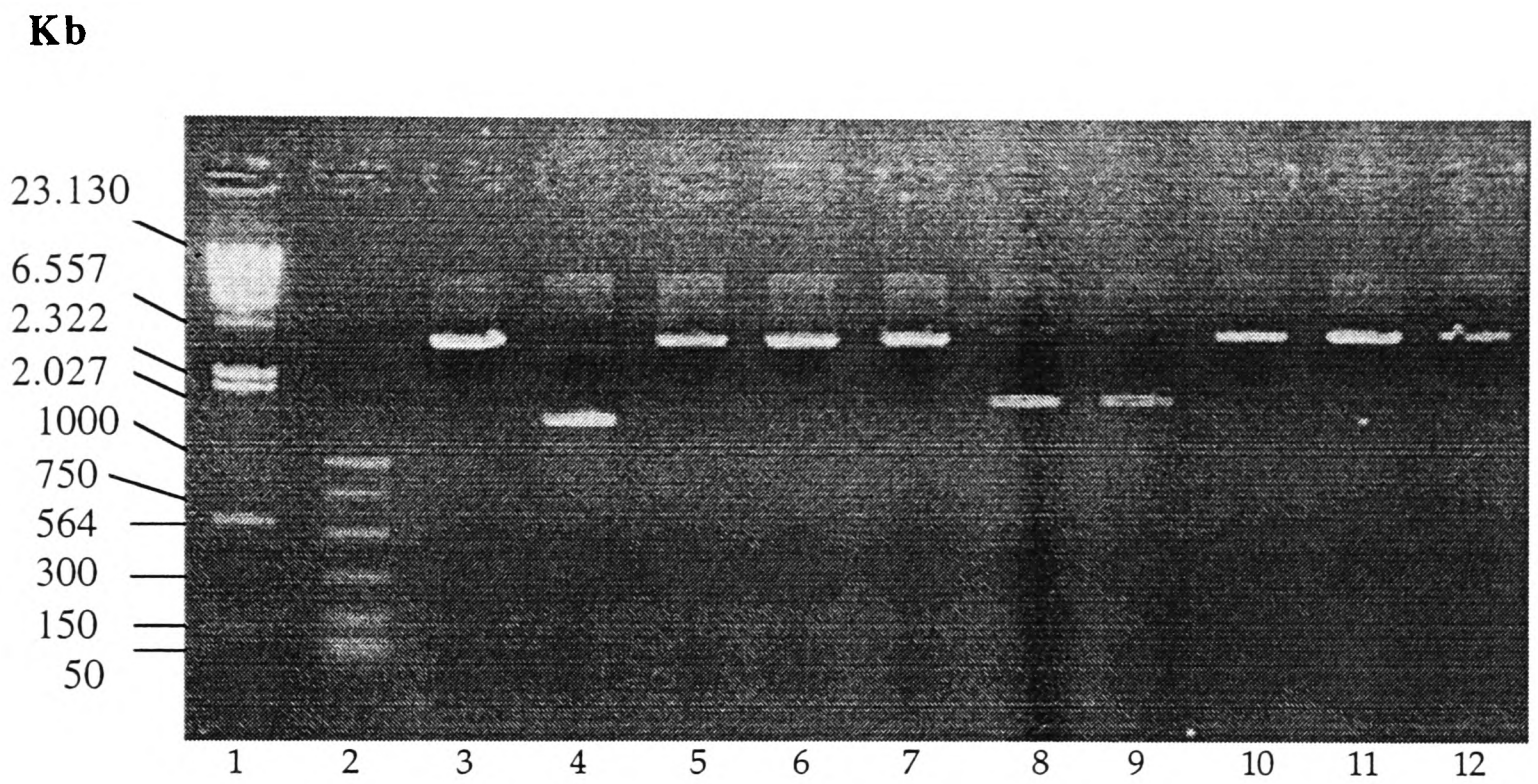


Figure 3.34 Agarose gel of pLEX/*bO* restriction digest cuts

Lane 1: λ DNA/*Hin*d III fragments from Gibco BRL™. Lane 2: PCR markers from Promega. Lane 3 to 12-restriction digest of different pLEX/*bO* constructs with *Nco* I which linearises vector. Samples in lanes 4, 8 and 9 did not contain the insert and were thus discarded.

3.2.2.7.2 Expression of *bO* in pLEX

The pLEX/*bO* expression vector was transformed into *E. coli* cell line GI724 to ensure proper regulation of expression. This bacterial strain has a *c*I repressor gene on its chromosome which is under the control of the trp promoter. Media used in the expression of pLEX/*bO* contained low levels of tryptophan to avoid leaky expression and growth was carried out at 30°C to eliminate the possibility of transcription that may be initiated through the P_L promoter. When an OD_{600} value of 1.0 was reached, tryptophan was added to a final concentration of 100 µg/ml and the cells grown for a further hour. The culture was harvested in the same way as for the other constructs and the expression levels were detected by immunoblotting.

3.2.2.7.3 Discussion on the expression of *bO* using the pLEX system

The P_L expression system allows the production of heterologous protein in *E. coli*. It has been especially noted to be useful in the expression of toxic proteins such as *bO*. This is because it uses a unique method of transcription so that no recombinant protein is produced until the inducer (typtophan) is added. The tightly regulated system allows significant amounts of *bO* to be produced before the host cell dies.

The *bO* gene was cloned into the multiple cloning site of the pLEX vector (3.2.2.7.1). Expression is driven by the tightly regulated P_L promoter from bacteriophage lambda (Buell and Panayotos. 1986; La Vallie *et al.*, 1992) which in turn is controlled by the *c*I repressor protein. The *c*I repressor gene was engineered

into the bacterial chromsome under the control of the trp promoter. The addition of tryptophan induces expression (figure 3.35).

1 In the absence of trytophan, expression of the *cI* repressor is driven by the trp promoter.
2 The *cI* repressor protein binds to the operator region up-stream of the P_L promoter and prevents transcription of the gene of interest.
3 Tryptophan in added to the medium and a tryptophan-trp repressor complex is formed. This complex binds tightly to the trp operator blocking expression of the *c*I repressor.
4 The cI repressor falls off the P_L operator, allowing transcription of the gene of interest

Figure 3.35 Regulation of expression using the pL expression system
(figure adapted from the P_L expression systems instruction manual by Invitrogen)

Reasons for choosing the pLEX expression system

- The strong P_L promoter is tightly regulated and offers high level expresison of the recombinant protein.
- The lambda *c*II ribosome binding site and ATG initiation allow efficient translation of recombinant protein.
- The *E. coli asp*A transription terminator results in the efficient transcription termination of mRNA.

- The ampicillin resistance allows selection and maintainance in *E. coli.*
- The ColE1 origin allows maintainance of the vector in bacteria as well as a high copy number.
- The polylinker region contains many sites for cloning the desired gene. The gene must be in-frame with the lambda *c*II initiation ATG in order to utilise the P_L promoter and ribosome binding site for high level expression. Insertion at the *Nde* I site prevents the fusion of non-native amino acids to the protein.

The host strain GI724 (2.1.6) contains the cI repressor under the control of the trp promoter. Cells containing plasmids must not be grown in LB medium (2.1.9) or at temperatures above 30°C. The large amount of tryptophan in LB medium allows transcription from the P_L promoter. If the temperature is above 30°C, except when inducing, the P_L promoter can allow low level transcription. When no plasmid is present in the cell, it can be grown in LB medium at 37°C.

Our data indicated that the pLEX expression system was the best option for the expression of the *bO* gene. To maximise the yield of protein produced, various parameters were investigated in an effort to optimise this system. We observed the yield of *bO* production by immunoblotting when growing the cells to both OD_{600} 0.6 and 1.0 (figure 3.36 and 3.37). The cells were then induced with tryptophan to a final concentration of 100 µg/ml. Samples were taken after half and hour and then at hourly intervals to determine the optimum time of induction required to produce the highest yield of *bO*. Results showed that cells grown to OD_{600} 0.6 did not produce as much protein as those cells that were induced at OD_{600} 1.0. Results also indicated that one hour was sufficient to obtain a good yield of *bO* as leaving

them for a longer length of time did not increase the level of protein produced (figures 3.36 and 3.37)

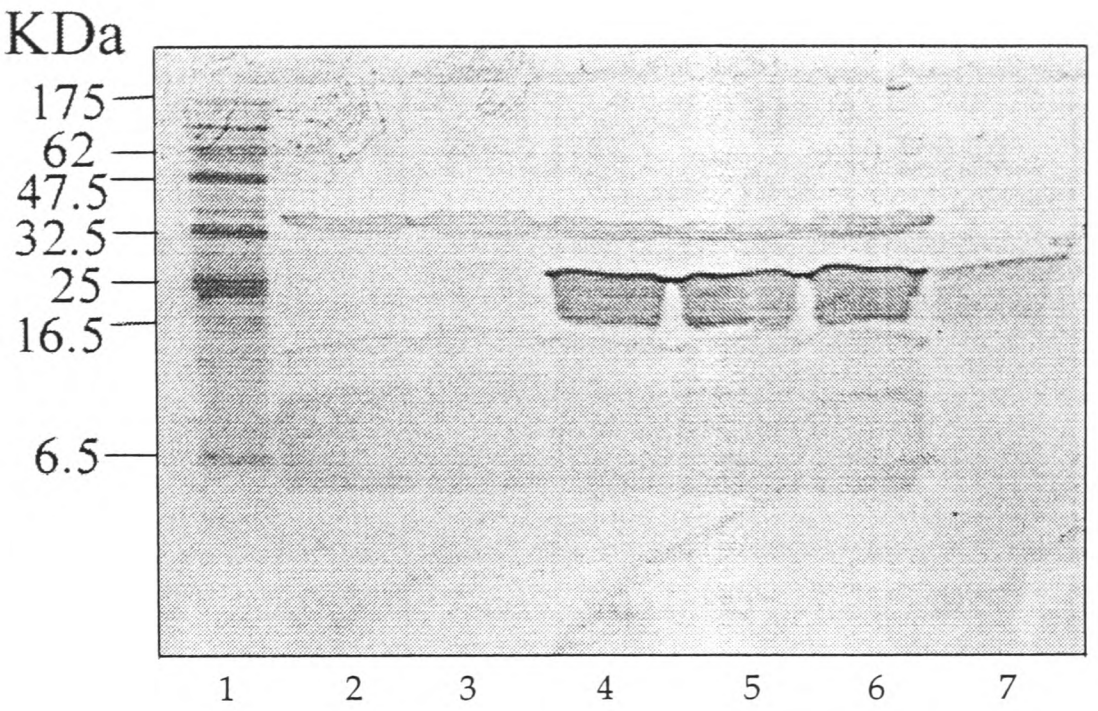


Figure 3.36 Induction of pLEX/*bO* at OD_{600} ~0.6
Lane 1:prestained protein markers, broad range from NEB. Lane 2: uninduced cells. Lane 3: cells after 30 minutes induction. Lane 4: cells after 1 hour induction. Lane 5: cells after 2 hours induction. Lane 6: cells after 3 hours induction. Lane 7: purple membrane standard (0.3 µg)

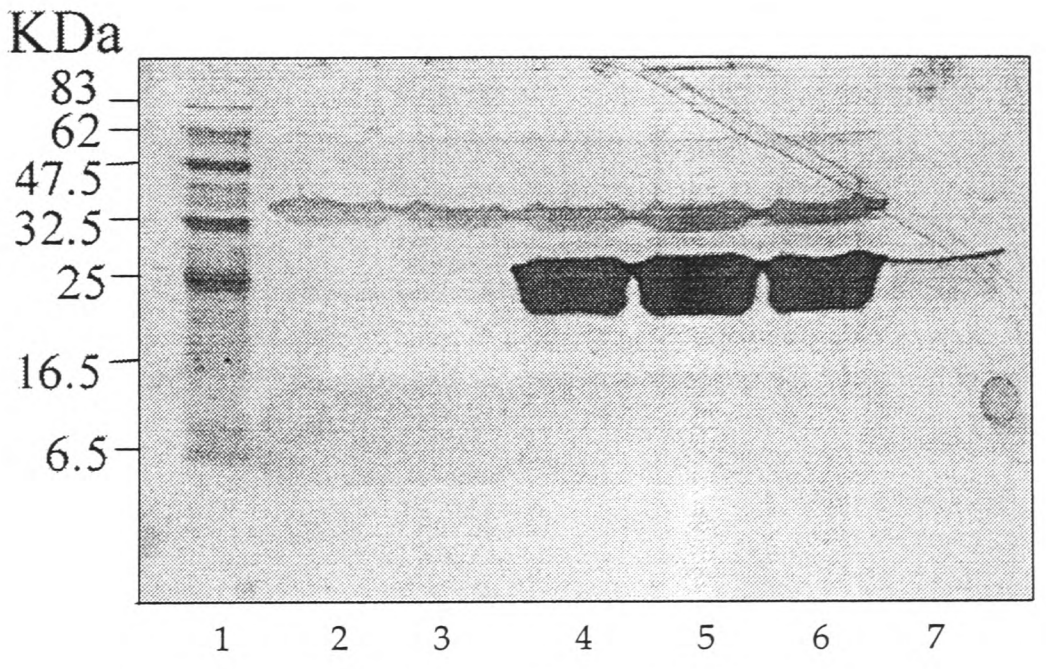


Figure 3.37 Induction of pLEX/*bO* at OD_{600} ~1.0
Lane 1:prestained protein markers, broad range from NEB. Lane 2: uninduced cells. Lane 3: cells after 30 minutes induction. Lane 4: cells after 1 hour induction. Lane 5: cells after 2 hours induction. Lane 6: cells after 3 hours induction. Lane 7: purple membrane standard (0.3 µg)

3.3 Conclusion

The results of this chapter are summarised in table 3.1 below.

	GI724	HMS174	BL21	BL21 pLysS	B834	C600
pPL1						0.3 mg/L
pRH1090						no expression
pT7-7		no expression	no expression	no expression	no expression	no expression
pET16b		0.1 mg/L v. leaky	no expression	no expression	no expression	no expression
pCYB2		no expression	no expression	no expression	no expression	no expression
pT7-IMPACT		0.6 mg/L	0.1 mg/L v. leaky	0.1 mg/L v. leaky	0.1 mg/L v.leaky	no expression
pLEX	> 1 mg/L v. good					

Table 3.1 Expression of *bO* in selected expression vectors and bacterial cell lines

Expression of the *bO* gene in a variety of plasmids and host strains has shown that production of a protein in *E. coli* depends very much on the expression system concerned. Some features however can be optimised in an attempt to improve heterlogous protein production levels. Some of these are summarised in table 3.2 (Weickert *et al.*, 1996).

Table 3.2 General techniques employed in the optimisation of heterologous protein production in E. coli (table taken from Weickert, 1996).

The phenotypes and genotypes of the selected bacterial strains are described in section 2.1.6. [*Molecular Cloning - A Laboratory Manual* (Sambrook J., Fritsch E. F., Maniatis T. (1989) Cold Spring Harbor, Cold Spring Laboratory Press]. Our results showed that the expression levels of *bO* vary in a particular host.

C600 allowed the expression of 0.3 mg/L of *bO* using the pPL1 expression system (figure 3.7) but did not allow the expression of *bO* in any of the other plasmids. It was originally thought that some *bO* was expressed using the pRH1090 plasmid in C600 but N-terminal sequence analysis (figure 3.12) as well as DNA sequencing results (appendix 2) showed this not to be the case.

No host was able to produce successful results for the expression of *bO* using pT7-7 vector (Data not shown). It was thought that the *bO* gene exerted a toxic effect on the host.

The same results were obtained for the expression of *bO* in pCYB2.

The toxicity of the *bO* gene was masked when expressed as a fusion to the Int/CBD region as in the IMPACT system. Although the pT7-7 vector did not allow expression under T7 promoter, the fusion helped reduce toxicity. The fusion itself however was not enough to produce *bO* as no expression was detected in the pCYB2 vector which would have expressed the same fusion protein but under the control of a pL promoter. Thus a combination of both a strong promoter and masking the toxic gene as a fusion protein allowed the production of *bO* (figure 3.29). Unfortunately, problems were encountered in the cleaving the *bO* protein from the fusion due to the conditions required by the chitin column.

Expression using the pET16b vector with HMS174 produced a small amount of protein but the system was noted to be very leaky (figure 3.19).

Expression of *bO* using the pLEX system allowed production of *bO* under a strong P_L promoter from bacteriophage lamdba. The system uses a unique method of transcription so that no protein is produced until the inducer (tryptophan) is added (figure 3.37). Growing to an OD_{600} 1.0 allows significant amount of this toxic protein to be expressed before the host cell dies.

Chapter 4:

Purification of Expressed Bacterioopsin

Chapter 4:Purification of expressed bacterioopsin

4.1 Introduction to the purification of *bO*

Since the expression levels of the *bO* gene in *E. coli* are low, purification is a considerable problem. With the isolation and characterisation of a monoclonal antibody to the COOH-terminal region of bR, BR114 (Kimura *et al.*, 1982), the expression of bR could be detected at low levels. The BR114 antibody also allowed the purification of *bO* from crude *E. coli* membranes using a one-step immunopurification procedure (Dunn *et al.*, 1987; Schneider *et al*, 1982). By creating an immunoaffinity matrix (Schneider *et al*, 1982) using the BR114 antibody and protein-A-sepharose, bO was purified by passing the membrane extract (which had been previously solubilised in Triton and PBS) through the immunoaffinity column overnight at 4°C. They eluted and precipitated the *bO* before washing and redissolving it in 1% SDS. Although 70-80% *bO* was purified using this procedure, a lot of material was lost in the process and the procedure was limited to a small scale.

An alternative method to purify large amounts of *bO* from *E. coli* membranes was developed by Braiman *et al.*, (1987). They made the observation that denatured *bO* is quite soluble in mixtures of chloroform, methanol and water at high or low pH. Subsequently, they developed an organic extraction procedure as well as chromatographic techniques which are similar to those used to purify proteolipids from various sources (Folch and Lees, 1951; Helynck *et al.*, 1983; Fillingame, 1976). Folch and Lees (1951) first described the solubility of lipid and protein complexes in a 2:1 chloroform: methanol mixture. The amount of *bO* expressed in *E. coli* is too low to be observed by SDS-PAGE. Organic extraction however allows *bO* to

be purified to 40% of membrane proteins compared to only 1% before the extraction. Similar results are obtained whether the extraction is carried out on whole cells or membrane. This was measured by staining the SDS-PAGE gel with coomassie blue and then recording the intensities of the bands using scanning gel densitometry which showed results identical to within 1%. Thus the procedure recovers nearly all of the *bO* in the cells. Further purification choices are: preparative SDS-PAGE, ion exchange chromatography on DEAE trisacryl (see section 4.4.2) or hydroxylapatite adsorption chromatography in organic solvents. In 1991, Miercke *et al.* developed a purification scheme which allowed bR and mutants to be purified to ~100% using a single detergent system. Despite the purity and yield of this procedure, it is extremely laborious. Further advancements in the purification of membrane proteins have been limited due to their extreme hydrophobic nature in comparison to the purification of soluble proteins.

The expression systems that have been investigated in this thesis require different purification procedures. The purification of *bO* will be discussed from pET16b which contains a polyhistidine tag, the T7IMPACT system which uses a chitin column and the pLEX system which utilises the organic extraction procedure to purify *bO* from *E. coli* membranes (Braiman *et al.*, 1987).

4.2 Results and Discussion

4.2.1 pET16b/*bO* purification using a His column

The main purpose of choosing the pET16b expression vector to express the *bO* gene was that it fuses a hexahistidine tag onto the N-terminus of the protein. This His-Tag binds to divalent cations such as nickel which are immobilised on a metal

chelation resin. By washing away all the unbound proteins, the *bO* is eluted using various concentrations of imidazole. The system is ideal for the purification of membrane proteins because it allows harsh denaturing conditions to be employed which are ideal to solubilise extremely hydrophobic proteins as well as inclusion bodies. Although yields for the expression of *bO* using the pET16b vector were low, we attempted to purify the protein using this procedure. A lot of systems fail at this stage because although expression levels seem high up until this point, the purification procedure may cause a substantial loss of sample and so the system may not be as valuable as first thought.

For the *bO* gene in the pET16b vector, the best expression levels were observed using the HMS174 cell strain. This section discusses the purification of this protein using the His column. The procedure outlined in the pET system manual by Novagen was followed. The purification was carried out under denaturing conditions by using 6M guanidine HCl in the buffers (see section 2.3.3.4).

4.2.1.1 Preparation of His column resin

The polypropylene column was packed with His-bind resin under gravity (one column volume). It was then charged and equilibrated as described in section 2.3.3.4.1.

4.2.1.2 Preparation of the cell extract

Before the cell extract was loaded onto a column, it was analysed by SDS-PAGE which allowed the detection of a very small amount of *bO* (figure 4.1, lane 3). Note that the bands for *bO* are observed at a slightly higher apparent molecular weight compared to the native protein due to the presence of the His tag.

The cell extract was then loaded and purified using the His column as described and the protein eluted in 1 ml fractions using various imidazole concentrations (100 mM, 200 mM, 300 mM and 400 mM). From the 1 ml fractions, a 100 µl was taken and precipitated by adding 10 volumes of acetone: methanol 1:1 and leaving them at -20°C. The samples were centrifuged and the supernatant carefully removed and discarded. The pellets were resuspended in SDS-PAGE loading buffer and visualised by SDS-PAGE. Unfortunately, the samples were too weak and thus all of the 1 ml fraction was precipitated. Due to the large volume of solvent required for this procedure, a lot of sample was lost in the process and since the amount of starting material was limited due to low expression levels, this posed a serious problem. By repeating the whole purification procedure again for a fresh sample, *bO* was eluted using 1 M imidazole to avoid spreading the elution over several fractions. The samples were precipitated in the same way and analysed by SDS-PAGE (figure 4.2). Increasing concentrations of the sample were run to observe the *bO* yield. A very small amount of *bO* was detected which was considered not worth further purifying.

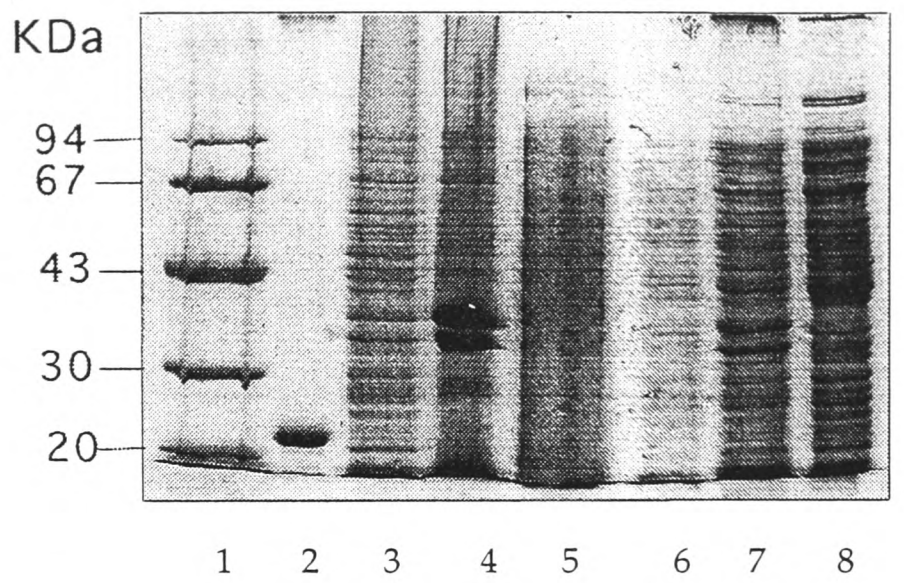


Figure 4.1 Analysis of the stages of sample preparation before loading onto the His column

Lane 1: Molecular weight markers from Pharmacia Biotech. Lane 2: purple membrane standard. Lane 3: supernatant containing *bO* ready to load onto the His column. The *bO* band is not very visible at this stage because it is very dilute amongst the other *E. coli* proteins However, after washing some of the contaminating proteins and eluting using the His column, the *bO* band is easier to detect as can be seen in figure 4.2. Lane 4: pellet after resuspending the sample in guanidine-HCl which should contain any insoluble material and no *bO*. Lane 5: pellet resuspended in guanidine-HCl which contains *bO*. Lane 6: supernatant after the second homogenisation and centrifugation step (not expected to contain any *bO*). Lane 7: The pellet after resuspending in binding buffer is expected to contain *bO*. Lane 8: supernatant after first homogenisation and centrifugation step.

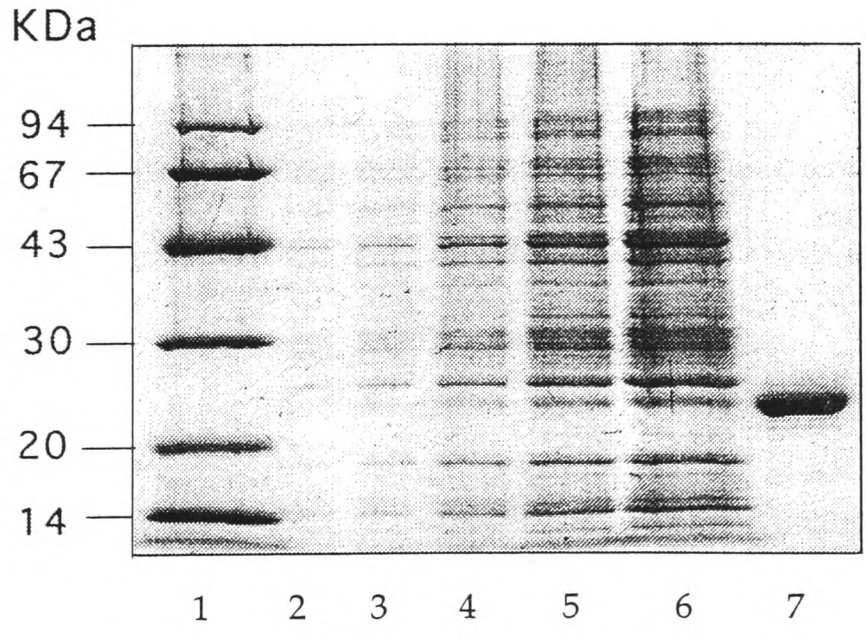


Figure 4.2 SDS-PAGE gel of pET16B/*bO* HMS174 eluted with 1 mM imidazole

1 mM *bO* fractions were eluted into 1 M imidazole containing guanidine-HCl and precipitated using 10 volumes of acetone: methanol 1:1 at -20°C. The samples were spun down and loaded onto the SDS-PAGE gel in sample buffer. Lane 1: molecular weight markers from Pharmacia Biotech. Lanes 2, 3, 4, 5 and 6: increasing concentrations of the precipitated sample eluted from the His column. Lane 6 contains detectable levels of *bO*. Lane 7: Purple membrane standard.

4.2.2 pET16b/*bO* purification using a solvent extraction procedure

As a comparison, the *bO* expressed using pET16b/*bO* in HMS174 was also purified using the organic extraction procedure (Braiman *et al.*, 1987). This is discussed in greater length for the pLEX vector (see section 4.2.4) but the results obtained using this procedure are described below. Figure 4.3 shows the various stages of purification using the organic extraction procedure. It appears by SDS-PAGE that the more *bO* is purified using this procedure than the His column purification. This may be due to the loss of protein during precipitation after the column procedure as was shown earlier above.

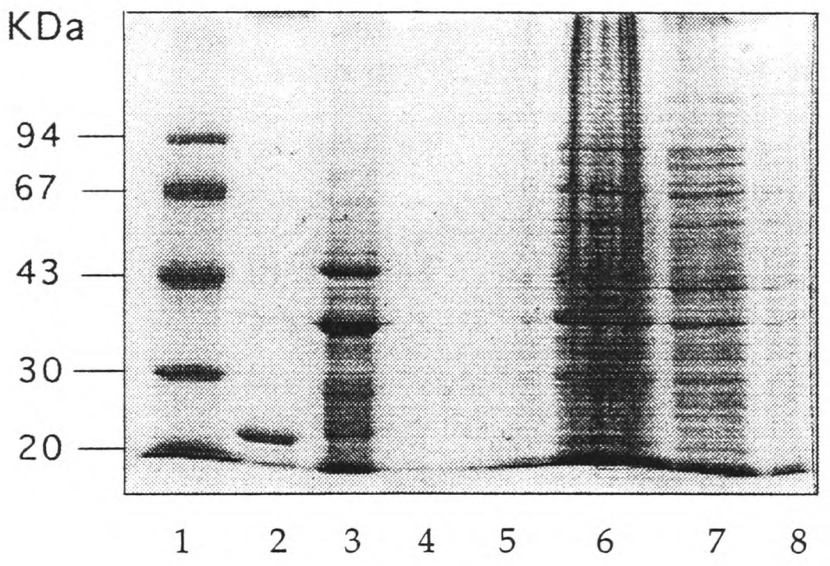


Figure 4.3 Purification of pET16b/*bO* in HMS174 by solvent extraction

Lane 1: molecular weight markers from Pharmacia Biotech. Lane 2: purple membrane standard. Lane 3: lipidacious *bO* pellet resuspended in loading buffer. Lane 4: bottom layer from phase separation. Lane 5: top layer from phase separation. Lane 6: pellet left behind after solvent extraction x 3. Lane 7: supernatant after centrifugation of sample resuspended in PBS. Lane 8: homogenised sample containing all proteins.

4.2.3 Purification of T7IMPACT/*bO*

The main advantage of using this expression system is that it should only involve a one-step purification procedure. Expression of the *bO* gene as a three part fusion with intein and CBD allows its purification *via* a chitin column. This is possible due to the presence of the 5 KDa CBD region from *Bacillus circulans* on the C-terminus of the intein which binds to the chitin beads on the column. When

crude cell extract is passed through the column, the fusion protein binds to the chitin beads and all the other unbound contaminants are washed through the column. Cleavage of *bO* is induced by adding DTT overnight at 4°C which causes the intein to undergo self-cleavage. The *bO* is then released from the three part fusion to leave only the intein and CBD region bound to the column. A wide range of pH and salt concentrations can be used for the chitin binding and cleavage reactions. NEB investigated several cleavage reagents which are shown in table 4.1. The concentration of cleavage buffer is usually between 30-50 mM although it is not crucial.

Reagent	Concentration	Cleavage efficiency
DTT (pH 5.5-9)	15-30 mM 1 mM	80-90% <5%
β-mercaptoethanol	15-30 mM 5 mM	80-90% <10%
NaCl	50-1000 mM	80-90%
EDTA	0.1-1 mM	90%
Triton X-100	0.1%	90%
Tween 20	0.1%	90%

Table 4.1 Conditions for binding and on column cleavage
*Standard cleavage reactions were tested using maltose binding protein as a target protein and incubated at 4°C overnight in 20 mM Hepes (pH 7.6), 0.5 M NaCl in the presence of 30 mM DTT unless otherwise noted. (Table taken from NEB IMPACT™1 instruction manual).

Very little *bO* was eluted after the cleavage step. Any *bO* that was expressed as a fusion protein and then cleaved on the chitin column may have been irreversibly bound to the column due to the nature of the elution buffer which did not contain any detergents. Various concentrations of salt were tried in the elution but even at very high concentrations, *bO* could not be eluted.

Thus, an alternative method was explored. A fresh sample of fusion protein was loaded onto the column, the column was washed of other contaminants, cleaved with DTT and then simply stripped using 1 x strip buffer containing 6 M guanidine-HCl. This cleaves the CBD region of the fusion from the chitin beads on the column but not the *bO* from the fusion resulting in the purification of the intact fusion protein. All the fractions were combined and precipitated using acetone: methanol 1:1 at -20°C and then centrifuged. The analysis of the purification of the fusion protein using the chitin column shows the presence of a small amount of *bO*. (figure 4.4).

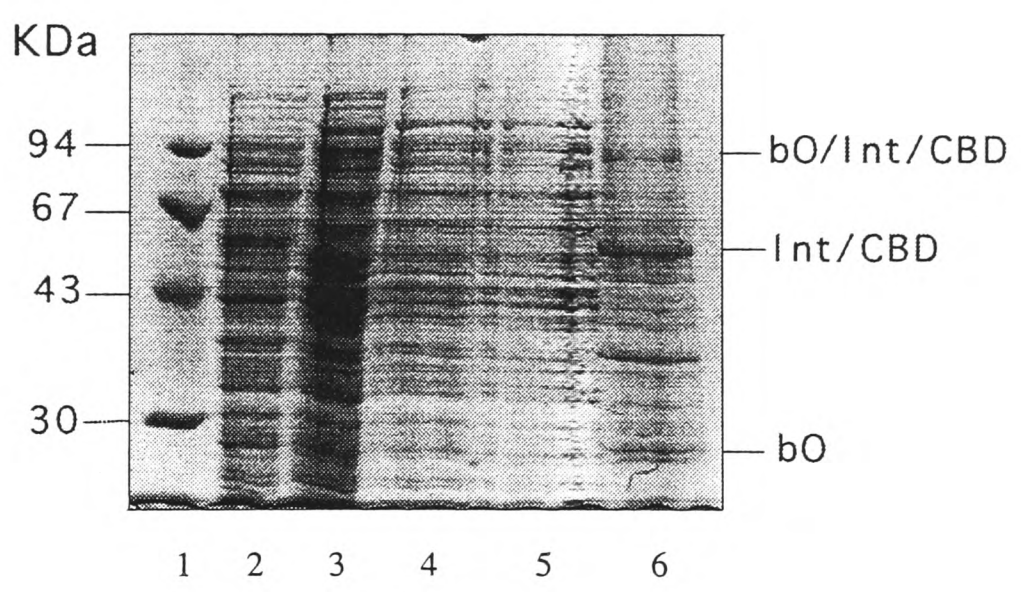


Figure 4.4 Purification of *bO* from pT7IMPACT using HMS174

Lane 1: Molecular weight markers from Pharmacia Biotech. Lane 2: supernatant from the pT7IMPACT vector only after cell lysis. Lane 3: supernatant from the pT7IMPACT/*bO* after cell lysis. Lane 4: first eluent from the chitin column containing unwanted proteins. Lane 5: second eluent fraction from the chitin column containing unwanted proteins. Lane 6: protein of interest eluted in 6 M guanidine-HCl and then precipitated using acetone: methanol, 1:1. Other bands suggest the various possible combinations of the three part fusion protein as suggested.

The results suggest that *bO* would have to be purified as a fusion protein and subsequently be cleaved and repurified. This was not considered to lead to any advantages over existing expression systems and the IMPACT system was therefore not pursued further.

4.2.4 Purification of pLEX/*bO*

The pLEX expression system was not actually chosen for its purification as it does not offer any of the advantages offered by other systems such as the His-tag in pET16b or the self-cleaving intein in pT7IMPACT. Instead, this vector was chosen for its tightly controlled expression, especially of toxic proteins, which turned out to yield the highest levels of *bO* when compared to all the other systems investigated. The procedure developed by Braiman *et al.* (1987) was used to purify *bO*. This is a two step process where the first step involves a solvent extraction procedure and the second involves a ion-exchange column.

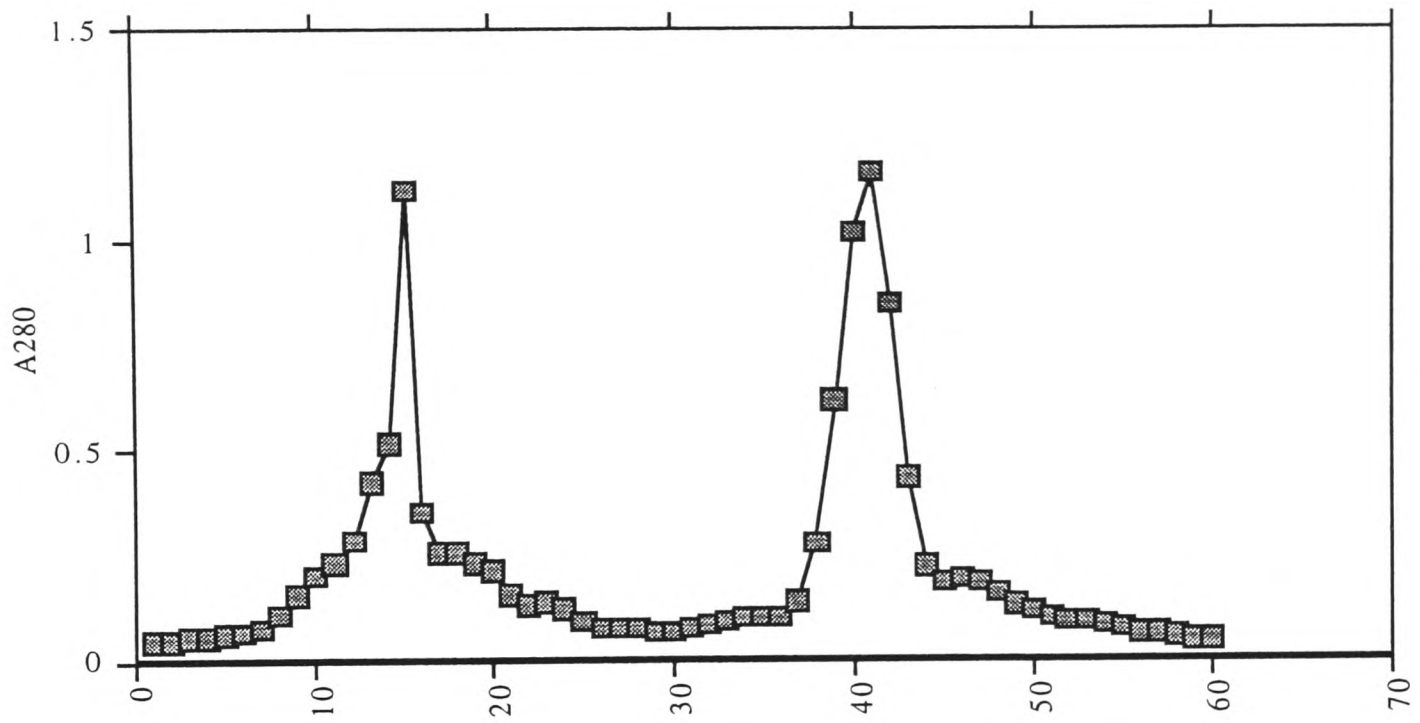
4.2.4.1 Membrane extraction to purify *bO*

This was carried out using a solvent extraction procedure as descibed in section 2.3.3.5 (Braiman *et al.*, 1987). The comparison between lanes 3 and 7 in figure 4.6 show that the solvent extraction step separates *bO* from most of the other proteins expressed in *E. coli* when using the pLEX expression vector.

4.2.4.2 Ion exchange chromatography to purify *bO*

Further purification of *bO* using ion exchange chromatography was carried out to yield pure protein (see section 2.3.3.6). The *bO* eluted with a salt concentration of 100-110 mM (graph 4.1). Fractions containing protein were investigated further by SDS-PAGE analysis which showed the presence of *bO*. (figure 4.5). Samples of *bO* free from any other contaminating bands (40, 41, 42 and 43) were pooled together and precipitated by adding 3 x the volume of dH_2O which also helps to dilute the salt. The sample was centrifuged at 5000 rpm for 15 min and the interface collected.

The concentration of protein was determined by measuring the OD_{280} where 1 OD_{280} ~0.42 mg/ml) and SDS was added to give a final ratio of 1:5 [protein:SDS]. The sample was aliquoted and dried overnight. Using this method, approximately 1 mg of pure protein was obtained from a litre of cell culture.



Graph 4.1 Purification of pLEX/*bO* by ion-exchange chromatography

The peak observed at fraction 10-16 appears to be a solvent peak and not a protein as no bands were observed by SDS-PAGE analysis. The peak observed at fraction 38-45 however, is *bO* which is confirmed by SDS-PAGE analysis as shown in figure 4.5 below.

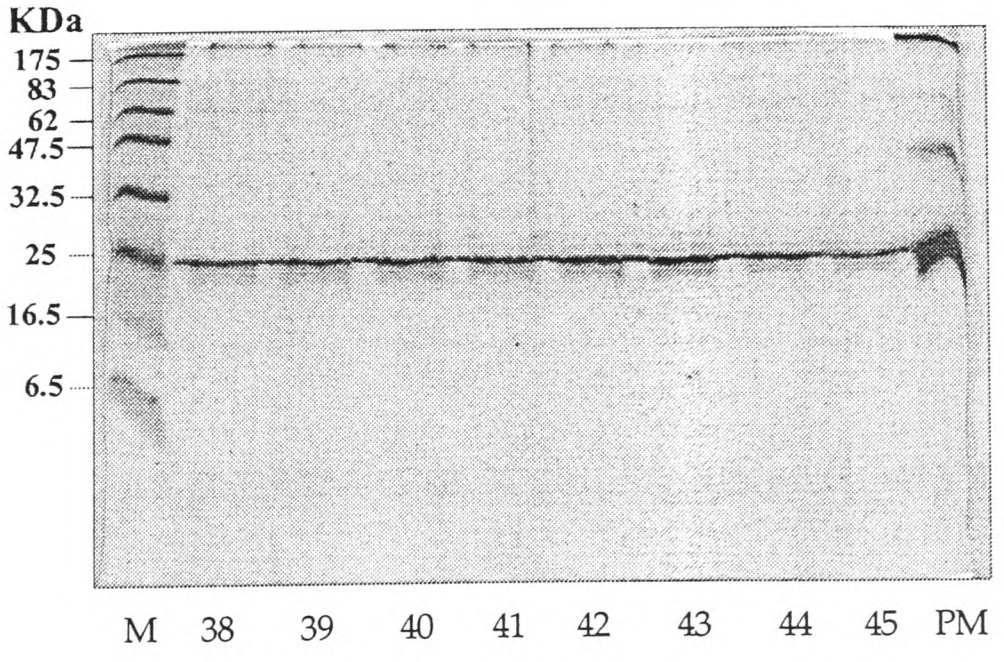


Figure 4.5 Analysis of eluted fractions as presented in graph 4.1

The gel is labelled to correspond to the fraction number as indicated in Graph 4.1. Lane marked 'M' is prestained protein markers, broad range from NEB and'PM' is purple membrane standard.

The complete expression and purification of *bO* using the pLEX vector in G1724 is summarised using SDS-PAGE analysis as shown in figure 4.6.

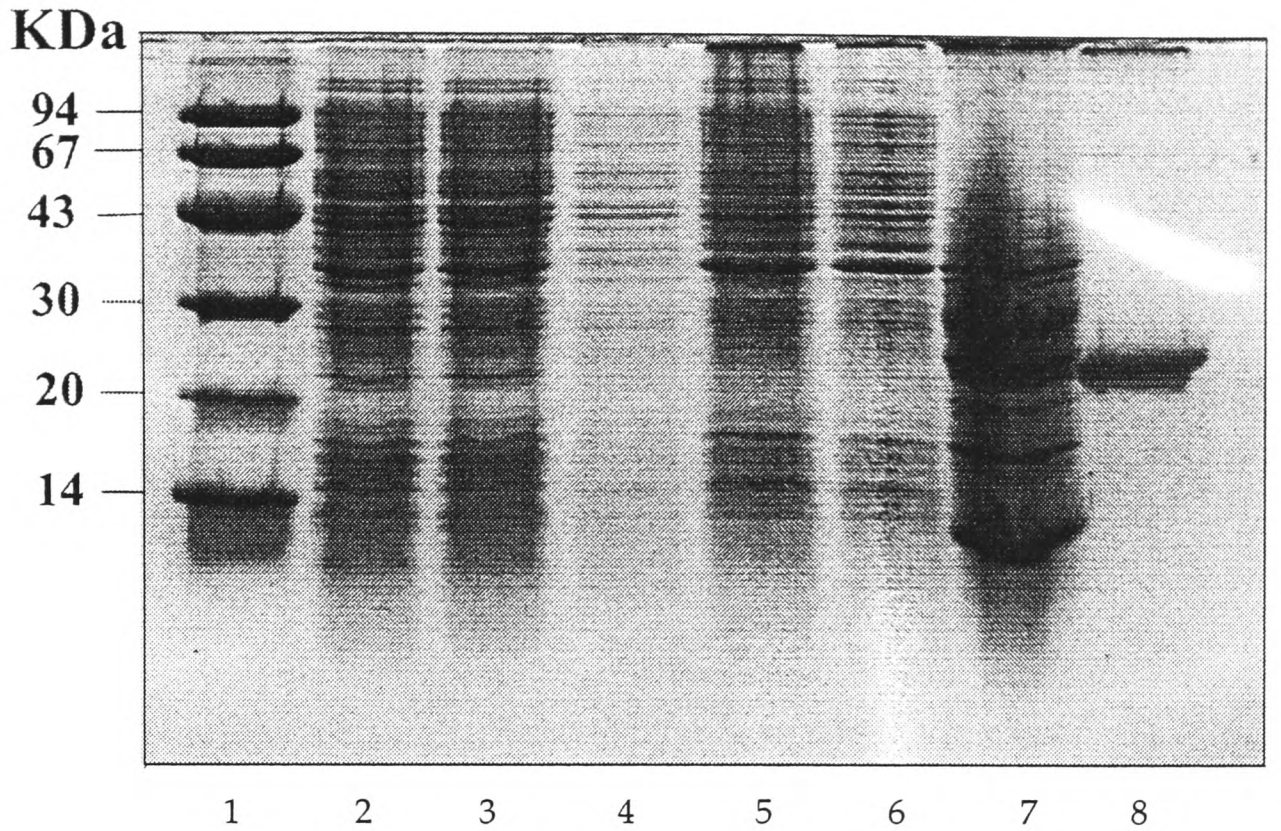


Figure 4.6 Various stages of expression and purification of *bO* in pLEX/ G1724

Lane 1: Molecular weight markers Pharmacia Biotech. Lane 2: uninduced cells. Lane 3: induced cells. Lane 4: supernatant after whole cell lysis. Lane 5: pellet after whole cell lysis. Lane 6: supernatant after organic extraction procedure. Lane 7: pellet after organic extraction procedure. Lane 8: purified *bO* after ion-exchange column.

Both expression and purification studies revealed that the pLEX system was the best choice for the expression of the *bO* gene and due to the nature of this expression system, the organic extraction procedure as well as the ion-exchange column were the most suitable for its purification.

After the expression and purification of *bO*, its functionality was tested by refolding the protein and measuring its refolding yield which was 70% (see section 2.3.4.6). The spectra could not be printed due to technical problems.

4.3 Conclusion

The purification of *bO* from pET16b using the His column (figure 4.2) was difficult to judge due to low expression yields (figure 3.16). However, in comparison to the solvent organic extraction carried out also on pET16b/*bO*, the purification was noted to be lower by SDS-PAGE.

Purification using the chitin column for the pT7IMPACT/*bO* system again posed a solubility problem due to the hydrophobic nature of *bO*. It was thought that some of the cleaved *bO* may have irreversibly bound to the column and thus the entire fusion protein had to be stripped from the column.

bO expressed using pLEX/*bO* was successfully purified using the organic extraction step and subsequent ion exchange chromatography. The complete purification process is shown in figure 4.4. The refolding yield was 70%. Is this due to a contaminating protein that has co-purified with *bO* which possesses the same electrophoretic mobility or is it because some of the *bO* has been damaged during the purification procedure or is it simply a combination of both? Crude cell extract analysis by SDS-PAGE showed the presence of other *E. coli* proteins that posses similar electrophoretic mobility to *bO*. However, SDS-PAGE analysis after ion exchange chromatography showed the other proteins eluted at a lower ionic strength. These other proteins did not react with the bR antibody either. This suggests that purified *bO* is not contaminated with other *E. coli* proteins but instead has been damaged by the purification procedure which does not therefore permit complete regeneration. This is also the case when pure bR obtained from

the natural host is subjected to the same purification procedure and also shows some loss in its ability to bind retinal (Braiman *et al.*, 1987).

Using the pLEX expression vector, the *bO* gene was successfully expressed, purified and refolded.

Chapter 5:

Construction of proline mutants and subsequent kinetic investigations

Chapter 5 : Construction of proline mutants and subsequent kinetic investigations

5.1 Introduction

Studies on the folding kinetics of bR (Booth *et al.*, 1995; Booth *et al.*, 1997) suggest that the rate limiting step may be due to *cis-trans* isomerisation about proline amide bonds (Brandts *et al.*, 1975; Nall *et al.*, 1994). Proline isomerisation has been shown to be responsible for the slow phase in folding studies of several proteins. These include RNAse A (Mayr *et al.*, 1996; Houry and Scheraga, 1996), RNAse T1 (Odefey *et al.*, 1995) and FKBP (Veeraraghavan *et al.*, 1996).

To investigate further whether *cis-trans* isomerisation is responsible for the slow phase in bR folding, three prolines present in the membrane embedded α-helices: Pro_{50}, Pro_{91} and Pro_{186}, have been individually replaced with alanine. Mutagenic PCR of the synthetic bacterio-opsin gene was used for this single amino acid substitution. The mutations were confirmed by DNA sequencing. They were subsequently expressed and purified using the pLEX expression vector (see section 3.2.2.7). Refolding yields were assayed using chromophore regeneration and kinetic studies were carried out by transferring the protein to mixed detergent lipid micelles (SDS/DMPC/CHAPS) in the presence of retinal and measuring changes in protein fluorescence using stopped-flow fluorimetry.

5.1.1 Proline residues in bacteriorhodopsin

There are eleven proline residues present in bacteriorhodopsin and three of them are embedded in the membrane as shown in Fig 5.1: Pro_{50} (helix B), Pro_{91} (helix C) and Pro_{186} (helix F).

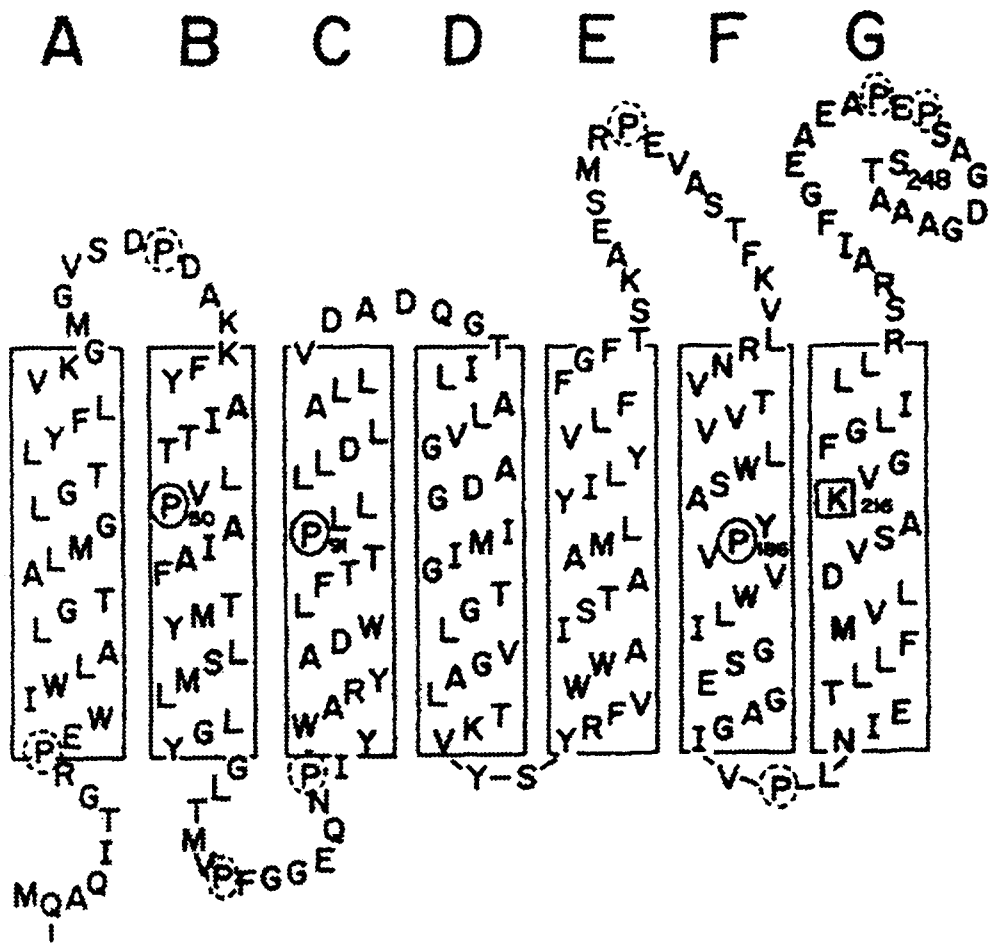


Figure 5.1 Position of membrane embedded proline residues in bacteriorhodopsin

5.2 Results and discussion

5.2.1 Construction of proline mutants

The three nucleotide alterations were introduced *in vitro* by means of mutagenic PCR to generate plasmids Pro_{50}/bO, Pro_{91}/bO and Pro_{186}/bO.

When designing oligonucleotides for the proline mutants, table 5.1 was consulted to ensure selected codons resulted in efficient translation in *E. coli*. Although virtually all organisms use the same genetic code, each organism has a bias towards preferred codons.

A high proportion of unfavoured codons may result in difficulties for the host cells tRNA in translating the gene and consequently reducing the amount of protein synthesised. On the basis of the codon usage table, we chose to replace the proline codon with the most optimum codon for alanine (GCT).

aa	Codon	No.[1]	/1000[2]	Fraction[3]
Gly	GGG	13	1.89	0.02
Gly	GGA	3	0.44	0.00
Gly	GGU	365	52.99	0.59
Gly	GGC	238	34.55	0.38
Glu	GAG	108	15.68	0.22
Glu	GAA	394	57.20	0.78
Asp	GAU	149	21.63	0.33
Asp	GAC	298	43.26	0.67
Val	GUG	93	13.50	0.16
Val	GUA	149	21.20	0.26
Val	GUU	289	43.26	0.51
Val	GUC	38	5.52	0.07
Ala	GCG	161	23.37	0.26
Ala	GCA	173	25.12	0.28
Ala	GCU	212	30.78	0.35
Ala	GCC	62	9.00	0.1
Arg	AGG	1	0.15	0.00
Arg	AGA	0	0.00	0.00
Ser	AGU	9	1.31	0.03
Ser	AGC	71	10.31	0.20
Lys	AAG	111	16.11	0.26
Lys	AAA	320	46.46	0.74
Asn	AAU	19	2.76	0.06
Asn	AAC	274	39.78	0.94
Met	AUG	170	24.68	1.00
Ile	AUA	1	0.15	0.00
Ile	AUU	70	10.16	0.17
Ile	AUC	345	50.09	0.83
Thr	ACG	25	3.63	0.07
Thr	ACA	14	2.03	0.04
Thr	ACU	130	18.87	0.35
Thr	ACC	206	29.91	0.55

aa	Codon	No.[1]	/1000[2]	Fraction[3]
Trp	UGG	55	7.98	1.00
stop	UGA	0	0.00	(stop)
Cys	UGU	22	3.19	0.49
Cys	UGC	23	3.34	0.51
stop	UAG	0	0.00	(stop)
stop	UAA	0	0.00	(stop)
Tyr	UAU	51	7.40	0.25
Tyr	UAC	157	22.79	0.75
Leu	UUG	18	2.61	0.03
Leu	UUA	12	1.74	0.02
Phe	UUU	51	7.40	0.24
Phe	UUC	166	24.10	0.76
Ser	UCG	14	2.03	0.04
Ser	UCA	7	1.02	0.02
Ser	UCU	120	17.42	0.34
Ser	UCC	131	19.02	0.37
Arg	CGG	1	0.15	0.00
Arg	CGA	2	0.29	0.01
Arg	CGU	290	42.10	0.74
Arg	CGC	96	13.94	0.25
Gln	CAG	233	33.83	0.86
Gln	CAA	37	5.37	0.14
His	CAU	18	2.61	0.17
His	CAC	85	12.34	0.83
Leu	CUG	480	69.69	0.83
Leu	CUA	2	0.29	0.00
Leu	CUU	25	3.63	0.04
Leu	CUC	38	5.52	0.07
Pro	CCG	190	27.58	0.77
Pro	CCA	36	5.23	0.15
Pro	CCU	19	2.76	0.08
Pro	CCC	1	0.15	0.00

[1] Number of occurences of the codon in the genes from which the table is compiled.
[2] Expected number of occurences per 1000 codons in genes whose codon usuage is identical to that compiled in the frequency table
[3] Fraction of occurences of the codon in its synonymous codon family.

Table 5.1 Efficient translation codons in *E. coli*

The following nucleotides were thus designed and ordered:

PROLINE MUTANT OLIGONUCLEOTIDES

Mutant Pro 50

5′ GAT CCG GAT GCG AAA AAA TTC TAC GCT ATC ACC ACC CTG GTG **GCT** GCT ATC G 3′ FOR

Bam H1 *Bsm 1*

Mutant Pro 91

5′ CGT TAC GCT GAC TGG CTG TTC ACC ACC **GCT** CTG CTG CTG CTG CTA 3′ FOR

Apa1 *Bgl 11*

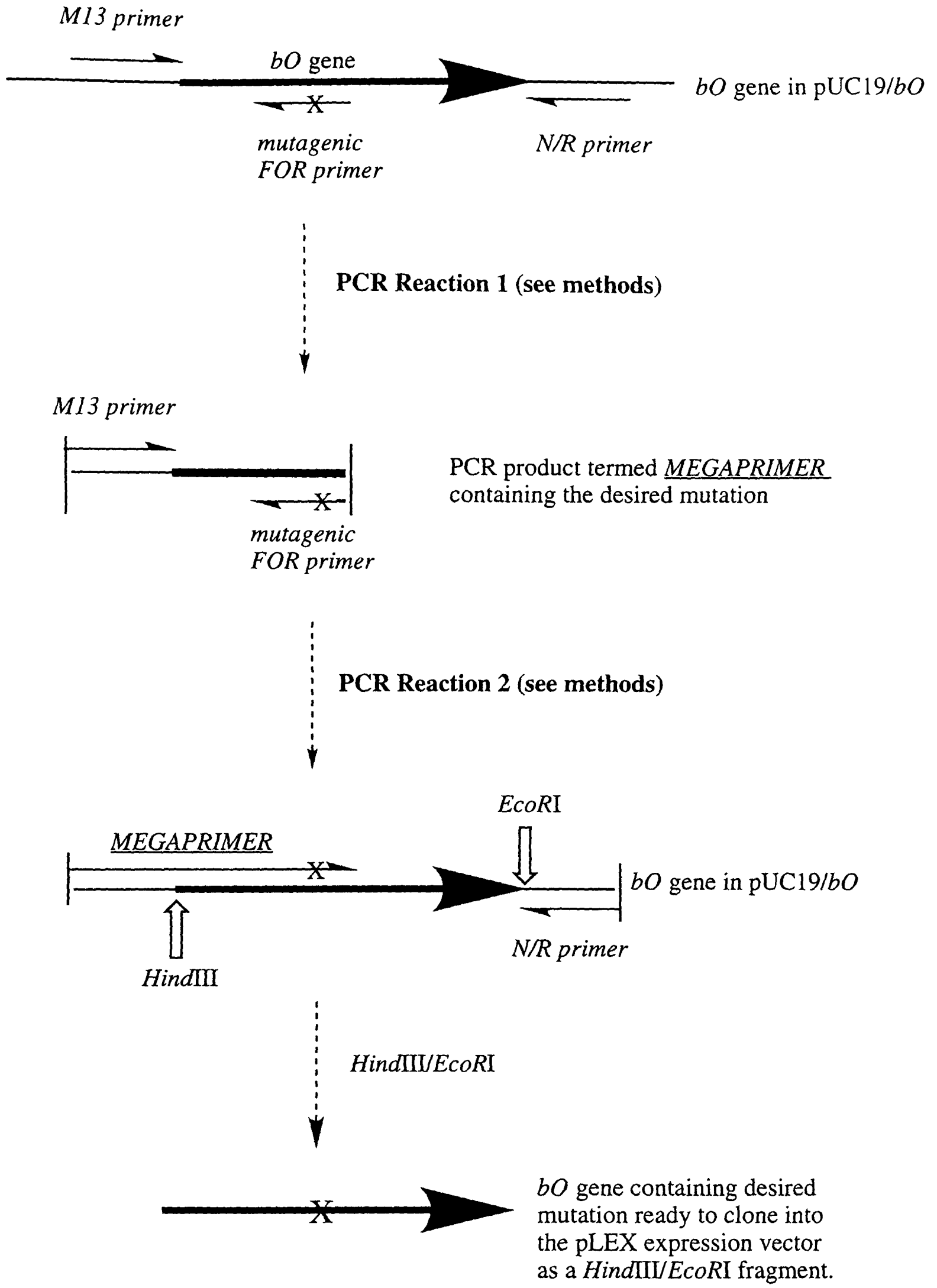
Mutant Pro 186

5′ G TCC GCT TAC **GCT** GTT GT 3′ FOR

Ava 11 *Bst X1*

The template for all PCR reactions was plasmid pUC19/*bO*, a pUC19 clone containing the synthetic *bO* originally present in pSB02 (kind gift from Prof. H. G. Khorana). The other primers used in this procedure were the pUC19 sequencing primers, M13 and NR (see section 3.2.1).

The overall strategy employed to introduce the Pro to Ala mutation in the *bO* gene using PCR is shown in figure 5.2. The first PCR reaction was used to amplify the region between the oligonucleotide primer using the Pro FOR and M13 sequencing primer (2.3.2.7). The PCR products were gel purified and used as 'megaprimers' to carry out a second PCR reaction (figure 5.3). The megaprimers, which should have incorporated the mutant, are used as one of the primers to PCR the full gene along with the NR primer for pUC19 in the second PCR reaction (2.3.2.7). The products of the second PCR reaction were gel purified and excised with *Hind*III and *Eco*R1 ready to ligate into pUC19 which was also digested with the same enzymes (figure 5.4). The resulting plasmids were called Pro_{50}/bO, Pro_{91}/bO and Pro_{186}/bO respectively.



X-highlights the Pro to Ala mutation as shown in the proline oligonucleotides

Figure 5.2 Strategy outlined for the introduction of the Pro to Ala mutation in the *bO* gene using PCR

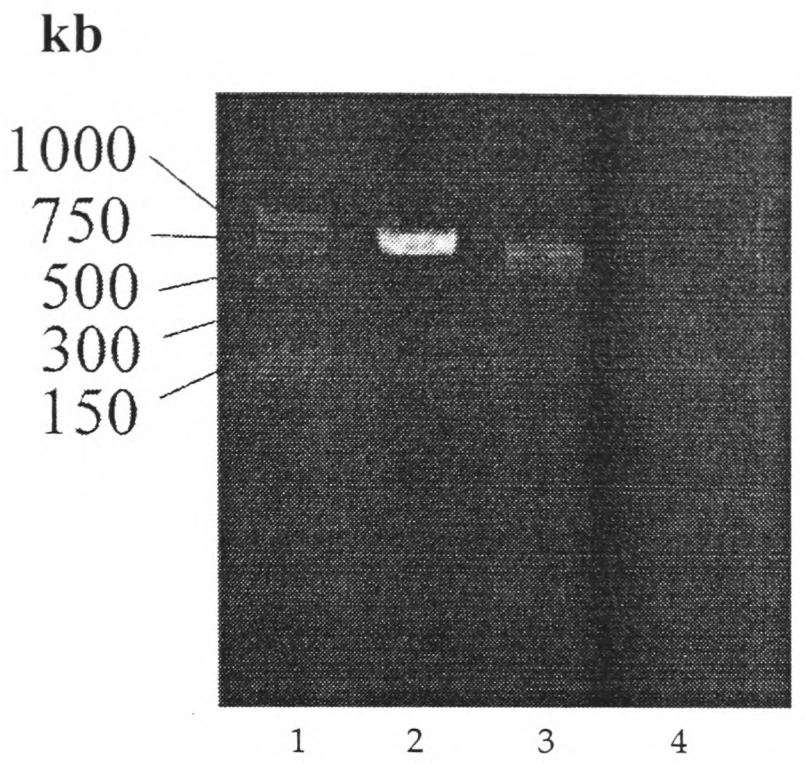


Figure 5.3 Purified PCR products after reaction 1 to be used as megaprimers in PCR reaction 2.

Lane 1: PCR markers from Promega as indicated. Lane 2: Pro_{50} megaprimer ~685 bp. Lane 3: Pro_{91} megaprimer ~550 bp. Lane 4: Pro_{186} megaprimer ~250 bp.

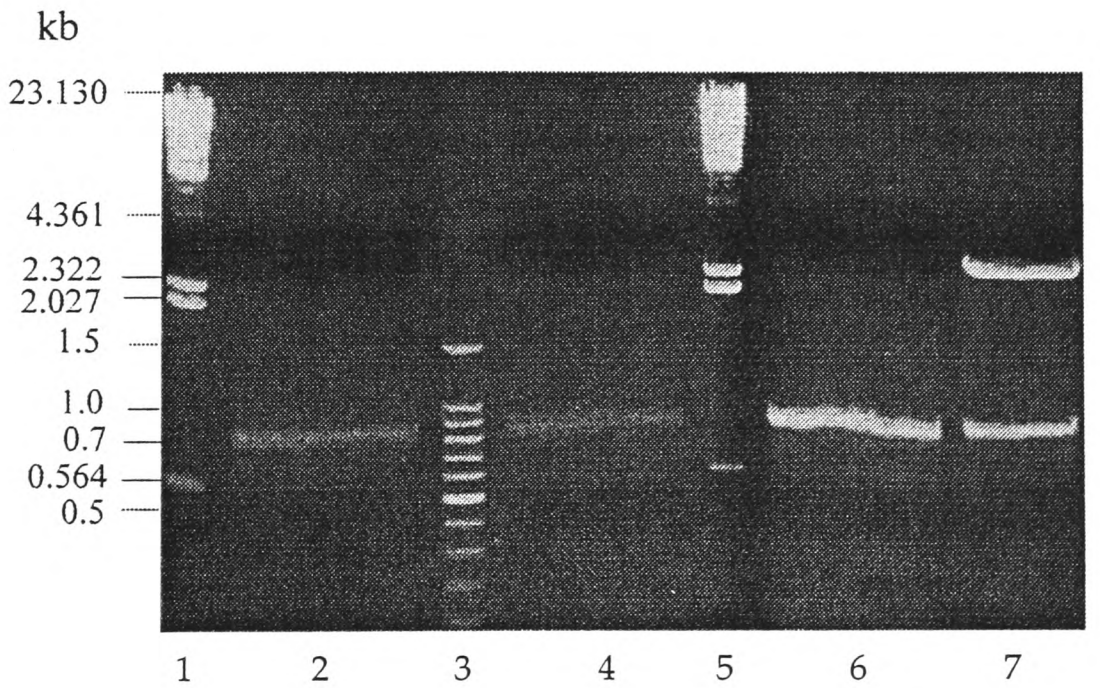


Figure 5.4 *bO* gene containing the three desired Pro to Ala mutations ready to clone into pUC19

Lane 1: λ DNA/ *Hind* III fragments from Gibco BRL™ as indicated. Lane 2: *bO* gene containing Pro_{50} to Ala mutation. Lane 3: 100 kb ladder from NEB. Lane 4:*bO* gene containing Pro_{91} to Ala mutation. Lane 5: λ DNA/ *Hind* III fragments from Gibco BRL™. Lane 6: *bO* gene containing Pro_{186} to Ala mutation. Lane 7: pUC19 cut *Hind* III/*Eco*R I for cloning mutant *bO* genes.

5.2.2 Sequence analysis of proline mutants

The cloned PCR products were sequenced to check the accuracy of PCR amplification. The nucleotide sequence corresponds precisely to the template DNA except for the engineered base changes.

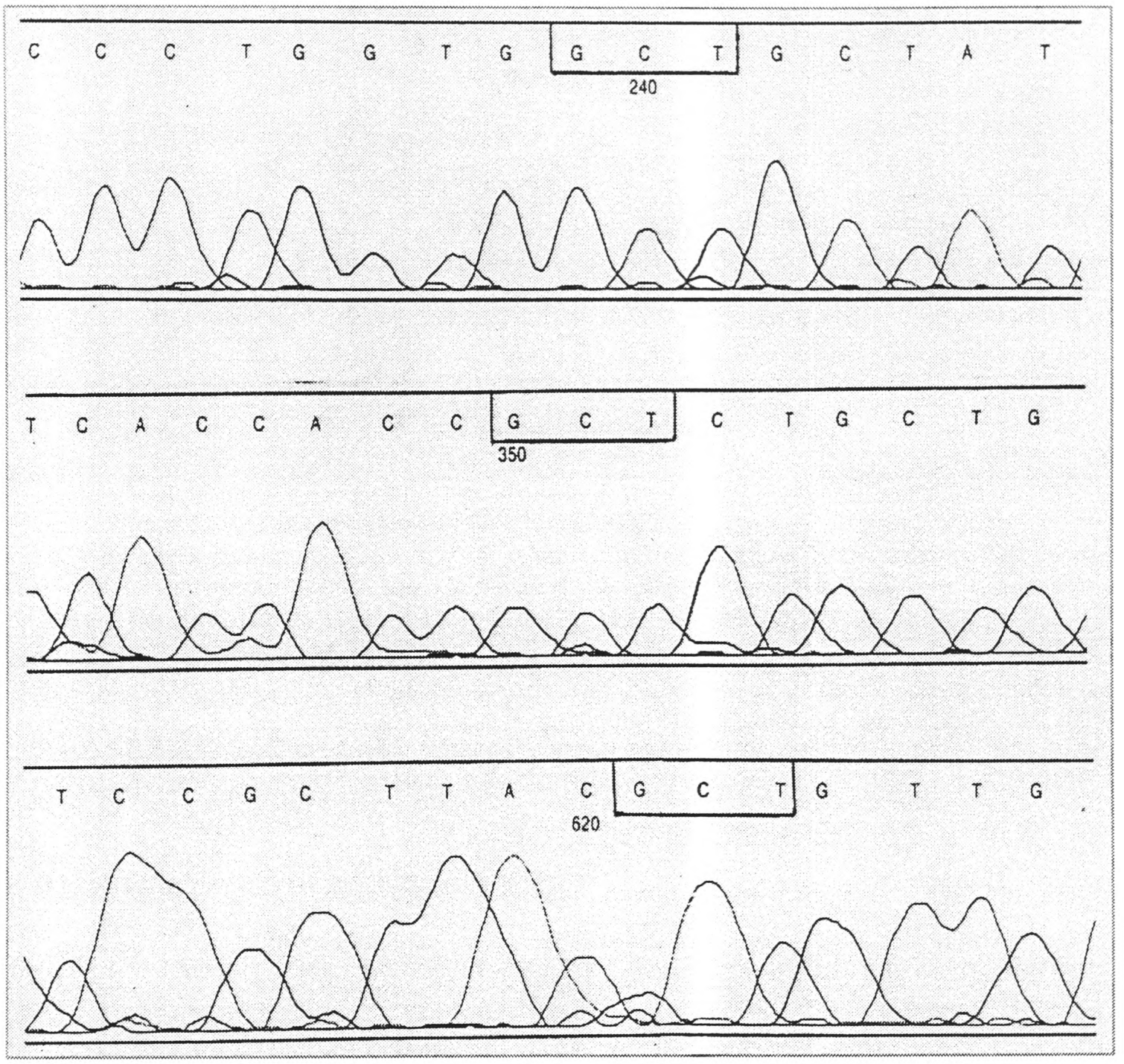


Figure 5.5 Sequence data for the three Pro to Ala mutation of the *bO* gene

The top row shows the Pro_{50} to Ala muataion as highlighted by the box. The second and third rows show the mutation of Pro_{91} and Pro_{186} to Ala respectively.

5.2.3 Expression studies on the proline mutants using the pLEX expression system

The mutant *bO* genes were cloned into the pLEX expression vector (3.2.2.7) and expressed (3.3.1.5). Expression yields for Pro_{91}/bO and Pro_{186}/bO were approximately 0.6 mg/ml and 0.8 mg/ml respectively. Immunoblotting results showed expression of Pro_{50}/bO in pLEX to be very leaky and subsequently, the protein proved lethal to the cells (figure 5.6).

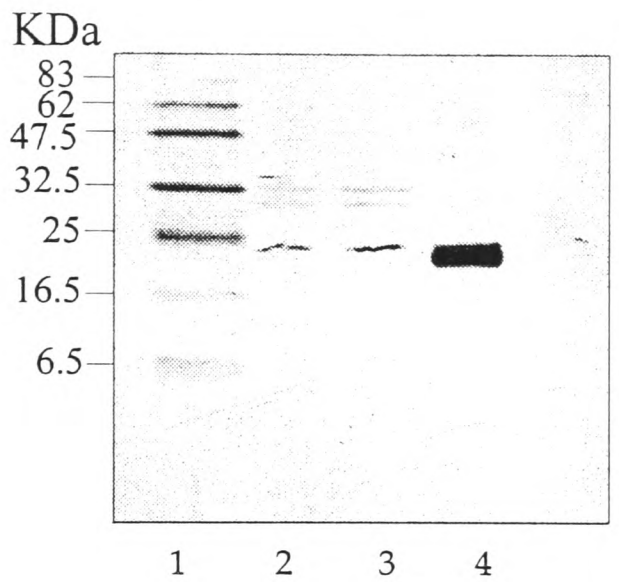
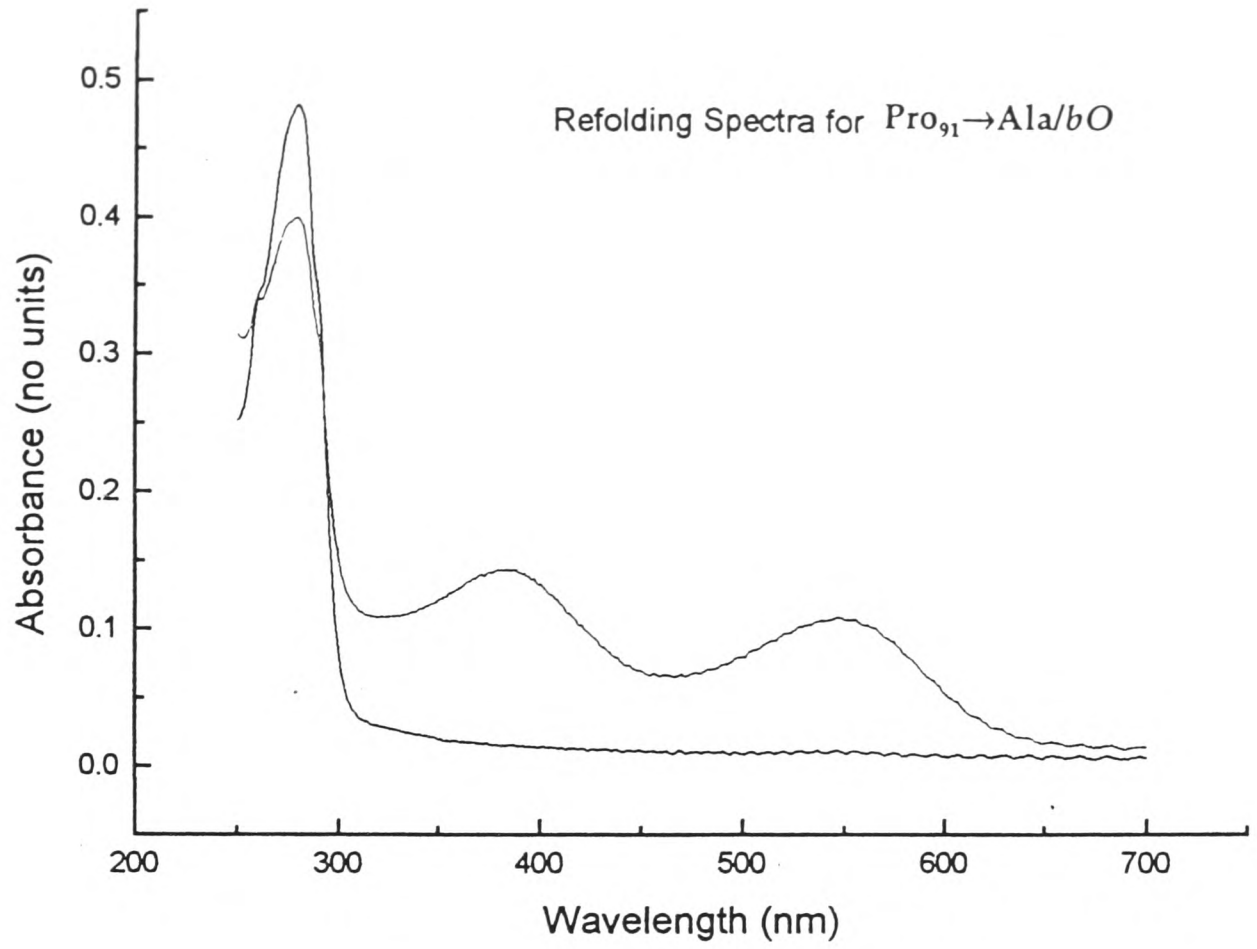


Figure 5.6 Leaky expression of Pro_{50}/bO in pLEX as shown by immunoblotting
Lane 1: Prestained protein markers, broad range from NEB. Lane 2: uninduced cells show the production of a small amount of *bO*. Lane 3: induced cells show the production of the same amount of *bO* suggesting the system is very leaky, Lane 4: purple membrane sample.

The expressed mutant proteins were purified using the organic extraction procedure (4.2.4.1) and ion exchange chromatography (4.2.4.2).

5.2.4 Refolding yields of proline mutants

Refolding studies on $Pro_{91}\rightarrow Ala/bO$ and $Pro_{186}\rightarrow Ala/bO$ (2.3.4.6) were carried out in the dark. The yields were calculated to be 45% and 40% respectively. The spectra for both are shown in figure 5.7. The λ_{max} values measured from the spectra were 548 nm for $Pro_{91}\rightarrow Ala/bO$ and 557 nm for $Pro_{186}\rightarrow Ala/bO$. These values are in very close agreement with results published by Mogi *et al.*, (1989) which quote the absorption maximum of $Pro_{91}\rightarrow Ala/bO$ to be 548 nm and $Pro_{186}\rightarrow Ala/bO$ to be 553 nm. The absorption maximum for dark adapted wild-type protein is 551 nm.



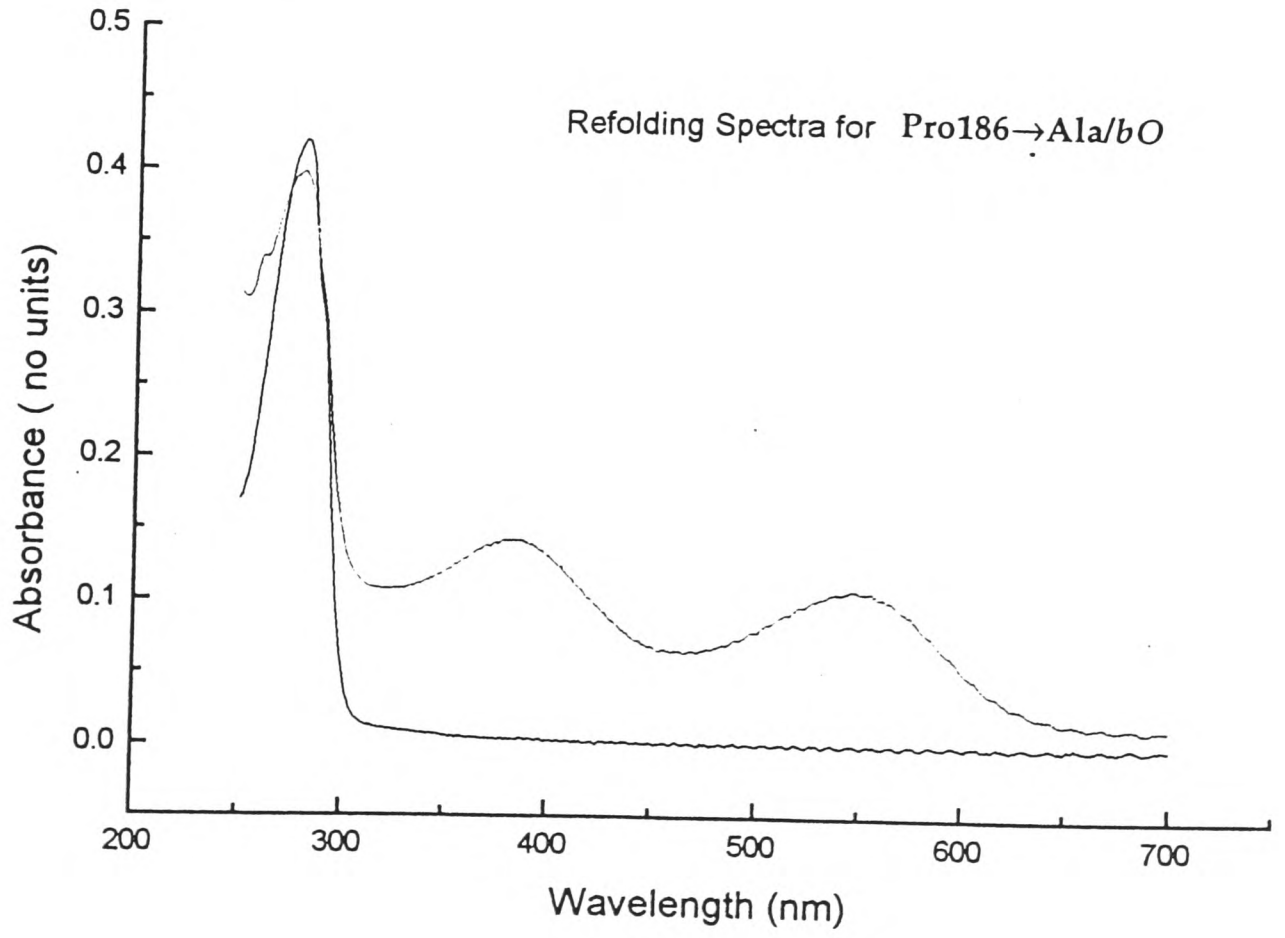
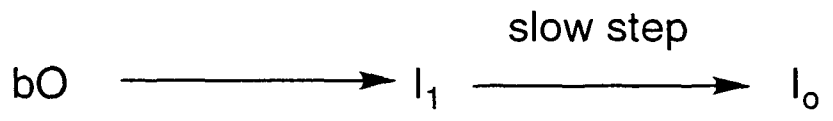


Figure 5.7 Refolding Spectra for Pro_{91}→Ala/*bO* and Pro186→Ala/*bO*

Yields were calculated to be 45% and 40% respectively.

5.2.5 Kinetic studies on proline mutants using stopped-flow fluorimetry

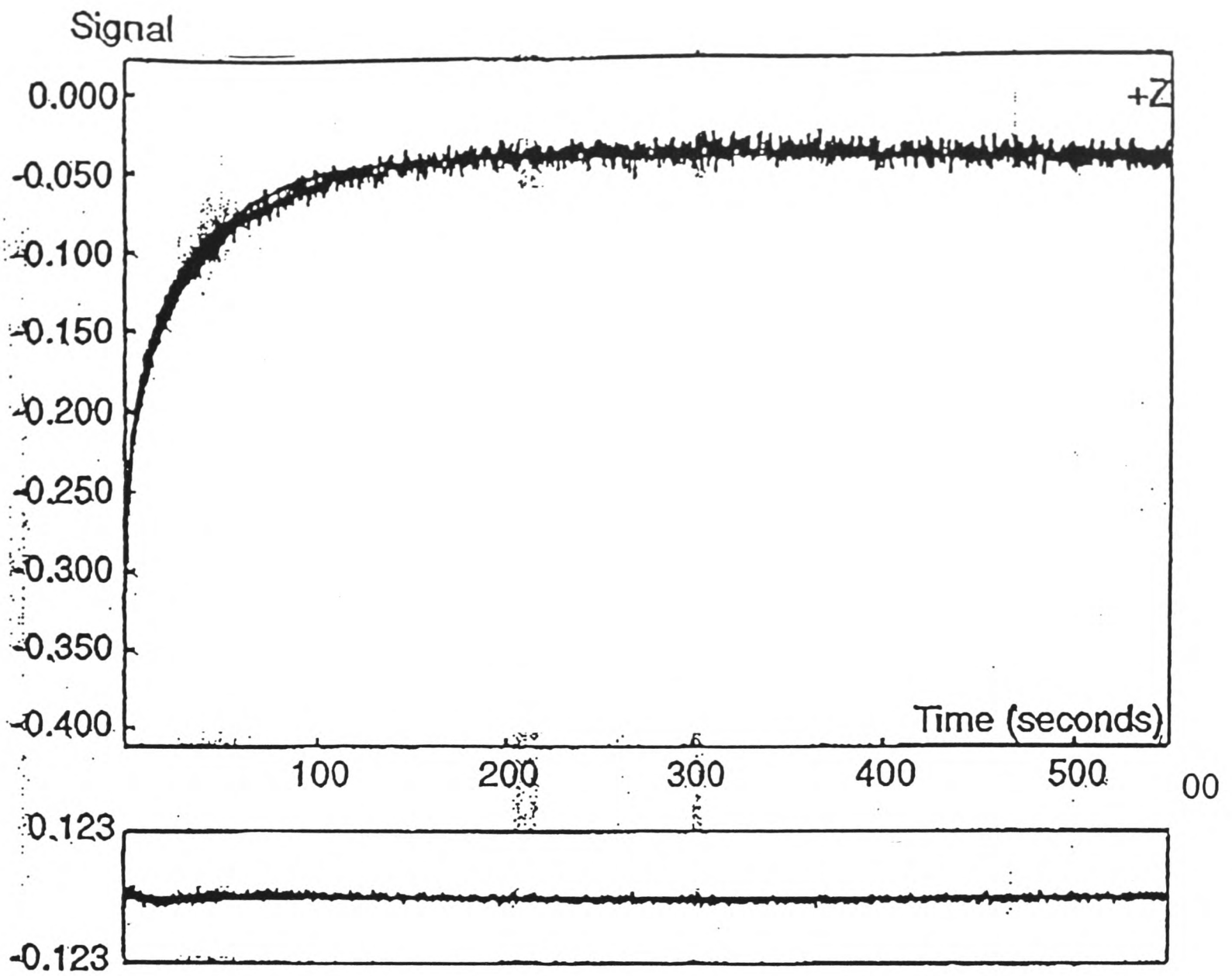
Proline mutants were constructed in an aim to determine whether the slow I_1 to I_o step in the folding pathway of bR is due to *cis-trans* isomerisation of the membrane embedded proline residues.



By mutating Pro_{91} and Pro_{186} to Ala, we can compare the folding kinetics of the mutant proteins to that of the *bO* gene alone using stopped-flow fluorimetry.

This work was carried out with the assistance of Dr. Hui Lu at Imperial College of Science and Technology in London. The refolding of *bO* can be initiated by a rapid-mixing, stopped-flow method which monitors changes in protein fluorescence. Folding *bO* in DMPC/CHAPS micelles in the absence of retinal allows us to monitor the kinetics for the bO→Io stage. We know from previous CD studies that Io possesses native secondary structure (London and Khorana, 1982). Studies in the presence of retinal were not performed as this corresponds to the second stage of the folding pathway which is not under investigation here. Experiments were done at 22°C and protein fluorescence was measured at 295 nm after excitation at 280 nm.

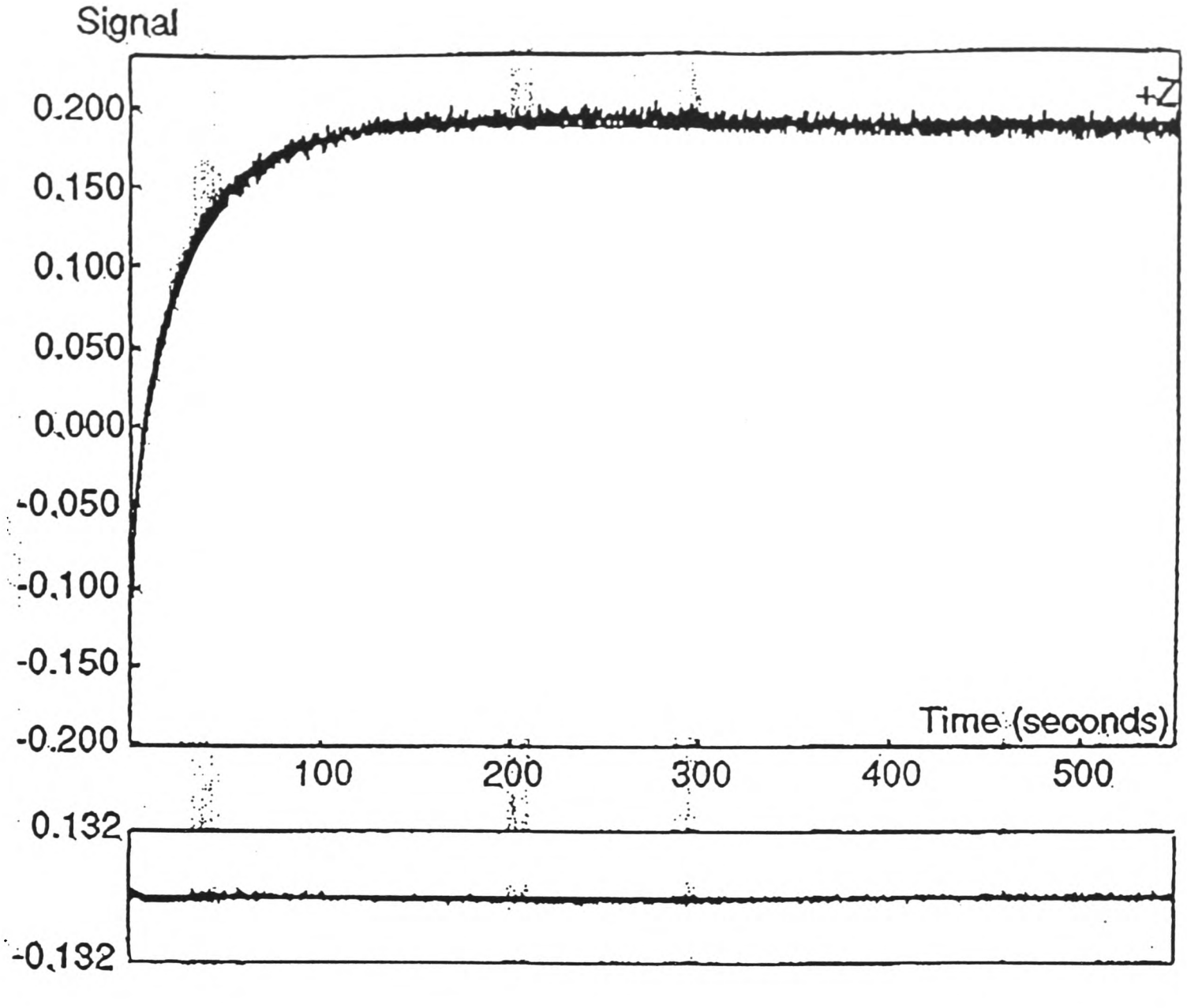
Data for Pro_{91}→Ala, Pro_{186}→Ala and *bO* without any mutations is shown in figure 5.8, 5,9 and 5.10 respectively.



Split Timebase=50.0s / 500s Ch:#1 Temp=22.0°C
Wlength=295nm PM volts=999.0v Offst=2.00v Fltr=1.00ms
REGRESSION RESULTS:
Function name: Double exponential, floating end point
Library: Standard
Formula: P(1)*EXP(-P(2)*X)+P(3)*EXP(-P(4)*X)+P(5)

Parameter:		Value:	Std.Error:
P(1)	Amp 1	-1.93E-1	1.84E-3
P(2)	Rate 1	8.01E-1	1.29E-2
P(3)	Amp 2	-1.85E-1	5.22E-4
P(4)	Rate 2	2.85E-2	1.38E-4
P(5)	Endpt	-4.01E-2	1.51E-4
Norm. Variance:		3.86E-5	Fit range: 1-4000

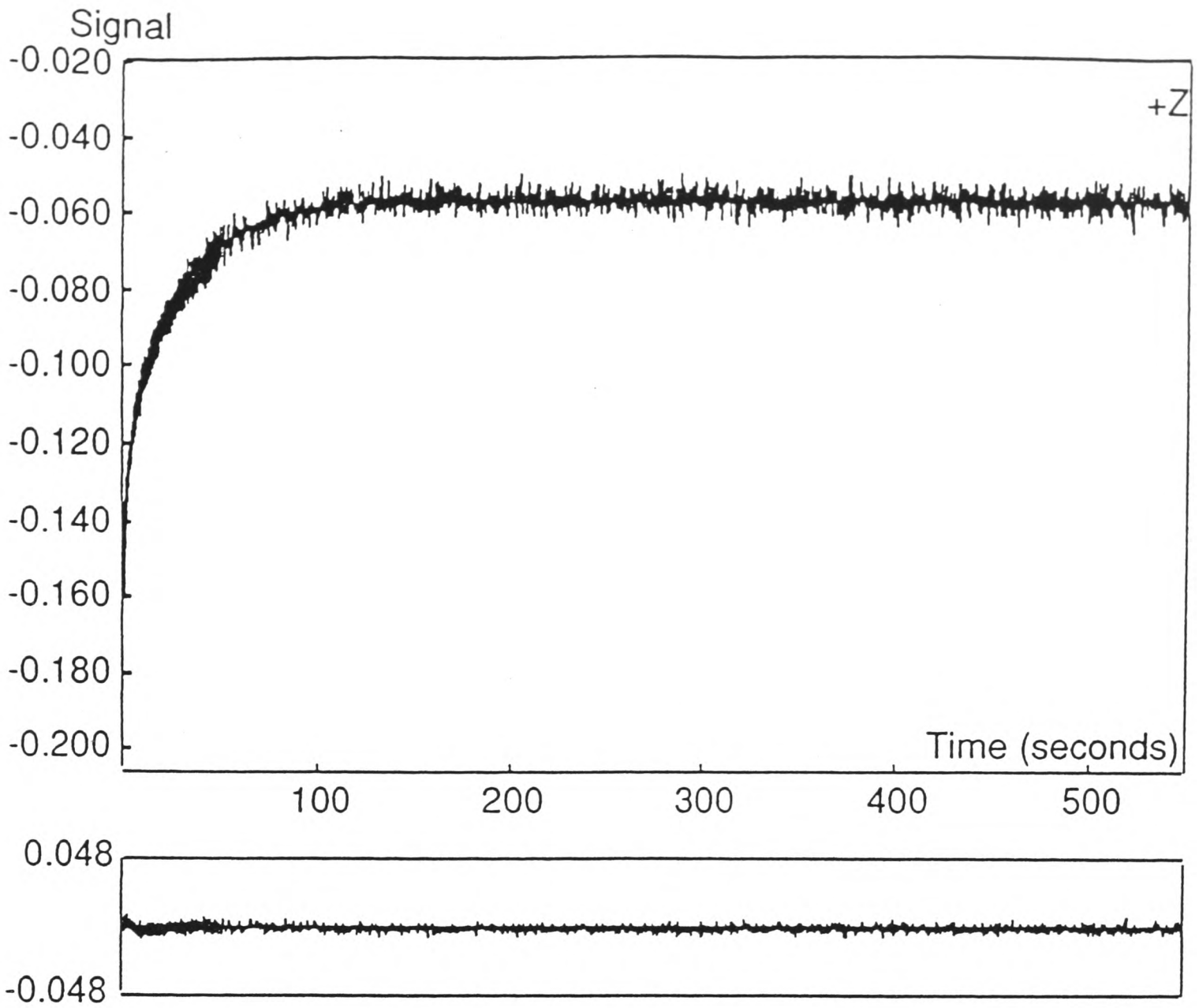
Figure 5.8 Refolding kinetics for Pro_{91} to Ala



Split Timebase=50.0s / 500s Ch:#1 Temp=21.9°C
Wlength=295nm PM volts=999.0v Offst=1.80v Fltr=1.00ms
REGRESSION RESULTS:
Function name: Double exponential, floating end point
Library: Standard
Formula: P(1)*EXP(-P(2)*X)+P(3)*EXP(-P(4)*X)+P(5)

Parameter:		Value:	Std.Error:
P(1)	Amp 1	1.96E-1	1.71E-3
P(2)	Rate 1	1.04E0	1.46E-2
P(3)	Amp 2	2.50E-1	4.79E-4
P(4)	Rate 2	3.57E-2	1.08E-4
P(5)	Endpt	1.89E-1	1.25E-4
Norm. Variance:		2.83E-5	Fit range: 1-4000

Figure 5.9 Refolding kinetics for Pro_{186} to Ala



Split Timebase=50.0s / 500s Ch:#1 Temp=22.0°C
Wlength=295nm PM volts=999.0v Offst=0.90v Fltr=1.00ms
REGRESSION RESULTS:
Function name: Double exponential, floating end point
Library: Standard
Formula: P(1)*EXP(-P(2)*X)+P(3)*EXP(-P(4)*X)+P(5)

Parameter:		Value:	Std.Error:
P(1)	Amp 1	-1.04E-1	8.54E-4
P(2)	Rate 1	9.66E-1	1.34E-2
P(3)	Amp 2	-7.31E-2	2.41E-4
P(4)	Rate 2	3.62E-2	1.88E-4
P(5)	Endpt	-5.65E-2	6.22E-5
Norm. Variance:		6.99E-6	Fit range: 1-4000

Figure 5.10 Refolding kinetics for *bO* alone

The rates constants (rate 1 and rate 2) shown in figures 5.8, 5.9 and 5.10 correspond to the k_1 and k_2 steps in the initial folding pathway of *bO* (figure 5.11). These rate constants are summarised in table 5.2

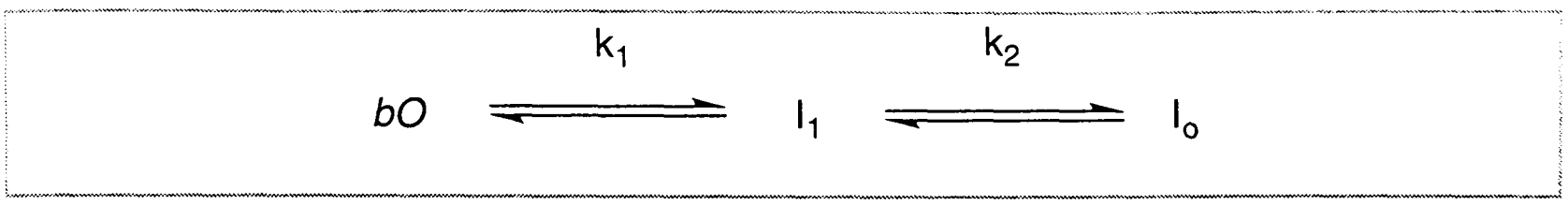


Figure 5.11 Initial bR folding pathway. Refer to figure 1.7 for further deatils.

	Pro_{91}→Ala/*bO*	Pro_{186}→Ala/*bO*	*bO* wild type
k_1	8.01E-1	1.04E0	9.66E-1
k_2	2.85E-1	1.89E-2	3.62E-1

where E is the exponential

The initial rapid rise in fluorescence as can be seen in figures 5.8, 5.9 and 5.10 correspond to the mixing of SDS and DMPC/CHAPS micelles (Booth *et al.*, 1997). During this 4 ms period, perhaps some of the α-helices are also formed. In the case of soluble proteins, it has been suggested that most α-helix formation takes place within 2 ms (Kim and Baldwin, 1982). This 'folding' event may represent formation of the first intermediate I_1.

5.3 Conclusion

Formation of the I_o intermediate from I_1 is very slow step (figure 1.7). This step was initially thought to perhaps be due to proline isomerisation. Our kinetic data however, did not show any difference between the mutant proteins and *bO*. This suggests that the kinetics are not affected by the mutation of the membranous

Pro_{91} and Pro_{186} residues to Ala. Thus proline isomerisation does not seem to account for this slow phase. Studies by Booth *et al.* (1997) suggest that the rate limiting step of bR is affected by the refolding conditions and lateral pressure the lipids impose on the protein as well as pH (Riley *et al.*, 1997). See section 1.1 for further details. Further studies on monitoring helix packing by changes in the near-UV CD signal of monmeric bR may provide further insight into this area (currently under investigation).

Chapter 6:

Mass spectrometry in membrane protein analysis

Chapter 6: Mass spectrometry in membrane protein analysis

6.1 Introduction

The last decade has experienced the increasing use of mass spectrometry as an integral part of biological research. Mass spectrometry differentiates and detects gas phase ions. Traditional methods employed to generate these ions tend to be too harsh for some non-volatile and thermolabile molecules. Fast atom bombardment (FAB) was introduced in 1981 which allowed non-volatile, thermolabile compounds to be studied. The technique however was limited to small peptides up to 4000 amu (Fenselau and Cotter, 1987) and offered only nanomole sensitivity. The recent development of two ionisation techniques: matrix-assisted laser desorption ionisation mass spectrometry (MALDI-MS) (Karas and Hillenkamp, 1988; Chait and Kent, 1992; Stults, 1995; Tanaka *et al.*, 1988) and electrospray ionisation mass spectrometry (ESI-MS) (Fenn *et al.*, 1989; Siuzdak, 1994) has made a major breakthrough in this area. Both these techniques offer picomole-to-femtomole (Siuzdak, 1994) sensitivity of a much wider range of biomolecules including peptides, proteins, carbohydrates, oligonucleotides, natural products and drug metabolites (Chait and Kent, 1992; Smith, 1990; Yates *et al.*, 1993; Burlingame, 1996; Stults, 1995; Mann and Wilm, 1995). They are able to achieve this by employing 'gentler' techniques to generate gas phase ions using non-destructive vaporisation and ionisation techniques (Siuzdak, 1996). The whole process means a wide range of biological systems can be studied under or near physiological conditions with minimum sample preparation, thus maintaining

their solution biochemistry. Although their ionisation techniques are quite different, the end result is essentially the same.

6.1.1 Matrix assisted laser desorption ionisation mass spectrometry (MALDI-MS)

MALDI was first commercially available in 1991. As the name suggests, this techniques uses a chromophoric matrix for the analysis of samples. Reviews on the mechanism are provided by Cornelius *et al.*, 1997; Feneslau, 1997; Fitzgerald and Siuzdak, 1996; Stults 1995. The sample is introduced onto a sample slide along with the matrix. The solvent is evaporated and the slide transferred to the vacuum system of the MS. The matrix not only separates the sample molecules from each other but allows the absorption of photons from a pulsed laser beam, usually 337 nm N_2 laser (Hillenkamp *et al.*, 1991). This causes the matrix to expand into the gas phase along with the sample molecules. Ionisation occurs in the plume above the sample slide by proton transfer between excited matrix and analyte molecules. Analytes with a molecular weight <20 KDa are typically ionised with only one or two charges whereas large ones (>20 KDa) can accumulate three to five charges depending upon the desorption conditions such as the matrix, ratio of matrix to analyte, laser power, wavelength, pulse width and whether positive or negative ions are recorded. Biolpolymers >30 3000 amu can be ionised with limited or no fragmentation (Cornelius *et al.*, 1997). The ions are then transferred to the mass analyser which measures the m/z and plots it against the abundance. Different compounds require different matrices for analysis.

A major advantage of the technique is the ability to analyse a mixture of components without chromatography as each one produces only one signal. It is

also tolerant of contaminants such as buffers and salts. The high sensitivity of the technique requires samples in the 100 fmol to 2 pmol range and even then, only a small percentage is used. The combination of MALDI with Time-of-Flight (TOF) and Fourier Transform (FT) allows the technique to be optimised to analyse compounds with higher masses. Compounds below 600 Da are difficult to detect due to intense matrix signals. Resolution can also be limited and on-line coupling with Liquid chromatography (LC) is difficult.

The technique itself has been extremely successful for the analysis of water soluble proteins and polypeptides up to 300 KDa (Beavis and Chait, 1990). The potential for membrane proteins is still being investigated (Schey *et al.*, 1992; Lewis *et al.*, 1993). Membrane proteins pose a problem due to organic solvents used to extract them, their tendencies to aggregate and their inability to completely fold even in the presence of detergents such as SDS (Gennis, 1989). These features result in spectra of a poor quality with regard to the signal to noise ratio and mass resolution. Despite this, the technique has been employed to study several membrane proteins. Ghaim *et al.* (1997) used MALDI to investigate molecular weights of hydrophobic subunits: cytochrome *bd* oxidase, cytochrome b_{O3}, cytochrome *bc* oxidase and cytochrome *c* oxidase. Other membrane proteins analysed include OmpF porin, cholesterolesterase and our protein of interest, bacteriorhodopsin (Rosinke *et al.*, 1995). They also investigated several matrices, the influence of detergents and various combinations of the two on the desorption of the protein and/or subunits. Some improvements on the matrix include a thinner layer which is able to detect subfemtomole levels (Vorm *et al.*, 1994) and a fast evaporating matrix which provides a uniform surface for desorption/ionisation allowing greater mass accuracy (Vorm and Mann, 1994).

The technique may not provide as much structural detail as nmr or crystallography but it allows the use of subnanomolar, dilute aqueous material on large protein and protein complexes to be studied (Lui and Smith, 1994; Zhang *et al.*, 1996).

6.1.2 Electrospray ionisation mass spectrometry (ESI-MS)

A vast array of literature documents the use of ESI-MS for the analysis of a wide range of biological samples. Soon it will become an integral part of any pharmaceutical or biological lab. The idea of electrospray originated from Dole's work over 20 years ago in an attempt to generate macromolecular ions (Wilm and Mann, 1994). The mechanism of this technique has been best described by Hop and Bakhtiar (1997), Fenn *et al.* (1989), Fitzgerald and Suidzdak (1996) and Siuzdak (1994). Ions are formed from a dilute solution which enters a hypodermic needle maintained at ≈4000V. The strong electric field at the tip of the needle causes the emerging liquid to form of a fine spray of highly charged droplets that are dispersed by coulombic forces (Fenn *et al.*, 1989). The strong electric field guides the charged droplets to the inlet of the glass capillary tube. Dry gas and/or heat is applied here to evaporate all the solvent until its diameter is really small. The decrease in droplet diameter results in an increase in its charge density until the Rayleigh limit[#] is reached at which point the droplets is broken down to even smaller droplets until they are so small that they transfer to the gas phase. From here they enter the mass analyser through a series of lenses where the m/z ratio is recorded. The number of charges retained by an analyte depends on

[#] Rayleigh limit this is the name given to the point at which the coulombic repulsion between the droplets becomes of the same order as the surface tension.

several factors such as pH of electrospray solvent and chemical nature of the sample (Chait and Kent 1992; Smith *et al.*, 1990). Large peptides and proteins tend to generate a series of multiply charged species whereas smaller ones usually generate singly charged ions. The several different charge states can allow the mass of a analyte to determined with great precision such that errors can be ≤ 0.05%. Compounds up to 1000 000 amu units can be analysed using this technique. The technique can also be coupled with LC and capillary electrophoresis (CE) as well as interfaced with other instruments such as quadrupole, ion trap and Fourier Transform. The disadvantages of the technique include its inability to analyse a mixture of analytes as to many signals would be generated corresponding to each sample molecule. Therefore, one would have to use either LC or CE to separate them first which adds an extra step to the analysis. The presence of salts, buffers and detergents can also reduce the sensitivity and so all need to be selected with care (volatile buffers) which may not always be suitable for the analyte. The disadvantages of the technique are greatly outweighed by its advantages as shown by recent literature reports. Maier *et al.* (1997) used ESI-MS to study the conformational properties of cytochrome *c* using hydrogen/ deuterium exchange (HDX) (6.1.3) as well as by Jaquinod *et al.* (1995) who carried out similar studies on cytochrome c_2 and some mutants. More recently, Chung *et al.* (1997) investigated HDX (6.1.3) properties of native and denatured proteins.

Differences between MALDI and ESI-MS have been both tabulated and diagramatically outlined by Fitzgerald and Siuzdak (1996). These adapted figures are shown in figure 6.1 and table 6.1.

Figure 6.1 Schematic representation of MALDI and ESI-MS
Adapted from Fitzgerald and Siuzdak, 1996.

Table 6.1 Capabilities, limitations and recent improvements of MALDI and ESI-MS.
Adapted from Fitzgerald and Siuzdak, 1996.

6.1.3 Hydrogen/deuterium exchange (HDX) studies on proteins using MS

Structural studies on proteins can be performed by measuring rates of hydrogen deuterium exchange (HDX) (Katta and Chait, 1991; Miranker *et al.*, 1993; Robinson *et al.*, 1994; Robinson and Radford, 1995; Yi and Baker, 1996). This method requires the protein under investigation to posses exchangeable protons at amides and certain side chains. The rates at which hydrogens in a protein undergo isotopic exchange is known to vary over eight orders of magnitude (Neubert *et al.*, 1997). It can take from seconds to several months depending upon the secondary and tertiary structure as well as experimental variables such as pH and temperature. The pH strongly influences the intrinsic rates of amide exchange with a minimum at pH 2 to 3 (Englander and Kallenbach, 1984). Each unit decrease in pH results in a ten fold decrease in the exchange rate until the minimum is reached at which point it increases again due to the presence of excess hydronium ion (Englander and Kallenbach, 1984). High salt concentrations can also influence the intrinsic exchange rates (Kim and Baldwin, 1982).

Amide exchange is dramatically slowed in a folded protein compared to a denatured, random coil structure. Rapid exchange suggests sites present on the protein surface where they are accessible to water whereas slow exchange suggests sites involved in secondary structure such as α-helices and β-sheets, intermolecular H-bonding (Englander and Kallenbach, 1984), steric hindrance and/or sites buried from the solvent (Woodward *et al.*, 1982). The number of deuteriums incorporated and the average rates of exchange can be easily measured by the increase in molecular masses. Thus kinetic and thermodynamic of proteins

can be studied using this technique (Woodward *et al.*, 1982; Bai*et al.*, 1995). The nature of folding intermediates in folding pathways could be determined this way as well as information regarding the cooperativity of folding events (Hooke *et al.*, 1995). Combining the technique with pulse labelling (Baldwin et al., 1993) allows *in vitro* folding pathways to be studied. This technique was used to investigate transient protein folding population during the refolding of lysozyme (Miranker *et al.*, 1993). For non native proteins, partially folded proteins or very large proteins, proton exchange and back exchange can be studied provided the protein can undergo acid proteolysis.

Many of the HDX studies have been carried out on soluble proteins where it is possible to use the method in conjunction with other techniques such as 2D nmr which allows one to monitor a set of individual protons at known locations throughout the protein structure. However, this is not possible in the case of membrane proteins due to their hydrophobic nature, although progress is being made in this area.

The aim of our studies is to investigate the rates at which hydrogens in bacteriorhodopsin undergo isotopic exchange. For bR we would expect the exchangeable protons present in the loop regions to exchange faster than those buried with in the membrane due to solvent. Thus by performing a series of deuteriation expts, we hope to further our knowledge on the structure and folding of bR.

6.2 Results and Discussion

6.2.1 HDX studies using MALDI

The use of MALDI was first investigated for our HDX studies on bR. The requirement was a high quality spectra of bR to set a reference whereby subsequent exchange could be monitored. In collaboration with Brian Wigham at the University of Edinburgh, experimental procedures were as previously described by Rosinke *et al.* (1995) for the analysis of bR using MALDI. The detergent was exchanged by dialysis using standard dialysis membranes with CHAPS, SDS, TRITON X-100 and LDAO. They also found ionic detergents to be detrimental to MALDI analysis as they may interfere with sample crystallisation although the quaternary ammonium salt LDAO appears to be an exception to the rule. In our laboratory, the use of octyl polyethylene (o-POE) allowed us to obtain a spectra for *bO* which is shown in figure 6.1, measuring a mass of 27071 for *bO*

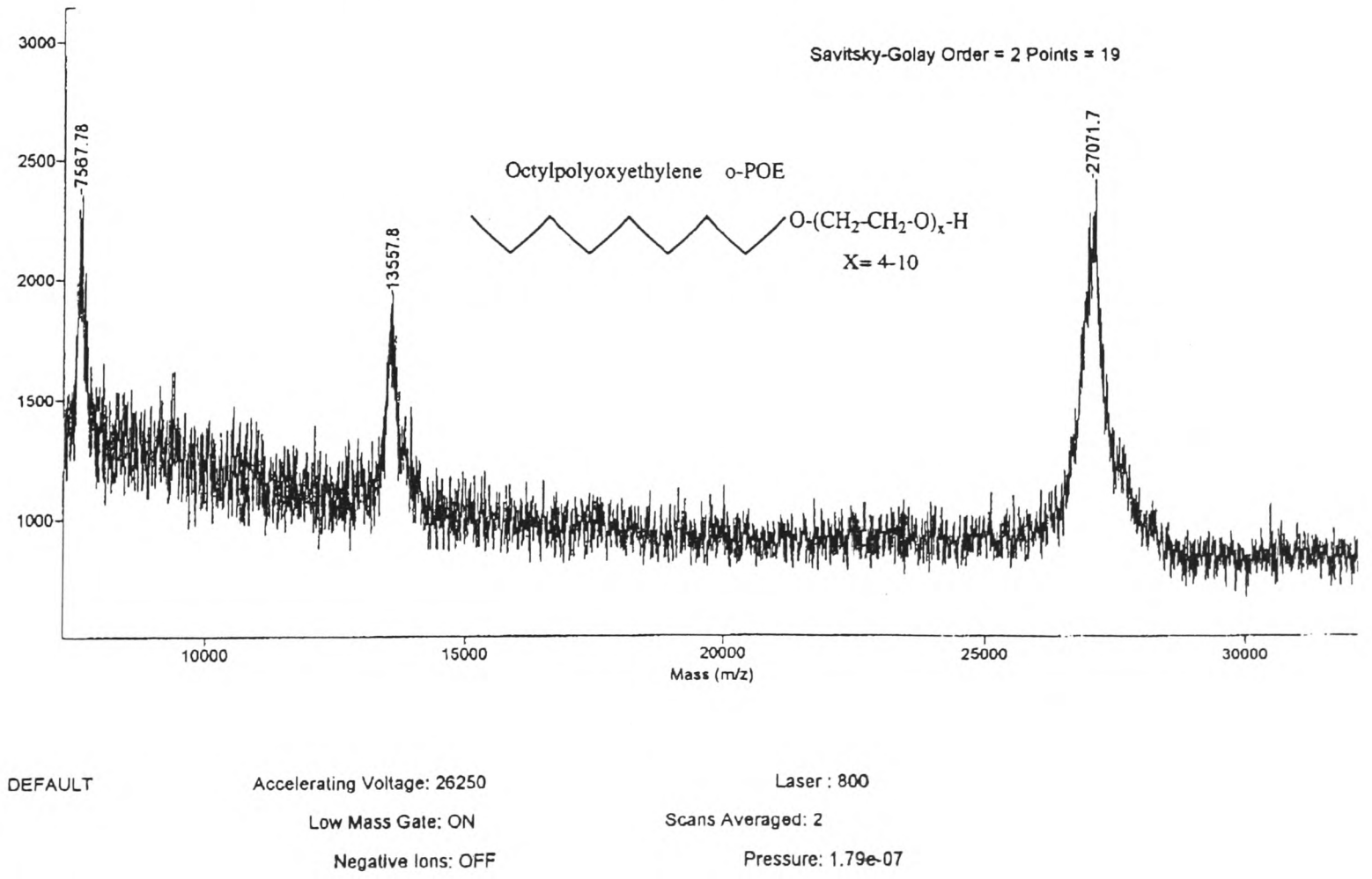


Figure 6.2 MALDI spectra of bR

Unfortunately, the resolution was not sufficient to allow us to continue with further experiments concerning H-D exchange. A higher resolution was technically out of our reach and thus we decided to continue our HDX studies focusing on ESI-MS.

6.2.2 HDX studies using ESI-MS

Samples for ESIMS were prepared using the phase separation method developed by Barnidge and coworkers (2.3.5.1). The deuteration step (2.3.5.3), led to little loss of sample as was monitored by SDS-PAGE (figure 6.4).

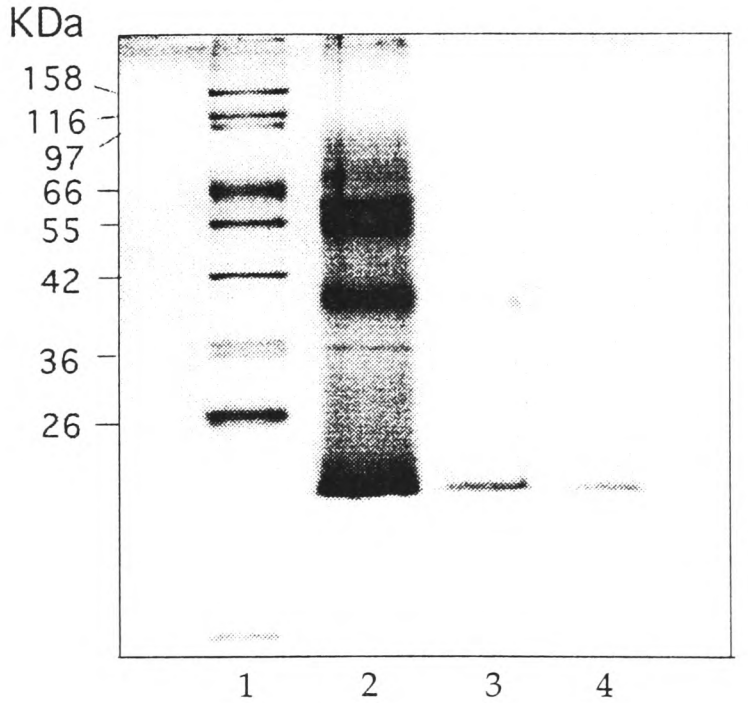


Figure 6.4 Protiated and deuterated samples of bR

Lane 1: Protein markers, broad range as marked. lane 2: Purple membrane 25 μg (degradation shows the presence of 3 bands), lane 3: bR sample in H_20 after phase extraction procedure, approximately 5 mg. lane 4: bR sample in D_2O after phase extraction procedure.

Using conventional electrospray, spectra were obtained for both the protiated and deuterated samples as shown in figure 6.3.

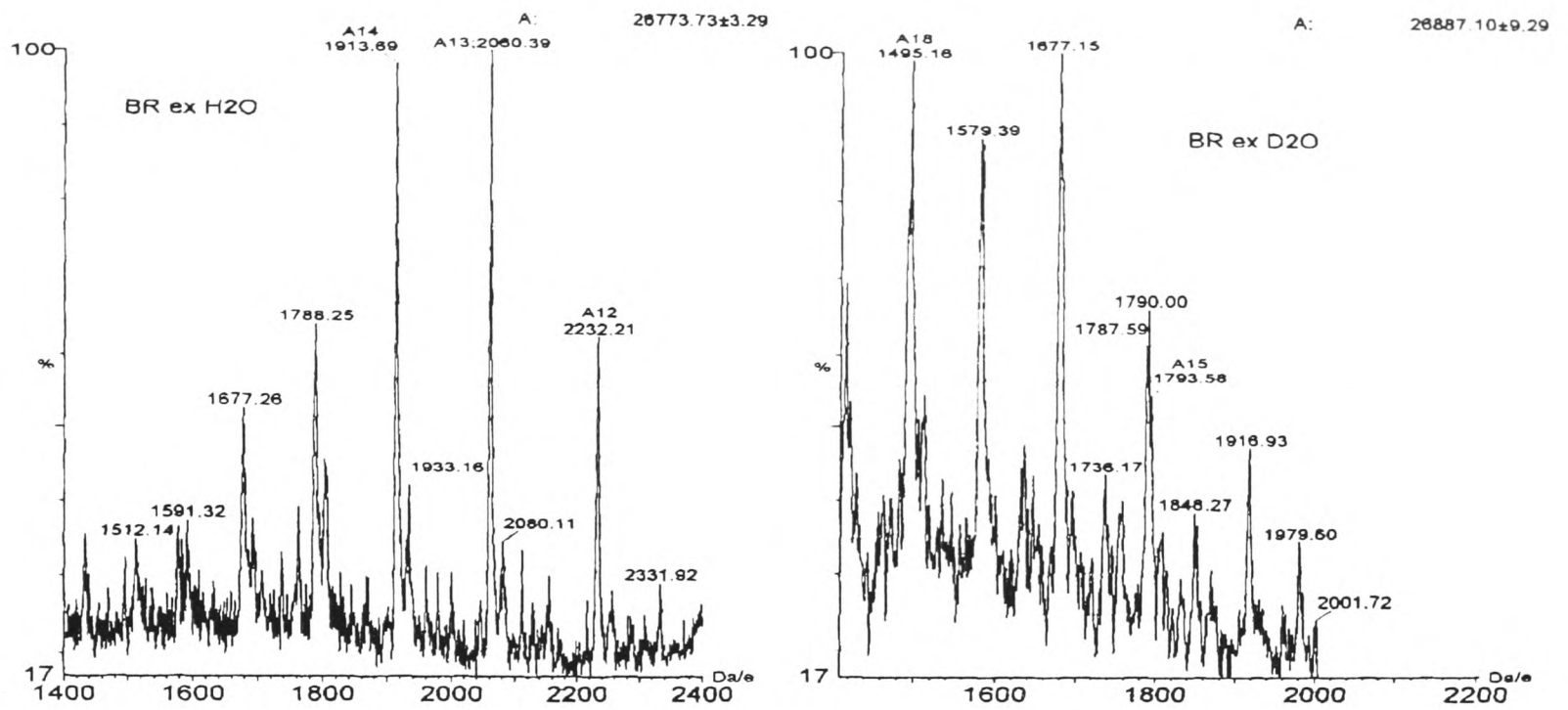


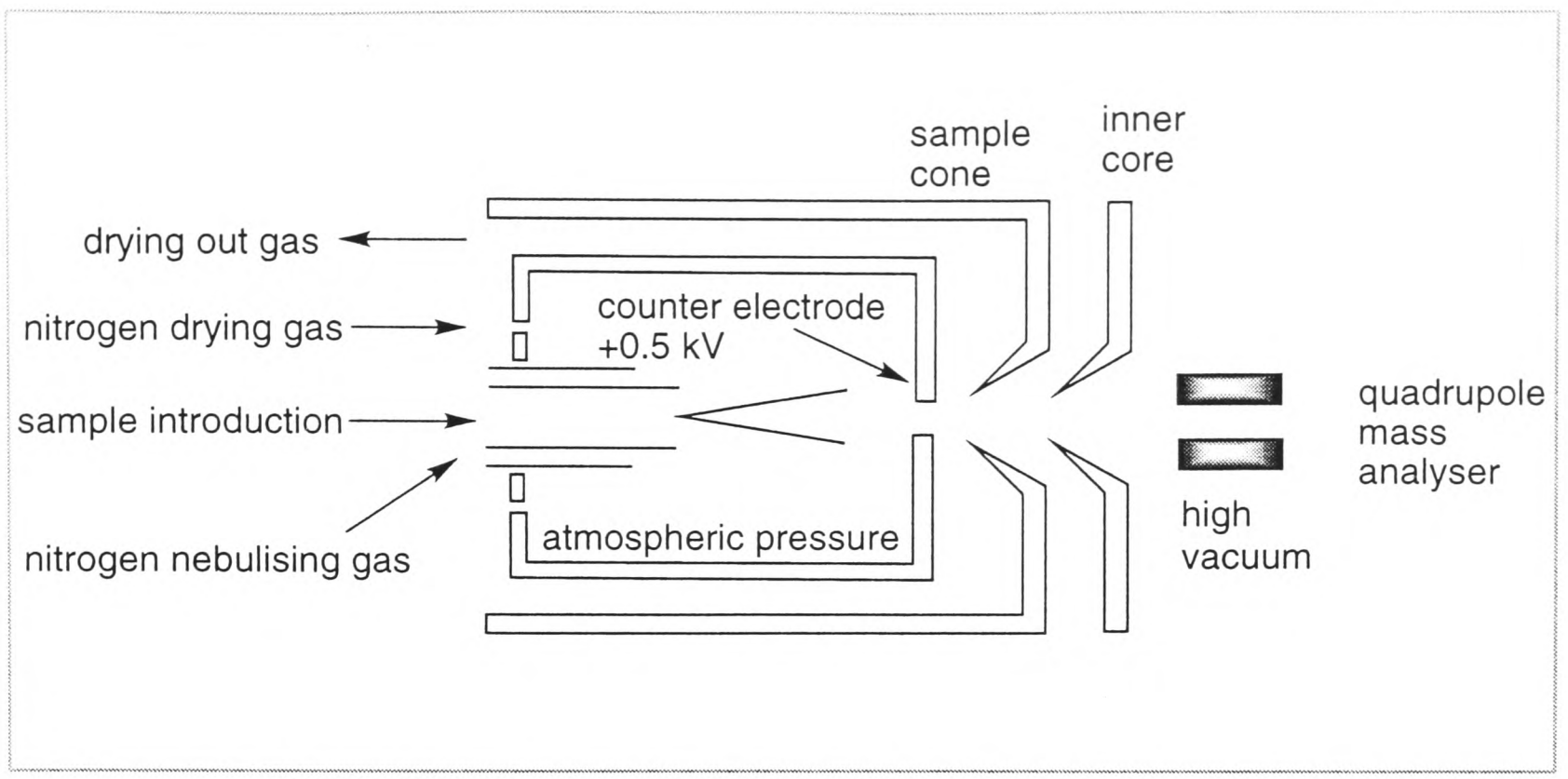
Figure 6.3 Spectra for protiated and deuterated bR

The deuterated spectra shown on the right hand side of figure 6.3 has shifted to a higher mass from 26773.73 to 26887.10. By calculating the difference in mass between the protiated and the deuterated sample (113 units) the amount of deuteration was estimated to be 35%. The total number of exchangeable sites in bR was calculated to be 326 using the amino acid sequence (appendix 1).

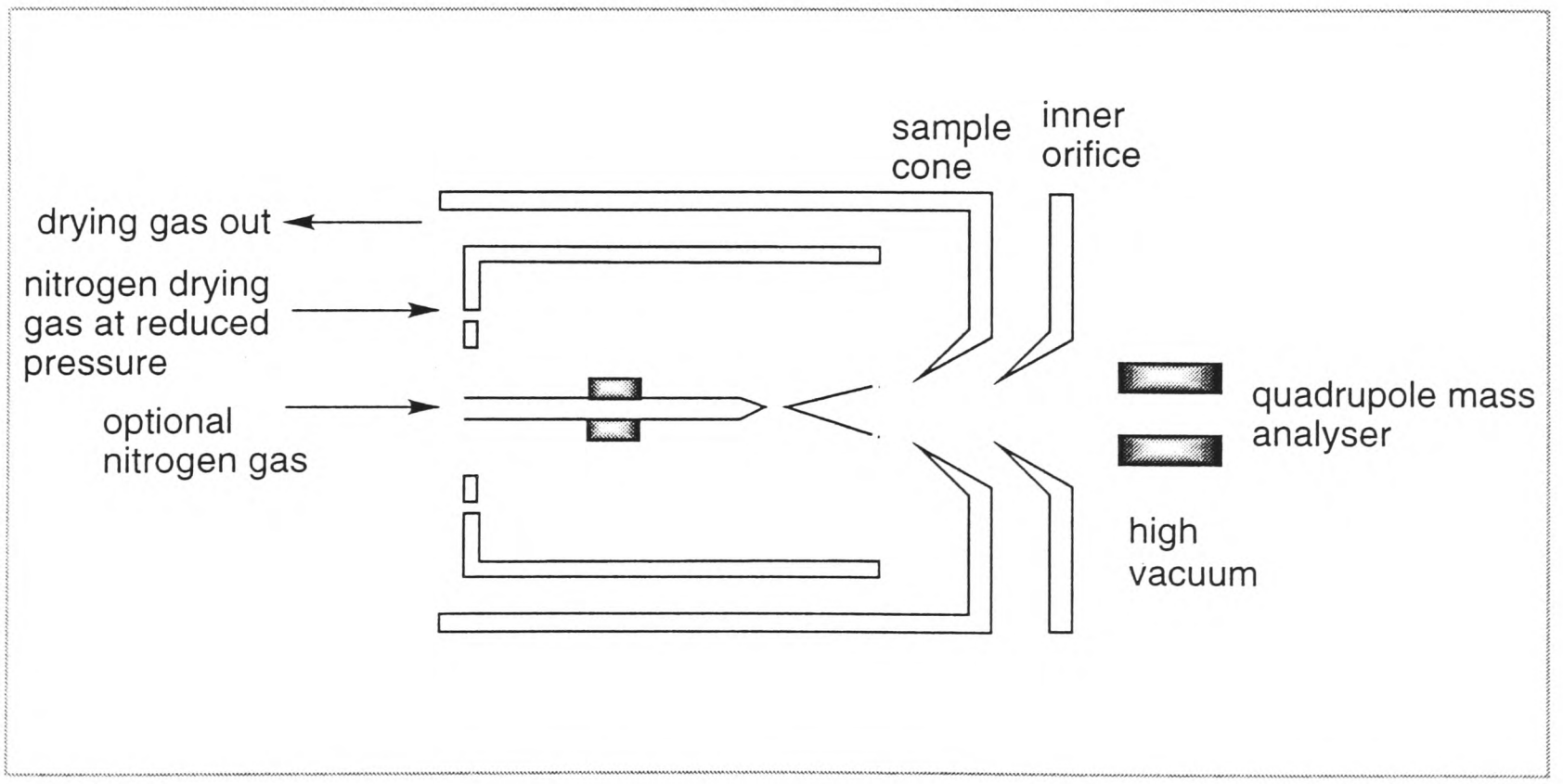
The simultaneous introduction of microelectrospray MS to the department (also known as nanoflow electrospray MS) allowed the possibility of a spectra for bR with a higher resolution. The differences between microspray and conventional electrospray are described in table 6.2 and diagramatically outlined in figure 6.6.

	Conventional	Nanoflow
Capillary	stainless steel capillary inner diameter 100 µm	gold coated borosilicate glass capillary orifice 1-3 µm, length 3 cm
Introduction of sample	via 10 µl rheodyne	<5 µl filled in capillary
Delivery solvent	pressurised for constant flow	not required
Flow rate of sample	1-10 µl/min	1-2 µl/hour
Initial size of droplets	>1 µm in diameter	≤ 200 nm in diameter
Drying and evaporation of droplets	by counter flow of N_2 gas at ~300 L/hr	by counter flow of N_2 gas at ~10 L/hr
Nebulising gas	N2 gas at ~60 L/hr	not essential but back pressure to ensure proper flow of liquid
Distance between capillary source	1 cm in front of counter electrode	2-5 mm in front of sample cone
Voltage required	3 kV on capillary, 0.5 kV on counter electrode, 80 V on sample cone	1 kV on capillary, 80 V on sample cone
Transmission coefficient (total number of detected ions in comparison to the number of peptide molecules introduced	orders of magnitude less	$8x10^{-4}$ for a peptide
Size of emission cone	larger i.e. electric field density smaller	smaller i.e. electric field density greater

Table 6.2 Comparison between conventional and nanoflow electrospray



Conventional electrospray interface



Nanoflow electrospray interface

Figure 6.6 Difference at the interface between conventional and nanoflow electrospray

Advantages of nanoflow electrospray over conventional electrospray

- Smaller initial volume droplets increases desorption/ionisation efficiency (taking into account loses due to detection systems, transmission of quadrupole etc., desorption/ ionisation efficiency and transfer into vacuum has an efficiency of 10%, orders of magnitude higher than that of conventional electrospray), leading to an overall increase in sensitivity.
- Desorption and evaporation from small droplets lead to a higher degree of sample molecule separation and reduces the probability of non-specific clustering.
- Small capillary orifice gives slow flow rate in a stable manner. Measurements are thus less dependent on solvents and buffers. Measurement time also increases.
- Sharpness of pointed capillary and small distance from needle tip to inner orifice mean a reduction in the applied voltage while maintaining the same field strength. Nitrogen gas at reduced pressure also minimises collision-induced dissociations. Non-covalent complexes can be better preserved.
- Being a more stable ion source, nanoflow electrospray can tolerate a salt level (NaCl in aqueous solution) of up to 100 mM^{-1} compared to ~0.25 mM with the conventional source.

Comparisons made between the two available systems as described above provided hope of a higher resolution bR spectra. Thus a sample of bR was prepared using the phase extraction procedure and injected into a nanoflow mass spectrometer. The spectra recorded is shown in figure 6.7.

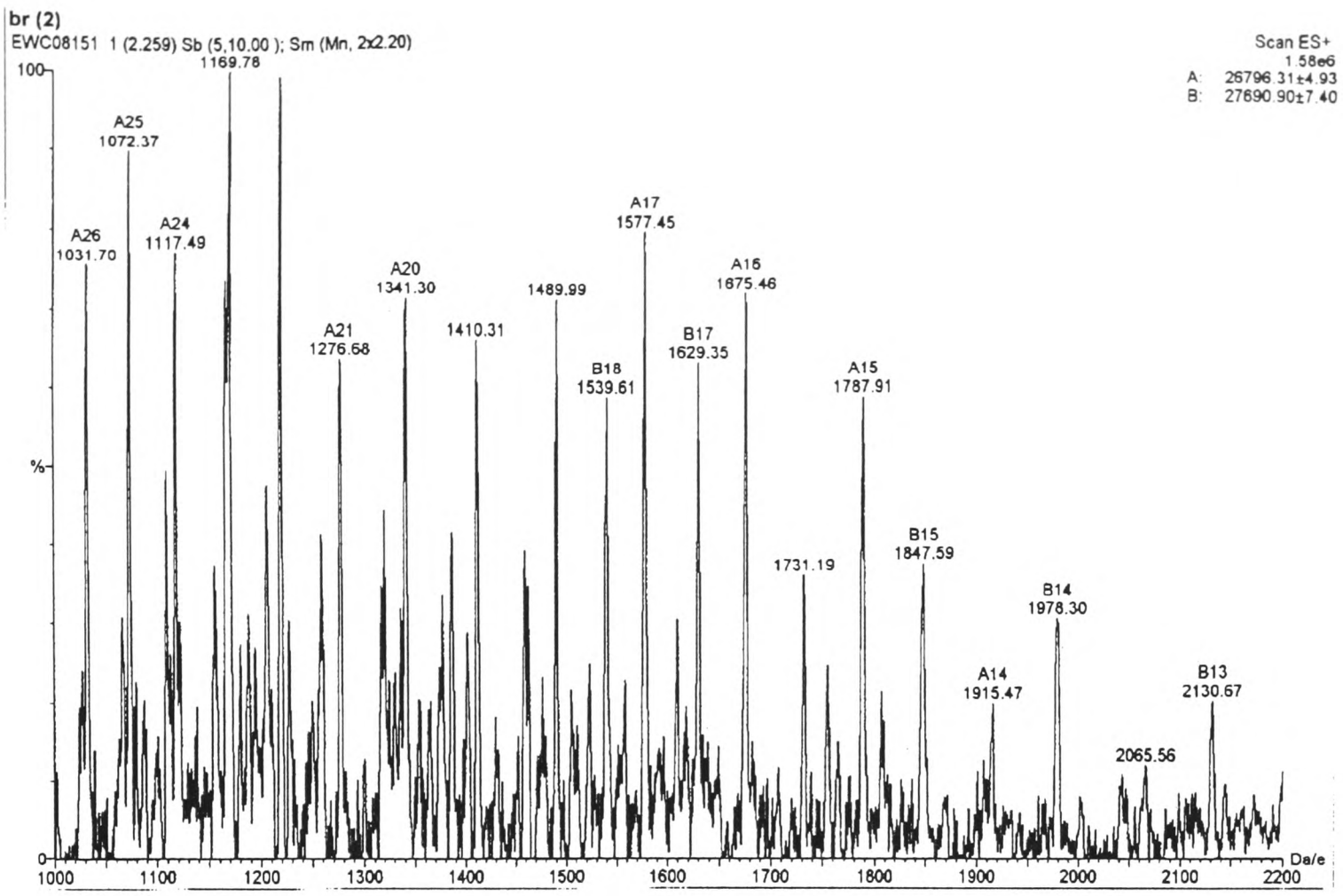


Figure 6.7 Microelectrospray of bR

Although very high resolution was achieved using this method, we encountered a few problems in reproducing the results. This may have arisen due blockage of the very fine needles that were used to spray the sample which, if broken, sprayed the sample but the peaks did not last long enough to be recorded.

We thus decided to try a range of other solvent systems to improve the sample spraying. On researching the literature, we found bR to be soluble in the following solvent systems.

Chloroform:methanol:water:triethylamine 10:10:3.5:0.1 v/v	CMWT
Chloroform:methanol:water:acetic acid 10:10:2.5:0.4 v/v	CMWA
88% formic acid:water 3:7 v/v	FW

bR samples were prepared as previously described and the lower chloroform rich layer removed. The solvent was evaporated using a Büchi Rotavapor R-114. The samples were then resuspended in 200µl of each of the above solvent systems as well as in chloroform only. A standard sample of bR was also preapred for comparison (2.3.5.1). The samples were analysed by ESI-MS and the results are summarised in table 6.3. Spectra obtained for samples CMWA and CMWT are shown in figure 6.8 and figure 6.9.

Samples	Results
Standard	sprayed in the 900m/z - 2300m/z region but not too well. The bR series could however be detected from the many weak and strong peaks obtained. Results were found to be reproducible.
Chloroform	bR did not appear too soluble in this system and did not spray very well.
F/W	Only clusters were observed- close peaks which are a constant number of daltons apart or some peaks that did not produce a series at all.
CMWT	Did not spray well at all and no bR peaks were observed (figure 6.8).
CMWA	Produced excellent results (figure 6.9) . Very clear distinct bR peaks were recorded in the m/z range slightly higher than 'O' 1400m/z - 2500m/z. In the lower m/z region from 900m/z -1400m/z an unknown polymer of 146 Da showed up. Luckily, this is out of the bR range so we can discard it. Results were found to be reproducible in this solvent.

Table 6.3 Investigation of Various Solvent Systems to resuspend bR

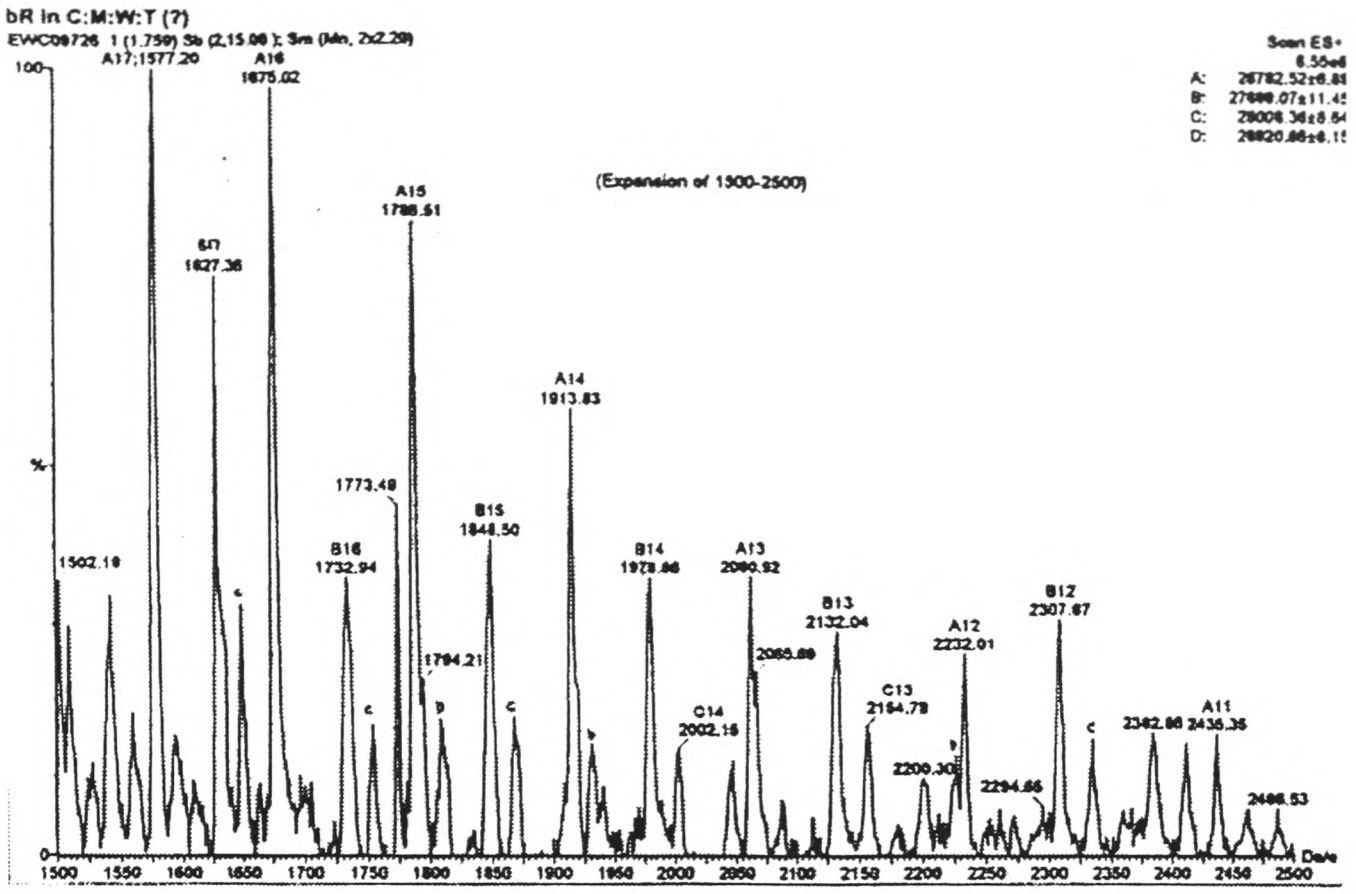


Figure 6.8 Microelectrospray of bR resuspended in CMWA

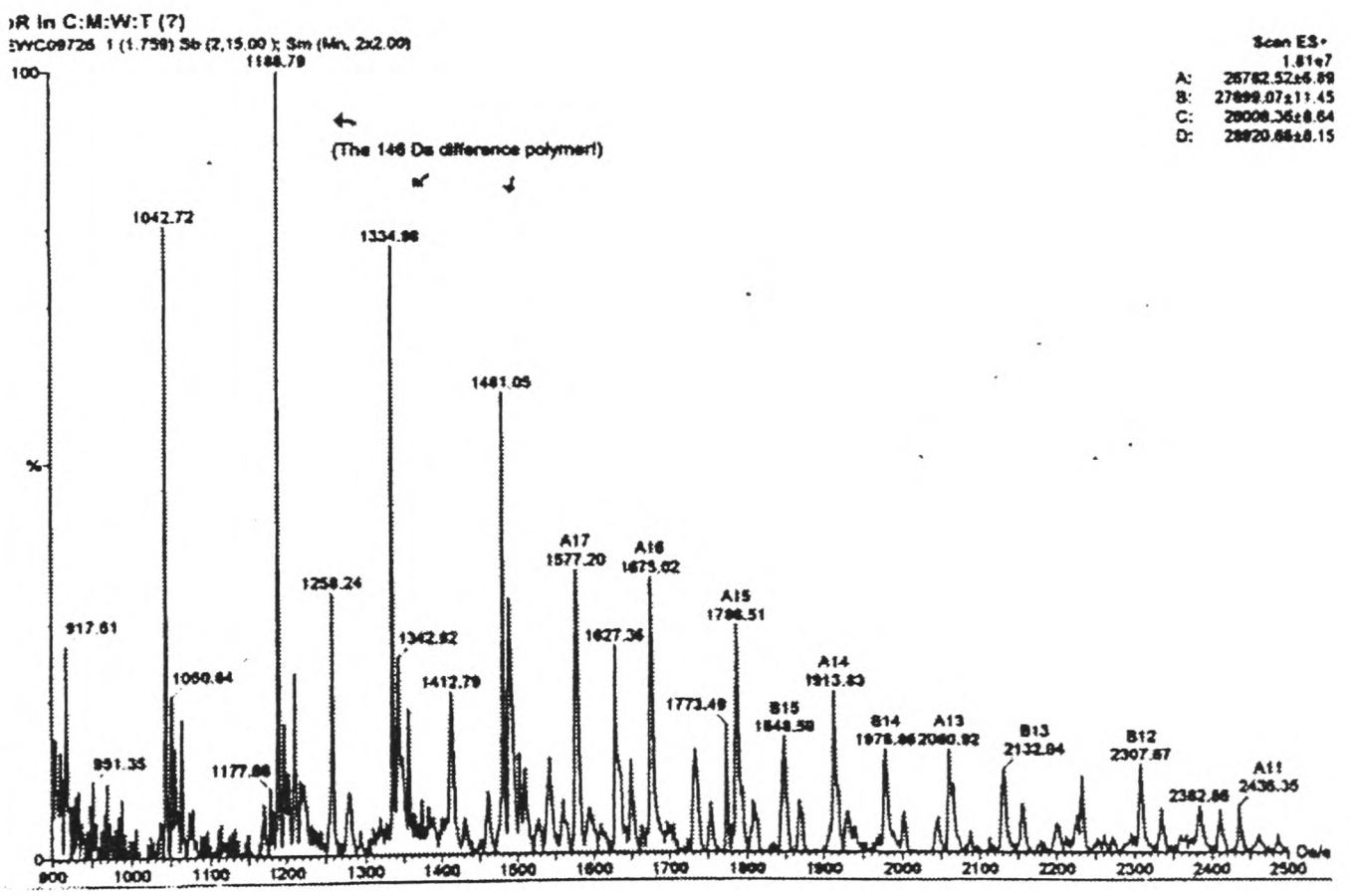


Figure 6.9 Microelectrospray of bR resuspended in CMWT

Thus we have so far optimised our sample preparation conditions as well as the solvent system (CMWA) to achieve good bR spectra.

A variety of experiments were designed to allow us to observe deuteration of bR at various stages during the solvent extraction process in order to investigate H/D exchange during sample preparation. Thus D_2O was used instead of H_2O either in the purple membrane or the extraction procedure or during the resuspending step resulting in six different experiments.

Expt. no.	Purple membrane	Phase extraction	resuspend
1	D_2O	D_2O	D_2O
2	D_2O	D_2O	H_2O
3	D_2O	H_2O	H_2O
4	H_2O	D_2O	D_2O
5	H_2O	H_2O	D_2O
6	H_2O	H_2O	H_2O

Table 6.4 bR deuteration experiments

The results for the 6 experiments in table 6.4 are summarised below and the spectra shown in figure 6.10.

Sample 1 no bR was detected

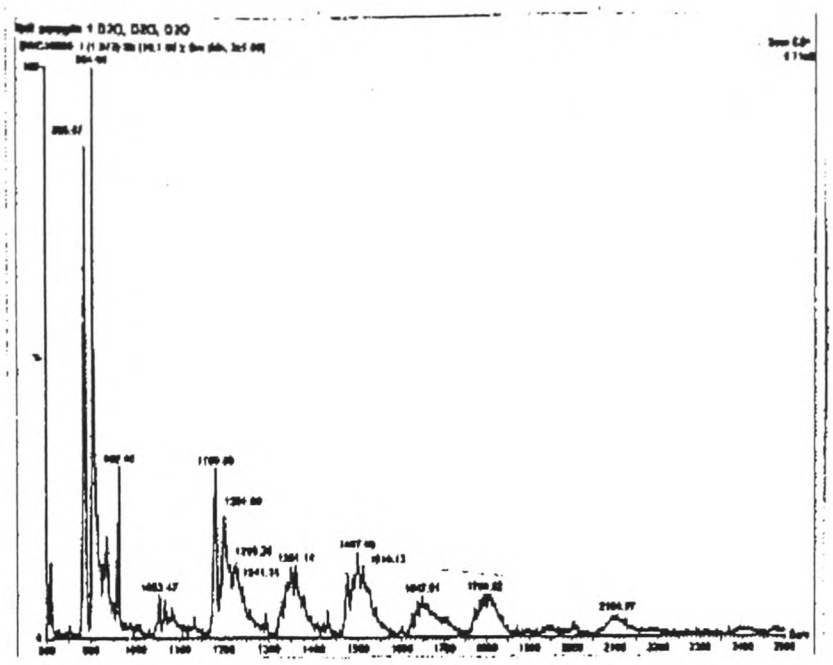
Sample 2 no bR was detected

Sample 3 a polymer was observed and thus a wider range was acquired to investigate this further- see spectra

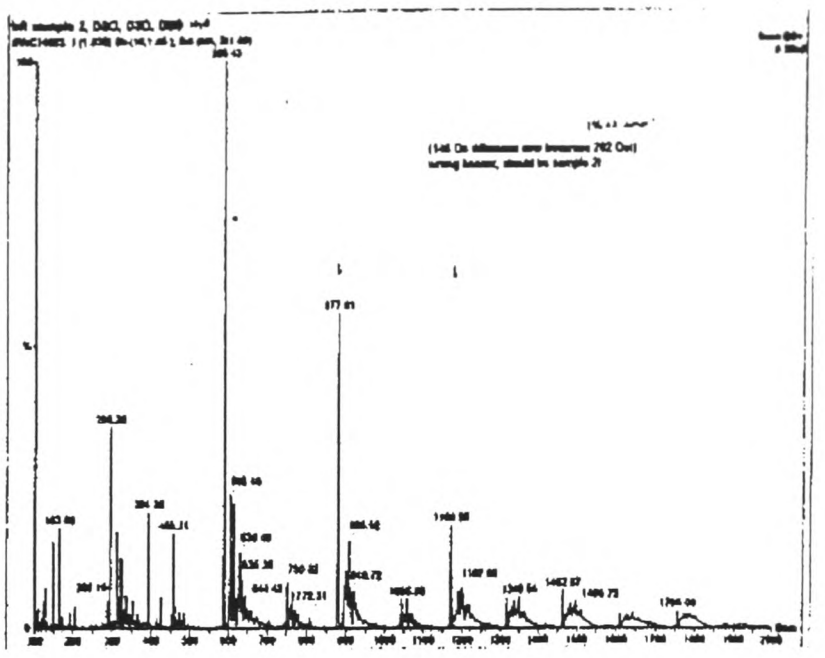
Sample 4 some bR peaks were observed although they do not appear to be convincing - see spectra

Sample 5 no peaks were observed for bR - see spectra

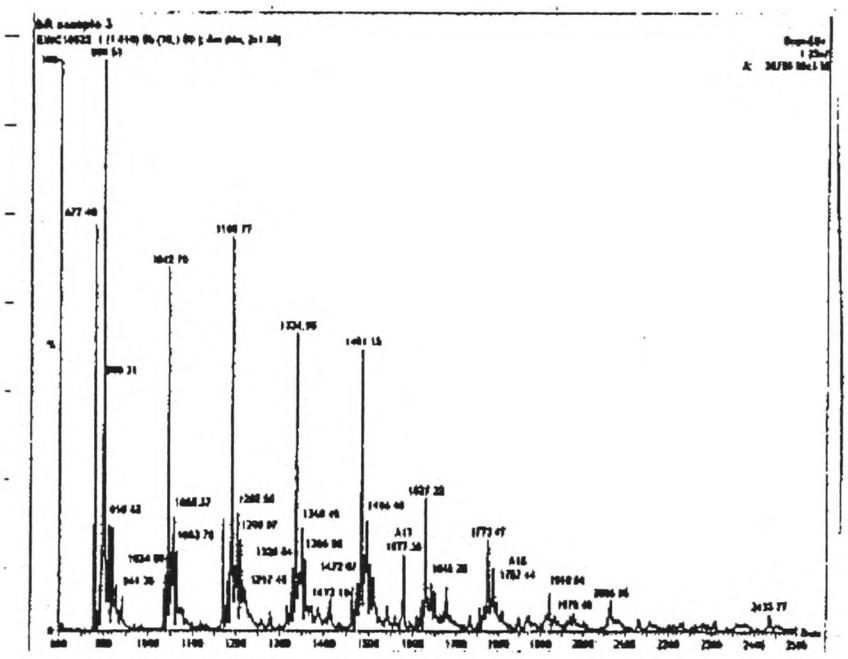
Sample 6 same quality of spectra as shown in figure 6.8.



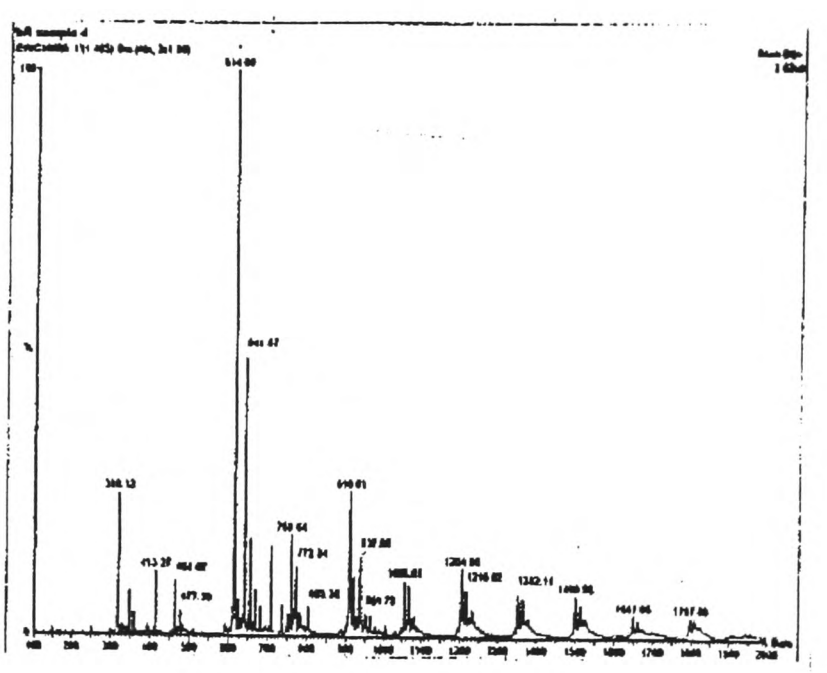
Experiment 1, see table 6.4



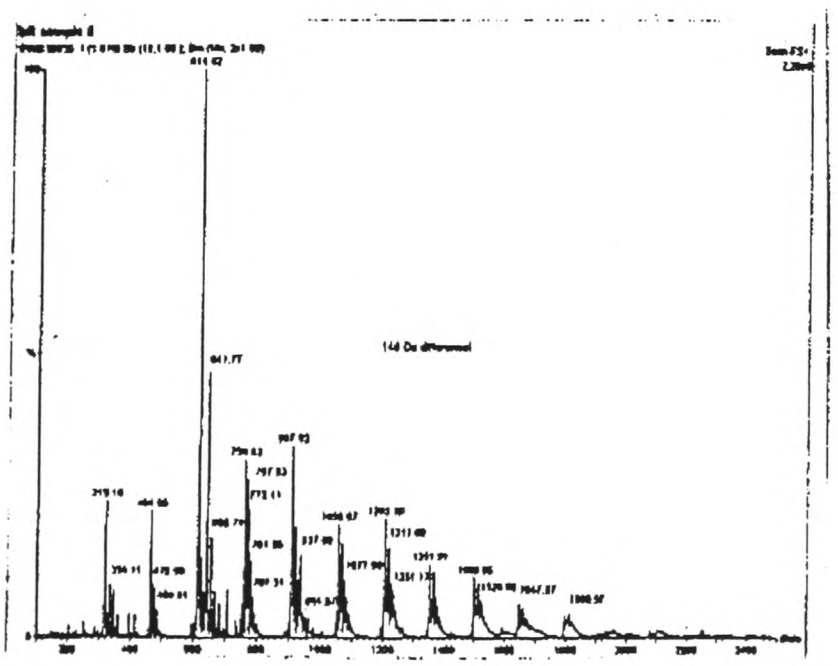
Experiment 2, see table 6.4



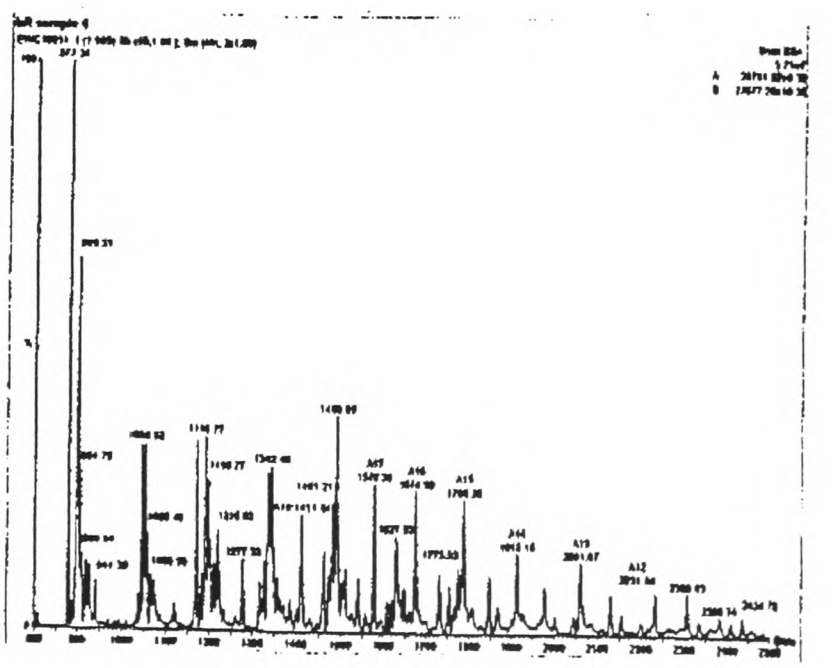
Experiment 3, see table 6.4



Experiment 4, see table 6.4



Experiment 5, see table 6.4



Experiment 6, see table 6.4

Figure 6.10 Deuteration experiments at various stages of sample preparation

It appears from the above results that the samples extracted with deuterated solvents have provided poor spectra with regard to bR. One possibility for this result may be due to the pH of the extraction step where deuterated solvents allow less bR to be extracted in comparison to the protonated solvents.

6.3 Conclusion

We have successfully used MALDI and ESI-MS to obtain spectra for bR. Our studies found ESI to provide a spectra of greater resolution for bR when compared to that obtained by MALDI. It was thus suggested to pursue our studies using ESI-MS. An even higher resolution spectra for bR was obtained using microelectrospray which also offers many other advantages such as greater sensitivity. A few problems were encountered in reproducing the results due to needle blockage and it was thus decided to investigate a range of solvent systems to resuspend bR prior to spraying into the electrospray. It was found that a combination of chloroform:methanol:water:acetic acid 10:10:2.5:0.4 v/v (CMWA) provided the best results. We were also successfully able to deuterate a fully protonated sample of bR as determined by mass spectrometry. Unfortunately, time did not permit the further investigation of HDX during sample preparation.

Chapter 7:
General Conclusions

7.1 A summary of the principal findings of this project

Detailed studies on the folding pathway of bacteriorhodopsin requires access to large amounts of purified protein. Previous expression systems for the *bO* gene in *E. coli* have reported low yields and difficulty with the purification procedures. This thesis investigated alternative expression systems for the *bO* gene in *E.coli* using a variety of plasmids and bacterial cell strains. Our studies found the pPL1 vector originally used by Karnik *et al.* (1987) to express *bO* in *E. coli* to produce 0.3 mg/L of protein (figure 3.7). The pRH1090 vector was previously used by colleagues to express large amounts of protein and was thus investigated for the production of *bO*. Initial studies estimated the expression of approximately 400 mg/L of protein. However, N-terminal sequencing analysis showed the expressed gene not to be *bO* (figure 3.12). Further analysis of the protein sequence suggested that the *bO* gene may have been expressed but with interrupted regions from the kanamycin gene. The toxicity of the *bO* gene limited the use of both pT7-7 and pCYB2 as no expression was detected for either of these two vectors (data not shown). The pET16b vector was chosen as it fuses a hexahistidine tag to the target protein which aids subsequent purification procedures. Expression was detected in only the HMS174 cell strain which was still quite poor (figure 3.19). The *bO* gene appeared to be less toxic to the host when expressed as a fusion protein under the control of a strong promoter. These results were obtained using the pT7IMPACT system although subsequent cleavage of the protein using a chitin column proved difficult (figures 3.26, 3,27, 3.28 and 3.29). The pLEX vector was also investigated as it is noted to be particularly effective in the production of toxic proteins. Results obtained using this procedure produced 1 mg/L of protein.

The temperature and OD_{600} value of the cells were optimised in an attempt to achieve the best expression levels (figures 3.36 and 3.37)

The protein was purified from the pLEX vector using the solvent extraction method (Braiman *et al.*, 1987) and ion exchange chromatography (figure 4.6).

With the production of sufficient amounts of protein, site directed mutagenesis was performed to mutate three membranous proline residues to alanine. By studying the folding kinetics of these mutant proteins using stopped flow fluorimetry, we hoped to investigate whether a slow step in the folding pathway of bR was due to proline isomerisation (figure 1.7). Two of the mutants Pro_{91} and Pro_{186} were successfully expressed and purified to yield 0.6 mg/L and 0.8 mg/L respectively. The expression of Pro_{50} in pLEX however was leaky and proved lethal to the host cells (figure 5.6).$Pro_{91}\rightarrow$Ala and $Pro_{186}\rightarrow$Ala refolded to 66% and 63% respectively. Kinetic studies on both mutants did not differ from results obtained for the wild type protein. This suggests that proline isomerisation is not responsible for the slow step in the folding pathway of bR. More recent studies have indicated the slow step is affected by refolding conditions and lateral pressure the lipids impose on the protein. Further studies on monitoring helix packing by changes in the near-UV signal of monomeric bR may further insight into this area which is currently being investigated.

Separate structural studies using mass spectrometry aimed to study the rates of isotopic exchange of amide and side chain protons in bR. Low resolution results obtained using MALDI-MS (figure 6.1) prompted the investigation of ESI-MS.

Even higher resolution was offered by microelectrospray and this was thus further investigated. By deuterating a sample of purple membrane and measuring the mass against a fully protiated sample, deuteration was estimated to be around 80%. A few problems were encountered in reproducing the results due to needle blockage. A variety of solvent systems were therefore investigated in an aim to prevent this (table 6.3). It was found that a combination of chloroform:methanol:water:acetic acid (10:10:2.5:4) v/v produced the best results (figure 6.8). A variety of deuteration experiments were designed to allow us to observe deuteration of bR at various stages during the solvent extraction procedure in order to investigate HDX during sample preparation. Samples extracted with deuterated solvents produced poor results (figure 6.10). One possibility for this result may be due to the pH of the extraction step where deuterated solvents allow less bR to be extracted in comparison to protiated solvents.

References:

Bai, J.; Qian, M. G.; Liu, Y.; Liang, X.; Lubman, D. M. (1995) Peptide mapping by CNBr degradation on a nitrocellulose membrane with analysis by matrix-assisted laser desorption ionization mass spectrometry. *Anal. Chem.* **67**:1705-1710.

Barnidge, D. R.; Dratz, E.; Sunner, J. (1990) Analysis of hydrophobic protein samples containing detergents by electrspray ionization mass spectrometry. 42nd ASMS Conference on mass spectrometry.

Beavis, R. C.; Chait, B. T. (1990) Rapid, sensitive analysis of protein mixtures by mass-spectrometry. *Proc. Natl. Acad. Sci. USA* **87**:6873.

Betlach, M. C.; Shand, R. F.; Leony, D. M. (1989) Regulation of the bacteriorhodopsin gene of a halophillic archaebacterium. *Can. J. Microbiol.* **35**:134-140.

Blaurock, A. E.; Stoeckenius, W. (1971) Structure of the purple membrane. *Nat. New. Biol* **233**:152-154.

Booth, P. J.; Flitsch, S. L.; Stern, L. J.; Greenhalgh, D. A.; Kim, P. S.; Khorana, H. G. (1995) Intermediates in the folding of the membrane protein bacteriorhodopsin. *Nature, Structural Biology* 2.

Booth, P. J. (1997) Folding α-helical membrane proteins: kinetic studies on bacteriorhodopsin. *Folding & Design* **2**:R85-R92.

Booth, P. J.; Riley, M. L.; Flitsch, S. L.; Templer, R. H.; Farooq, A.; Curran, A. R.; Chadborn, N.; Wright, P. (1997) Evidence that bilayer bending affects membrane protein rigidity. *Biochemistry* **36:**197-203.

Bormann, B.; Engelman, D. M. (1992) Intramembrane helix-helix association in oligomerization and transmembrane signalling. *Annu. Rev. Biophys. Biomol. Struct.* **21**:223-242.

Bowie, J. U. (1997) Helix packing in membrane proteins. *The Journal of Molecular Biology* **272**:780-789.

Braiman, M.; Mathies, R. (1982) Resonanace raman spectra of bacteriorhodopsins primary photoproduct- evidence for a distorted 13-cis retinal chromophore. *Proc. Natl. Acad. Sci. USA* **79**:403-407.

Braiman, M. S.; Stern, L. J.; Chao, B. H.; Khorana, H. G. (1987) Structure-function studies on bacteriorhodopsin IV. purification and renaturation of bacterio-opsin polypeptide expressed in Escherichia coli. *Journal of Biological Chemistry* **262**:9271-9276.

Brandts, J.; Halvorson, H. R.; Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerisation of proline residues. *Biochemistry* **14**:4953-4963.

Buell, G.; Panayotos, N. Mechanism and Practice. (1986) In maximizing gene expression (Reznikoff. W. and Gold L., eds) Butterworth Publishers, Boston, MA.

Burlingame, A. L.; Boyd, R. K.; Gaskell, S. J. (1996) Mass spectrometry. *Anal. Chem.* **66**:634R-683R.

Chait, B. T.; Kent, S. B. H. (1992) Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science* **257**:1885-1894.

Chang, A. C. Y.; Cohen, S. N. (1974) Genome construction between bacterial species in vitro: replication and expression of Staphylococcus plasmid genes in Escherichia coli. *Proc. Natl. Acad. Sci. USA* **71**:1030-1034.

Chen, G. Q.; Choi, I.; Ramachandran, B.; Gouaux, J. E. (1994) Total gene synthesis: Novel single-step and convergent strategies applied to the construction of a 779 base pair bacteriorhodopsin gene. J. *Am. Chem. Soc.* **116**:8799-8800.

Chen, G.-Q.; Gouaux, J. E. (1996) Overexpression of bacterio-opsin in *Escherichia coli* as a water-soluble fusion to maltose binding protein: Efficient regeneration of the fusion protein and selective cleavage with trypsin. *Protein Science* **5** 456-467.

Chong, S.; Shao, Y.; Paulus, H.; Benner, J.; Perler, F. B.; Xu, M. (1996) Protein splicing involving the Saccharomyces cerevisae VMA intein. *J. Mol. Biol.* **271**:22159-22168.

Chung, E. W.; Nettleton, E. J.; Morgan, C. J.; Groβ, M.; Miranker, A.; Radford, S. E.; Dobson, C. M.; Robinson, C. V. (1997) Hydrogen exchange properties of proteins in native and denatured states monitored by mass spectrometry and NMR. *Protein Science* **6:**113-1324.

Clarke, N. D. (1994) A proposed mechanism for the self-splicing. *Proc. Natl. Acad. Sci. USA* **91**:11084-11088.

Cline, S. W.; Doolittle, W. F. (1987) Efficient transfection of the archaebacterium Halobacterium halobium. *J. Bacteriol* **169**:1341-1344.

Cline, S. W.; Lam, W. L.; Charlebois, R. L.; Schalkwyk, L. C.; Doolittle, W. F. (1989) Transformation methods for halophilic archaebacteria. *Can. J. Microbiol.* **35**:148-152.

Cooper, A. A.; Chen, Y. J.; Lindorfer, M. A.; Stevens, T. H. (1993) Protein splicing of the yeast TFP1 intervening protein sequence- a model for self excision. I **12**:2575-2583.

Cooper, A. A.; Stevens, T. H. (1995) Protein splicing: self-splicing of genetically mobile elements at the protein level. *TIBS* **20**:351-356.

Das, A. (1990) Overproduction of proteins in *E. coli* vectors, hosts and strategies. *Methods in Enzymology* **182**:93-112.

Davis, E. O.; Sedgwick, S. G.; Colston, M. J. (1991) Novel structure of the recA locus of mycobacterium tuberculosis implies proccessing of the gene product. *J. Bacteriol.* **173**:5653-5662.

Davis, E. O.; Thangaraj, H. S.; Brooks, P. C.; Colston, M. J. (1994) Evidence for selection of protein introns in the recas of pathogenic mycobacterium. *EMBO J.* **13**:699-703.

Donnelly, D.; Findlay, J. B. C. (1994) Seven-helix receptors: structure and modelling. *Current Opinion in Structural Biology* **4**:582-589.

Dunn, J. J.; Studier, F. W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**:477.

Dunn, R. J.; Hackett, N. R.; McCoy, J. M.; Chao, B. H.; Kimura, K.; Khorana, H. G. (1987) Structure function studies on bacteriorhodopsin I. Expression of the bacterioopsin gene in Escherichia coli. *Journal of Biological Chemistry* **262**:9246-9254.

Eisele, J. L.; Rosenbusch, J. P. (1990) In vitro folding and oligomerization of a membrane protein. *J. Biol. Chem* **265**:10217-10220.

Engelman, D. M.; Steitz, T. A. (1981) The spontaneous insertion of proteins into and across membranes: The helical hairpin hypothesis. *Cell* **23**:411-422.

Englander, S. W.; Kallenbach, N. R. (1984) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Quart. Rev. Biophys.* **16**:521-655.

Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**:64-71.

Fenselau, C.; Cotter, R. J. (1987) Chemical aspects of fast atom bombardment. *Chem. Rev.* **57**:501-512.

Fenselau, C. (1997) MALDI MS and strategies for protein analysis. *Analytical Chemistry News & Features* 661-665.

Ferrando, E.; Schweiger, U.; Oesterhelt, D. (1993) Homologous bacterio-opsin-encoding gene expression via site-specific vector integration. *Gene* **125**:41-47.

Fillingame, R. H. (1976) Purification of the carbodiimide-reactive protein component of the ATP energy-transducing system of Escherichia coli. *J. Mol. Biol.* **251**:6630-6637.

Fitzgerald, M. C.; Siuzdak, G. (1996) Biochemical mass spectrometry: worth the weight? *Chemistry and Biology* **3**:707-715.

Folch, J.; Lees, M. (1951) Proteolipids: a new type of tissue lipoproteins-their isolation from brain. *J. Biol. Chem.* **191**:807-817.

Gennis, R. B. Molecular structure and function. Springer Verlag, New York. (1989)

Ghaim, J. B.; Tsatsos, P. H.; Katasonouri, A.; Mitchell, D. M.; Salcedo-Hernandez, R.; Gennis, R. B. (1997) Matrix-assisted laser desorption ionisation mass spectrometry of membrane proteins: demonstration of a simple method to determine subunit molecular weights of hydrophobic subunits. *Biochimica et Biophysica Acta* **1330**:113-120.

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

Golomb, M.; Chamberlain, M. (1974) Characterization of T7-specific ribonucleic acid polymerase. *The Journal of Biological Chemistry* **249**:2858-2863.

Gouaux, E. (1998) It's not just a phase: crystallization and X-ray structure determination of bacteriorhodopsin in lipidic cubic phases. *Structure* **6**:5-10.

Grigoreiff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin. *Journal of Molecular Biology* **259**:393-421.

Gruner, S. M. (1985) Intrinsic curvature hypothesis for biomembrane lipid composition: a role for nonbilayer lipids. *Proc. Natl. Acad. Sci. USA* **82**:3665-3669.

Gu, H. H.; Xu, J.; Gallager, M.; Dean, G. E. (1993) Peptide splicing in the vacuolar ATPase subunit-a from Candida tropicalis. *J. Biol. Chem.* **268**:7372-7381.

Hackett, N. R.; Stern, L. J.; Chao, B. H.; Kronis, K. A.; Khorana, H. G. (1987) Structure-function studies on bacteriorhodopsin V. Effects of amino acid substitutions in the putative helix F. *The Journal of Molecular Biology* **262**:9277-9284.

Hargrave, P. A. (1991) Seven-helix receptors. *Current Opinion in Structural Biology* **1**:575-581.

Helynck, G.; Luu, B.; Nussbaum, J.-L.; Picken, D.; Skalidis, G.; Trifilieff, E.; Dursselaer, A. V.; Seta, P.; Sandeaux, R.; Gauach, C.; Heitz, F.; Simon, D.; Spach, G. (1983) Brain proteolipids. Isolation, purification and effect on ionic permeability of membranes. *Eur. J. Biochem.* **133**:689-695.

Henderson, R.; Unwin, P. N. T. (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**:28-32.

Henderson, R.; Shotton, D. (1980) Crystallization of purple membrane in three dimensions. *J. Mol. Biol* **139**:99-109.

Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cyro-microscopy. *J. Mol. Biol.* **213**:899-929.

Hildebrandt, V.; Ramezani-Rad, M.; Swida, U.; Wrede, P.; Grzesiek, S.; Primke, M.; Büldt, G. (1989) Genetic transfer of the pigment bacteriorhodopsin into the eukaryote Schizosaccharomyces pombe. *FEBS Letters* **06710**:243 137-140.

Hildebrandt, V.; Polakowski, F.; Büldt, G. Purple fission yeast: overexpression and processing of the pigment bacteriorhodopsin. (1991) *Photochemistry and Photobiology* **54**:1009-1016.

Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* **63**:1193A-1203A.

Hirata, R.; Ohsumi, Y.; Nakano, A.; Kawasaki, H.; Suzuki, K.; Anraku, Y. (1990) Molecular structure of a gene, VMA1 encoding the catalytic subunit of H+-translocating adenosine triphosphate from vacuolar membranes of Saccharomyces cerevisae. *J. Biol. Chem.* **265**:6726-6733.

Hodges, R. A.; Perler, F. B.; Noren, C. J.; Jack, W. E. (1992) Protein splicing removes intervening sequnces in an archae DNA polymerase. *Nucleic Acids Res.* **20**:6253-6157.

Holm, L. (1986) Codon usage and gene expression. *Nucleic Acids Research* **14**:3075-3087.

Hooke, S. D.; Eyles, S. J.; Miranker, A.; Radford, S. E.; Robinson, C. V.; Dobson, C. M. (1995) Cooperative elements in protein folding monitored by electrospray mass spectrometry. J. *Am. Chem. Soc.* **117**:7548-7549.

Hop, C. E. C. A.; Bakhtiar, R. (1997) An Introduction to Electrospray Ionization and matrix-assisted laser desorption/ionization mass spectrometry: essential tools in a modern biotechnology environment. *Biospectroscopy* **23**:4, 259-280.

Houry, W. A.; Scheraga, H. A. (1996) Nature of the unfolded state of ribonuclease A: effect of cis-trans X-Pro peptide bond isomerisation. *Biochemistry* **35**:11719-11733.

Huang, K.-S.; Bayley, H.; Liao, M.-J.; London, E.; Khorana, H. G. (1981) Refolding of an integral membrane protein. *The Journal of Biological Chemistry* **256**:3802-3809.

Ito, H.; Fukuda, Y.; Murata, K.; Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**:163-168.

Jaquinod, M.; Halgand, F.; Caffrey, M.; Saint-Pierre, C.; Gagnon, J.; Fitch, J.; Cusanovich, M.; Forest, E. (1995) Conformational properties of Rhodobacter capsulatus cytochrome c2 wild-type and site-directed mutants using hydrogen/deuterium exchange monitored by electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **9**:1135-1140.

Jay, E.; Rommens, J.; Pomperoycloney, L.; MacKnight, D.; Lutzewallace, C.; Wishart, P.; Harrison, D.; Lui, W. Y.; Asundi, V.; Dawood, M.; Jay, F. (1984) High level expression of a chemically synthesized gene for humaninteferon-gamma using a prokaryotic expression system. *Proc. Natl. Acad. Sci. USA* **81**:2290-2294.

Kane, P.; Yamashiro, C. T.; Wolczyk, D. F.; Neff, N.; Gobel, M.; Stevens, T. H. (1990) Protein splicing converts the yeast TFPI gene product to the 69 kD subunit of the vacuolar H+- adenosine triphosphate. *Science* **250**:651-657.

Karas, M.; Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10 000 Daltons. *Anal. Chem.* **60**:2299.

Karnik, S. S.; Nassal, M.; Doi, T.; Jay, E.; Sgaramella, V.; Khorana, H. G. (1987) Structure-function studies on bacteriorhodopsin II improved expression of the bacterio-opsin gene in Escherichia coli. *Journal of Biological Chemistry* **262**:9255-9263.

Karnik, S.; Doi, T.; Molday, R.; Khorana, H. G. (1990) Expression of the archaebacterial bacterio-opsin gene with and without signal sequences in *Esherichia coli:. Proc. Natl. Acad. Sci. USA* **87**:8955-8959.

Katta, V.; Chait, B. T. (1991) Conformational changes in proteins probed by hydrogen exchange electrospray ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.* **5**:214.

Keller, S. L.; Bezrukov, S. M.; Gruner, S. M.; Tate, M. W.; Vodyanoy, I.; Parsegian, V. A. (1993) Probability of Alamthicin conductance states varies with nonlamellar tendency of bilayer phospholipids. *Biophys. J.* **65**:23-27.

Khorana, H. G. (1993) Bacteriorhodopsin, a membrane protein that uses light to translocate protons. *J. Biol. Chem.* **263**:7439-7442.

Kim, P. S.; Baldwin, R. L. (1982) Influence of charge on the rate of amide proton exchange. *Biochemistry* **21**:1-5.

Kimura, K.; Mason, T. L.; Khorana, H. G. (1982) Immunological probes for bacteriorhodopsin. Identification of three distinct antogenic sites on the cytoplasmic surface. *J. Biol. Chem.* **257**:2859-2867.

Kimura, Y.; Vassylyev, D. G.; Miyazawa, A.; Kidera, A.; Matsushima, M.; Mitsuoka, K.; Murata, K.; Hirai, T.; Fujiyoshi, Y. (1997) Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* **389**:206-211.

Koyama, K.; Ymaguchi, N.; Miyasaka, T. (1994) Antibody-mediated bacteriorhodopsin orientation for molecular device architectures. *Science* **265**:762-765.

Krebs, M. P.; Hauss, T.; Heyn, M. P.; RajBhandary, U. L.; Khorana, H. G. (1991) Expression of the bacterioopsin gene in Halobacterium halobium using a multicopy plasmid. *Proc. Natl. Acad. Sci USA* **88**:859-863.

Krebs, M. P.; Khorana, H. G. (1993) Mechanism of light dependent proton translocation by bacteriorhodopsin. *Journal of Bacteriology* **175**:1555-1560.

Kurland, C.; Gallant, J. (1996) Errors of heterologus protein expression. *Current Opinion in Biotechnology* 7:489-493.

Laemmeli, U. K. (1970) Cleavage of the structural proteins during the assembly of the head of the bacteriophage T7. *Nature* **227**:680-685.

Lam, W. L.; Doolittle, W. F. (1989) Shuttle vectors for the archaebacterium Halobacterium volcanii. *Proc. Natl. Acad. Sci USA* **86**:5478-5482.

Landau, E. M.; Rosenbusch, J. P. (1996) Lipidic cubic phases: a novel concept for the cystallisation of membrane proteins. *Proc. Natl. Acad. Sci. USA* **93**:14532-14535.

Landau, E. H.; Rummel, G.; Cowan-Jacob, S. W.; Rosenbusch, J. P. (1997) Crystallisation of a polar protein and small molecules from the aqueous compartment of lipidic cubic phases. *The Journal of Physical Chemistry* **101**:1935-1937.

Lanyi, J. K. (1995) Bacteriorhodopsin as a model for proton pumps. *Nature* **375**:461-463.

LaVallie, E. R.; DiBlasio, E. A.; Kovacic, S.; Grant, K. L.; Schendel, P. F.; McCoy, J. M. A (1992) Thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology* **11**:187-193.

Leucke, H.; Richter, H. T.; Lanyi, J. K. (1998) Proton transfer pathways in bacteriorhodposin at 2.3 angstrom resolution. *Science* **280**:1934-1937.

Lewis, S.; Korsmeyer, K. K.; Correia, M. A. (1993) Matrix-assisted laser desorption mass spectrometry of cytochromes-P450. *Rapid Commun. Mass Spectrum* **7**:16.

Lindblom, G.; Brentel, I.; Sjölund, M.; Wikander, G.; Wieslander, A. (1986) Phase equilibria of membrane lipids from Acholeplasma laidlawii: Importance of a single lipid forming nonlamellar phases. *Biochemistry* **25**:7502-7510.

Lindblom, G.; Rilfors, L. (1989) Cubic phases and isotropic structures formed by membrane lipids- possible biological relevance. *Biochimica et BIophysica Acta* **988**:221-256.

London, E.; Khorana, H. G. (1982) Denaturation and renatuartion of bacteriorhodopsin in detergents and lipid-detergent mixtures. *The Journal of Biological Chemistry* **257**:7003-7011.

Lui, Y.; Smith, D. L. (1994) Probing high order structure of proteins by fast atom bombardment mass spectrometry. *J. Am. Chem. Soc. Mass Spectom.* 5.

Luzzati, V. (1997) Biological significance of lipid polymorphism: the cubic phases. *Current Opinion in Structural Biology* **7**:661-668.

Maier, C. S.; Kim, O.-H.; Deinzer, M. L. (1997) Conformational properties of the A-state of cytochrome c studied by hydogen/deuterium exchange and electrospray mass spectrometry. *Anal. Biochem.* **252**:127-135.

Maina, C. V.; Riggs, P. D.; Grandea, A. G.; Slatko, B. E.; Moran, L. S.; Tagliamonte, J. A.; McReynolds, L. A.; Guan, C. D. (1988) An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* 74:56-373.

Mann, M.; Wilm, M. (1995) Electrospray mass spectrometry for protein characterisation. *Trends Biochem. Sci.* **20**:219-224.

Matthews, C. R. (1993) Pathways of protein folding. *Annual Review of Biochemistry* **62**:653-83.

Mayr, C. M.; Odefey, C.; Schutkowski, M.; Schmid, F. X. (1996) Kinetic analysis of the unfolding and refolding of ribonuclease T1 by a stopped-flow double-mixing technique. *Biochemistry* **35**:5550-5561.

Miercke, L. J. W.; Betlach, M. C.; Mitra, A. K.; Shand, R. F.; Fong, S. K.; Stroud, R. M. (1991) Wild-type and mutant bacteriorhodopsin D85N, D96N AND R82Q: Purification to homogeniety, pH dependence of pumping and electron diffraction. *Biochemistry* **30**:3088-3098.

Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. A.; Dobson, C. M. (1993) Detection of transient protein folding population by mass spectrometry. *Science* **262**:896-899.

Mogi, T.; Stern, L. J.; Chao, B. H.; Khorana, H. G. (1989) Structure-function studies on bacteriorhodopsin. VIII. Substitutions of the membrane-embedded prolines 50, 91, and 186: The effects are determined by the substituting amino acids. *The Journal of Biological Chemistry* **264**:14192-14196.

Nall, B. Mechanisms of protein folding (edited by Pain R. H.) Oxford University Press, New York. (1994)

Nassal, M.; Mogi, T.; Karnik, S. S.; Khorana, H. G. (1987) Structure-function studies on bacteriorhodopsin III. total synthesis of a gene for bacterio-opsin and its expression in Escherichia coli. *The Journal of Biological Chemistry* **262**:9264-9270.

Neubert, T. A.; Walsh, K. A.; Hurley, J. B.; Johnson, R. S. (1997) Monitoring calcium-induced conformational changes in recoverin by mass spectrometry. *Protein Science* **6**:843-850.

Ni, B.; Chang, M.; Duschl, A.; Lanyi, J.; Needleman, R. (1990) An efficient system for the synthesis of bacteriorhodopsin in Halobacterium halobium. *Gene* **90**:169-172.

Odefey, C.; Mayr, C. M.; Schmid, F. X. (1995) Non-propyl cis-trans peptide bond isomerisation as a rate-determining step in protein unfolding and refolding. *J. Mol. Biol.* **245**:69-78.

Oesterhelt, D.; Stoeckenius, W. (1971) Rhodopsin-like protein from the purple membrane of Halobacterium halobium. *Nat. New Biol.* **233**:149-152.

Oesterhelt, D. (1982) Reconstitution of the retinal proteins bacteriorhodopsin and halorhodopsin. *Methods in Enzymology* **88**:10-17.

Paulsen, H.; Rumler, U.; Rudiger, W. (1990) Reconstitution of pigment containing complexes from light-harvesting chlorophyll a/b binding protein overexpressed in Escherichia coli. *Planta* **181**:204-211.

Pebay-Peyroula, E.; Rummel, G.; Rosenbusch, J. P.; Landau, E. M. (1997) X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* **277**:1676-1681.

Perler, F. B.; Gomb, D. G.; Jack, W. E.; Moran, L. S.; Qiang, B.; Kucera, R. B.; Benner, J.; Slatko, B. E.; Nwanko, D. O.; Hempstead, S. K.; Carlow, C. K. S.; Jannasch, H. (1992) Intervening sequences in an archae DNA-polymerase. *Gene* **89**:5577-5581.

Perler, F. B.; Davis, E. O.; Dean, G. E.; Gimble, F. S.; Jack, W. E.; Neff, N.; Noren, C. J.; Thorner, J.; Belfort, M. (1994) Protein splicing elements: inteins and exteins-a definition of terms and recommended nomenclature. *Nucleic Acids Res.* **22**:1125-1127.

Pietrokovski, S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Science* **3**:2340-2350.

Pompejus, M.; Friedrich, K.; Teufel, M.; Fritz, H.-J. (1993) High-yield production of bacteriorhodopsin via expression of a synthetic gene in Escherichia coli. *European Journal of Biochemistry* **211**:27-35.

Popot, J.-L.; Engelman, D. M. (1990) Membrane protein folding and oligomerization: The two stage model. *Biochemistry* **29**:4031-4037.

Pouwels, P. H. (1991) Essential Molecular biology: a practical approach. T. A. Brown ed. Oxford University Press, Oxford. 179-239.

Riley, M. L.; Wallace, B. A.; Flitsch, S. L.; Booth, P. J. (1997) Slow a helix formation during folding of a membrane protein. *Biochemistry* **36**:192-196.

Robinson, C. V.; Grob, M.; Eyles, S. J.; Ewbank, J. J.; Mayhew, M.; Hartl, F. U.; Dobson, C. M.; Radford, S. E. (1994) Conformation of GroEL bound a-lactalbumin probed by mass spectrometry. *Nature* **372**:646-372.

Robinson, C. V.; Radford, S. E. (1995) Weighing the evidence for structure electrospray ionisation mass spectrometry of proteins. *Structure* **3**:861-865.

Rosenbusch, J. P. (1990) The critical role of detergents in the crystallization of membrane proteins. *J. Struct. Biol.* **104**:134-138.

Rosinke, B.; Strupat, K.; Hillenkamp, F.; Rosenbusch, J.; Dencher, N.; Krüger, U.; Galla, H.-J. (1995) Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) of membrane proteins and non-covalent complexes. *J. Mass Spectrom.* **30**:1462-1468.

Saraste, M.; Walker, J. E. (1998) Membrane proteins; channels, pumps and charge separators. Editorial overview. *Current Opinion in Structural Biology* **8**:477-479.

Schertler, G. F. X.; Villa, C.; Henderson, R. (1993) Projection structure of rhodopsin. *Nature* **362**:770-772.

Schey, K. L.; Papac, I. D.; Knapp, D. R.; Crouch, R. K. (1992) Matrix-assisted laser desorption mass-spectrometry of rhodopsin and bacteriorhodopsin. *Biophys. J.* 63 1240.

Schmid, F. X. (1993) Propyl isomerase: enzymatic catalysis of slow protein-folding reactions. *Annu. Rev. Biophys. Biomol. Struct.* **22**:123-143.

Schneider, C.; Newman, R. A.; Sutherland, D. R.; Asser, U.; Greaves, M. F. A (1982) One-step purification of membrane proteins using a high efficiency immunomatrix. *J. Biol. Chem.* .**257**:10766-10769.

Shand, R. F.; Miercke, L. J. W.; Mitra, A. K.; Fong, S. K.; Stroud, R. M.; Betlach, M. C. (1991) Wild-type and mutant bacterioopsins D85N, D96N AND R82Q: High-level expression in Escherichia coli. *Biochemistry* **30**:3082-3088.

Shao, Y.; Kent, S. B. H. (1997) Protein splicing: occurence, mechanism and related phenomena. *Chemistry and Biology* **4**:187-194.

Siuzdak, G. (1994) The emergence of mass spectrometry in biochemical research. *Proc. Natl. Acad. Sci.* **91**:11290-11297.

Smith, R. D.; Loo, J.; Edmunds, C. G.; Baringa, C. J.; Udseth, H. R. (1990) New developments in biochemical mass spectrometry- electrospray ionisation mass spectrometry. *Anal. Chem* **62**:882-899.

Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* 60-89.

Stults, J. T. (1995) Matrix-assisted laser desoprtion/ionization mass spectrometry (MALDI-MS). *Current opionion in structural biology* **5**:691-698.

Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. (1988) Protein and polymer analyses up to m12 100,000 by laser ionization time-of high mass spectrometry. *Rapid commun mass spectom* **2**:151-153.

Veeraraghavan, S. (1996) Autocatalyzed protein folding. *Biochemistry* **35**:10601-10607.

Vorm, O.; Mann, M. (1994) Improved mass accuracy in matrix-assisted laser desorption ionization time-of-flight mass spectrometry of peptides. *J. Am. Soc. Mass Spectrom.* **5**:955-958.

Vorm, O.; Roepstorff, P.; Mann, M. (1994) Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Anal. Chem.* **66**:3281-3287.

Wallace, C. J. A. (1993) The curious case of protein splicing: mechanistic insights suggested by protein semisynthesis. *Protein Science* **2**:697-705.

Weickert, M. J.; Doherty, D. H.; Best, E. A.; Olins, P. O. (1996) Optimization of heterologous protein production in *Escherichia coli*. *Current Opinion in Biotechnology* **7**:494-499.

Wilm, M. S.; Mann, M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? (1994) *International Journal of Mass Spectrometry and Ion Processes* **136**:167-180.

Woodward, C.; Simon, I.; Tuschsen, E. (1982) Hydrogen exchange and the dynamic structures of proteins. *Mol. Cell Biochem.* **48**:135-160.

Xu, M.; Southworth, M. W.; Mersha, F. B.; Hornstra, L. J.; Perler, F. B. (1993) In vitro protein splicing of branched precursor and the identification of a branched intermediate. *Cell* **75**:1371-1377.

Xu, M. Q.; Comb, D. G.; Paulus, H.; Noren, C. J.; Shao, Y.; Perler, F. B. (1994) Protein splicing- an analysis of the branched intermediate and its resolution by succinamide formation. *EMBO J.* **13**:5517-5522.

Xu, M. Q.; Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15**:5146-5153.

Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**:397-408.

Yi, Q.; Baker, D. (1996) Direct evidence for a two-stable protein unfolding transition from hydrogen exchange and mass spectrometry. *Protein Science* **5**:1060-1066.

Zhang, Z.; Post, C. B.; Smith, D. L. (1996) Amide hydrogen exchange determined by mass spectrometry. Application to rabbit muscle aldolase. *Biochemistry* **35**:779-791.

Appendices:

Appendix 1: Amino acid sequence of bacteriorhodopsin

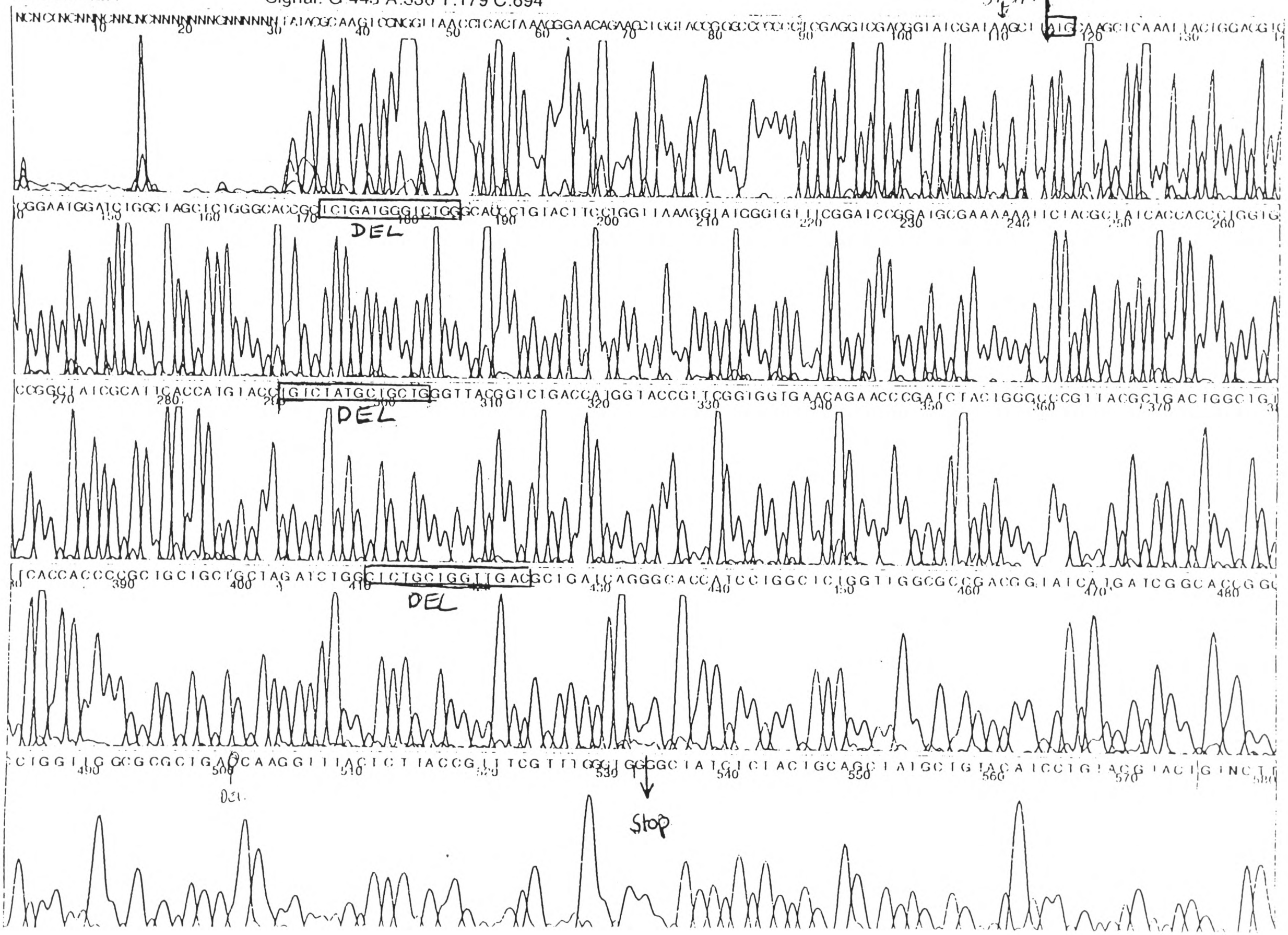
```
BO GENE SEQ [6 to 756] -> 1-phase Translation

DNA  sequence    761 b.p.   AGCTTATGCAAG ... TCTTGATGAGCT   linear

6    /    1                                  36   /   11
ATG CAA GCT CAA ATT ACT GGA CGT CCG GAA TGG ATC TGG CTA GCT CTG GGC ACC GCT CTG
met gln ala gln ile thr gly arg pro glu trp ile trp leu ala leu gly thr ala leu
66   /   21                                 96   /   31
ATG GGT CTG GGC ACC CTG TAC TTC CTG GTT AAA GGT ATG GGT GTT TCG GAT CCG GAT GCG
met gly leu gly thr leu tyr phe leu val lys gly met gly val ser asp pro asp ala
126  /   41                                 156  /   51
AAA AAA TTC TAC GCT ATC ACC ACC CTG GTG CCG GCT ATC GCA TTC ACC ATG TAC CTG TCT
lys lys phe tyr ala ile thr thr leu val pro ala ile ala phe thr met tyr leu ser
186  /   61                                 216  /   71
ATG CTG CTG GGT TAC GGT CTG ACC ATG GTA CCG TTC GGT GGT GAA CAG AAC CCG ATC TAC
met leu leu gly tyr gly leu thr met val pro phe gly gly glu gln asn pro ile tyr
246  /   81                                 276  /   91
TGG GCC CGT TAC GCT GAC TGG CTG TTC ACC ACC CCG CTG CTG CTG GTA GAT CTG GCT CTG
trp ala arg tyr ala asp trp leu phe thr thr pro leu leu leu leu asp leu ala leu
306  /  101                                336  /  111
CTG GTT GAC GCT GAT CAG GGC ACC ATC CTG GCT CTG GTT GGC GCC GAC GGT ATC ATG ATC
leu val asp ala asp gln gly thr ile leu ala leu val gly ala asp gly ile met ile
366  /  121                                396  /  131
GGC ACC GGC CTG GTT GGC GCG CTG ACC AAG GTT TAC TCT TAC CGT TTC GTT TGG TGG GCT
gly thr gly leu val gly ala leu thr lys val tyr ser tyr arg phe val trp trp ala
426  /  141                                456  /  151
ATC TCT ACT GCA GCT ATG CTC TAC ATC CTG TAC GTA CTG TTC TTC GGT TTC ACC TCT AAA
ile ser thr ala ala met leu tyr ile leu tyr val leu phe phe gly phe thr ser lys
486  /  161                                516  /  171
GCT GAA AGC ATG CGT CCG GAA GTT GCG TCG ACC TTC AAA GTA CTG CGT AAC GTT ACC GTT
ala glu ser met arg pro glu val ala ser thr phe lys val leu arg asn val thr val
546  /  181                                576  /  191
GTT CTG TGG TCC GCT TAC CCA GTT GTT TGG CTG ATC GGT TCT GAA GGT GCC GGC ATT GTT
val leu trp ser ala tyr pro val val trp leu ile gly ser glu gly ala gly ile val
606  /  201                                636  /  211
CCG CTG AAT ATT GAA ACC CTG CTG TTC ATG GTT CTA GAC GTT TCT GCT AAA GTT GGT TTC
pro leu asn ile glu thr leu leu phe met val leu asp val ser ala lys val gly phe
666  /  221                                696  /  231
GGT CTG ATC CTG CTG CGT TCT CGA GCT ATC TTC GGT GAA GCT GAA GCT CCG GAA CCG TCC
gly leu ile leu leu arg ser arg ala ile phe gly glu ala glu ala pro glu pro ser
726  /  241
GCG GGT GAC GGT GCG GCC GCT ACC TCT TGA
ala gly asp gly ala ala ala thr ser OPA
```

Applied Biosystems
Model 373A
Version 1.2.1

Sample 10
DyeTerminator{AnyPrimer}
Lane 10
Signal: G:443 A:336 T:179 C:694

Points 687 to 8350 Base 1: 687
matrix 654 5fw
KAW 5

Mon, Mar 6, 1995 5:36 PM Page 1 of 2
X: 0 to 11564 Y: 0 to 1200
Spacing: 9.00



Applied Biosystems
Model 373A
Version 1.2.1

KAW 4
DyeTerminator(AnyPrimer)
Lane 26
Signal: G:248 A:141 T:56 C:49

Points 853 to 8350 Base 1: 853
matrix 654 5lw
KAW 4

Tue, Mar 21, 1995 6:53 PM Page 1 of 2
X: 0 to 11265 Y 0 to 1200
Spacing: 9.01