XYLOSE TRANSPORT IN ESCHERICHIA COLI

by

Elaine O. Davis

A dissertation submitted to the University of Cambridge
in candidature for the degree of Doctor of Philosophy

October, 1985

Corpus Christi College and
Department of Biochemistry,
Cambridge.

PREFACE

I would like to express my gratitude to Dr. P.J.F. Henderson, my supervisor, for his guidance and encouragement. I would also like to thank everyone who provided assistance in any way, particularly Dr. M.C. Jones-Mortimer for helpful discussions and the provision of strains and Mr. M.C.J. Maiden for the communication of unpublished results and discussion. In addition, I would like to thank Dr. J.E. Walker of the MRC Laboratory of Molecular Biology for performing the N-terminal protein sequence analysis. Finally I would like to thank my husband, Nigel, for his continued encouragement and support.

I declare that this thesis is the result of my own work, except where indicated otherwise, and includes nothing which is the outcome of work done in collaboration.

Financial support was provided by the award of an MRC Studentship.

SUMMARY

1. Mud(Ap^R*lac*)I insertions in the genes coding for xylose-proton symport (*xylE*) and a component of the binding protein dependent xylose transport system (*xylG*) were isolated. The ampicillin resistance determinants of the prophages were used to map the positions of these genes on the *E. coli* chromosome. The *xylE* gene was located at 91.4min between *pgi* and *malB*, while the *xylG* gene mapped near the xylose metabolic genes, *xylAB*, at 80min.

2. Xylose-proton symport was shown to be susceptible to inhibition by the sulphydryl reagent N-ethylmaleimide. This inhibition was alleviated by the substrate, xylose, or by the substrate analogues 6-deoxyglucose and glucose. This protective effect of 6-deoxyglucose was exploited to label the XylE protein in membrane vesicles with radioactive N-ethylmaleimide. Comparison of protected versus unprotected samples for XylE$^+$ and XylE$^-$ strains by SDS-polyacrylamide gel electrophoresis identified the XylE protein to have an apparent molecular mass in the range 36 000-41 000Da.

3. A hybrid XylE-LacZ protein produced by a Mud(Ap^R*lac*)II insertion in *xylE* was shown to be membrane bound. The band on an SDS-polyacrylamide gel corresponding to this protein was identified. After testing various methods to purify the protein, a purification scheme was devised involving gel filtration of Triton X-100 solubilized membrane proteins. The N-terminal amino acid sequence of the hybrid protein corresponded to that deduced from the DNA sequence of the cloned *xylE* gene.

4. A restriction map of the *xylE-malB* region of the *E. coli* chromosome was made using specialised transducing phage derived from λ*plac*Mu insertions in *xylE* and the nearby *malK* gene. A 2.7kb HincII fragment was identified as containing the intact *xylE* gene by restoration of xylose-proton symport activity to a strain lacking both xylose transport systems. The DNA sequence of this fragment was determined by the Sanger dideoxy chain termination method. The *xylE* gene was found to be 1 473bp in length, corresponding to 491 amino acids and a protein molecular mass of 53 607Da. Comparison of the amino acid sequences revealed considerable homology of XylE with AraE (arabinose-proton symport), but no significant homology with LacY (lactose-proton symport) or MelB (melibiose-sodium cotransport).

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

AMP	adenosine monophosphate
ATP	adenosine triphosphate
CCCP	carbonyl cyanide m-chlorophenylhydrazone
DEAE	diethylaminoethyl
DNP	2,4-dinitrophenol
EDTA	ethylenediaminetetraacetic acid
G6P	glucose-6-phosphate
Hol	histidinol
IPTG	isopropyl-β-D-thiogalactoside
MES	2-(N-morpholino)ethanesulphonic acid
ONPG	ortho-nitrophenyl-β-D-galactoside
NADH	nicotinamide adenine dinucleotide (reduced form)
NEM	N-ethylmaleimide
pmf	proton motive force
PMS	phenazine methosulphate
POPOP	1,4-bis(5-phenyl-2-oxazolyl)-benzene
PPO	2,5-diphenyloxazole
PTS	phosphotransferase system
SDS	sodium dodecyl sulphate
TDG	thiodigalactoside
TEMED	NNN'N'-tetramethyl-1,2-diaminoethane
TMG	methyl-β-D-thiogalactoside
XGal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

The cell membrane defines boundaries between different aqueous environments by providing a separate hydrophobic phase. It acts as a permeability barrier, hindering the diffusion of hydrophilic solutes between these environments. The translocation of specific solutes is mediated by carriers, allowing regulation of the intra-cellular environment and concentration of fuel sources. This translocation may be passive, permitting the passage of a substance down its electrochemical gradient, or active, accumulating a substance against its electrochemical gradient by utilising some other source of energy and thereby resulting in an overall decrease in free energy.

In Gram-negative micro-organisms there are, in addition to the cytoplasmic membrane, the barriers of the murein layer and the outer membrane. The murein layer in the periplasmic space, between the inner (cytoplasmic) and outer membranes, provides no hindrance to the diffusion of substances required by the cell. The outer membrane contains non-specific porins for the passage of substances of molecular weight less than approx. 600Da into the periplasm (Nakae, 1976).

1.1 Carrier Mediated Transport Systems in Bacteria

There are five recognised types of carrier mediated sugar transport systems in bacteria, which are summarised in Figure 1.1 (from Dills *et al.*, 1980), where an example is given of each type for *Escherichia coli*. The classification is largely based on the energy source for the process and any known structural differences.

1.1.1 Facilitated diffusion

Facilitated diffusion is dependent on a carrier protein equilibrating a hydrophilic solute across the membrane, but is not coupled to energy release. The only known example in the inner membrane of *E. coli* is for glycerol. Accumulation results from the rapid phosphorylation of the glycerol inside the cell by α-glycerokinase, resulting in the intracellular concentration of

1. INTRODUCTION

The cell membrane defines boundaries between different aqueous environments by providing a separate hydrophobic phase. It acts as a permeability barrier, hindering the diffusion of hydrophilic solutes between these environments. The translocation of specific solutes is mediated by carriers, allowing regulation of the intra-cellular environment and concentration of fuel sources. This translocation may be passive, permitting the passage of a substance down its electrochemical gradient, or active, accumulating a substance against its electrochemical gradient by utilising some other source of energy and thereby resulting in an overall decrease in free energy.

In Gram-negative micro-organisms there are, in addition to the cytoplasmic membrane, the barriers of the murein layer and the outer membrane. The murein layer in the periplasmic space, between the inner (cytoplasmic) and outer membranes, provides no hindrance to the diffusion of substances required by the cell. The outer membrane contains non-specific porins for the passage of substances of molecular weight less than approx. 600Da into the periplasm (Nakae, 1976).

1.1 Carrier Mediated Transport Systems in Bacteria

There are five recognised types of carrier mediated sugar transport systems in bacteria, which are summarised in Figure 1.1 (from Dills *et al.*, 1980), where an example is given of each type for *Escherichia coli*. The classification is largely based on the energy source for the process and any known structural differences.

1.1.1 Facilitated diffusion

Facilitated diffusion is dependent on a carrier protein equilibrating a hydrophilic solute across the membrane, but is not coupled to energy release. The only known example in the inner membrane of *E. coli* is for glycerol. Accumulation results from the rapid phosphorylation of the glycerol inside the cell by α-glycerokinase, resulting in the intracellular concentration of

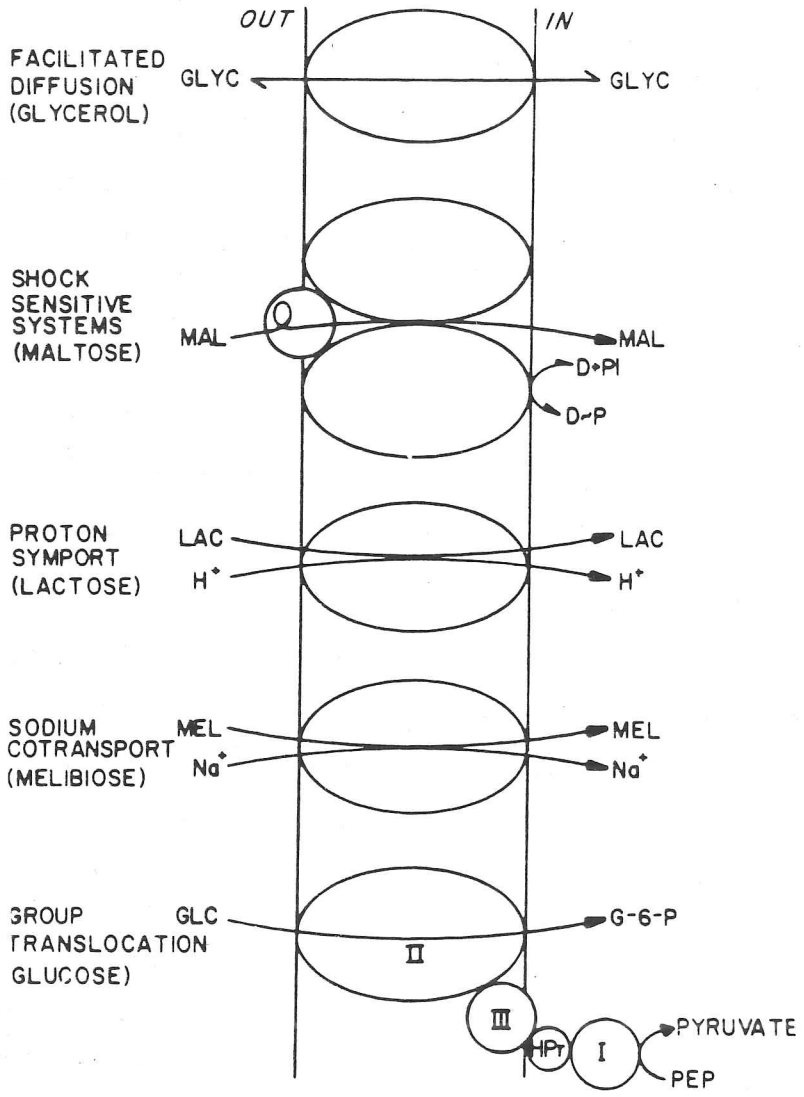


Figure 1.1 Carrier mediated transport mechanisms (from Dills et al., 1980)

Representative examples (in parentheses) are given for E. coli. Abbreviations: GLYC, glycerol; MAL, maltose; LAC, lactose; MEL, melibiose; GLC, glucose; G-6-P, glucose-6-phosphate; PEP, phosphoenolpyruvate; $D+P_i$, nonphosphorylated donor plus inorganic phosphate; D~P, high-energy phosphorylated donor; I, enzyme I; II, enzyme II; III, enzyme III.

free glycerol being kept low (Hayashi and Lin, 1965). The presence of a facilitator can be deduced from saturable kinetics of uptake, inducibility, substrate specificity, competitive inhibition by substrate analogues, and genetic lesions.

The existence of the glycerol facilitator was demonstrated by the rapid recovery from plasmolysis of cells exposed to hypertonic solutions of glycerol if they had been grown on glycerol (Sanno et al., 1968); this response was much slower in cells grown on glucose. This result has been confirmed by the more sensitive technique of stopped-flow spectrophotometry (Alemohammad and Knowles, 1974). The inducer is in fact L-α-glycerophosphate, as no induction was observed for glycerokinase negative mutants.

The requirement of the facilitator for effective utilization of glycerol has been demonstrated (Richey and Lin, 1972). Cells carrying the gene, $glpF^+$, coding for the glycerol facilitator grew much faster than glpF strains at glycerol concentrations of less than 5mM.

Substrate specificity is shown by the glycerol facilitator. Erythritol, pentitols and hexitols could enter on the facilitator in addition to glycerol, but the analogous sugars were not transported (Heller et al., 1980). However, there was little or no competitive inhibition between the two substrates glycerol and xylitol. This, together with the observation that xylitol transport was insensitive to low temperatures was taken to suggest that the facilitator functions as a membrane channel.

1.1.2 Group translocation

Group translocation involves chemical modification of the substrate during transport. In the case of the sugar phosphotransferase system (PTS) phosphorylation of the substrate passing across the membrane occurs, which is dependent on phosphoenolpyruvate (PEP) via a series of components (Kundig et al., 1964). The system includes both cytoplasmic and membrane bound components, and requires magnesium ions. The scheme for the reactions involved is shown in Figure 1.2 (from Roseman, 1977).

Enzyme I and HPr are soluble components which initiate the transfer of phosphate from PEP. Enzyme I and HPr are common to the uptake of all sugars

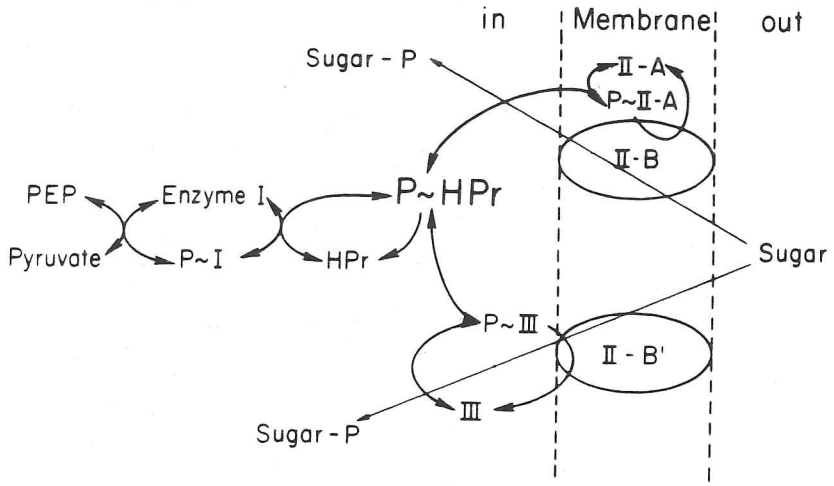


Figure 1.2 Sugar translocation and phosphorylation by the phosphoenol-pyruvate sugar phosphotransferase system (from Roseman, 1977)

The product of vectorial translocation is the sugar phosphate, and the reaction is catalysed by the sugar binding protein in the membrane (II-B or II-B'), the phosphoryl group being derived from the other sugar-specific component, a soluble protein (phospho-III) or a membrane protein (phospho-II-A). The sugar-specific proteins are phosphorylated by sequential transfer of the phosphoryl group from phosphoenolpyruvate through the general (non sugar-specific) proteins of the system, enzyme I and HPr.

transported by the PTS. Mutations in the genes coding for their synthesis, ptsI and ptsH respectively, prevent the uptake of all PTS sugars, except for fructose which has its own equivalent of HPr called FPr (Waygood, 1980). The active form of enzyme I is a dimer (Misset et al., 1980). The activity requires a divalent metal ion (Mg^{2+} or Mn^{2+}) bound to the enzyme and this also increases the stability of the dimers (Hoving et al., 1982). Enzyme I has also been reported to be active as a trimer (Grenier et al., 1985). Two classes of sulphydryl group have been found, one of which is essential for catalytic activity (Grenier et al., 1985). Studies with 3-bromopyruvate causing alkylation of cysteine residues indicated that one residue per enzyme dimer is alkylated without loss of activity, whereas alkylation of the second residue leads to complete inactivation (Hoving et al., 1984), suggesting that there is only one active site per dimer. HPr (histidine containing protein) is phosphorylated on the N-1 atom of a histidine residue (Anderson et al., 1971) to give phospho-HPr, a true high-energy compound with an apparent standard free energy of hydrolysis of almost twice the value for the pyrophosphate bonds in ATP (Roseman, 1977).

The membrane bound components of the PTS, enzymes II and III, provide the sugar specific part of the system. A pair of proteins is required for the phosphorylation of a given sugar, one of which, enzyme II-B, is an integral membrane protein, while the other may be either an integral membrane protein, enzyme II-A, or a peripheral membrane protein, enzyme III. Glucose may be transported by either type of system: a II-A/II-B system called PtsM, which also transports mannose, glucosamine and N-acetylglucosamine; or a III/II-B system called PtsG, which also transports methyl-α-D-glucoside, although PtsG is the main transport system for normal growth on glucose in wild type E. coli (Curtis and Epstein, 1975; Stock et al., 1982). It is the enzyme II-B which serves as the sugar recognition site and has a strict specificity for both the sugar and the sugar phosphate (Roseman, 1977). The enzyme II-A or enzyme III is phosphorylated by phospho-HPr, then passes the phosphate on to the sugar, phosphorylating carbon-6, via enzyme II (Peri et al., 1984). Evidence for the interaction of enzyme III^{glc} with HPr has been obtained using enzyme III^{glc} derivatised with a fluorescent label (Jablonski et al., 1983).

There is a functional association of the soluble PTS proteins with the membrane. The PEP-dependent sugar phosphorylation is catalysed more

efficiently by the membrane bound complex than by a mixture of the individual enzymes (Dills et al., 1980). However, the rates of phophoryl exchange reactions catalysed by enzyme I and the enzyme II complexes were essentially the same for the associated and dissociated forms of the complexes. This functional association can be destroyed by detergents and by sonication, indicating that enzyme I and HPr are peripheral membrane constituents associated with the integral membrane protein enzyme II (Dills et al., 1980). In addition, cytoplasmic membrane pellets obtained by differential centrifugation contain larger quantities of HPr and enzyme I than would be expected for a cytoplasmic protein with no affinity for the membrane. In sucrose density gradient centrifugation enzyme I, whilst being found mainly at the top of the gradient as expected for cytoplasmic proteins, tailed into the gradient towards the position of the cytoplasmic membranes; this would be readily explained if the enzyme I was initially bound to the membranes (Misset and Robillard, 1982). Perhaps the most striking piece of evidence is that vesicles prepared by an osmotic shock procedure causing release of soluble proteins are able to transport and phosphorylate methyl-α-D-glucoside when loaded with PEP (Kaback, 1968), whilst vesicles prepared from a mutant strain lacking enzyme I could not accumulate the sugar.

The proton motive force (see section 1.1.4) has been shown to affect transport by the PTS, although not being the primary source of energy. The transport of methyl-α-D-glucoside was found to be inhibited by the addition of respiratory substrates in a manner independent of the ATP concentration in both the wild type and ATPase mutants (del Campo et al., 1975). The inhibitory effect could be abolished by the uncoupler CCCP, and was attributed to the energy rich state of the membrane, with ATP playing no direct role. The use of a cytochrome deficient mutant indicated that methyl-α-D-glucoside transport could be inhibited by the formation of an energised state of the membrane from ATP hydrolysis (Singh and Bragg, 1976). Inhibition was observed on the addition of galactose, but not on the addition of D-lactate which the mutant was unable to oxidise, and was reversed by uncouplers and inhibitors of ATPase activity. This was explained by glycolysis of galactose yielding ATP which was then hydrolysed by the ATPase to form an energised membrane. Support for this theory included the stimulation of proline uptake (a sodium symport system) by galactose. Sugar transport by the PTS was also found to be

inhibited by the energised state of the membrane in both intact cells and membrane vesicles by Reider et al.(1979). Uncouplers relaxing the proton gradient stimulated uptake by the PTS in aerobic cells; in anaerobic cells the membrane was of low energisation and PTS transport occurred at the maximum rate. The inhibitory effect of D-lactate as a respiratory substrate in vesicles could be reversed by cyanide or prevented by oxamate, a specific inhibitor of lactate dehydrogenase. The effect was also absent in a strain lacking cytochromes and thus unable to energise the membrane via the respiratory chain. A similar effect on PTS transport has been demonstrated by the addition of oxidising agents as well as by substrate oxidation, and in either case this effect could be blocked or reversed by dithiothreitol (Robillard and Konings, 1981). By studying phosphorylation in the absence of transport with inside out vesicles, the level of inhibition was identified as the binding of sugar to enzyme II. A sulphydryl group was implicated as undergoing reversible oxidation, resulting in the conversion of enzyme II from a reduced, high-affinity form to an oxidised, low-affinity form with an affinity for substrate reduced by 100-1000 fold. An identical change in affinity was observed due to the generation of a proton motive force. Robillard and Konings proposed that the proton motive force regulates the PTS via the redox potential in the membrane such that enzyme II is converted to the oxidised, low-affinity state in its presence. However, it has subsequently been shown that methyl-α-D-glucoside may be phosphorylated by PtsG or PtsM with different affinities and that these two systems are differentially sensitive to oxidising agents (Grenier et al., 1985). The low K_m system, II^{glc}, was strongly inhibited by oxidising agents, whilst the high K_m system, II^{man}, was relatively insensitive to oxidising agents. Thus the apparent K_m change reported by Robillard and Konings (1981) can be explained by this differential sensitivity of the two systems.

The transport of sugars, particularly glucose, by the PTS affects the uptake of other non-PTS sugars. In addition, competition between separate PTS enzyme IIs for phospho-HPr may result in one PTS sugar inhibiting the uptake of another. Mutants lacking enzyme I (ptsI) or HPr (ptsH) are not only defective in growth on all PTS sugars but, in addition, are unable to grow on many non-PTS substrates such as lactose, melibiose, glycerol and maltose (Saier et al., 1976). This has been attributed to two effects (Postma and

Roseman, 1976; Saier, 1977): (i) inability of the mutant cells to accumulate certain non-PTS compounds which are required inside the cell as inducers (inducer exclusion); (ii) lowering of adenylate cyclase activity resulting in intracellular cyclic AMP levels too low for the expression of certain cyclic AMP dependent operons. Leaky enzyme I mutants could utilise these sugars but only in the absence of a PTS sugar, and were more sensitive than the wild type to repression by the PTS substrates. An additional mutation in an enzyme II relieved the repression by the substrate of that particular enzyme II in a leaky enzyme I mutant although repression by other PTS sugars was not affected. However, neither appreciable uptake nor metabolism of the PTS sugars was required for these compounds to effect repression.

A single secondary mutation (*crr*) allows growth on the non-PTS compounds in these different classes of PTS mutants (Saier and Roseman, 1976a), and maps near to the *ptsI* and *ptsH* genes but outside the *pts* operon. This second mutation is not a reversion of the original *pts* mutation, since the double mutant remained unable to utilise PTS sugars. The *crr* mutation also rendered the wild type strain resistant to PTS mediated repression. The only biochemical defect detected in *crr* mutants was the absence or reduced levels of enzyme III^{glc} activity. This could be explained in two ways: (i) *crr* is the structural gene for enzyme III^{glc}, and the mutants contain normal levels of an altered protein; (ii) *crr* regulates the synthesis of enzyme III^{glc}, and the mutants contain reduced levels of the normal protein.

Evidence that *crr* is the structural gene for enzyme III^{glc} has been obtained from comparison of the protein from crr^+ and *crr* strains, and from cloning of the crr^+ gene (Meadow *et al.*, 1982). These workers found approximately normal levels of a protein which cross-reacted with an anti-enzyme III^{glc} antibody in both wild type and *crr* strains. The purified enzymes III^{glc} from these strains were distinguishable by some differences in behaviour on gel filtration and in polyacrylamide gel electrophoresis under certain conditions, and the greatly reduced ability of the mutant enzyme (approx. 2-3% of the wild type) to phosphorylate methyl-α-D-glucoside *in vitro*. A plasmid carrying the crr^+ gene, cloned on a small fragment, restored PTS-mediated repression of the utilization of maltose, lactose, glycerol and melibiose when transformed into a *crr* strain. The plasmid also directed the synthesis of enzyme III^{glc} as determined from maxicells and by

rocket immunoelectrophoresis (Meadow et al., 1982).

The extent of repression of metabolic enzyme synthesis was found to correlate with the degree of inhibition of transport of the inducing non-PTS sugar (Saier and Roseman, 1976b). These workers also found a similar inhibition of transport for fully induced transport systems of non-PTS sugars, indicating these non-PTS systems to be subject to control by the PTS, and that this is responsible for inducer exclusion in the uninduced cells. This PTS regulation was shown to affect uptake, but not efflux, of the non-PTS sugar. The inhibition caused by a PTS sugar was shown to be reversible and the possibility of inhibition being caused by the PTS sugars binding to the non-PTS transport proteins was excluded. It was also found that mutations in the structural genes for the non-PTS transport proteins could result in resistance to regulation by the PTS for the relevant system.

It has been suggested (Roseman, 1977; Saier, 1977) that enzyme III^{glc} in its non-phosphorylated form (i.e. in the presence of PTS sugar, or ptsI or ptsH mutations) directly binds to the various non-PTS transport systems and inhibits transport activity, thus preventing inducer entering the cell. In addition, the phosphorylated form of enzyme III^{glc} is proposed to activate adenylate cyclase. Evidence in favour of this model comes from the demonstration of an interaction between enzyme III^{glc} and membranes containing the lactose carrier (Osumi and Saier, 1982). Substrate was required for this interaction, and PEP in the presence of the other phosphate transfer components of the PTS prevented it, indicating that the phosphorylated form of enzyme III^{glc} did not interact with the lactose carrier. A direct interaction between purified enzyme III^{glc} and the lactose carrier was demonstrated by Nelson et al.(1983). This binding required the presence of the non-phosphorylated form of enzyme III^{glc} and the substrate of the lactose carrier. The binding of enzyme III^{glc} to the lactose carrier resulted in the inhibition of galactoside translocation both in membrane vesicles and liposomes reconstituted with the purified lactose carrier.

Interactions in vivo between enzyme III^{glc} and the glycerol and maltose transport systems of Salmonella typhimurium have also been demonstrated (Nelson and Postma, 1984). Here it was shown that the presence of a second PTS-sensitive uptake system and its substrate rendered maltose or glycerol

uptake wholly or partially resistant to PTS inhibition, whereas a PTS-insensitive system and its substrate had no effect. Increasing the intracellular levels of enzyme III^{glc} by use of a plasmid carrying *crr*$^+$ prevented this protective effect of the second PTS-sensitive uptake system, and rendered the maltose and glycerol systems hypersensitive to inhibition by PTS substrates. The protective effect is attributed to competition for enzyme III^{glc} effectively lowering its intracellular concentration.

The site of interaction of enzyme III^{glc} with the maltose transport system has been identified as the MalK protein, for mutants insensitive to the inhibitory effect of glucose have been mapped in *malK* (M.H. Saier and M. Schwartz, unpublished results quoted in Hengge and Boos, 1983). In the case of glycerol repression, enzyme III^{glc} has been shown to act on glycerol kinase and not the glycerol facililtator (Postma *et al.*, 1984). Only the unphosphorylated form of enzyme III^{glc} inhibited glycerol kinase. No inhibition of transport by methyl-α-D-glucoside was observed as measured from swelling of cells using a spectrophotometer. Previously, measurements of accumulation of radiolabel from glycerol had been made, but such accumulation is dependent on kinase activity as unmodified glycerol is lost during the washing procedures inherent in measuring accumulation due to its rapid equilibration across the membrane. It was noted that this action of enzyme III^{glc} is still one of inducer exclusion as α-glycerophosphate is the inducer of the glycerol regulon. Mutants resistant to inhibition by enzyme III^{glc} have been found to map in *glpK*, the gene coding for glycerol kinase (Novotny *et al.*, 1985). It was confirmed using crystalline glycerol kinase that free, but not phosphorylated, enzyme III^{glc} was inhibitory, and it was shown that inhibition and binding of enzyme III^{glc} to the kinase was strongly pH dependent. Glycerol kinase is also subject to allosteric regulation by fructose-1,6-bisphosphate. Although there were many similarities in the inhibition caused by enzyme III^{glc} and by fructose-1,6-bisphosphate, the presence of two distinct regulatory sites on the kinase was demonstrated. Mutants were isolated which were resistant to inhibition by either enzyme III^{glc} or fructose-1,6-bisphosphate but fully sensitive to the other regulatory agent.

The proposal that a different gene product, *iex*, interacts with the non-PTS transport systems (Parra *et al.*, 1983; Britton *et al.*, 1983) has been clarified

by the finding that the *iex* mutation results in an altered enzyme III^{glc} (Nelson *et al.*, 1984a). The mutant enzyme III^{glc} is heat labile, and, although it functions normally in glucose and methyl-α-D-glucoside transport, cannot bind to the lactose carrier. Thus in the *iex* strain, non-PTS systems are resistant to inhibition by PTS sugars. The introduction of the crr^+ allele on a plasmid restores the iex^+ phenotype to an *iex* strain.

It has been shown that the inhibition of lactose transport in *E. coli* by glucose or methyl-α-D-glucoside is not immediate even at very high concentrations of sugar (Koch, 1985). Thus, competition for some membrane bound component as a mechanism of inhibition is excluded, and production of the dephosphorylated form of enzyme III^{glc}, which then binds to and inhibits the carrier protein, is further supported.

1.1.3 Binding protein dependent active transport

Binding protein dependent transport systems were discovered as a result of loss of transport activity by cold osmotic shock (Neu and Heppel, 1965). This is due to the loss of binding proteins from the periplasmic space, and a corollary of this is that binding protein dependent systems are inactive in subcellular vesicles (Lombardi and Kaback, 1972). These systems are encoded by more than one gene, coding for membrane components in addition to the periplasmic binding protein. A well studied example is the maltose transport system, where *malE* codes for the binding protein (Kellerman and Szmelcman, 1974), *malF*, *malG*, and *malK* code for inner membrane components, and, unusual among such systems, a further gene, *lamB*, codes for an outer membrane component which is involved in transport, as well as being the receptor for phage λ.

The binding proteins have a high affinity for their particular substrates with dissociation constants of the order of 1μM. In addition to their role in transport, many of these binding proteins are receptors for chemotaxis (Adler *et al.*, 1973; Adler, 1975). The sequence of the maltose binding protein has been reported and two main regions of homology with the arabinose, galactose, and ribose binding proteins were identified (Duplay *et al.*, 1984); these are predicted to have similar molecular structures (Argos *et al.*, 1981). Use of mutants deleted for *malE* indicated that the maltose binding protein is

essential for detectable translocation of maltose across the cytoplasmic membrane (Shuman, 1982). Transport may be reconstituted in a malE deletion mutant by calcium induced permeabilization of the outer membrane enabling import of exogenous maltose binding protein into the periplasm (Brass et al., 1983; Bukau et al., 1985). The requirement for a direct interaction between the binding protein and one of the membrane components (the P protein) of the histidine transport system in S. typhimurium has been demonstrated; a mutation in the binding protein leaving the histidine binding site unaltered can be at least partially compensated for by a secondary mutation in the P protein (Ames and Spudich, 1976).

Of the remaining components the malF gene product has been shown to be an inner membrane component of apparent molecular mass approx. 40 000Da by immunoprecipitation with antibodies raised against a malF-lacZ hybrid protein (Shuman et al., 1980). The malK product has been identified as an inner membrane protein of molecular mass 40 000-43 000Da (Bavoil et al., 1980). However, in mutants lacking the MalG protein the malK product is found in the cytoplasm, suggesting that the MalK protein is associated with the inner surface of the cytoplasmic membrane via an interaction with the MalG protein (Shuman and Silhavy, 1981). In addition, the nucleotide sequences of the genes malF (Froshauer and Beckwith, 1984), malG (Dassa and Hofnung, 1985) and malK (Gilson et al., 1982a) have been determined. The molecular mass of the MalG protein was deduced to be 32 188Da from the DNA sequence although SDS-polyacrylamide gel electrophoresis of in vitro synthesised protein indicated an apparent molecular mass of 22 000Da. This discrepancy was attributed to anomalous binding of SDS to the hydrophobic protein as has been found for the lactose carrier protein (see section 1.1.4). Homology between the malK gene and ndh coding for NADH dehydrogenase was found, suggesting that the MalK protein may play a role in energisation, possibly involving an oxido-reduction reaction.

The binding protein dependent transport systems are sensitive to inhibition by arsenate (Berger, 1973; Berger and Heppel, 1974) which lowers intracellular ATP levels (Klein and Boyer, 1972). In addition, ATPase mutants defective in oxidative phosphorylation could not energise binding protein dependent transport systems with the respiratory substrate D-lactate or the artificial electron donor ascorbate plus PMS, unlike shock resistant

transport systems (see section 1.1.4). Glucose stimulated binding protein dependent transport via substrate level phosphorylation (Berger and Heppel, 1974). This has been taken as evidence that the binding protein dependent transport systems are energised by ATP. It has also been suggested that acetyl-phosphate is the immediate source of energy for binding protein dependent transport from experiments with mutants defective in phosphotransacetylase and using metabolic inhibitors to inactivate acetate kinase (Hong et al., 1979).

More recently evidence has been presented that the arginine-ornithine binding protein involved in transport may be found in a phosphorylated or a non-phosphorylated form (Celis, 1985). The non-phosphorylated protein could be phosphorylated in vitro by incubation with ATP, magnesium ions, and a phosphokinase enzyme from the periplasmic fluid. The dissociation constant of the phosphorylated protein was 5.0μM compared with 0.1μM for the unmodified protein. It was suggested that the modified protein was the active form responsible for release of the ligand for its translocation through the cytoplasmic membrane. The phosphoryl group was identified as most probably being linked covalently to a carboxyl group on the protein. This was taken as evidence that ATP is the direct source of energy for this active transport system and that phosphorylation of the binding protein was involved in utilization of that energy.

The presence of a nucleotide binding site on the HisP and maybe the HisM proteins of the histidine transport system in S. typhimurium was demonstrated by the use of the photoaffinity reagent 8-azido-ATP (Hobson et al., 1984). The HisP protein is highly homologous to the MalK protein of the E. coli maltose transport system (Gilson et al., 1982b). The labelling of HisP by 8-azido-ATP was reduced in the presence of ATP whilst other nucleotides had less effect. This indicated that ATP or a closely related molecule is the natural substrate for the site. Evidence for a nucleotide binding site on the HisP, MalK and OppD proteins of the histidine, maltose and oligopeptide transport systems has also been presented (Higgins et al., 1985). Extensive amino-acid sequence homology exists between these three proteins, and a consensus nucleotide-binding sequence was identified in the same relative position in each of them. Experimental evidence for OppD included binding of the protein to a Cibacron Blue affinity column, which is known to retain many ATP-binding

proteins, and elution by ATP, and specific labelling by the nucleotide affinity analogue 5'-p-fluorosulphonylbenzoyladenosine in a reaction which is inhibited by ATP. These results were taken as evidence of direct energisation of binding protein dependent transport systems by ATP or a closely related nucleotide, and that the HisP, MalK and OppD proteins are responsible for this energy coupling.

An inhibitor of α-ketoacid dehydrogenases, 5-methoxyindole-2-carboxylic acid, has been shown to be a potent and specific inhibitor of transport by binding protein dependent systems; it had no effect on lactose transport by proton symport or on glucose transport by the PTS (Richarme, 1985a). This inhibition was much more rapid and complete than the effect of 5-methoxyindole-2-carboxylic acid on reducing the ATP pool. The binding protein interaction with its substrate was shown to be unaffected. 5-Methoxyindole-2-carboxylic acid strongly inhibits lipoamide dehydrogenase, although not affecting NADH dehydrogenase or lactate dehydrogenase activities in E. coli. Subsequently it was shown using a mutant deficient in the synthesis of lipoic acid that binding protein dependent transport systems were inhibited under conditions of lipoic acid deprivation, whereas lactose and glucose transport were unaffected (Richarme, 1985b). The ATP levels under conditions of lipoic acid deprivation and supply were similar. The lipoic acid dependent transport was arsenate sensitive, but the transport activity remaining in lipoic acid deprivation was not inhibited by arsenate. It was suggested that arsenate inhibition may result from in vivo reduction to arsenite which then acts as an inhibitor of thiol functions, and that a lipoic acid dependent reaction might be the main target for arsenate inhibition of binding protein dependent transport.

The mechanism of transport by the binding protein dependent systems can be envisaged in two ways (Figure 1.3 from Ames and Higgins, 1983). In the 'pore' model the interaction of the binding protein-substrate complex with the membrane components causes a conformational change resulting in the formation of a pore within or between the proteins in the membrane and thus allowing entry of the substrate into the cell. In the 'binding site' model a substrate binding site is found on some or all of the membrane components, and interaction of the binding protein-substrate complex would sequentially 'activate' one or more of the intramembrane binding sites resulting in the

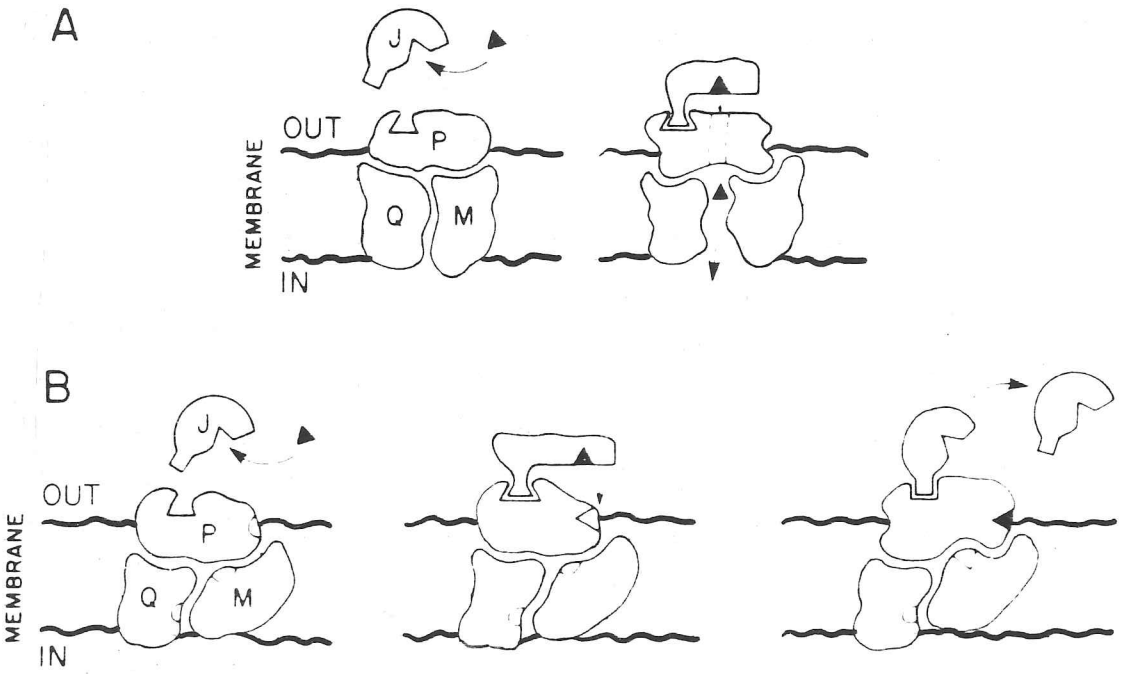


Figure 1.3 Models for transport by a binding protein dependent system (from Ames and Higgins, 1983)

The figure represents the components of the histidine transport system in S. typhimurium: J, binding protein; M, P, Q, membrane components. The two possible models for transport, A, the 'pore' model, and B, the 'sequential binding site' model, are described in the text.

substrate being transported through the membrane.

Evidence in favour of the second model comes from mutants which can transport maltose in a strain with a non-polar internal deletion of malE. One such strain has been shown to be mutated in the malF and malG genes coding for membrane components of the maltose transport system. This transport is specific for maltose, and the transport activity is retained in spheroplasts confirming the independence of any binding protein (Shuman, 1982). A further series of similar mutants were found to be mutated either in malF or in malG. In some of these strains maltose binding protein inhibited maltose transport (Treptow and Shuman, 1985). It thus appears that the cytoplasmic membrane components in the mutants possess a substrate recognition site. It is likely that this site exists in the wild type cells but is only available to substrate molecules bound to the maltose binding protein.

The lamB product, maltoporin (or λ receptor), is found as a trimer in the outer membrane (Palva and Westermann, 1979; Maezawa et al., 1983). Its importance in maltose transport was demonstrated in λ resistant strains by reduced maltose transport and reduced growth at low maltose concentrations (Szmelcman and Hofnung, 1975). Antibodies against purified maltoporin added to a wild type strain had a similar effect on maltose transport at low concentrations. The dissociation constant for maltose binding to maltose binding protein was found to be approximately 100-fold higher in a λ resistant strain than in the wild type, where the value (1μM) was equal to the in vitro value, whilst the maximum rate of transport remained unaltered (Szmelcman and Schwartz, 1976). In addition, the lamB strain was unable to transport maltotriose, whilst even in the wild type full access of maltotriose to the binding protein was not achieved, the dissociation constant being approximately 13-fold higher than the in vitro value for maltotriose binding, indicating that maltoporin was necessary for the transport of higher oligomers of glucose.

Maltoporin may also facilitate the diffusion of other sugars, for in a strain lacking other porin proteins the transmembrane diffusion of glucose and lactose was accelerated in the presence of maltoporin, although there was no such effect for histidine or 6-aminopenicillanic acid (von Meyenburg and Nikaido, 1977). The presence of a binding site within maltoporin was

demonstrated by the selection of mutants in *lamB* with enhanced affinity for maltose but not starch, or for starch but not maltose (Clune *et al.*, 1984).

Conflicting results have been presented on the role of the maltose binding protein in the general porin activity of maltoporin. Addition of maltose binding protein gave steps in a conductance measurement experiment indicating the opening and closing of channels (Neuhaus *et al.*, 1983). It was suggested that the association of the maltose binding protein with maltoporin altered the equilibrium in favour of closed channels. However, in a porin deficient strain measurement of lactose uptake under conditions where transport was limited by diffusion through maltoporin indicated no difference in kinetics irrespective of whether the strain was *malE*$^+$ or carried a *malE* deletion (Brass *et al.*, 1985).

Interaction of the maltose binding protein with maltoporin in relation to the transport of maltodextrins has been demonstrated (Wandersman *et al.*, 1979). The wild type grows equally well on maltose and on maltodextrins of up to seven glucose residues. Some mutants which were unable to grow on maltodextrins higher than maltotriose but could still grow on maltose were affected in *malE* (the maltose binding protein). The phenotype was the same as for a *lamB* mutant. The *malE* mutants were altered in substrate binding properties but in such a way that the highest affinity was for the maltodextrins on which the strain could not grow. The *malE* mutants were resistant to inhibition of maltose transport by maltotetraose in a way analogous to a *lamB* mutant, indicating that the tetraose could not reach the periplasm. Thus these *malE* mutants were impaired in their ability to transport maltodextrins across the outer membrane. A direct physical interaction between maltose binding protein and maltoporin has been demonstrated using maltose binding protein linked to Sepharose (Bavoil and Nikaido, 1981). On passing an extract of the outer membrane through such a column, maltoporin was quantitatively and specifically adsorbed, whereas no binding was observed using activated Sepharose alone or Sepharose linked to the histidine binding protein of *S. typhimurium*.

1.1.4 Proton symport

Proton symport systems belong to the class of transport systems which are resistant to cold osmotic shock (Neu and Heppel, 1965). A more useful distinction is that such shock resistant systems remain active in subcellular vesicles, indicating that no cytoplasmic or periplasmic components are required for transport. The use of vesicles has provided a means of studying transport by these systems under more closely defined conditions than is possible with intact cells.

Proton symport is one way that can be envisaged to utilise the electrochemical potential, or proton motive force (pmf), across the membrane to energise transport (Mitchell, 1961). The pmf is generated by the electron transport chain from oxidative respiration, or from ATP hydrolysis by the F_0F_1-ATPase complex under non-oxidative conditions. In either case protons are expelled from the cell, setting up a chemical gradient of protons across the membrane and so a pH difference (inside alkaline) and also a potential difference (inside negative). These two effects sum to give the electrochemical potential, $\Delta\mu_{H^+}$, or proton motive force, Δp (Mitchell, 1970 & 1976; Rosen and Kashket, 1978):

$$\Delta\mu_H = F\Delta\psi - 2.3RT\Delta pH$$

$$\Delta p = \Delta\psi - Z\Delta pH$$

This proton extrusion may be viewed as primary transport, and the sugar transport resulting from the coupling of the pmf produced to uptake as secondary transport (Rosen and Kashket, 1978).

The most studied proton symport system is that for lactose in E. coli. Evidence that the system operates by proton symport has been provided by a variety of experiments. An alkaline pH change is observed on addition of substrate (West, 1970; West and Mitchell, 1972). Uncouplers stimulate the downhill transport of substrate in energy depleted cells by dissipating the initial unfavourable pmf developed by the transport (Cecchini and Koch, 1975). An artificially produced pmf in cells depleted of endogenous energy reserves provided energy for transport of galactosides and this could be achieved by either a ΔpH or a $\Delta\psi$ (Flagg and Wilson, 1977). Transport activity in vesicles could be driven by various electron donors with an effectiveness which correlated with their ability to generate a pmf (Ramos et al., 1976). By

using valinomycin to abolish $\Delta\psi$ or nigericin to abolish ΔpH at pH5.5 lactose transport was found to depend on both components of the pmf; at pH7.5 only $\Delta\psi$ was available to energise transport (Ramos and Kaback, 1977a&b).

A stoichiometry of 1:1 for protons:lactose was initially measured at pH7.0 from parallel uptakes of lactose and protons in cells and extrapolation to zero time (West and Mitchell, 1973). Extrapolation to zero time is required because as time increases protons leak out of the cell due to the pH difference which builds up across the membrane, restoring the pmf at equilibrium. However, it has been claimed that in subcellular vesicles the pmf is insufficient to account for the accumulation of substances e.g. lactose which is found to occur at pH7.5, assuming a stoichiometry of 1:1 for protons:lactose (Ramos and Kaback, 1977b). Further work suggested that the stoichiometry varied with pH from 1:1 at pH5.5 to 2:1 for protons:lactose at pH7.5 (Ramos and Kaback, 1977c). This was in agreement with a model put forward by Rottenberg (1976) based on the assumption that the inward moving complex of carrier, proton(s) and substrate must be neutral. However, the values for the stoichiometry put forward by Ramos and Kaback were determined using the equation:

$$Z\log([S]_i/[S]_o) = (n+m)\Delta\psi - nZ\Delta pH$$

which only holds if substrate is accumulated until it reaches thermodynamic equilibrium with the pmf. Kaback's conclusions are disputed by Booth et al. (1979,1980) who claim that accumulation reaches a kinetic steady state and who present evidence that a thermodynamic steady state is not reached. At no point over a pH range of 5.9-8.7 was the lactose accumulation and the pmf in equilibrium but throughout this range the pmf was in excess of what was required for the transport observed, indicating that the accumulation was governed by kinetic factors. These workers have found a value of 1:1 for the proton:lactose stoichiometry in intact cells throughout the pH range 6.5-7.7, which is supported by a stoichiometry of 1:1 found in the pH range 6-8 by Zilberstein et al. (1979). In both these cases the value was obtained by direct determination from measuring proton and lactose uptake in parallel experiments. Stoichiometries and possible kinetic models are reviewed by West (1980).

The lactose carrier protein is susceptible to inhibition by N-ethylmaleimide and has been labelled using radioactive N-ethylmaleimide

with prior protection by thiodigalactoside (Fox and Kennedy, 1965). This allowed the identification of the protein by SDS-polyacrylamide gel electrophoresis as having an apparent molecular mass of 30 000Da (Jones and Kennedy, 1969). This apparent molecular mass was also observed for lactose carrier labelled with the photoaffinity reagent 4-nitrophenyl-α-D-galactoside which was specific for the lactose carrier (Kaczorowski et al., 1980).

Cloning of the lacY gene allowed overexpression of the lactose carrier protein and again an apparent molecular mass of 30 000Da was observed by SDS-polyacrylamide gel electrophoresis (Teather et al., 1978). The DNA sequence of the lacY gene revealed the true molecular mass of the LacY protein to be 46 502Da (Buchel et al., 1980). The in vitro lacY gene product was identical to the in vivo lactose carrier isolated from cytoplasmic membranes as determined by apparent molecular mass and N-terminal sequence analysis, and the amino acid composition of the in vivo product was in good agreement with that predicted by the DNA sequence (Ehring et al., 1980). This indicated that the lower molecular mass of the protein compared with that predicted by the DNA sequence was not the result of post-translational modification.

It has been suggested that the discrepancy between the molecular mass deduced from the DNA sequence and that observed by SDS-polyacrylamide gel electrophoresis is due to abnormally high binding of SDS by the lactose carrier (Beyreuther et al., 1980). The resulting increased charge-to-mass ratio causes the protein to run ahead of its true molecular mass. This effect can be overcome at sufficiently high gel concentrations where the molecular sieving factor predominates, and a value of 46 000Da is found by SDS-polyacrylamide gel electrophoresis on gels of 20% or more. The molecular mass of the lactose carrier has also been determined by gel filtration in hexamethylphosphoric triamide giving a value of 47 500Da (Konig and Sanderman, 1982) in agreement with the value determined from the DNA sequence. The cysteine residue which is labelled by N-ethylmaleimide has been identified as Cys^{148} (Beyreuther et al., 1981).

The lactose carrier has been purified and reconstituted into proteoliposomes which are active in counterflow and energised transport (Newman et al., 1981; Viitanen et al., 1984; Wright and Overath, 1984). These

results indicate that the carrier protein is the only polypeptide species required for active galactoside transport. However, the maximum rate of active transport was considerably lower than in cells, although similar to the value found for cytoplasmic vesicles.

The secondary structure of the lactose carrier protein has been proposed to consist of at least 12 α-helical segments spanning the membrane (Foster et al., 1983). This is based on circular dichroism measurements of the purified protein, which indicate 85% of the amino acids to be in helical secondary structures, and the hydropathic profile of the protein. However, until sufficient membrane proteins have their structures determined by physical methods any such predictions cannot be considered to be reliable.

The translocator has been found to be intrinsically symmetrical, since both right-side out and inside out vesicles exhibit similar rates of flux by efflux, counterflux and potassium/valinomycin induced active transport (Teather et al., 1977). In addition, using inverted vesicles counterflow has been demonstrated in either direction, the efflux of internal lactose was driven by respiration or ATP hydrolysis dependent on the F_oF_1-ATPase, and accumulation of lactose against a concentration gradient was achieved by the formation of an artificial pmf with $\Delta\psi$ negative inside or ΔpH basic inside (Lancaster and Hinkle, 1977).

It has been proposed that the pmf imposes asymmetry to the system such that the carrier has a high affinity for its substrate on the outside, and a low affinity on the inside, of the membrane. It has been suggested from a comparison of the K_t of exit and entry in energised cells and cells poisoned with metabolic inhibitors that the effect of energy coupling is to reduce the affinity of the carrier for its substrate on the inner surface of the membrane, since no apparent change in K_t of entry was observed whilst K_t for exit decreased by approximately two orders of magnitude (Winkler and Wilson, 1966). It has also been reported that the affinity of the galactoside binding site on the carrier is much lower (approx. 60-fold) than the half saturation constant for active transport; this increase in affinity has been considered in terms of a change in the affinity of binding or an increase in mobility of the loaded carrier in the presence of the pmf (Wright et al., 1981).

The generation of a pmf was also observed to decrease the apparent K_m for lactose by at least 100-fold compared with the value under de-energised conditions by Robertson et al.(1980). This low apparent K_m could be elicited by either ΔpH or $\Delta\psi$. a decrease in K_m, and partial dissipation of the pmf resulted in biphasic kinetics of transport with one component characteristic of active transport and the other of facilitated diffusion. This was evidence that the pmf alters the distribution of the lactose carrier protein between two kinetic states. It was tentatively suggested that the carrier protein may exist in two forms, the monomer responsible for catalysing facilitated diffusion and the dimer responsible for catalysing active transport, and that the pmf promotes the aggregation of monomers to dimers. It was also acknowledged that a change in the rate limiting step for transport by the pmf could account for these results without a change in the structural state of the carrier.

Evidence in support of the lactose carrier being active as a dimer came from the discovery of negative dominant mutations in the lacY gene (Mieschendahl et al., 1981). The presence of dimers or oligomers as the active form of the carrier is the easiest way to explain such mutants. However, measurements of the rate of galactoside transport in diploid strains did not verify the negative dominant mutant phenotype (Wright et al., 1983). In addition, purified lactose carrier was shown to be a monomer by sedimentation equilibrium in detergent micelles of dodecyl-maltoside, a nonionic detergent not expected to disrupt protein-protein interactions and in which galactoside is bound with a 1:1 stoichiometry (Wright et al., 1983). It was also noted that the affinity change observed with lactose on energisation of the membrane was not found for all substrates, e.g. TDG. As a result it was suggested that the membane embedded carrier functions as a monomer in transport.

Studies by radiation inactivation analysis and determination of the functional molecular mass by target theory suggested that the pmf may cause an alteration in subunit interactions and indicated that the functional molecular mass of the lactose carrier might be a dimer (Goldkorn et al., 1984). After irradiation the protein was extracted with octyl-β-D-glucoside, reconstituted into proteoliposomes, and assayed for transport activity. For non-energised vesicles the functional molecular mass was determined to be

45 000-50 000Da, whereas for energised vesicles the functional molecular mass of the lactose carrier was 85 000-100 000Da, whilst no change in the target size of the control lactate dehydrogenase was observed. When the vesicles were energised under conditions in which the pmf was collapsed the target size of the lactose carrier was also 45 000-50 000Da. Two possibilities were suggested to explain these results: (i) in the non-energised membrane the lactose carrier is a monomer and dimerisation occurs in the presence of the pmf; (ii) the membraneous lactose carrier is always a dimer and generation of the pmf causes an increased interaction between the subunits.

Purified lactose carrier in the non-ionic detergent dodecyl octaethylene glycol monoether, a detergent that appears to maintain the trimeric state of maltoporin, was used to determine the molecular mass of the carrier by analytical ultracentrifugation and gel filtration (Houssin et al., 1985). These experiments indicated that the lactose carrier existed mainly as a dimer in this detergent, and it was suggested that this may reflect the situation that prevails in the membrane. During centrifugation to determine the sedimentation equilibrium it was noticed that some of the dimer dissociated to monomer. It was suggested that such dissociation may be more pronounced in dodecyl maltoside explaining the results of Wright et al.(1983) where only a monomer was found.

Proteolysis experiments with right-side out and inside out vesicles containing photoaffinity labelled lactose carrier protein demonstrate directly that the polypeptide spans the membrane (Goldkorn et al., 1983). Studies with antibodies have yielded further information on the structure and function of the lactose carrier protein. Site directed polyclonal antibodies against the C-terminal decapeptide indicated that the C-terminus is directed towards the cytoplasmic surface of the membrane (Seckler et al., 1983). Binding of these antibodies did not inhibit substrate binding or translocation.

One monoclonal antibody, 4B1, has been produced which inhibits transport, but only those reactions involving net proton translocation; the antibody has little effect on equilibrium exchange and no effect on the ability of the vesicles to generate a pmf from reduced PMS, or on the ability of the carrier to bind a high-affinity ligand, p-nitrophenyl-α-D-galactoside (Carrasco

et al., 1984a). This antibody, and another monoclonal antibody, 5F7, bind to distinct non-overlapping epitopes on the lactose carrier protein. It has been shown that these antibodies bind to spheroplasts and right-side out vesicles, but only slightly to inside out vesicles (Herzlinger et al., 1984). This demonstrates that the lactose carrier has an asymmetric orientation within the cytoplasmic membrane and that the epitopes for both these antibodies are located on the periplasmic surface. It also appears that, with regard to the 4B1 epitope at least, the lactose carrier protein in reconstituted proteoliposomes has the same orientation as in the native membrane, since the antibody bound to proteoliposomes with a stoichiometry very close to that observed for right-side out vesicles.

However, although antibodies directed against a C-terminal dodecapeptide located the C-terminus of the carrier on the cytoplasmic surface of the membrane in vesicles, these antibodies bound stoichiometrically in the same manner as antibody 4B1 to proteoliposomes. This indicated that in proteoliposomes the 4B1 epitope and the C-terminus of the carrier were present on the exterior surface (Carrasco et al., 1984b). Furthermore, digestion with carboxypeptidase A and B reduced the binding of the anti-C-terminal antibodies by more than 80%, while the binding of antibody 4B1 and transport activities were essentially unaffected. It was suggested that during reconstitution the lactose carrier protein undergoes intramolecular dislocation of the carboxyl terminus with no significant effect on catalytic activity. The epitope of another site directed polyclonal antibody, present on the cytoplasmic surface of the membrane in vesicles, was inaccessible in proteoliposomes, providing evidence that a significant number of molecules were not scrambled in the reconstitution system.

The orientation of the lactose carrier in native and reconstituted vesicles has also been investigated by Seckler and Wright (1984) using antibodies against the C-terminal decapeptide and carboxypeptidase. All accessible external C-termini were digested with carboxypeptidase A followed by detection of any remaining internal C-termini with labelled antibody after electrophoresis and blotting. Alternatively, external C-termini were labelled with the antibody. Spheroplasts and right-side out vesicles were thus confirmed to contain the carrier molecule in a single orientation. In proteoliposomes only 48% of the carrier molecules were orientated in the same

et al., 1984a). This antibody, and another monoclonal antibody, 5F7, bind to distinct non-overlapping epitopes on the lactose carrier protein. It has been shown that these antibodies bind to spheroplasts and right-side out vesicles, but only slightly to inside out vesicles (Herzlinger et al., 1984). This demonstrates that the lactose carrier has an asymmetric orientation within the cytoplasmic membrane and that the epitopes for both these antibodies are located on the periplasmic surface. It also appears that, with regard to the 4B1 epitope at least, the lactose carrier protein in reconstituted proteoliposomes has the same orientation as in the native membrane, since the antibody bound to proteoliposomes with a stoichiometry very close to that observed for right-side out vesicles.

However, although antibodies directed against a C-terminal dodecapeptide located the C-terminus of the carrier on the cytoplasmic surface of the membrane in vesicles, these antibodies bound stoichiometrically in the same manner as antibody 4B1 to proteoliposomes. This indicated that in proteoliposomes the 4B1 epitope and the C-terminus of the carrier were present on the exterior surface (Carrasco et al., 1984b). Furthermore, digestion with carboxypeptidase A and B reduced the binding of the anti-C-terminal antibodies by more than 80%, while the binding of antibody 4B1 and transport activities were essentially unaffected. It was suggested that during reconstitution the lactose carrier protein undergoes intramolecular dislocation of the carboxyl terminus with no significant effect on catalytic activity. The epitope of another site directed polyclonal antibody, present on the cytoplasmic surface of the membrane in vesicles, was inaccessible in proteoliposomes, providing evidence that a significant number of molecules were not scrambled in the reconstitution system.

The orientation of the lactose carrier in native and reconstituted vesicles has also been investigated by Seckler and Wright (1984) using antibodies against the C-terminal decapeptide and carboxypeptidase. All accessible external C-termini were digested with carboxypeptidase A followed by detection of any remaining internal C-termini with labelled antibody after electrophoresis and blotting. Alternatively, external C-termini were labelled with the antibody. Spheroplasts and right-side out vesicles were thus confirmed to contain the carrier molecule in a single orientation. In proteoliposomes only 48% of the carrier molecules were orientated in the same

way as in the cell. Cycles of freeze-thawing or sonication resulted in a reshuffling of the carrier molecules between the inside out and right-side out populations, while maintaining 41% in the right-side out orientation. Digestion of the C-terminus was again found to have no effect on transport properties.

It has also been shown that the first eight N-terminal amino acids of the LacY protein can be replaced by varying amounts of the N-terminus of β-galactosidase without impairing lactose carrier activity qualitatively (Bocklage and Muller-Hill, 1983). However, the N-terminal region from amino acid 9 to 34 is required. These LacZ-LacY hybrids were shown to transport in vivo and have a similar affinity for TDG to the wild type although a reduced maximum velocity; immunoblotting indicated that the hybrid was present in the membrane (Griesser et al., 1983).

Histidine residue(s) have been implicated in the response of the lactose carrier to the pmf (Padan et al., 1979; Garcia et al., 1982). Diethylpyrocarbonate inactivated lactose transport and counterflow in vesicles but this effect could be blocked by substrates of the carrier protein. Lactose counterflow was regenerated by subsequent exposure to hydroxylamine which reacts to regenerate a histidine imidazole ring. The variation of the effect of diethylpyrocarbonate with pH was very similar to that for reaction with histidine, and it was determined that acylation of a single site in the transport system was sufficient for inactivation. However, the ability of the carrier to bind ONPG or to catalyse facilitated diffusion of lactose was unaffected. Treatment with diethylpyrocarbonate was shown to confer biphasic kinetics on the lactose transport system, analogous to the biphasic kinetics observed on partial dissipation of the pmf, but with no reduction in the pmf. It was concluded that histidine residues are involved in the response of the lactose carrier to the pmf. Diethylpyrocarbonate was observed to cause inactivation of lactose-induced proton influx in vesicles. The rate of lactose facilitated diffusion was enhanced in a manner similar to the effect of the proton uncoupler CCCP, but without increasing the permeability of the membrane (Patel et al., 1982). It was suggested that the altered lactose carrier catalysed lactose influx without the symport of protons. The hypothesis was put forward that acylation of a histidine residue(s) in the lactose carrier dissociates lactose influx from proton

influx, and that this residue plays an important role in the pathway of proton translocation.

However, diethylpyrocarbonate is reported to inhibit completely the binding of the substrate p-nitrophenyl-α-D-galactoside to the carrier, and with partial inhibition loss of transport was found to parallel loss of binding sites (Neuhaus and Wright, 1983). The modified carrier molecules were completely inactive, while the remaining active carrier molecules exhibited normal transport and binding parameters. Substrate binding could not be regenerated by treatment with hydroxylamine. Thus it was suggested that there is no direct evidence for the involvement of histidine residues in coupling transport to the pmf.

Deuterium oxide has no effect on the initial rate of accumulation in vesicles, but that of efflux is retarded over three fold in deuterium (Viitanen et al., 1983). Also, counterflow at saturating external lactose concentrations is independent of the solvent, but at subsaturating concentrations the efficiency of counterflow is stimulated by deuterium. These observations indicated that protonation or deprotonation was not rate determining for uphill lactose transport in the presence of a pmf.

Experiments reconstituting lactose carrier transport activity in proteoliposomes composed of different phospholipids have indicated that the amino group of amino phospholipids (e.g. phosphatidylethanolamine, phosphatidylserine) is required for the full function of the lactose carrier (Chen and Wilson, 1984). Blockage of the amino group by either trinitrophenylation or methylation greatly reduces the function of the carrier. Partial restitution of transport activity was obtained by detergent extraction of the carrier from inactive proteoliposomes and reconstitution into proteoliposomes containing normal E. coli lipid.

The N-ethylmaleimide reactive sulphydryl group has been shown not to be freely accessible from the external medium using various other maleimides including hydrophobic and impermeant ones (Cohn et al., 1981). Also, the substrate binding site of the lactose carrier has been implicated to lie within a segment(s) of the protein which is embedded in the lipid bilayer, since proteolysis of vesicles had no effect on substrate binding (Goldkorn

et al., 1983). The location of the galactoside binding site was further identified as being within the carrier with respect to both the lipid and aqueous phases, and the suggestion was made that this binding site communicates with the aqueous phase through a pore (Mitaku et al., 1984). These workers determined the position of fluorophores specifically bound to the lactose carrier by the use of various collisional quenchers. The fluorophores were the pyrenyl residue of N-(1-pyrenyl)-maleimide attached to Cys^{148}, which is presumed to be at or near the galactoside binding site, and the dansyl moieties of a series of fluorescent substrate molecules. The position of these fluorophores within the lipid phase was determined using nitrosyl-labelled fatty acids and phospholipids carrying the quencher at different positions in the acyl chain, and the accessibility of the fluorophores from the aqueous phase was ascertained using a water-soluble quencher, the N-methylpicolinium ion.

Site directed mutagenesis of Cys^{148} has been used to convert this residue to glycine (Trumble et al., 1984) and to serine (Neuhaus et al., 1985). In both cases the mutant carrier was much less sensitive to inhibition by N-ethylmaleimide, and TDG provided no protection against the residual slow inactivation, confirming that Cys^{148} is the residue labelled by N-ethylmaleimide. However, although the rate of transport was found to be reduced, both mutants were capable of active transport. Thus it would appear that Cys^{148} is not essential for lactose-proton symport. Neuhaus et al. suggested that the inhibition caused by N-ethylmaleimide may result from alkylation of the side chain prohibiting access of the substrate to the binding site by steric hindrance.

Mutants of the lacY gene have been described which transport maltose (not normally a substrate for the lactose carrier) independently of all normal maltose transport system components in a strain constitutive for lacY expression (Shuman and Beckwith, 1979). The altered lactose transport system still transported lactose, had a decreased ability to transport melibiose, and had an increased affinity for ONPG and TMG. Similar mutants able to grow on maltose have been described by Mieschendahl et al.(1981). A malB strain having a wild type lacY gene could apparently grow on 0.1M maltose plus IPTG (a gratuitous inducer of the lac operon) but not on normal concentrations (5mM) required for growth in a $malB^+$ strain, indicating a low affinity for

maltose in the wild type lactose carrier. The DNA sequences of some of these mutants have been determined (Markgraf et al., 1985). A mutant in lacY able to grow on 5mM maltose in the absence of the maltose transport system, but only able to grow on elevated concentrations of lactose (0.1M), was shown to have a single base change resulting in Thr^{266} becoming an isoleucine residue. Another mutant only able to grow on high concentrations of lactose but unable to grow on maltose, was shown to have a single base change such that Tyr^{26} became an aspartate residue. A further 18 mutants in the lactose carrier able to transport maltose have been described and their DNA sequences determined (Brooker and Wilson, 1985). In these cases single base pair changes caused replacement of Ala^{177} with valine or threonine, or of Tyr^{236} with phenylalanine, asparagine, serine or histidine. The mutants at position 177 retained the ability to transport galactosides such as lactose or melibiose at rates similar to the wild type. However, the mutants at position 236 were defective in the ability to transport galactosides. The residues Ala^{177} and Tyr^{236} are located on adjacent hydrophobic segments of the lactose carrier protein that are predicted to span the membrane, further suggesting that the substrate recognition site of the lactose carrier is located within the lipid bilayer.

An energy uncoupled mutant of lacY was isolated and found to have a greatly reduced ability to accumulate lactose analogues (approximately 10% of the capacity of the parental strain) despite an increase in total membrane carriers (Wong et al., 1970). Exit of TMG occurred faster than in the parental strain, explaining the inability of the mutant to maintain high intracellular concentrations of galactosides. Growth was normal at high concentrations of lactose but the mutant was unable to grow at low concentrations. The defect was inferred to be in the coupling of metabolic energy to lactose transport, and was consistent with the hypothesis that the energy requiring step in transport is a change from high affinity to low affinity for substrate at the inner surface of the membrane (Winkler and Wilson, 1966).

Vesicles of the uncoupled mutant described by Wong et al.(1970) have been shown to be defective in the ability to accumulate TMG in the presence of a pmf, and exhibited reduced TMG-induced proton influx (Herzlinger et al., 1985). However, vesicles of the mutant strain bound p-nitrophenyl-α-D-galactoside and the monoclonal antibody 4B1, and catalysed facilitated

diffusion and equilibrium exchange, as well as the wild type vesicles. The effect on counterflow at saturating and subsaturating substrate concentrations was similar to that of deuterium oxide (Viitanen et al., 1983), and the isotope had no effect on the mutant vesicles.

It was suggested that the mutation results in a lactose carrier molecule with a higher pK_a, limiting the rate of deprotonation or altering the equilibrium between protonated and deprotonated forms (Herzlinger et al., 1985). A monoclonal antibody against a cytoplasmic epitope which is partially related to the C-terminus of the lactose carrier exhibited reduced binding to inside out mutant vesicles. In view of the finding that carboxypeptidase digestion of the C-terminus of the wild type carrier does not inhibit transport (Carrasco et al., 1984b), it was suggested that the mutation probably causes a conformational change in the C-terminal region rather than actually occurring there. This view was supported by the finding that polyclonal antibodies directed against specific hydrophilic segments elsewhere in the protein exhibited altered binding to the mutant vesicles compared with the wild type, indicating a significant alteration in the conformation of the protein.

Another mutant was described as being energy uncoupled, having a severe defect in active transport of TMG with no reduction in the total membrane carriers (Wilson et al., 1970). Again TMG exit occurred faster than in the parent, and accumulated to only 17% of the level observed for the parental strain.

The effective inflow of protons caused by the addition of TMG was much reduced in both mutants described compared with the parental strains (West and Wilson, 1973). In both mutants the defect in accumulation was most striking at about pH5.8 and became progressively less marked at higher pH. It was shown that there was no loss of the ability to establish a pH difference across the membrane. The suggestion was made that there was a defect in energy coupling in these mutants such that the stoichiometry for protons to substrate was decreased several fold. The mutation was shown to map at the extreme C-terminus of the lacY gene or to be a deletion between the lacY and the adjacent lacA genes by Hobson et al.(1977).

However, this mutant has subsequently been shown to possess normal coupled symport and its phenotype has been attributed to the lower rate of transport it exhibits (Wright and Seckler, 1985). Both the maximum velocity for active transport and the level of accumulation of TDG were lower in the mutant at pH6, although no different from the parental strain at pH8-9. At pH6 and low galactoside concentrations the symport stoichiometry was shown to be close to 1:1 for both the mutant and the parental strains. The mutant carrier was shown to be similar to the wild type in terms of apparent molecular mass, C-terminal sequence, and level of incorporation into the membrane, using a C-terminal directed antibody.

1.1.5 Melibiose-Na^+ cotransport

The cotransport of sodium ions and TMG by the melibiose transport system of Salmonella typhimurium was first identified by Stock and Roseman (1971). A second permease for TMG in E. coli was discovered which was inducible by melibiose or galactinol, and in E. coli K12 was temperature sensitive (Prestidge and Pardee, 1965). Using strains deleted for the lactose transport system it was shown that the transport of TMG by this melibiose transport system of E. coli was dependent on sodium ions (Tsuchiya et al., 1977). An inwardly directed electrochemical gradient of sodium ions resulted in accumulation of TMG. Conversely, addition of TMG gave a transient acidification of the medium in the presence of sodium ions. In the absence of sodium ions no pH change was observed. This acidification was attributed to sodium ions moving inwards causing a charge imbalance such that protons moved out of the cell to compensate. These effects were specific to sodium or lithium ions; a pH gradient caused no uptake of TMG.

Use of a sodium electrode provided direct evidence for sodium cotransport. On addition of TMG the sodium ion concentration in the medium decreased. The stimulation of TMG uptake by sodium and lithium ions was due to an increase in the affinity of the carrier for the substrate; the K_m decreased while the V_{max} remained essentially unchanged (Lopilato et al., 1978).

The decrease in external sodium ion concentration on addition of TMG was only observed for induced cells, and ocurred with melibiose, TDG, methyl-α-D-galactoside, methyl-β-D-galactoside, and galactose in addition to TMG. In the

absence of sodium or lithium ions, an alkalinization of the medium indicating sugar-proton cotransport was observed but only for melibiose and methyl-α-D-galactoside (Tsuchiya and Wilson, 1978). However, when sodium ions were present monitoring the pH revealed an acidification on addition of melibiose, as with TMG, consistent with the entry of sodium ions with the sugar.

Although both lithium and sodium ions stimulated TMG uptake by the melibiose transport system, lithium ions inhibited the transport of melibiose (Tsuchiya et al., 1978). Thus, the cation specificity for the melibiose carrier varies with different substrates. By measuring radioisotope-labelled melibiose transport it was found that in the presence of lithium ions the K_m of melibiose transport is reduced to approximately the same value as by sodium ions, but with lithium ions the V_{max} is also severely reduced (Tanaka et al., 1980). It was also shown that in a sodium free medium melibiose or methyl-α-D-galactoside stimulated the rate of proton entry in response to an acid pulse with the same characteristics as the melibiose carrier (Tsuchiya et al., 1980), indicating an obligatory coupling between protons and melibiose translocation in the absence of sodium ions.

Use of a lithium ion selective electrode has shown that lithium influx occurs with TMG, methyl-α-D-galactoside, methyl-β-D-galactoside and galactose in the presence of a functional melibiose carrier (Tsuchiya et al., 1983). Melibiose itself gave a small lithium uptake indicating inefficient melibiose-lithium cotransport and maybe providing an explanation of the inhibitory effect of lithium ions on melibiose transport.

Various mutants in the cation coupling of melibiose transport have been described. In one case the mutants had lost the ability to utilise protons, had retained the ability to utilise sodium ions, and had gained the ability to utilise lithium ions for cotransport with melibiose (Niiya et al., 1982). In addition, TMG and methyl-β-D-galactoside were no longer substrates (Tsuchiya et al., 1983). Comparison of five such independent mutants with the wild type nucleotide and amino-acid sequences of the melB gene (Yazyu et al., 1984) indicated that in each case Pro^{122} was replaced with serine (Yazyu et al., 1985). A second mutation in these strains in the sodium-proton exchanger was shown to be unlinked (Shiota et al., 1984). Another category of mutants had become resistant to lithium ions, but could not utilise lithium in

cotransport with melibiose; these mutants had also lost coupling of melibiose transport to protons although retaining coupling to sodium ions (Shiota et al., 1985).

Experiments using vesicles have indicated that energisation causes a large increase in the V_{max} of TMG-sodium cotransport with little or no apparent change in K_m (Cohn and Kaback, 1980). A similar result has been found for melibiose-sodium cotransport, although for melibiose-proton cotransport imposition of the membrane potential affected the apparent affinity constant selectively (Bassilana et al., 1985).

The gene product of melB has been indicated to have an apparent molecular mass of 30 000Da from SDS-polyacrylamide gel electrophoresis of proteoliposomes reconstituted from induced and uninduced preparations (Hanatani et al., 1984). Further experiments to substantiate this identification, for example, overexpression of the cloned gene, would be of value. The molecular mass from the DNA sequence is 52 202Da (Yazyu et al., 1984). This discrepancy is attributed to a similar effect as for LacY, an increased binding of SDS to the hydrophobic protein. The carrier has also been solubilised and reconstituted into proteoliposomes (Tsuchiya et al., 1982; Wilson et al., 1985).

1.2 Xylose

Xylose enters the cell by an active transport system and is isomerised to D-xylulose then phosphorylated to xylulose-5-phosphate which enters the pentose phosphate pathway. The enzymes of xylose metabolism and transport in E. coli have been shown to be coordinately controlled and induced by xylose (David and Weismeyer, 1970). In addition the induction of all three enzymes (transport, xylose isomerase, and xylulokinase) was sensitive to actinomycin D and chloramphenicol and was, therefore, attributed to de novo enzyme synthesis. The enzymes were also all subject to catabolite repression by glucose and by glycerol. Pleiotropic negative mutants exhibited a reversion frequency comparable with that of mutants deficient in only one of the three enzymes, indicating a common controlling element.

The equivalent systems in Salmonella typhimurium have also been shown to be xylose inducible (Shamanna and Sanderson, 1979a). The transport of xylose was found to be inhibited by D-xylitol and to a lesser extent by L-arabinose, and it was assumed that this was a result of competition. Osmotic shock reduced the transport by 50% and a binding protein was detected. Xylose transport was claimed to be by only one system having a K_m of 0.41mM, although it was reported that E. coli possessed two transport systems with K_ms of 110 and 24μM; it should be noted that 24μM is an unusually high K_m for a binding protein dependent system. The transport in S. typhimurium was induced by L-arabinose in addition to D-xylose. The genetics of xylose metabolism and transport in S. typhimurium indicated that the genes involved were clustered at 78min on the chromosome map, and under positive control by the regulatory gene xylR (Shamanna and Sanderson, 1979b). The presence of nonsense mutations which were pleiotropic negatives, and the high frequency of pleiotropic negative but absence of constitutive mutants, provided evidence of positive control; F'$xylR^+$/xylR partial diploids were Xyl^+ indicating that the pleiotropic negatives could not be due to a super-repressor for negative control as this would be dominant over $xylR^+$.

Evidence for a xylose-proton symport system in E. coli has been provided by Lam et al. (1980). An alkaline pH change which was susceptible to the action of uncouplers was elicited on addition of xylose to energy-depleted cells. Transport of [^{14}C]-xylose occurred in energised vesicles and uptake in both cells and vesicles was inhibited by uncouplers. Xylose transport was not susceptible to inhibition by fluoride or arsenate, however. It was concluded that transport was energised by the proton motive force rather than directly by phosphorylation or by the PTS. In addition, both the induction of transport and the transport activity were shown to be specific for xylose of the sugars tested (D-ribose, L-arabinose, fucose, IPTG).

The presence of a binding protein dependent xylose transport system has been demonstrated by the isolation of a D-xylose binding protein (Ahlem et al., 1982). The binding protein had an apparent molecular mass of 37 000Da from SDS- polyacrylamide gel electrophoresis and amino acid analysis, and was specific for xylose (out of a wide range of sugars tested) with a K_d of 0.6μM. Two transport systems for xylose were indicated by uptake studies, with reported K_m values of 5 and 25μM. Osmotically shocked cells showed

preferential reduction of the high affinity system, in line with the general nature of binding protein dependent systems.

It would, therefore, appear that E. coli possesses both a proton symport system and a binding protein dependent system for the transport of D-xylose.

The gene coding for xylose isomerase, xylA, has been cloned by several groups, and the DNA sequence has been determined (Schellenberg et al., 1984). A sequence resembling a rho-independent transcriptional terminator was identified downstream of the xylA gene and preceding the region coding for the xylB gene. The DNA sequence reported by Briggs et al. (1984) appears to contain some errors and as a result the protein sequence deduced from it terminates prematurely. However, contrary to the finding of Schellenberg et al., cloning and identification of promoters indicated that the xylA and xylB (xylulokinase) genes were organised as an operon with xylA promoter proximal (Rosenfeld et al., 1984). The cloned E. coli genes also complemented S. typhimurium mutants in xylose isomerase, xylulokinase, and a xylose regulatory gene (Maleszka et al., 1982).

1.3 Bacteriophages Mud(Ap^{R}lac) and λplacMu

1.3.1 Bacteriophage Mu

Bacteriophage Mu may lysogenise on infecting E. coli, but is unusual among temperate phages in that it may integrate at random anywhere around the chromosome. If there is a specific host sequence for an integration site it must be very short. Mu, or the mutator phage, gives approx. 2% lysogens with a new nutritional requirement, mutagenesis being caused by prophage insertion into the relevant gene. Within a lysogenic population the mutation frequency of any particular gene is 50-100 times greater than its spontaneous mutation frequency (Howe and Bade, 1975), although this frequency may be reduced if the gene concerned is transcribed at a high rate during Mu infection. In addition Mu-induced mutations are very stable with a reversion frequency of less than 10^{-9}-10^{-10} per colony forming unit. The low frequency of curing may be due to lack of excision, lack of rejoining of the ends of the chromosome after excision, or death of the cell due to lytic development of the excised phage. All these factors make Mu a useful mutagen.

The DNA of Mu exists in a double-stranded linear form when isolated from phage heads and contains E. coli DNA particularly at one end, the variable end, as 3-4% of the total Mu DNA. The prophage DNA is colinear with the DNA from particles, i.e. it is not a circular permutation, but the variable end is not found in the prophage. It is thought that in lytic cycle replication of Mu DNA occurs in circles of alternating phage and host DNA, and then packaging of headfuls results in the inclusion of host DNA, mainly at the variable end.

The formation of lysogens is not efficient but the proportion of lysogens among survivors may be increased by prolonged incubation (e.g. overnight) of the phage-cell mixture. Multiple cycles of infection ensure the selective survival of lysogens, which are immune to superinfection.

Mutations of Mu include the temperature inducible c^{ts} mutation. Phage containing this mutation can lysogenise at low temperature (30°C) but not at high temperatures (42°C), and on exposure to high temperature lysogens enter lytic phase.

1.3.2 Bacteriophage Mud(Ap^Rlac)I

Casadaban has used phages Mu and λ to fuse the lac structural genes to selected promoters to study gene regulation by assaying β-galactosidase (Casadaban, 1976a&b). More recently the lac structural genes have been incorporated into the genome of a Mu already carrying a gene from Tn3 coding for ampicillin resistance to form a specialised transducing phage (see Figure 1.4) (Casadaban and Cohen, 1979). The properties of this Mud(Ap^Rlac)I with respect to integration are normal for Mu (section 1.3.1). Lysogens of Mud(Ap^Rlac)I can be readily selected by ampicillin resistance. There are no transcriptional termination sites between the beginning of the lac genes and the end of Mu nearest to this, either in Mu itself or in the trp genes preceding the lac genes. When integration has occurred in the orientation of transcription of the gene involved, the lac structural genes will be expressed solely under the control of that promoter, forming an operon fusion mutant. Thus the activity of that gene may be followed by monitoring β-galactosidase levels. The E. coli strain infected must have the lac operon deleted otherwise recombination may occur at that region of homology, and there will be a background level of β-galactosidase activity. This Mu phage

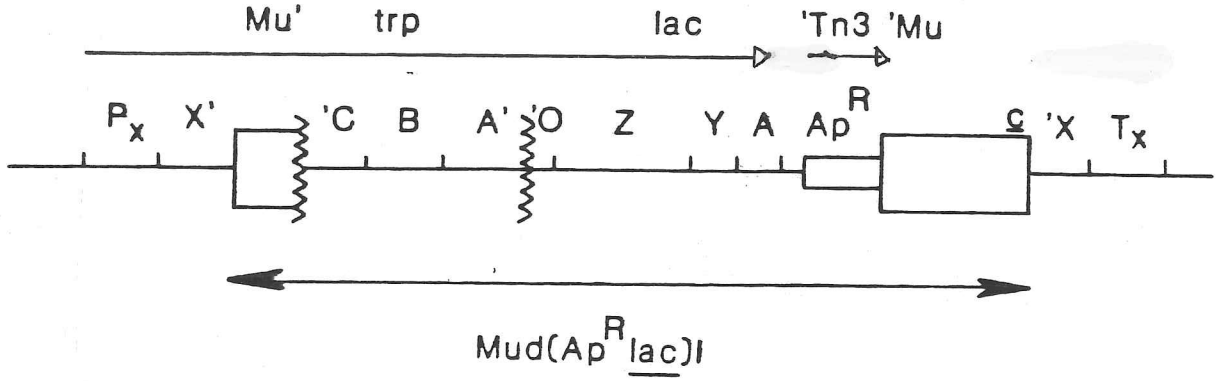


Figure 1.4 Mud(Ap^Rlac)I insertion into a gene X in the orientation that fuses transcription from the promoter P_x to lac

The insertion must be between the promoter and the terminator (T) of the gene but it need not be within the structural gene. A prime next to a genetic symbol indicates that it is deleted or interrupted on the side on which the prime is written. (Not drawn to scale.) The open arrows indicate directions of transcription.

has been used to study the regulation of the L-arabinose transport operons (Kolodrubetz and Schleif, 1981), where it was shown that the Mud(Ap^R*lac*)I insertion permitted a reliable measurement of the expression of a gene by assaying β-galactosidase activity.

1.3.3 Bacteriophage Mud(Ap^R*lac*)II

A further derivative of Mu, Mud(Ap^R*lac*)II, having no transcriptional or translational termination signals between the end of Mu and the *lac* structural genes has been constructed (Casadaban and Chou, 1984). This phage is lacking the first eight codons of *lacZ*, which are not necessary for β-galactosidase activity. Up to 26 amino acids can be removed and replaced by apparently any number of amino acids from another protein without eliminating enzymatic activity (Brickman *et al.*, 1979). Insertion of Mud(Ap^R*lac*)II into a gene in the correct orientation and reading frame results in gene fusion and the formation of a hybrid protein, having the N-terminus of the gene into which the phage is inserted fused to β-galactosidase.

1.3.4 Bacteriophage λ

Phage λ is also a temperate phage of *E. coli*, but it differs from Mu in that it only inserts into the chromosome at a specific site, *att*, between *gal* and *bio*. In the event of this site being deleted insertion may occur at a few other secondary attachment sites, but not at random. The DNA in the prophage state is a circular permutation of the DNA in the phage particle, through the cohesive ends, or *cos* sites. Thus, in the prophage the two ends of the phage DNA are defined by the integration sites, whereas in the phage particle the ends are defined by the *cos* sites.

The gene *cI* is responsible for the maintenance of the lysogenic state, by repressing the expression of the genes involved in lytic development. A mutation in this gene, *cI*^{857}, codes for a temperature sensitive repressor protein. This allows induction of the lytic state by raising the temperature at which the lysogen is growing. Another useful λ mutation is the *s*^{7} mutation which prevents the cell from being lysed by the phage inside it.

On entering the lytic pathway, early replication in the θ form proceeds to relication by means of a rolling circle for a gam$^+$ phage. Oligomeric DNA is required for packaging. Therefore, for a gam phage the recA$^+$ function of the host is required for homologous recombination to occur between single copies of the phage DNA in order to allow packaging, unless the host is recBC. In a recBC host a gam phage can replicate by the rolling circle method as gam$^+$ normally acts to inhibit RecBC$^+$ nuclease, which otherwise prevents the transition from the θ form to the rolling circle form of DNA replication.

Phage λ has the advantage that when a lysogen is induced to excise, excision may occur at the wrong site such that the DNA carried in the phage particle is partly phage DNA and partly bacterial DNA. The lysogen may pick up DNA from either side of its point of integration, but the amount of DNA carried by the phage is limited by packaging constraints. This DNA will then be replicated and transferred as part of the phage DNA, forming a specialised transducing phage.

Phage λ cannot grow on a strain lysogenic for phage P2 (Spi$^+$, sensitive to P2 inhibition), unless it has lost the red and gam functions (Spi$^-$) (Hershey, ed. 1971; Hendrix et al., ed. 1983). However, the correct amount of DNA is required for packaging, so chromosomal DNA must be picked up from the other end of λ. Thus the Spi$^-$ selection yields phage carrying DNA from only one side of its point of integration.

1.3.5 Bacteriophage λplacMu

The advantages of both Mu and λ have been combined in phage λplacMu (Bremer et al., 1984). This phage is a λ plaque-forming phage which contains sequences from Mu enabling it to integrate into the chromosome at random by the Mu transposition system. Phage λplacMu also carries the lacZ and lacY genes adjacent to the terminal Mu sequence as in Mud(Ap^{R}lac)II, with no transcriptional or translational termination signals (Figure 1.5 from Bremer et al., 1984). Thus insertion of λplacMu in the correct orientation and reading frame creates protein fusions. Induction of such a lysogen and selection for Spi$^-$ phage yields specialised transducing phage carrying the entire gene fusion and any adjacent gene(s) (Figure 1.5).

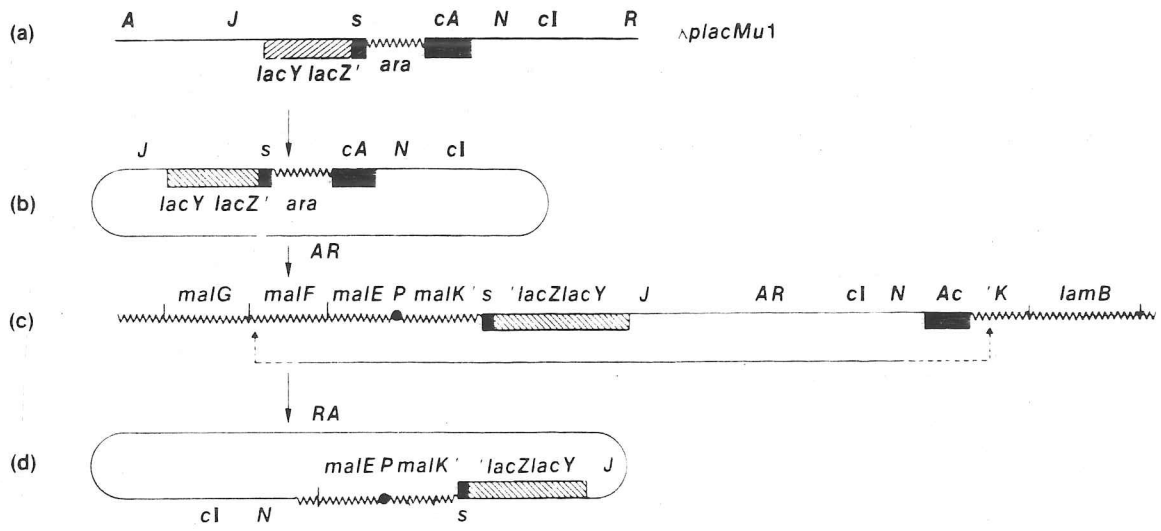


Figure 1.5 Insertion of λplacMu into malK and isolation of transducing phages (from Bremer et al., 1984)

Upon introduction of the mature λplacMu chromosome (a) into the cell, it circularises (b) and inserts into the malK gene (c), resulting in a Lac^+ Mal^- malK-lacZ gene fusion. Upon u.v. induction, the prophage is excised by an illegitimate recombination event (d), leading to a Lac^+ malK-lacZ specialised transducing phage which also carries the malE gene. The black dot represents the divergent promoters (P) of the malB region.

1.4 Aims of this Work

At the start of this study little was known about xylose transport in *E. coli* except that there were two systems, a proton symport system and a binding protein dependent system (section 1.2). The positions of the genes coding for these two transport systems on the chromosome were unknown. Therefore, the initial research was aimed at making Mud(Ap^R*lac*) insertion mutants in the genes coding for the xylose transport systems which would then enable the genes to be located on the chromosome. This knowledge would permit experiments aimed at cloning the gene(s). In addition, it was hoped to identify the XylE protein by a method similar to that used to identify the LacY, AraE and GalP proteins using radioisotope labelled N-ethylmaleimide.

The main aim of this study was to clone and sequence the DNA of the gene coding for xylose-proton symport. It was hoped that comparison of the sequence obtained with those of other proton symport systems would identify residues which might be important in substrate binding or the transport process. The effect of altering such residues by, for example, site-directed mutagenesis, would provide a means of exploring the roles played by these residues. In addition, the availability of the cloned gene should enable overexpression of the XylE protein to be achieved, making experiments to purify the protein worthwhile. Ultimately, structural studies on the purified protein could be undertaken to probe stucture-function relationships.

CHAPTER 2

MATERIALS AND METHODS

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

The chemicals used were of Analar grade wherever possible. D-[U-^{14}C]-xylose, [D-glucose-1-^{14}C]-lactose, N-ethyl-[2,3-^{14}C]-maleimide, adenosine 5'-α-[^{35}S]-thiotriphosphate, 'SmaI cut and phosphatased' M13mp10 and restriction enzymes were from the Radiochemical Centre, Amersham, Bucks. N-[ethyl-2-^{3}H]-maleimide was from New England Nuclear, Swindon, Wilts. DNA polymerase I (Klenow fragment) and T4 DNA ligase were the kind gifts of Dr. T. Hunt; 'SmaI cut and phosphatased' M13mp8 was the kind gift of Mr. M.C.J. Maiden. Dideoxynucleotide triphosphates and deoxynucleotide triphosphates were from BCL-The Boehringer Corporation (London) Ltd., Lewes, East Sussex, and M13 sequencing primer was from New England Biolabs (distributer C.P. Laboratories Ltd., Bishop's Stortford, Herts.). Affinity matrix p-aminophenyl-β-D-thiogalactoside-agarose was from Sigma Ltd., Poole, Dorset; DEAE-Toyopearl-650M was from TSK Toyo Soda MFG Co. Ltd., Tokyo, Japan; and Bio-gel A-5m was from Bio-Rad Laboratories Ltd., Watford, Herts. Anti-β-galactosidase monoclonal antibody-Sepharose-4B resin was the kind gift of Dr. H. Chase. Bacto Yeast Extract was from Difco Laboratories, Detroit, Michigan, USA. Agar, Nutrient Broth, Tryptone, Casein Hydrolysate, and Nuflow cellulose acetate membrane filters (pore size 0.45μm) were obtained from Oxoid Ltd., Basingstoke, Hants. Millipore cellulose nitrate membrane filters (pore size 0.45μm) were from Millipore SA, Molsheim, France.

2.1.2 Organisms and media

The strains of Escherichia coli K12 used are listed in Table 2.1. The genetic terminology is as defined by Bachmann (1983) or is defined in the text. Strains were maintained on nutrient agar slopes, or in glycerol deeps at -80°C, and streaked out to obtain single colonies before use.

The media used are defined in Table 2.2. Minimal medium was supplemented with the appropriate amino-acids at 80μg ml^{-1} and a carbon source at a

Table 2.1 Strains

AB264	(Muc$^+$)	
AB1157	thr leu his arg pro lac ara mtl xylA gal	
MAL103	Δ(proAB lacIZYA)X111 araD139 rpsL araB::Muc^{ts} (Mud(Ap^{R}lac)I)	
NB1	zwf pgi zjb::Tn10	
PB13	recA srl::Tn10	
PO85	Hfr malBΔ107(malF-lamB) his	
SP2	(his-gnd)$^\Delta$ lac$^\Delta$ araD ptsF ptsM argH ppc (xylE-lamB)$^\Delta$ xylF	
JM559	(his-gnd)$^\Delta$ fda^{ts} Hfr Cavalli	
JM2087	(his-gnd)$^\Delta$ lac$^\Delta$ araD fda ptsF ptsM rpsL	
JM2235	leu thr argH mtl mal pheA cysA his trp gal lac rpsL	
JM2336	as JM2087 but fda$^+$ xylE::Mud(Ap^{R}lac)II	
JM2349	argH metA aceA (Muc$^+$)	
JM2365	(his-gnd)$^\Delta$ lac$^\Delta$ araD ptsF ptsM (metA-xylE)$^\Delta$	
JM2390	(his-gnd)$^\Delta$ lac$^\Delta$ araD ptsF ptsM argH ppc (xylE-malB)$^\Delta$	
JM2433	araAB araD (his-gnd)$^\Delta$ lac$^\Delta$ rpsL	
EJ13	as JM2087 but (Mud(Ap^{R}lac)I)	JM2087 → Xyl^R(Gly)
EJ14	as JM2087 but xylE::Mud(Ap^{R}lac)I	JM2087 → Xyl^R(Gly)
EJ15	as JM2087 but fda$^+$	P1.K10 x JM2087
EJ17	as EJ13 but fda$^+$	P1.K10 x EJ13
EJ18	as EJ14 but fda$^+$	P1.K10 x EJ14
EJ20	as JM2235 but xylE::Mud(Ap^{R}lac)I	P1.EJ18 x JM2235
EJ21	as EJ20 but arg$^+$	Hfr JM559 x EJ20
EJ22	as JM2349 but xylE::Mud(Ap^{R}lac)I	P1.EJ21 x JM2349
EJ23	as PO85 but (Muc$^+$)	
EJ25	his malB (Mud(Ap^{R}lac)I) (Muc$^+$)	P1.EJ18 x EJ23
EJ26	his malB zjb::Tn10 (Muc$^+$)	P1.NB1 x EJ25
EJ32	as EJ14 but Δ(ace pgi malB) λ^S	EJ14 → Th^R & (Xyl + Lac)^R
EJ38	his zhj::Tn10	
EJ39	as AB1157 but zhj::Tn10	P1.EJ38 x AB1157

continued

Table 2.1 continued

EJ40	as SP2 but *zhj*::Tn10	P1.EJ38 x SP2
EJ54	as JM2336 but *xylA zhj*::Tn10	P1.EJ39 x JM2336
EJ61	as JM2087 but *xylABR*::λ*plac*Mu	JM2087 → Xyl^R (Gly)
EJ62	as JM2087 but *xylE*::λ*plac*Mu	JM2087 → Xyl^R (Gly)
EJ63	as EJ62 but *zjb*::Tn10	P1.NB1 x EJ62
EJ64	as EJ62 but $xylE^+$ *zjb*::Tn10	P1.NB1 x EJ62
EJ65	as EJ15 but *xylE*::λ*plac*Mu	P1.EJ63 x EJ15
EJ68	as JM2433 but *xylA zhj*::Tn10	P1.EJ39 x JM2433
EJ70	as JM2365 but *xylF zhj*::Tn10	P1.EJ40 x JM2365
EJ71	as EJ32 but fda^+	P1.EJ15 x EJ32
EJ80	as JM2390 but *xylG*::Mud(Ap^R*lac*)I	JM2390 → Xyl^R (Ace)
EJ81	as EJ80 but $xylE^+$ *zjb*::Tn10	P1.EJ64 x EJ80
EJ82	as EJ39 but (Muc^+)	
EJ83	as EJ68 but (Muc^+)	
EJ84	as JM2433 but *malK*::λ*plac*Mu	JM2433 → λ^R
EJ93	as EJ71 but (Muc^+)	
EJ94	as EJ93 but *xylG*::Mud(Ap^R*lac*)I	P1.EJ80 x EJ93
EJ95	as EJ94 but *recA srl*::Tn10	P1.PB13 x EJ94
EJ96	as EJ95 but srl^+ Tc^S	P1.EJ94 x EJ95
EJ97	as EJ94 but (λ(*xylE*)φ(*malK'-lacZ*))	

Table 2.2 Media

2 x Nutrient Broth:	26g Oxoid No.1 nutrient broth
	2ml 2% cysteine in 0.5M HCl
	2ml 6% thymine in 0.5M NaOH
	made up to 1l
2 x Tryptone Broth:	20g tryptone
	10g NaCl
	2ml 1M $MgSO_4$
	0.5ml 0.2mg ml^{-1} thiamin
	2ml 2% cysteine in 0.5M HCl
	2ml 6% thymine in 0.5M NaOH
	made up to 1l and pH7
2 x Tryptone-Yeast Extract:	10g tryptone
	10g yeast extract
	5g NaCl
	made up to 1l
CY Medium:	10g Casamino acids
	5g yeast extract
	3g NaCl
	2g KCl
	made up to 1l and pH7
2 x minimal medium:	3.81g KH_2PO_4
	10.23g Na_2HPO_4
	2.70g NH_4Cl
	0.5ml 0.2mg ml^{-1} thiamin
	made up to 1l

2ml trace salts added per 100ml medium before use.

Trace salts:	0.4ml 1M HCl
	0.54g $CaCl_2.2H_2O$
	1.60g $MgSO_4.7H_2O$
	0.08g $FeSO_4.7H_2O$
	0.08g $MnCl_2.4H_2O$
	made up to 200ml

concentration of 10mM for monosaccharides or 5mM for disaccharides.

2.1.3 Preparation of subcellular vesicles

Spheroplasts were prepared by the procedure of Witholt et al. (1976), and these were used to make vesicles by the method of Kaback (1971). The principles behind these procedures are to form spheroplasts by digestion of the cell wall with lysozyme. Osmotic shock of these spheroplasts results in lysis followed by the resealing of membrane fragments into vesicles with the concomitant loss of cytoplasm and periplasm. Vesicles prepared by this method have been shown mainly to have the same orientation as the membrane of the intact cell (Stroobant and Kaback, 1975). The vesicles may be stored at -80°C with retention of transport activity; they are devoid of any transport activity dependent on binding proteins as these are removed by the osmotic shock procedure.

2.1.4 French press membrane preparation

Cultures (10l) were grown in minimal medium (1l per 2l flask) appropriately supplemented overnight at 30°C and 220 revolutions per minute. The cells were harvested at 3 800g, 15min, 4°C in a Sorvall GS3 rotor, and washed in 2.5l 50mM Tris-HCl/1mM magnesium sulphate/50mM 2-mercaptoethanol, pH7.2. The cells were then resuspended to a final volume of 70ml in the same buffer with the addition of protease inhibitors (0.1mM PMSF, 1mM benzamidine), and passed through the French Pressure cell in two batches at 20 000 psi. The resulting suspension was centrifuged at 20 000g, 20min, 4°C in a Sorvall SS34 rotor to remove cell debris. The supernatant was then centrifuged at 145 000g, 60min, 4°C in a MSE SS65 10x10ml titanium angle rotor (with the addition of fresh PMSF) to obtain the membranes. The reddish pellets were resuspended in 30ml buffer plus protease inhibitors and re-centrifuged at 145 000g. The resulting pellets were resuspended in a final volume of 10ml and stored in aliquots at -80°C. Some preparations were of half this size.

2.2 Genetical Techniques

2.2.1 Mutagenesis by phage Mud(Ap^{R}lac)I

An overnight culture in Tryptone Broth (see Table 2.2) supplemented with 10mM sodium gluconate of strain MAL103, the Mu lysogen (Casadaban and Cohen, 1979), was diluted 50-fold into more of the same medium and incubated, with shaking, at 30°C for 2h. The culture was transferred to 42°C (and bubbled with air) for 20min to induce the prophage. It was then transferred to 37°C and incubated for 1h, with shaking, to allow lysis. Chloroform (5-10 drops) was added and the mixture whirlimixed briefly to kill any surviving bacteria. The chloroform was allowed to settle out by standing for 10min, and then half the supernatant was pipetted into a centrifuge tube and cell debris removed by centrifugation (MSE bench centrifuge, full speed, 10min).

A mixture of 0.1ml of 0.05M calcium chloride/0.1M magnesium chloride (required for phage attachment), 0.1ml phage suspension, and 1.0ml of an overnight culture of the strain to be infected was incubated at 30°C without shaking for 20min (to allow attachment and infection by phage).

Nutrient broth (10ml) was added, and the resulting suspension dispensed into 10 test-tubes and incubated overnight at 30°C to allow expression of Mu genes and genotypic and phenotypic segregation of the bacteria. Separation at this stage ensures that mutants obtained from each tube are independent of those from other tubes, as the 20min allowed for infection is insufficient for the bacteria to have divided. Multiple infection is unlikely to have occurred owing to an immune system set up once a cell has been infected with a phage (Howe and Bade, 1975).

A sample (0.1ml or 0.2ml) of each culture was plated onto the selection medium. Single colony isolates were then patched out and screened on various media.

2.2.2 Propagation and titration of phage λ

Propagation and titration of phage λ was performed according to the method of Miller (1972).

2.2.3 Mutagenesis by phage λplacMu

A mixture of 1.0ml overnight culture in Tryptone Broth (see Table 2.2, containing 10mM magnesium sulphate required for infection, and with 5mM maltose added to induce the λ receptor) of the strain to be infected and sufficient of λplacMu3 (Bremer *et al.*, 1984) stock to give a multiplicity of infection of two and of λpMu507.3 (Bremer *et al.*, 1984) stock to give a multiplicity of infection of five, was incubated at 37°C without shaking for 30min. Where necessary segregation was allowed as described for Mud(Ap^R*lac*) in section 2.2.1, before plating onto selection media.

2.2.4 Phage P1 mediated generalised transduction

Generalised transduction by phage P1 was carried out as described by Miller (1972).

2.2.5 Conjugation

The Hfr strains JM559 and KL16 were tested for fertile colonies as follows. Single colony isolates were patched in squares on nutrient agar, and incubated at 37°C for 4-6 hours, until the patches were just visible. The recipient strain (0.1ml per plate) was spread onto a selection plate for the nearest marker to the origin for each Hfr. The patched out colonies were then replica-plated to the relevant lawn, and incubated at 30°C overnight. Fertile colonies gave dots of growth within the patch on the selection plate. Unfortunately strain KL16 gave no fertile colonies, so the conjugation was performed with strain JM559 alone, and fortunately the *xylE* gene was found in the half of the chromosome satisfactorily covered by this Hfr (Chapter 3).

An overnight culture of the Hfr strain was diluted 50-fold into nutrient broth and grown at 37°C to exponential phase (such that the absorbance approximated that of a 10-fold dilution of the original culture). An overnight culture of the F^- strain was diluted 2-fold and incubated at 37°C for the same length of time (approx. 2h). Nutrient broth (2ml), 1ml of the culture of the F^- strain, and 1ml of that of the Hfr strain were mixed and incubated without shaking at 37°C for 2-3 hours.

One, 10^{-1}, 10^{-2}, 10^{-3} dilutions of the mating mixture (0.1ml) were plated onto the selection media, and 10^{-5}, 10^{-6}, 10^{-7} dilutions onto nutrient agar plates to indicate the number of cells. Control plates of the F^- and Hfr strains alone were also plated onto the selection media.

2.2.6 Lysogenisation with phage Muc$^+$

The cells from an overnight culture of AB264 (a Muc$^+$ lysogen) were removed by centrifugation. The supernatant was decanted onto chloroform, whirlimixed and the chloroform allowed to settle out. A soft (0.4%) agar overlay containing 0.2ml of this supernatant and 0.1ml of an overnight culture of a sensitive strain was poured on a nutrient agar plate, and incubated at 37°C for 7 hours. This was to amplify and concentrate the phage suspension. The phage were harvested as for phage P1 (section 2.2.4), an overlay poured using 0.1ml of this concentrated phage suspension, and this was incubated overnight at 37°C.

In addition, 3 drops of the concentrated phage suspension were dropped onto a Tryptone soft agar lawn of the strain to be lysogenised. This was not spread. If this did not give a satisfactory area of lysis this stage could be repeated with phage from the second concentrating step.

The soft agar from the area of clean lysis was cut out, suspended in 2ml nutrient broth and incubated at 37°C for 6 hours. This was to allow surviving bacteria to grow out of the agar. Single colony isolates were made by streaking out directly from the liquid culture; 40 colonies were patched out, with a control non-lysogenic strain. After 6-8 hours incubation at 37°C this was replica-plated onto a soft agar lawn of a sensitive strain, which was incubated overnight at 37°C. Muc$^+$ lysogens gave rise to plaques in the lawn.

2.2.7 Induction of λplacMu lysogens by ultra-violet light

An overnight culture of the lysogenic strain was diluted 50-fold into nutrient broth and incubated for 2 hours, to an absorbance at 680nm of 0.18-0.25. The cells were harvested by centrifugation, resuspended in an equal volume of 10mM Tris-HCl/10mM $MgSO_4$/0.1mg ml^{-1} gelatin, pH7.4, and u.v. irradiated at a distance of 15cm giving an intensity of 250J m^{-2} s^{-1} for 4min.

After 10min in the dark an equal volume of CY (Table 2.2) was added and the induction mixture divided into 10 aliquots. These were incubated at 37°C for approx. 5 hours, then lysed with chloroform and the debris removed by centrifugation.

Serial dilutions of these phage lysates were used in spot tests on lawns of the relevant indicator bacteria (a wild type strain and a P2 lysogen) to determine the quantity to be plated out. Using the appropriate dilution of phage lysate, Spi^- phage were selected using the P2 lysogenic strain RB341, and phage with inducible β-galactosidase were screened for using XGal + inducer in a lawn of a wild type strain, JM2433.

Suitable plaques were picked, propagated, and screened for inducible β-galactosidase (by stabbing into a lawn of JM2433 containing XGal and a similar lawn containing XGal + inducer), for being Spi^- (by stabbing into a lawn of RB341) and, for phage from the malK-lamB insertion, for carrying xylE (by spot tests on a lawn of a $xylE^{\Delta}$ xylFG strain, EJ70 or EJ94). Spi^- phage with inducible β-galactosidase, and an indication of complementing xylE if relevant, were plaque purified and rescreened. One in particular giving a positive result in the screen for xylE was plaque purified and rescreened a second time to be sure of having pure phage for DNA preparation.

2.3 β-galactosidase Assays

2.3.1 Plate assay for β-galactosidase

The plates to be tested were inverted over approx. 2ml chloroform on a watchglass for 10min, to render the cells permeable to subsequent reagents. A soft agar overlay containing 4mM ortho-nitrophenyl-β-D-galactoside was poured evenly over the plates. The presence of β-galactosidase gave rise to yellow colonies.

2.3.2 Quantitative assay for β-galactosidase

Samples were assayed and their activity calculated according to the method of Miller (1972), using the value of 380 for the extinction coefficient of nitrophenol.

2.3.3 Preparation to determine if β-galactosidase is membrane bound

Cultures were grown and harvested as described in section 2.6.1. The pellets were resuspended in Z buffer (Miller, 1972) to give an absorbance at 680nm of 20-30. Samples (8ml) were sonicated (MSE sonicator, probe diameter 9mm) on ice with four bursts of 30s on, 30s off, at 9-12μm. Samples were taken for β-galactosidase assay, and the remainder was centrifuged at 145 000g for 1 hour at 4°C as described in section 2.1.4. The supernatants were decanted, and the pellets resuspended in an equal volume of Z buffer. Samples of each were assayed for β-galactosidase, and the results expressed as a percentage of the total.

2.4 Protein Assay

Protein assays were performed by the method of Schaffner and Weissmann (1973) using bovine serum albumin standards. The principles of this method are to (i) precipitate the protein with trichloroacetic acid, (ii) filter the precipitate onto Millipore 0.45μm pore size membrane filters, (iii) stain it with naphthalene black, (iv) dissolve the protein-dye complex in alkaline ethanol, and (v) read the absorbance at 630 nm. The advantages of this method over others (e.g. Biuret, Folin) are that the results are less variable, few reagents interfere with the assay, and this technique has greater sensitivity.

2.5 Gel Electrophoresis

2.5.1 SDS-polyacrylamide gel electrophoresis

SDS-PAGE was performed according to the basic procedure of Laemmli (1970). The composition of the running gel was as given in Table 2.3 for the relevant percentage gel. The solution was degassed before the addition of 100μl 10% APS and 10μl TEMED immediately prior to pouring. The gel was overlayed with water-saturated n-butanol while it was setting to exclude air and to obtain a level surface. The butanol was carefully washed away with distilled water and the region above the running gel dried off before pouring the stacking gel. The stacking gel was made up as given in Table 2.3, with the addition of 50μl 10% APS and 5μl TEMED after degassing and immediately before pouring and

insertion of the comb.

Table 2.3 Composition of polyacrylamide gels

	Gel Concentration (%)					
	5%	7.5%	10%	12.5%	15%	Stacking Gel
30% Acrylamide (ml)	5.0	7.5	10.0	12.5	15.0	1.67
1% Bisacrylamide (ml)	7.8	5.8	3.9	3.1	2.6	1.3
1.5M Tris-HCl, pH8.7 (ml)	7.5	7.5	7.5	7.5	7.5	-
1.5M Tris-HCl, pH6.8 (ml)	-	-	-	-	-	1.25
Water (ml)	9.2	8.7	8.0	6.4	4.4	4.35

The running buffer consisted of 6g Tris-HCl, 28.8g glycine, 10ml 10% SDS in 1l; electrophoresis was carried out at 25 mA for 4-5 hours or at 60 V overnight. The samples were prepared by the addition of one third volume of 4 x SDS-dissolving buffer, and solubilised at 100°C for 2min or at 37°C for 30min as indicated in the text. The 4 x dissolving buffer consisted of: 1.26g glycerol, 1ml 10% SDS, 0.3ml water, 0.1ml 2-mercaptoethanol, 0.1ml 1M Tris-HCl, pH7.2 and 60µl 0.5% bromophenol blue. The gels were stained in Coomassie Brilliant Blue (0.1% in 45% methanol/10% acetic acid/45% water) at room temperature overnight or at 55°C with agitation for 30min, and destained in 5% acetic acid/7% methanol at 37°C or at 55°C. Alternatively, they were fixed in 4% sulphosalicylic acid/10% acetic acid at 55°C for 30min, soaked overnight in 50% methanol, then silver stained by the method of Wray et al. (1981).

2.5.2 Agarose gel electrophoresis

Agarose gel electrophoresis was performed as described by Maniatis et al. (1982). The agarose gels were run in TAE buffer consisting of: 242g Tris-HCl, 18.6g disodium EDTA, brought to pH7.7 with glacial acetic acid in a final volume of 1l. The gels were routinely stained after running in 1µg ml^{-1} ethidium bromide in the running buffer, except for minigels which were run in buffer containing 0.5µg ml^{-1} ethidium bromide. The DNA was then visualised using a transilluminator at 254nm, and photographed with a polaroid camera.

2.5.3 DNA sequencing gel electrophoresis

The reactions were run on a 4% polyacrylamide wedge gel (0.2-0.7 mm) containing 7M urea cast by the sliding plate method of LKB (Olssen et al., 1984). The gel was run at a circulating temperature of 55°C in 100mM Tris-HCl/83mM boric acid/0.5mM disodium EDTA, pH8.6. When using the 36 track comb the applied potential was 2 000V, with the 44 track comb it was 1 500V. These conditions were found to give optimal resolution of the sequencing ladder, allowing 300-400 bases per clone to be read routinely.

The gel was fixed and urea washed out by soaking in 10% acetic acid for 30-45min, dried on the glass plate at 90-100°C and autoradiographed directly.

The sequence data was transferred to disc using the program GELIN written by Dr. M. Bishop on a BBC microcomputer with Torch discdrive and GP7 sonic digitiser (made by Science Accessories Corporation). The data was processed on an IBM3081 using the DBUTIL and DBAUTO programs of Dr. R. Staden.

2.6 Sugar Transport Assays

2.6.1 Uptake of radioisotope-labelled sugar into intact cells

Cultures (200ml) in minimal salts medium (Henderson et al., 1977) supplemented with appropriate amino-acids (80 µg ml^{-1}), glycerol or succinate (20mM), and, where relevant, xylose (10mM) were grown from a 2ml inoculum in 250ml flasks overnight (approx. 16 hours) in a Gallenkamp orbital incubator at 30°C with shaking at 220 revolutions per minute. The cells were harvested, and depleted by incubation for 1 hour at 220 revolutions per minute and 30°C in an equal volume of 150mM KCl/5mM MES/1mM 2-mercaptoethanol, pH6.5. They were then washed in 150mM KCl/5mM MES, pH6.5 and finally resuspended in the same buffer to an absorbance at 680nm of 1.8-2.2. The method for measuring uptake was as described by Henderson et al. (1977) using 50µM labelled sugar, except for K_m evaluations when the concentration of sugar was varied. In cases where inhibitors were used these were added 3min before the addition of the labelled sugar.

2.6.2 Uptake of radioisotope-labelled sugar into vesicles

Vesicles were diluted to a final concentration of approx. 1.5mg ml^{-1} protein in 50mM potassium phosphate/10mM magnesium sulphate, pH6.6. At time t=0 a sample (250µl final volume) of this was placed at 25°C and bubbled with oxygen. At t=2.5min phenazine methosulphate was added to a final concentration of 0.1mM. At t=2.75min potassium ascorbate, pH6.5 was added to a final concentration of 20mM. At t=3min labelled sugar was added to a final concentration of 40µM. Samples (100µl) were withdrawn and filtered at 15s and 2min after addition of sugar, and washed with approx. 5ml 0.1M lithium chloride. Any inhibitors were added at time t=0, except for NEM when the procedure was as described below (section 2.8).

2.6.3 Measurements of sugar-promoted pH changes

Cells were grown as described in section 2.6.1. They were depleted and washed (using glycylglycine as buffer), and pH changes following the addition of substrate were measured, as decribed by Henderson et al. (1977). The measurements were made within 24 hours of preparing the cells.

2.7 Binding Protein Assay

Cultures were grown as described in section 2.6.1. The cells were washed with 0.2M Tris-HCl, pH8.0, then made into spheroplasts by the method of Witholt et al. (1976). The supernatants (60ml) were concentrated using an Amicon ultrafiltration cell with PM10 filter at 60psi. to a volume of approx. 5ml.

Samples (1ml) of the concentrated shock fluids were dialysed in a total volume of 10ml 10mM potassium phosphate, pH6.9 containing 1µM [^{14}C]-xylose at 89.9 mCi $mole^{-1}$ at 4°C for 18 hours (Ahlem et al., 1982). Samples (0.3ml) were taken from inside the dialysis bag and from the solution outside it, 15ml methoxyethanol/toluene/PPO/POPOP (500ml/2l/10g/0.5g) scintillant added, and the radioactivity measured using a Packard 3385 or a Beckman LS2800 scintillation counter. In addition the protein concentration within each bag was determined.

2.8 Use of N-ethylmaleimide for Inhibition and Labelling Studies

2.8.1 Inhibition of uptake into vesicles by N-ethylmaleimide and protection by sugars

Vesicles (in the absence of 2-mercaptoethanol) were diluted to a known final concentration of approx. 1.5mg ml^{-1} protein in 50mM potassium phosphate/10mM magnesium sulphate, pH6.6. Samples (250µl final volume) were equilibrated to 25°C. Unlabelled sugar was added where appropriate to a final concentration of 20mM (except where indicated otherwise) and the samples incubated for 5min at 25°C. NEM was added to a final concentration of 1mM (except where indicated otherwise) and incubation continued for a further 15min. The reaction was terminated by placing in ice followed by centrifugation at 4°C (Eppendorf microfuge, 5min). The vesicles were washed four times by resuspension in 50mM potassium phosphate/10mM magnesium sulphate, pH6.6 and centrifugation. The final resuspension was to 250µl in the same medium. The vesicles were equilibrated to 25°C and then PMS/ascorbate energised uptake of [^{14}C]-xylose was measured (as above).

2.8.2 Labelling of vesicle proteins with radioisotope-labelled N-ethylmaleimide

Fresh vesicles (stored overnight as a pellet at 4°C) were diluted to approx. 1.5mg ml^{-1} as above. Samples (0.6ml) were incubated at 25°C as described in section 2.8.1 with or without protecting sugar (20mM 6-deoxyglucose), and with 1mM unlabelled NEM. In this step non-protected sulphydryl groups should react with unlabelled NEM. After the series of washes the pellets were resuspended in 0.2ml of the same buffer, and incubated either with [^{3}H]-NEM or with [^{14}C]-NEM (1mM) for 1 hour at 25°C. In this step previously protected sulphydryl groups should be labelled by the radioactive NEM. The samples were washed three times in 0.6ml of the same buffer, and then resuspended in 50mM sodium phosphate/10mM magmesium sulphate, pH6.6. They were then combined in pairs (with washing to ensure complete transfer) such that a protected [^{3}H]-NEM labelled sample was mixed with an unprotected [^{14}C]-NEM labelled sample, or a protected [^{14}C]-NEM labelled sample was mixed with an unprotected [^{3}H]-NEM labelled sample. These joint samples were finally resuspended in 0.25ml of the sodium phosphate buffer, and stored at -80°C.

2.8.3 Separation of proteins and analysis of the dual isotope label

The radioisotope-labelled preparations (45µl per track) were fractionated on a 15% SDS-polyacrylamide gel at 25mA and 4°C. Pyronin was used to mark the intermediate tracks and the point of junction of the stacking gel and the running gel. The gel was electro-blotted onto nitrocellulose at 150mA and 4°C overnight, having pre-soaked the gel in the blotting buffer (25mM Tris-HCl/190mM glycine/0.1% SDS/20% methanol) for 30min to allow any size changes to occur.

The gel was stained with Coomassie Blue to ensure the transfer had been complete. Strips of the nitrocellulose blot bearing molecular mass standards were stained with naphthalene black and the positions of the standards marked onto the blot. The sample tracks on the blot were cut into 1.5mm slices, noting the positions of the standards, placed into scintillation vials and 10ml methoxyethanol/toluene/PPO/POPOP scintillant added. These were incubated at 4°C for 24 hours prior to counting.

For the preparation of EJ15 the gel itself was sliced (2mm slices) after being stained and dried. Hydrogen peroxide (0.4ml of 100 volume per slice) was added and the samples incubated at 50-55°C for at least 36 hours. After cooling to 4°C, 10ml Triton X-114/xylene/PPO (11/21/9g) scintillant were added and the vials were counted as before. The data was transferred from paper tape to an IBM3081 and processed using programs written by Macpherson (1982) and subsequently modified.

2.9 Identification of XylE-LacZ Hybrid Protein

The band corresponding to the XylE-LacZ hybrid protein on an SDS-polyacrylamide gel was identified as described below.

Cultures (400ml) were grown as described in section 2.6.1. After harvesting (3 700g, 8min) the cells were washed in 400ml per culture of 50mM Tris-HCl/1mM $MgSO_4$/50mM 2-mercaptoethanol, pH7.2 before resuspending to a final volume of 10ml in the same buffer, with the addition of the protease inhibitor PMSF to 0.1mM. This suspension was then passed through the French Pressure cell at 20 000psi. and subsequently centrifuged at 20 000g, 20min, 4°C in a

Sorvall SS34 rotor. A sample (2ml) of the supernatant was solubilized by the addition of octyl-glucoside to a final concentration of 1.2% (0.22ml of 12% stock solution) and incubation on ice for 15min. This mixture was then diluted 2-fold with the same buffer prior to centrifuging at 145 000g, 60min, 4°C, the supernatant providing a solubilized membrane preparation.

This solubilized membrane preparation was then reacted with a monoclonal antibody to β-galactosidase which was bound to a Sepharose-4B resin. A sample (1ml) of the supernatant was added to 0.4ml settled antibody-resin complex, and the mixture was incubated on ice for 2h with gentle agitation every 10-15min. The resin was pelleted by centrifugation (Eppendorf microfuge, 5min) and the supernatant removed with a pipette. The resin was then washed three times by incubation with 1ml of 100mM Tris-HCl/1M NaCl/0.6% octyl glucoside/0.1mM PMSF, pH8.7 for 15min per portion with gentle agitation. The proteins bound to the antibody-resin complex were then eluted by incubation with 0.3ml of 6M urea/20mM Tris-HCl/10mM NaCl/10mM EDTA/100mM 2-mercaptoethanol/0.6% octyl glucoside/0.1mM PMSF, pH7.2 for 20min with gentle agitation. The eluates obtained were dialysed overnight against 100ml 20mM Tris-HCl/10mM NaCl/10mM EDTA/100mM 2-mercaptoethanol/0.6% octyl glucoside/0.1mM PMSF, pH7.2 at 4°C.

Samples from each stage of the procedure were assayed for β-galactosidase. In addition 75µl samples of the dialysed eluates for each preparation were solubilized in SDS-dissolving buffer at 100°C for 2min and loaded onto a 10% SDS- polyacrylamide gel.

2.10 Solubilization of Membranes

2.10.1 Testing solubilization conditions

To test the conditions for solubilization of the hybrid protein membranes, buffer (50mM Tris-HCl, pH7.4), and detergent were mixed in a final volume of 200µl in an airfuge tube and incubated at room temperature for 15min. Samples were taken for assay of the whole mixture, then the remainder was centrifuged at approx. 100 000g (22psi) for 60min in an airfuge (Beckman). The supernatant was sampled for β-galactosidase assays.

The activities of whole mixtures over a concentration range of detergent indicated whether inhibition was caused by the detergent, while comparison of the activity in each supernatant with that in the corresponding whole mixture indicated the degree of solubilization achieved.

2.10.2 Solubilization in Triton X-100 for column chromatography

Except where indicated otherwise, membranes were washed twice in 50mM Tris-HCl/2mM EDTA, pH7.5 then solubilized in Triton X-100 as follows. For small samples the membranes were diluted to 1mg ml^{-1} protein and 1% Triton in 50mM Tris-HCl, pH7.4; for larger quantities they were diluted to 7.5mg ml^{-1} protein and 5% Triton in the same buffer. In either case the mixture was incubated at room temperature (approx. 20-24 °C) for 15min with occasional gentle mixing, and then centrifuged at 145 000g as described in section 2.1.4 to remove any non-solublized material.

2.11 Column Chromatography

For each column the fractions were assayed for β-galactosidase and protein, and where a salt gradient was used the conductivity of the fractions was measured.

2.11.1 Affinity column

A 4ml column of the β-galactosidase affinity matrix p-aminophenyl-β-D-thiogalactoside was poured in 50mM Tris-HCl/100mM NaCl, pH7.4 in a 5ml pipette. The column was equilibrated each time before use with 5-10 column volumes of 50mM Tris-HCl/100mM NaCl/1% Triton, pH7.4. After loading the sample, the column was washed through with 5 column volumes of the same buffer before eluting under various conditions as described in the Results. The final elution in each case was with 100mM sodium borate/1% Triton, pH10 over 5 column volumes. The pressure head used was 75cmH_2O, and the fraction size was 1ml.

2.11.2 Testing conditions for DEAE column chromatography

The conditions for running the DEAE column were checked in a test-tube experiment as follows. An aliquot (1ml suspension of DEAE resin equivalent to 0.5ml settled volume) was added to each test-tube. Each was equilibrated with 5ml buffer (10mM Tris-HCl/1% Triton) at the relevant pH or salt concentration four times. The final suspension was in 2ml, the controls being 2ml of buffer at each pH, or the extremes of the salt concentrations being used, with no DEAE resin. To each tube a sample of the hybrid protein (30µl pool from Bio-gel column) was added, and the resin prevented from settling out for 5min by gentle mixing. The resin was then allowed to settle, and samples (100µl and 200µl) of the supernatants were assayed for β-galactosidase activity.

2.11.3 DEAE ion-exchange column

An approx. 10ml column was poured in 10mM Tris-HCl, pH7.6, and equilibrated in 10mM Tris-HCl/1% Triton, pH7.6 or pH6.5 until the pH and conductivity of the buffer and the effluent were the same. After loading the sample the column was washed through with 3-4 column volumes of the same buffer. The sample was then eluted with a salt gradient from 0 to 0.3M NaCl in the same buffer over 4 column volumes, unless indicated otherwise, followed by more 0.3M NaCl in the same buffer. The column was finally washed with 1M NaCl in the same buffer. The pressure head used was approx. 30cmH_2O, giving a flow rate of approx. 15ml hr^{-1}, and the fraction size was 1ml.

2.11.4 Bio-gel A-5m gel filtration column

An approx. 500ml column was poured in 50mM Tris-HCl/50mM NaCl, pH7.5 as an approx. 75% degassed slurry. It was equilibrated and packed with 2 column volumes of 50mM Tris-HCl/50mM NaCl/1% Triton, pH7.5 from a pressure head of 150cmH_2O, giving a flow rate of approx. 25ml hr^{-1}. The column was calibrated with cytochrome c (2mg ml^{-1}) blue dextran (2mg ml^{-1}) and β-galactosidase (approx. 25 000 units) in a sample volume of 2ml. collecting 5ml fractions. After loading the hybrid protein sample (approx. 10^6 units in 20ml) the column was eluted with the equilibrating buffer at the same pressure head and 5ml fractions were collected.

2.11.5 Concentration of the eluted fractions

Fractions eluted from a gel filtration column are dilute in protein. The fractions from the Bio-gel column were, therefore, concentrated by loading onto a small (approx. 2ml) DEAE column equilibrated with 50mM Tris-HCl/1% Triton, pH7.5 from a pressure head of 20cmH_2O, and eluting with 0.5M NaCl in the same buffer, collecting 0.3ml fractions.

2.12 Preparation of Purified XylE-LacZ Hybrid Protein for N-terminal Sequence Analysis

A sample of concentrated purified XylE-LacZ hybrid protein was dialysed extensively against 10% ethanol to reduce the concentrations of salt and detergent, then freeze-dried. Approx. 800µg dried, purified protein was analysed on a gas-phase sequenator by Dr. J.E. Walker of the MRC Laboratory of Molecular Biology, Cambridge.

2.13 Preparation of DNA

2.13.1 Preparation of DNA from phage λ

Minipreps of λ DNA were performed as described by Davis *et al.*, 1980. For large scale preparations the following procedure was used. Phage λ particles were prepared and collected by centrifuging at 100 000g as described by Maniatis *et al.* (1982). The DNA was extracted with proteinase K/SDS and phenol (Maniatis *et al.*, 1982), then precipitated with isopropanol, spermine, and finally ethanol (Kieser, 1984).

2.13.2 Preparation of plasmid DNA

Plasmid DNA was prepared as described by Kieser (1984).

2.14 Manipulation of DNA

2.14.1 Restriction digestions and ligations

Restriction digestions and ligations were performed as described by Maniatis et al. (1982).

2.14.2 Subcloning

The vector was digested with the appropriate enzymes, then phophatased by the addition of 5µl (µg DNA)$^{-1}$ of a 5mg ml^{-1} solution of calf intestinal alkaline phosphatase in TE (10mM Tris-HCl/0.1mM EDTA, pH8.0) and continuing the incubation at 37°C for 30min. The plasmid DNA was then phenol and chloroform extracted and ethanol precipitated. The DNA to be subcloned was digested with the appropriate enzymes, then phenol and chloroform extracted, ethanol precipitated, and ligated into the prepared vector.

2.15 Transformations

Competent cells were prepared according to the method of Hanahan (1983). Aliquots (200µl) of competent cells were added to 5-10ng plasmid DNA in 1.5ml Sarstedt tubes and, after incubating on ice for approx. 45min, were heat shocked at 42°C for 60s. Tryptone-Yeast Extract (Table 2.2) containing 10mM $MgSO_4$/10mM $MgCl_2$/10mM glucose (800µl) was added to each tube and the sample incubated at 30°C or 37°C as appropriate for 1 hour to allow expression of antibiotic resistance genes, before plating out onto selective medium.

2.16 DNA Sequencing

2.16.1 Preparation of HincII fragment

Approx. 100µg of plasmid pEJ1 DNA was digested with 100-200 units HincII overnight in a total volume of 300-600µl. The digestion was checked on a minigel, concentrated by ethanol precipitation, and run on a 1% agarose gel (25cm in length) overnight at 60V. The gel was stained in TAE buffer containing 1µg ml^{-1} ethidium bromide and visualised on a transilluminator at 360nm. The gel above the desired band was cut away to remove contaminating

bands, and a slot cut below the desired band. DE81 paper, which had been soaked in 2.5M NaCl for several hours, then in TAE buffer during the staining of the gel, was inserted into this slot to collect the band. The gel was returned to the tank, run on at 150V for 30-45min to collect the required DNA, and then it was checked on the transilluminator that the DNA had entered the paper.

The DNA was eluted from the DE81 paper by vortexing in 0.4ml 20mM Tris-HCl/1mM EDTA/1.5M NaCl, pH7.5 then incubating at 37°C for 5 hours or at 4°C overnight. The paper was removed by centrifugation, then the solution was extracted twice with water-saturated n-butanol, followed by phenol extraction and ethanol precipitation.

2.16.2 Preparation of M13 clones

Purified HincII fragment was circularised by ligation, sonicated using a Heat Systems Ultrasonics W-375 cup-horn sonicator to produce random fragments, and end-repaired using DNA polymerase I (Klenow fragment) as described by Bankier and Barrell (1983). The random fragments were size fractionated using the DE81 paper method (section 2.16.1) on a 1.5% minigel, taking fragments in the range 300-700 base-pairs in length.

Size fractionated, end-repaired fragments were ligated into SmaI cut and phosphatased M13mp8 or M13mp10. This was used to transfect JM101 cells made competent as for transformations, plating out the heat-shocked preparation in a lawn of soft agar containing IPTG (250µg ml^{-1}) and XGal (250µg ml^{-1}) on Tryptone-Yeast extract plates. After incubation overnight at 37°C white plaques were picked, grown up and DNA prepared from them as described by Bankier and Barrell (1983); this was finally redissolved in 30µl TE.

2.16.3 DNA sequencing reactions

The solutions required for these reactions are defined by Bankier and Barrell (1983). The reactions were half the size of those used for 0.4mm Raven gels (Bankier and Barrell, 1983) in order to keep the concentrations of ions in the samples loaded onto the gels essentially the same. Electrophoresis was carried out as described in section 2.5.3.

The DNA and primer (0.2pmol per clone) were annealed by adding 4µl DNA to 4µl primer mix, (a 1:1:3 mix of primer:TM buffer:water) in a microtitre plate well, and incubating at 60°C for 1 hour with the plate wrapped in Saran wrap. Having centrifuged down condensation, the solution was aliquoted to 1µl per well for 4 wells, to each of which was added 1µl dideoxynucleotide mix (A,C,G, or T) and 1µl reaction mix. Then incubation was carried out at 33°C for 20min. Subsequently 1µl chase mix was added to each well and incubation was continued at 33°C for a further 20min. The reactions could be stored at -20°C at this stage. Immediately before running on a gel 1µl formamide dyes mix was added to each well, and the plate incubated at 80°C for 20min.

CHAPTER 3

MUTAGENESIS AND MAPPING OF THE XYLOSE TRANSPORT SYSTEMS

3. MUTAGENESIS AND MAPPING OF THE XYLOSE TRANSPORT SYSTEMS

Genetical techniques may be used to isolate each of two (or more) transport systems for a single substrate. There is evidence for both a proton-symport system and a binding protein system for xylose in E. coli K12. (Lam et al., 1980; Ahlem et al., 1982). The experiments described in this chapter were undertaken in order to obtain strains with the two systems isolated from each other, and to map the gene coding for proton symport as a preliminary to cloning experiments.

It has been shown that binding protein systems are much more sensitive to catabolite repression than proton symport systems (Wilson, 1974; Kolodrubetz and Schleif, 1981; Daruwalla et al., 1981). In addition, xylose transport and metabolism may be selected against under conditions where xylose is inhibitory to growth, as described in detail below. Thus, it should be possible to isolate mutants lacking the proton-symport system when the binding protein system is repressed. Subsequently it should be possible to make mutants lacking both transport systems. Once the map position of the proton symport gene is known, the symport system may be reintroduced selectively with an adjacent marker.

In the following the gene coding for xylose proton symport is designated xylE, analogous to the arabinose system (araE), and the genes coding for the binding protein dependent system are designated xylFG (cf. araFG). XylF represents the binding protein and XylG represents a membrane component, although it is recognised that this is probably an oversimplification. The binding protein dependent transport systems studied in most detail (e.g. maltose in E. coli, histidine in S. typhimurium) have been found to involve three cytoplasmic membrane components in addition to the binding protein (see Chapter 1, section 1.1.3).

3.1 Selection of Mud(Ap^{R}lac)I Insertion Mutants in xylE

A mutation in the gene coding for fructose bisphosphate aldolase (fda) prevents the normal metabolism of fructose-bisphosphate to glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate. Thus any sugars entering the cell

that are normally metabolised by this route lead to an accumulation of phosphorylated sugars, in particular fructose-bisphosphate, which prove toxic to the cell. Therefore, xylose will inhibit the growth of a strain harbouring an *fda* mutation (Figure 3.1), and such a strain can be used to select against xylose transport and metabolism.

Mud(Ap^R*lac*)I was inserted into strain JM2087 as described in Chapter 2 (section 2.2.1). Before the selection could be made it was necessary to allow genotypic and phenotypic segregation as described in Chapter 2 (section 2.2.1). The normal copy number of the chromosome in *E. coli* is approx. 2.5, so, in order to select for the absence of a gene, DNA replication and cell division must proceed to obtain a homogenous genotype in each cell. If this were not allowed the expression of the unaltered gene in the heterogenous situation would dominate. Similarly cell division ensures dilution of preformed protein, the activity of which might otherwise prevent resistance (in this case) from being achieved. The cells were then plated onto minimal agar supplemented with histidine (80µg ml^{-1}), glycerol (20mM), xylose (5mM) and ampicillin (25µg ml^{-1}). The selection was for both ampicillin resistance (i.e. the presence of Mud(Ap^R*lac*)I) and for xylose resistance in the presence of glycerol as carbon source.

Resistance may be acquired by the loss of the ability to convert xylose to a phosphorylated intermediate or to transport xylose, or by reversion of the original *fda* mutation. Under the conditions used resistance cannot be due to a mutation in the pentose phosphate pathway, through which xylulose-5-phosphate is metabolised, as this would lead to a requirement for aromatic amino-acids. The mutational events bringing about resistance may be the insertion of Mud(Ap^R*lac*)I into one of the xylose genes, or the spontaneous mutation of one of these genes with the simultaneous insertion of Mud(Ap^R*lac*)I elsewhere (Mud(Ap^R*lac*)I must be present as the bacteria are resistant to ampicillin). Thus the possible genotypes arising from the selection procedure may be summarised:

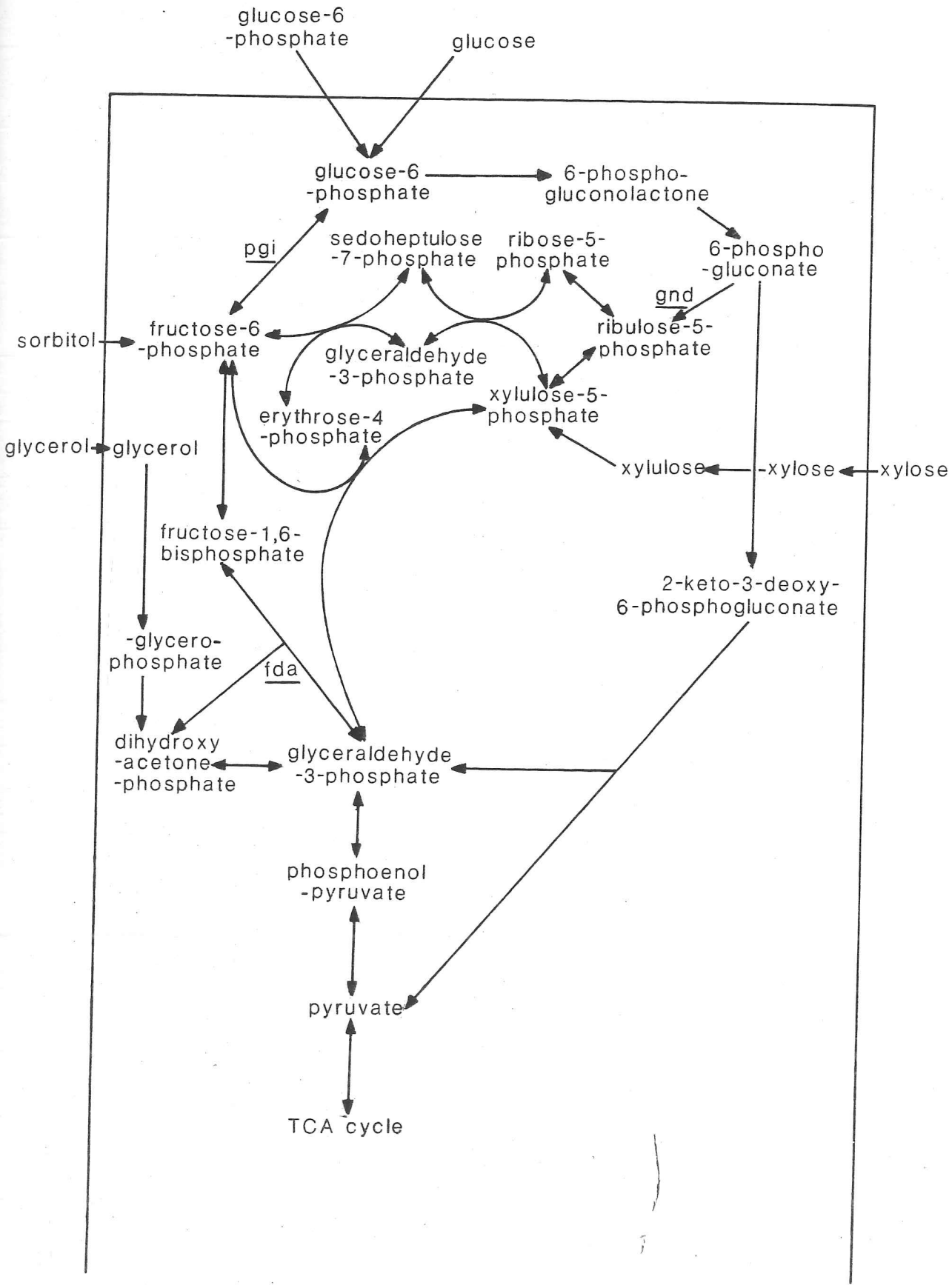


Figure 3.1 Metabolic pathways relevant to the selection of mutants in xylE

(a) Mu insertion into *xylR* (regulatory);

(b) Mu insertion into *xylA* (isomerase);

(c) Mu insertion into *xylB* (kinase);

(d) Mu insertion into *xylE* (proton-symport), *xylFG* (binding protein) being repressed by gycerol;

(e) A mutation in any of (a)-(d) occuring spontaneously and insertion of Mu elsewhere;

(f) An *fda* reversion to *fda*$^+$ and insertion of Mu elsewhere.

Of these (f) may be distinguished from (a)-(e) as it will grow on xylose as sole carbon source. In the other cases either xylose metabolism is blocked or the xylose gives rise to toxic phosphorylated sugars. In 50% of the cases where Mud(Ap^R*lac*)I is inserted into a xylose gene, (a)-(d), it will be oriented such that the directions of transcription of the relevant *xyl* promoter and the *lac* structural genes of the prophage are the same. These may be distinguished from those cases where Mud(Ap^R*lac*)I is inserted in the opposite orientation or in some other position, (e) & (f), since where the directions of transcription are the same the cells will have β-galactosidase activity inducible by xylose. The other cases will either have no β-galactosidase activity or have activity which is not inducible by xylose. Such colonies may be detected by the plate assay for β-galactosidase, comparing colonies replica-plated onto histidine, glycerol minimal medium and onto histidine, glycerol, xylose minimal medium as described in Chapter 2 (section 2.3.1).

Single colony isolates were, therefore, replica-plated to the following media:

Medium		Required Phenotype
nutrient agar, 42°C	-	(*Muc*^{ts})
histidine, xylose	-	(*fda*)
histidine, glycerol, xylose	+	β-galactosidase +
histidine, glycerol	+	β-galactosidase -
nutrient, ampicillin	+	

Five independent strains with xylose-inducible β-galactosidase were obtained from 10 independent selections.

To distinguish between the remaining possibilities, (a)-(d), it was necessary to make the strains obtained *fda*$^+$. This was done by P1 generalised transduction, as described in Chapter 2 (section 2.2.4). When in lytic cycle the phage P1 occasionally packages a random piece of host DNA into the phage heads instead of phage DNA. These particles remain infective and can introduce the acquired DNA into a new host cell, where recombination may occur (Ozeki and Ikeda, 1968).

P1 was grown on the wild type strain K10 and *fda*$^+$ transductants were selected by growth on sorbitol. Sorbitol enters glycolysis below phosphogluco-isomerase, so the progeny must be positive in fructose-bisphosphate-aldolase for growth on sorbitol to occur, and there can be no interference with the xylose systems.

In the *fda*$^+$ strains (d) should be distinguishable from (a)-(c) as in this case growth should occur on xylose as sole carbon source, xylose entering by the binding protein system, whereas in the other cases the mutants are blocked in xylose metabolism. Single colony isolates were patched out and replica-plated onto the following media:

Medium	Required Phenotype		Classes (a)-(c)
histidine, xylose	+		-
histidine, sorbitol	+	(*fda*$^+$)	+
histidine, glycerol, xylose	+	β-galactosidase +	+
histidine, glycerol	+	β-galactosidase -	+

Of the 5 independent mutants which were xylose-inducible for β-galactosidase, 2 appeared to be mutants in transport by the proton symport system from this screening: strains EJ17 and EJ18.

3.2 Biochemical Characterisation of the Mutation in xylE

The mutants obtained from the screening procedure were characterised by the biochemical criteria described below.

3.2.1 Xylose promoted pH changes

An alkaline pH change on the addition of substrate to a de-energised suspension of cells provides a convenient assay for proton symport (West, 1970; Henderson et al., 1977). Such a pH change was observed on addition of xylose for strain EJ15, as control, and for strain EJ17, but not for strain EJ18 (Table 3.1 and Figure 3.2). This indicated that strain EJ18 only, and not strain EJ17, was a mutant in xylose-proton symport. All three showed a subsequent acidification due to metabolism, confirming that transport of xylose did occur in strain EJ18 (but not by proton symport).

The alkaline pH changes observed with strains EJ15 and EJ17 were abolished in the presence of DNP (an uncoupler which renders the membrane permeable to protons) (Table 3.1), indicating that the pH changes were due to the movement of protons across the membrane. In addition strain EJ18 showed an alkaline pH change on the addition of TMG, a well characterised substrate for proton symport by the lacY gene product, in this case present due to expression of the prophage genes; strain EJ15 did not show TMG-proton symport. Thus strain EJ18 is capable of symport and the hypothesis that xylose symport was not observed due to an unspecific effect on the membrane is untenable. Strain EJ18 is XylE$^-$, where xylE is the gene for xylose-proton symport. The nature of the mutation in strain EJ17 has not been investigated further.

3.2.2 Sensitivity of [^{14}C]-xylose uptake to arsenate

Transport by a binding protein system is sensitive to inhibition by arsenate at pH6.5, whereas transport by a proton symport system is relatively insensitive (Berger, 1973; Berger and Heppel, 1974). This reflects the different modes of energisation, the binding protein system being energised directly (the levels of ATP being lowered by arsenate) as compared with the energisation of the proton symport system by the proton-motive-force, generated by respiration without the intermediate formation of ATP. Even in a strain with both systems present there is relatively little inhibition by arsenate (Daruwalla et al., 1981).

The transport of 180μM xylose in strains EJ15 and EJ18, and of 180μM lactose in strain EJ18, in the presence of varying concentrations of

Table 3.1 Xylose and TMG promoted alkaline pH changes

The cultures were grown on glycerol plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3).

Strain	Conditions	Rate ($nmolH^+ min^{-1} mg^{-1}$)	Extent ($nmolH^+ mg^{-1}$)	Number of Measurements*
EJ15	Xylose	4.80 ± 1.39	2.55 ± 1.76	13 (4)
EJ17	Xylose	3.34 ± 1.66	0.99 ± 0.27	3 (1)
EJ18	Xylose	0.11 ± 0.26	0.02 ± 0.02	6 (3)
EJ15	Xylose + DNP	0	0.03 ± 0.06	3 (2)
EJ17	Xylose + DNP	0	0.08	1 (1)
EJ15	TMG	0.46 ± 0.42	0.04 ± 0.04	6 (2)
EJ18	TMG	1.66 ± 1.42	4.58 ± 2.04	6 (2)

*: Number of measurements (number of preparations).

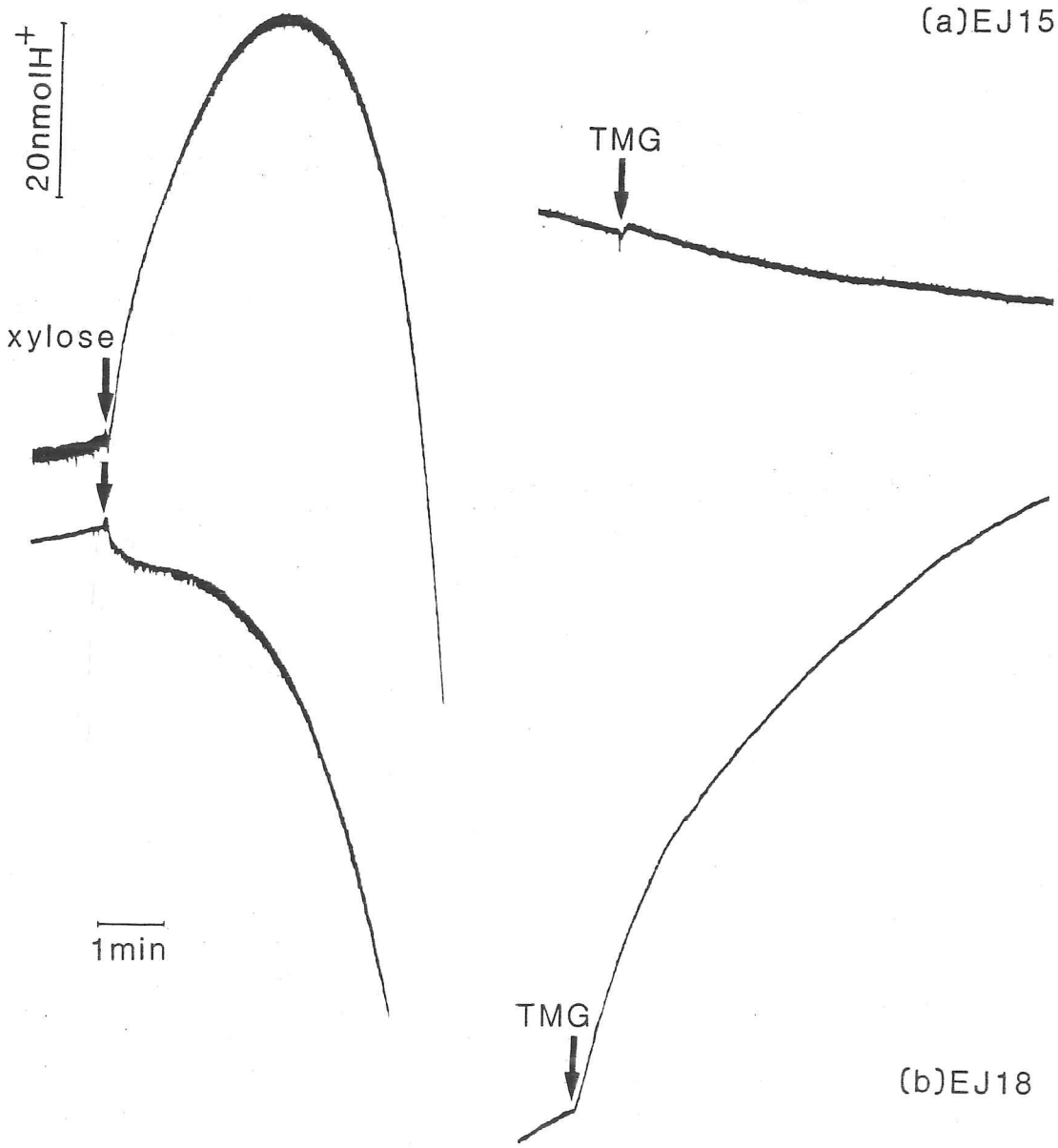


Figure 3.2 Xylose- and TMG-promoted alkaline pH changes in strains EJ15 (a) and EJ18 (b)

The cultures were grown on glycerol plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3). In each case the recording was calibrated by the addition of 3µl 0.01M NaOH immediately prior to the addition of substrate (20µl 0.5M).

potassium arsenate, pH6.5 was compared. The concentration range of the arsenate was 0-50mM, which was achieved by using varying volumes of 100mM arsenate and maintaining a fixed volume and ionic strength with 150mM potassium chloride. Strain EJ18 ($XylE^-$,$XylF^+$) was sensitive to inhibition by arsenate (Figure 3.3) compared with the control strain EJ15. The values shown in the diagram are for the transport at 2min. The increased transport of lactose (which is taken up by a proton symport system) may reflect an increase in respiration due to the uncoupling of ATP synthesis by arsenate. This could also account for the initial increase in xylose transport in strain EJ15 ($XylE^+$,$XylF^+$), which becomes outweighed at higher arsenate concentrations by the inhibition of the binding protein system. These effects were also apparent but to a lesser extent for the transport at 15s. Even at arsenate concentrations as high as 50mM there was relatively little inhibition of xylose transport in strain EJ15. This difference in arsenate sensitivity between strains EJ15 and EJ18 was also apparent, although less marked, when the xylose concentration was 48μM. These results support the conclusion that strain EJ18 is $XylE^-$.

3.2.3 Steady state kinetic analysis

The uptake of [^{14}C]-xylose at varying concentrations of xylose was determined for 15s time points for each of the strains EJ15 and EJ18. The concentration was varied by changing the volume of 2mM xylose added, the total volume being kept constant by the addition of distilled water as required. The 15s points were used as it was hoped to obtain a value for the initial rate while the velocity versus time plot was linear, and the Michaelis-Menten assumptions (that there is no significant reaction in the reverse direction and that the concentration of intermediate remains constant) might be valid. In fact at the lower concentrations used this was not achieved but it was not practicable to filter samples prior to 15s, or to use less cells and still obtain sufficient counts. The double reciprocal plot for strain EJ15 was biphasic (Figure 3.4), in accord with the expectation that two transport systems (a proton symport system and a binding protein system) would be present. An iterative calculation by computer to fit to the sum of two Michaelis-Menten equations (Cleland, 1967) yielded the values of 3-4μM and 53-169μM for the two K_m's, and 5-9nmol mg^{-1} min^{-1} and 15-21nmol mg^{-1} min^{-1}

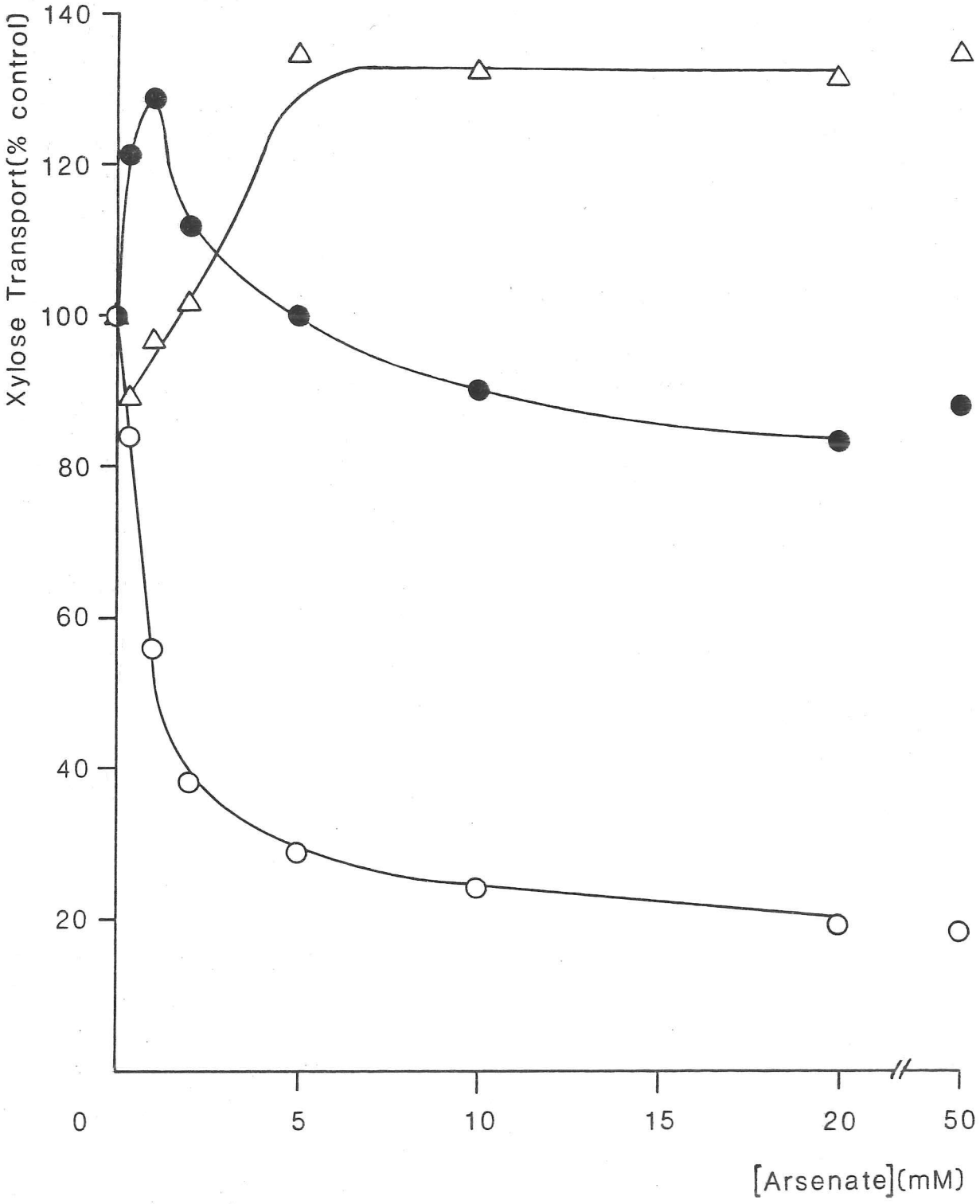


Figure 3.3 Sensitivity to arsenate of xylose transport in strains EJ15 and EJ18 and of lactose transport in strain EJ18

The cultures were grown on glycerol plus 10mM xylose as inducer, and prepared as described in Chapter 2 (section 2.6.1). Transport activity was assayed in the presence of varying concentrations of arsenate at pH6.5, added 3min prior to the addition of substrate. The final concentration of substrate was 180μM. The figure shows the means of duplicate measurements for samples taken 2min after the addition of substrate, expressed as a percentage of the control value: EJ15 + xylose, (●); EJ18 + xylose, (○); EJ18 + lactose, (△). The control values were: EJ15 + xylose, 25.0nmol mg^{-1}; EJ18 + xylose, 13.7nmol mg^{-1}; EJ18 + lactose, 14.5nmol mg^{-1}.

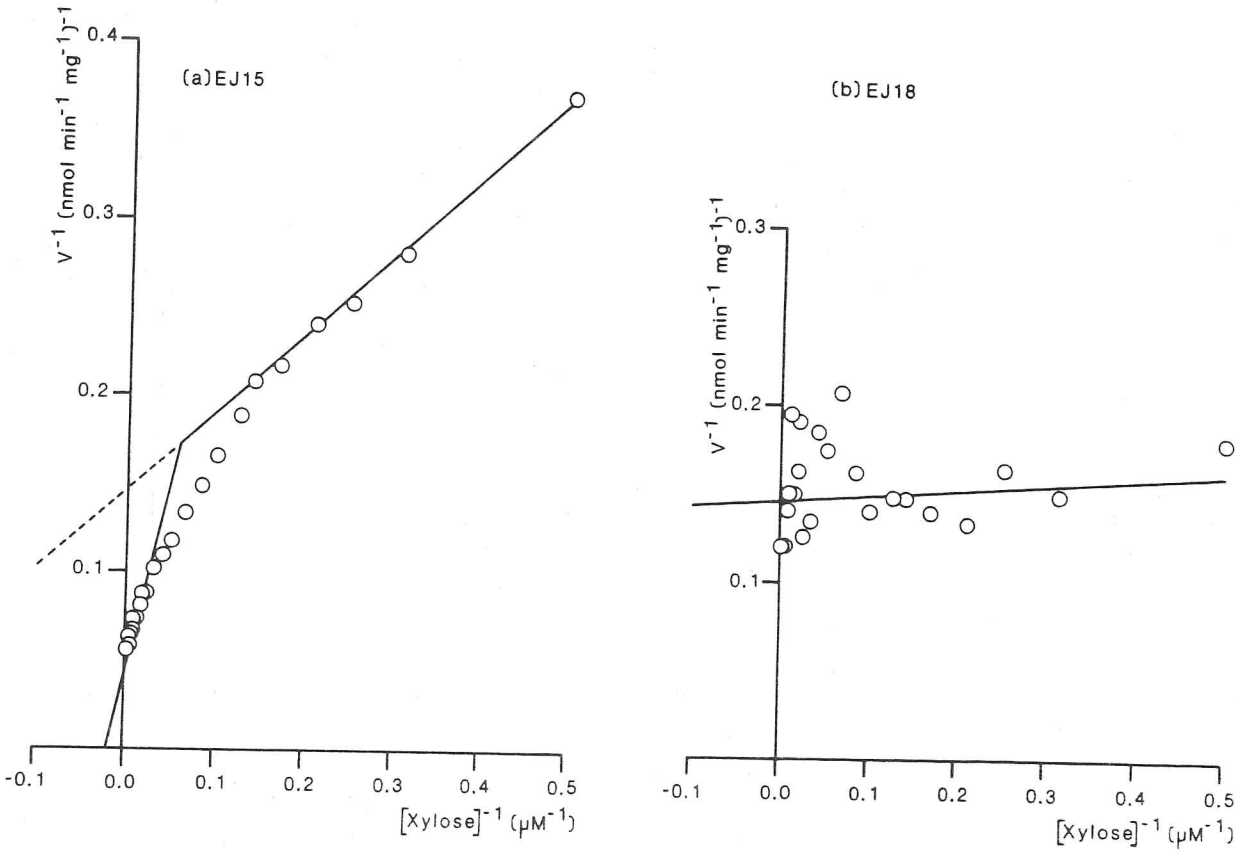


Figure 3.4 Steady state kinetic analysis of xylose transport in strains EJ15 (a) and EJ18 (b)

In each case the cells were grown on glycerol plus 10mM xylose as inducer, and prepared and transport measured for 15s time points as described in Chapter 2 (section 2.6.1). The figure shows the means of duplicate measurements. The line for EJ18 is a computer best fit by the least squares analysis of Cleland (1967). The K_m and V_{max} values are given in the text.

for the corresponding V_{max}'s from two separate experiments. These K_m values may be compared with those obtained by other workers: 5 and 25μM by Ahlem *et al.* (1982), and 24 and 110μM by Shamanna and Sanderson (1979).

The plot for strain EJ18 appeared monophasic (Figure 3.4), with a K_m of 0.2-1.5μM and V_{max} of 4.7-6.8nmol mg^{-1} min^{-1} from a computer least squares fit to a hyperbola for two separate experiments. A direct linear plot fit by computer yielded a K_m of 0.2-2.3μM and a V_{max} of 4.2-7.7nmol mg^{-1} min^{-1} from two separate experiments. The errors on this value cannot be significantly improved because sufficiently low concentrations to be in the region of the K_m are not possible for a transport assay. The results do, nevertheless, clearly demonstrate the loss of the higher K_m, i.e. lower affinity, transport system, which would correspond to the proton symport system, and thus concur with the other conclusions in this section.

3.2.4 Uptake of sugars in vesicles

In vesicles only the proton symport system should be active, as any periplasmic binding protein components are lost in the osmotic shock procedure during vesicle preparation. Thus it would be expected that strain EJ18 should not exhibit transport of xylose in vesicles. The uptake of both [^{14}C]-xylose and [^{14}C]-lactose was measured over a 10min time course in vesicles of strains EJ15 and EJ18 (Figure 3.5). As expected, xylose was transported by vesicles of EJ15 but lactose was not. However, transport of xylose did not occur in vesicles of strain EJ18 although lactose was taken up (the *lac* genes being present on the prophage), showing the vesicles to be capable of symport. Thus the hypothesis that the vesicles had been damaged in preparation is untenable. Additional evidence that this was due to a mutation resulting in loss of the transport protein came from the absence of a background level of uptake due to facilitated diffusion. This was observed for unenergised vesicles with lactose as substrate.

3.2.5 The xylose mutation is not in a maltose gene

During mapping (below) it became plausible that *xylE* might be a maltose gene.

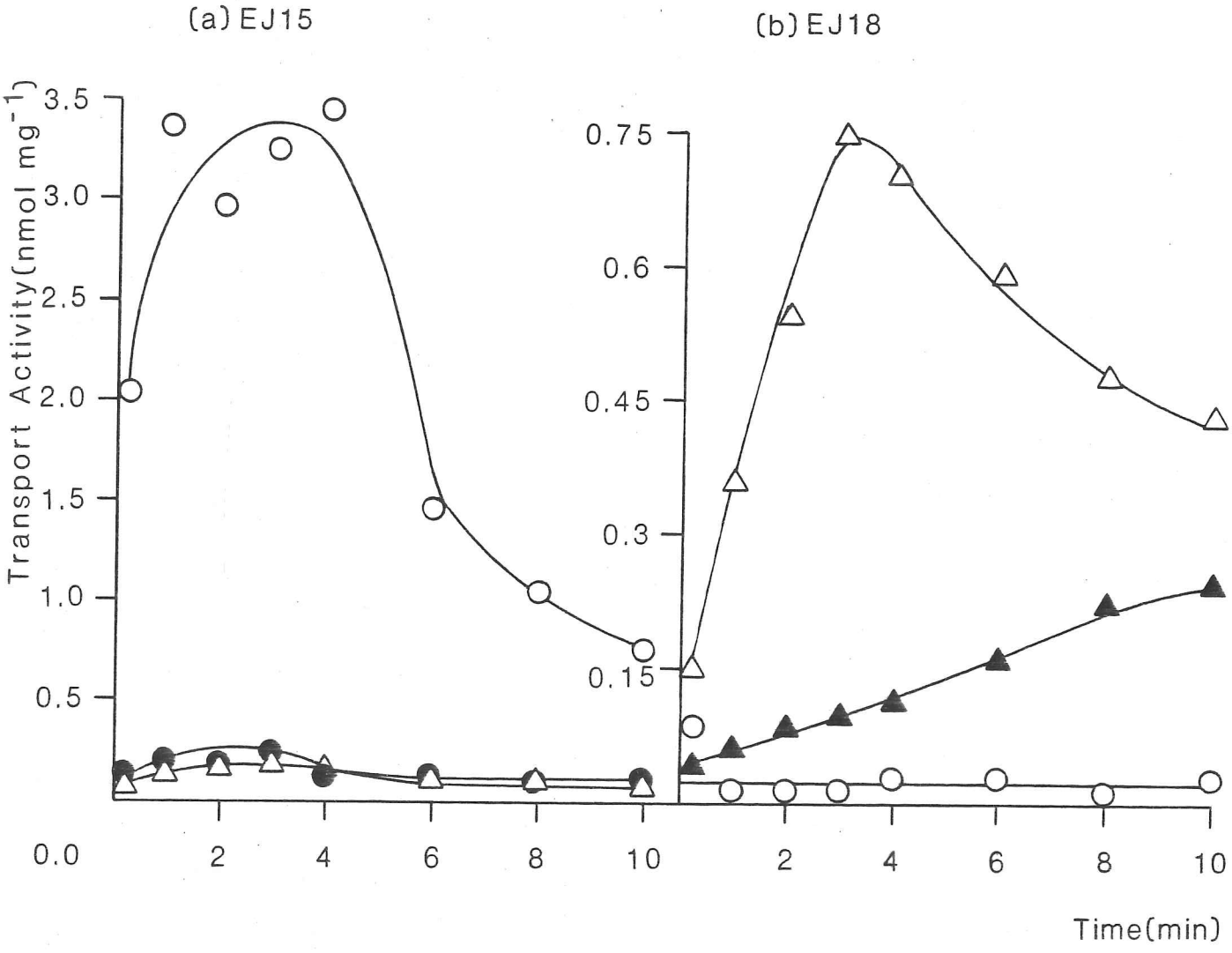


Figure 3.5 Transport of xylose and lactose in vesicles of strains EJ15 (a) and EJ18 (b)

PMS/ascorbate energised, and unenergised, transport was measured as described in Chapter 2 (section 2.6.2), but taking 50µl samples at intervals. The figure shows the means of duplicate measurements for EJ15 and of duplicate measurements from two independent determinations for EJ18: energised transport of xylose (○), energised transport of lactose (△), unenergised transport of xylose (●), unenergised transport of lactose (▲).

Xylose, lactose and maltose transport and β-galactosidase activity were measured in xylose-induced, uninduced and maltose-induced preparations of strains EJ15 and EJ18, samples from the same cultures used for measuring transport being sonicated and used to measure β-galactosidase activity. The results (Table 3.2) clearly showed that in EJ15 xylose transport, and in EJ18 β-galactosidase activity and lactose transport, were not induced significantly by maltose compared to the induction by xylose. Likewise, maltose transport was not induced by xylose in these strains.

These results demonstrate that xylE is not a maltose gene.

3.3 Mapping of the Mud(Ap^{R}lac)I Insertion in xylE

In order to find the region on the E. coli chromosome where xylE is situated it is convenient to find first its proximity to one (or more) of several genetic markers spaced at intervals round the entire chromosome. This may be conveniently done by conjugation: an F$^-$ strain which is a multiple auxotroph and contains the mutation in xylE is mated with a wild type Hfr strain. Once this technique has established the approximate location of the xylE gene more precise mapping can be obtained by P1 transduction using local markers spaced much closer together. In the experiments which follow it is actually the resistance to ampicillin which is being mapped, since that can be determined in any background, whereas the xylE mutation itself can only be scored in certain specific backgrounds.

3.3.1 Conjugation

The DNA from an Hfr strain is transferred to an F$^-$ strain from the same origin and proceeding in the same direction for each conjugation with a particular Hfr. The conjugation may break off at any time, resulting in a gene near the origin being transferred to the recipient at a higher frequency than one which is further away. In addition the two sets of DNA must recombine for incorporation to occur, and this may happen at any two points along the transferred section, so incorporation of a gene further from the origin does not necessitate incorporation of one nearer to the origin. Thus the relative frequencies at which genes are incorporated indicate the order in which they occur along the chromosome.

Table 3.2 Transport and β-galactosidase activities of strains EJ15 and EJ18

A single preparation of each strain was grown on glycerol plus the inducer indicated, harvested, and the activities measured in duplicate as described in Chapter 2 (sections 2.6.1 and 2.3.2).

Strain	Inducer	Xylose Transport (nmol min^{-1} mg^{-1})	Maltose Transport (nmol min^{-1} mg^{-1})	Lactose Transport (nmol min^{-1} mg^{-1})	β-galactosidase Activity (units mg^{-1})
EJ15	Xylose	14.9	0.7	0.4	1.5
EJ15	-	0.2	0.4	0.3	1.1
EJ15	Maltose	0.3	12.2	1.2	0.8
EJ18	Xylose	5.8	0.6	3.1	475.8
EJ18	-	0.2	0.7	0.4	2.3
EJ18	Maltose	0.2	9.6	1.1	4.6

These frequencies are determined by plating dilutions of the mating mixture onto a set of selection plates for each gene. The selection plates allow only recombinants to grow, and not the original Hfr or F^- strains. However, this method only works where a positive selection is possible. In the case of *xylE*::Mud(Ap^R*lac*)I in the F^- strain it is the gain of ampicillin sensitivity which must be determined. Therefore, the chosen markers are selected, then recombinants from each selection are patched out and tested for ampicillin sensitivity and for other nearby markers. This allows the relative position of *xylE*::Mud(Ap^R*lac*)I amongst the other genes to be determined. The *xylE*::Mud(Ap^R*lac*)I was used in the F^- strain rather than the Hfr strain as two Hfr's, with different origins were to be used. The reason for this was because no one Hfr gives a satisfactory number of recombinants towards the end of its chromosome transfer.

The *xylE*::Mud(Ap^R*lac*)I mutation was introduced into the multiply auxotrophic F^- strain JM2235 by P1-generalised transduction, selecting for ampicillin resistance. In order to prevent the transduced Mu from entering lytic cycle it was necessary for the recipient to be a Mu lysogen, so that Mu repressor would be present. The colonies obtained were screened for xylose-inducible β-galactosidase to ensure that Mud(Ap^R*lac*)I had not moved. One resultant strain was designated EJ20 (see Chapter 2, Table 2.1).

The conjugation was performed as described in Chapter 2 (section 2.2.5). The *xylE* gene was conveniently found in the half of the chromosome satisfactorily covered by the Hfr which had been found to be fertile, JM559.

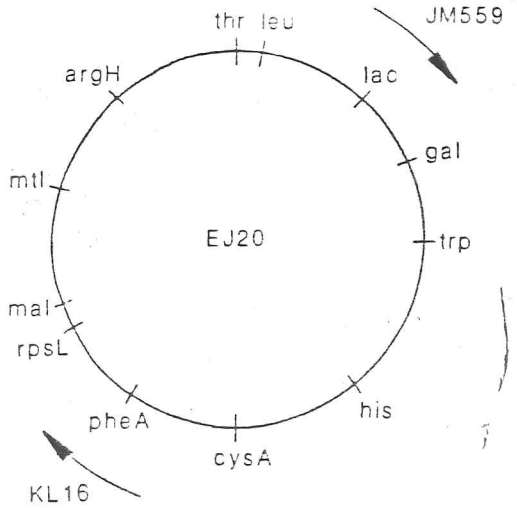


Figure 3.6 Strains used in the conjugation

The selected markers were as follows: leucine, arginine, mannitol + arginine, and maltose + arginine. Arginine was included in the selections for mannitol and maltose to eliminate the background due to spontaneous mutations allowing growth on these carbon sources. In addition streptomycin (100μg ml^{-1}) was used to select against the Hfr strain, except when the maltose marker was being selected. In this case the maltose selection might have been biased if streptomycin had been included due to the close proximity of *rpsL* (Str^R) to *malA*. Therefore in this case the Hfr was selected against by the use of histidinol rather than histidine to supplement the medium; the F^- strain can use histidinol to make histidine, but the Hfr cannot due to a different *his* mutation that lacks histidinol dehydrogenase. The other required amino acids were added to each plate.

Therefore, the media used were as in Table 3.3:

Table 3.3 Selection media for conjugation

Markers Selected	Amino-acid Supplements	Carbon Source	Anti-biotic
leu$^+$	- , Thr, Arg, Phe, Cys, His, Trp	Glc	Str
arg$^+$	Leu, Thr, - , Phe, Cys, His, Trp	Glc	Str
arg$^+$ *mtl*$^+$	Leu, Thr, - , Phe, Cys, His, Trp	Mtl	Str
arg$^+$ *mal*$^+$	Leu, Thr, - , Phe, Cys, Hol, Trp	Mal	-

Colonies from each selection medium were patched out on the medium on which they had been selected and screened for ampicillin sensitivity, with the results in Table 3.4:

Table 3.4 Ampicillin results for conjugation

Markers Selected	Numbers of Ampicillin Sensitives	Percentages of Ampicillin Sensitives
leu$^+$	10/79	12.7
arg$^+$	51/79	64.6
arg$^+$ *mtl*$^+$	64/80	80.0
arg$^+$ *mal*$^+$	60/79	75.9

This suggests that the Mud(Ap^R*lac*)I is near arginine (89.5min), there being no significant difference between the numbers of ampicillin sensitives when the selection was for arginine alone or for arginine + mannitol or arginine +

maltose. If Mud(Ap^R*lac*)I had been nearer, for example, to mannitol (89min), a much greater number of ampicillin sensitives would have been expected when the selection was for arginine + mannitol than when it was for arginine alone.

The patched colonies from the selections for arginine alone and for arginine + mannitol were further screened for *leu* and for *mtl* (in the first case only), in an attempt to obtain an indication as to which side of *arg* Mu lay. The results were as in Tables 3.5 and 3.6:

Table 3.5 Results from selection for arginine

	leu	Ap	Observed	Expected
donor type	+	S	34	29.7
	+	R	12	16.3
	-	S	17	21.3
recipient type	-	R	16	11.7
	mtl	Ap	Observed	Expected
donor type	+	S	14	10.3
	+	R	2	5.7
	-	S	37	40.7
recipient type	-	R	26	22.3

Table 3.6 Results from selection for arginine + mannitol

	leu	Ap	Observed	Expected
donor type	+	S	38	35.2
	+	R	6	8.8
	-	S	26	28.8
recipient type	-	R	10	7.2

Unfortunately no linkage was observed, so there was no indication as to which side of *argH* Mud(Ap^R*lac*)I, and therefore *xylE*, lay.

3.3.2 P1 transductions

A series of transductions was conducted to locate the position of *xylE*::Mud(Ap^R*lac*)I more exactly. The gene *argH* lies at 89.5min on the *E. coli* chromosome map; the initial transduction investigated the region towards 90-91min since strains with markers on that side of *argH* were readily available.

Transduction I

P1 was grown on a strain resulting from the conjugation, EJ21 (see Chapter 2, Table 2.1), which had become *arg*$^+$ but remained ampicillin resistant. This was used to transduce strain JM2349, which bears the markers *argH*, *metA*, *aceA*:

P1.EJ21	*xylE*::Mud(Ap^R *lac*)I	(Ap^R)	*arg*$^+$	*met*$^+$	*ace*$^+$
JM2349	*xylE*$^+$	(Ap^S)	*argH*	*metA*	*aceA*

Each marker was selected independently, and then screened for each of the other markers to establish the gene order from the frequencies obtained.

(a) When the selection was for *arg*$^+$, no cotransduction of ampicillin resistance was observed, although both *met* and *ace* were cotransduced (7/80 and 8/80 respectively). This indicated that Mu, and hence *xylE*, was further from *argH* than were either *met* or *ace*, but again there was no indication as to which side of *argH* *xylE* lies.

(b) When the selection was for *met*$^+$, no cotransduction of ampicillin resistance was again observed, although both *arg* and *ace* were cotransduced (26/80 and 66/80 respectively). In this case, however, more information could be obtained by examining the classes arising from the transduction (Table 3.7):

Table 3.7 Results from transduction I, selection for *met*$^+$

	ace	Ap	Observed
donor type	+	R	0
	+	S	66
	-	R	0
recipient type	-	S	14

These results indicated that *aceA* was nearer to *metA*, the selected marker, than was *xylE*, since there were 66 transductants which became *ace*$^+$ but none which became ampicillin resistant. If the small class other than the donor class was a four crossover class, the gene order would be *argH* *metA* *aceA* *xylE*. In a three point cross the smallest class will correspond to the situation when four crossovers, rather than two, have occurred, as this happens at a lower frequency. This gives rise to the exchange of two outer markers for the donor type while the central marker is unchanged. In this case it would be *ace* which remained of the recipient type while *met* and Ap became donor type.

(c) When the selection was for ampicillin resistance no cotransduction of *argH* was observed, but *met* and *ace* were cotransduced at frequencies of 6/80 and 11/80 respectively. The classes arising from this selection are listed in Table 3.8:

Table 3.8 Results from transduction I, selection for Ap^R

	met	ace	Observed
donor type	+	+	6
	+	-	0
	-	+	5
recipient type	-	-	69

In this case there was clearly a four crossover class which had become ampicillin resistant and met^+ but which remained *ace*. This indicated that *ace* lies between *metA* and *xylE* (Ap^R), i.e. the gene order is *argH metA aceA xylE*.

(d) When the selection was for ace^+ 2/67 colonies had become ampicillin resistant, 63/67 had become met^+ and 12/67 had become arg^+, consistent with the gene order, *argH metA aceA xylE*, proposed above.

Transduction II

Strain NB1 has the markers *pgi* and *zjb*::Tn10 in the region of interest on the *E. coli* chromosome map. These lie relative to other markers in the order (ace^+) *pgi* ($malB^+$) Tn10. P1 was grown on strain NB1 and used to infect strain EJ14:

P1.NB1		$xylE^+$ (Ap^S)	*pgi*	Tn10 (Tc^R)
EJ14	*xylE*::Mud(Ap^R*lac*) (Ap^R)		pgi^+	(Tc^S)

The transductants were selected for tetracycline resistance (10μg ml^{-1}) and screened for ampicillin sensitivity and *pgi*. Strain EJ14 is *fda* and *gnd* and in this background *pgi* is scored on glucose-6-phosphate such that growth occurs for *pgi* negative colonies but there is no growth for *pgi* positive colonies (Fraenkel and Vinopal, 1973). This is because when the cell is *pgi* as well as *gnd* glucose-6-phosphate cannot form fructose-bisphosphate (which is toxic due to the *fda* mutation), but the G-6-P can be metabolised to trioses via 6-phosphogluconate. The *gnd* mutation prevents synthesis of fructose-bisphosphate via the pentose phosphate pathway, and the *pgi* mutation prevents

isomerisation to fructose-6-phosphate and thence to fructose-bisphosphate (Figure 3.1). The synthesis of fructose-bisphosphate (which appears to be the most toxic of the phosphorylated sugars) by the aldolase in the condensing direction (Stribling and Perham, 1973) does not appear to occur fast enough for it to reach toxic levels. However, when the cell is *pgi*$^+$ fructose-bisphosphate is rapidly formed and inhibits cell growth.

The results from the screening are listed in Table 3.9:

Table 3.9 Results from transduction II, selection for Tc^R

	pgi	Ap	Observed
donor type	-	S	0
	-	R	1
	+	S	13
recipient type	+	R	66

These results indicated that *xylE* (Ap^R) was nearer to Tn10 than was *pgi*, since 13 transductants became ampicillin sensitive whereas only 1 became *pgi*. If the small class of 1 actually represented the four crossover situation then it would indicate that *xylE* (Ap^R) was between Tn10 (Tc^R) and *pgi*, i.e. the gene order would be *pgi xylE* Tn10. If this were not the case, the gene order would have to be *pgi* Tn10 *xylE*, but this is not compatible with the results from transduction I (c) and (d) where ampicillin resistance was found to cotransduce with *aceA*.

Transduction III

It was hoped to use phage P1 grown on strain NB1 to infect a strain with the genotype *xylE*::Mud(Ap^R*lac*)I *malB*. In order to make this strain the Mu insertion mutation was to be transduced into a strain with the *malB* marker.

Phage P1 grown on strain EJ18 was used to infect strain EJ23 (*malB* (Muc$^+$)), and transductants were selected for ampicillin resistance. These were screened for growth on maltose (*malB*) and xylose-inducible β-galactosidase. Of 230 ampicillin resistant colonies none were both *malB* and *xylE*::Mud(Ap^R*lac*)I. Either Mu remained inserted in *xylE* but they had become *mal*$^+$, or they remained *malB* but Mu had moved (β-galactosidase no longer being xylose-inducible in about 50% of the cases). This suggests that *xylE* is very

close to malB.

Transduction IV

Phage P1 was grown on strain EJ26 (malB zjb::Tn10) and used to transduce each of strains EJ14 and EJ18, selecting for tetracycline resistance. There was, therefore, no possibility of Mu having moved in this transduction.

P1.EJ26 xylE$^+$ (Ap^S) malB Tn10 (Tc^R)

EJ14 xylE::Mud(Ap^Rlac) (Ap^R) malB$^+$ (Tc^S)

The transductants were screened for malB and ampicillin sensitivity. In the case of strain EJ14 malB was scored on histidine, succinate, maltose minimal medium such that growth occured for malB negative colonies whereas no growth occured for malB positive colonies. This is because the strain is fda so if maltose enters the cell and is metabolised toxic fructose-bisphosphate accumulates and growth is inhibited. However, if the cell is malB this cannot occur so it is able to grow on the succinate.

Both markers were cotransduced into strain EJ14 with tetracycline resistance (i.e. Tn10) with the frequencies 62/160 for ampicillin sensitives (i.e. xylE$^+$) and 75/160 for malB, and into strain EJ18 with the frequency of 75/160 for both xylE$^+$ and malB. This gives rise to the following classes (Table 3.10):

Table 3.10 Results from transduction IV, selection for Tc^R

	malB	Ap	EJ14	EJ18
donor	-	S	62	75
	-	R	13	0
	+	S	0	0
recipient	+	R	85	85

The lack of the malB Ap^R class in the case of strain EJ18 may be due to the different screening procedure for malB necessitated in this case. Since strain EJ18 is fda$^+$ malB must be screened by growth on maltose as sole carbon source. Thus, some growth might have occurred if a colony were not pure or if the maltose were contaminated with some glucose.

The classes for strain EJ14 show that in 62/75 cases where malB is transferred so is ampicillin sensitivity, indicating that the two genes are close together. In 13 cases malB is transferred but not ampicillin sensitivity, but none have ampicillin sensitivity transferred and not malB. This latter case corresponds to the four crossover class indicating the gene order for these markers xylE malB Tn10. The results from strain EJ18 are consistent with this conclusion.

Map position from P1 transductions

Taking the results from all the transductions together gives the unambiguous gene order:

metA aceA pgi xylE malB zjb::Tn10.

Thus the gene for xylose-proton symport, xylE, maps between minutes 91.3 and 91.5 on the E. coli chromosome.

3.3.3 Deletion mapping

With a Mu insertion mutation the position of the gene on the chromosome may be determined by inducing the excision of the prophage. When Mu excises part of the adjoining chromosomal DNA is also removed and by screening for the loss of nearby genes the location of the prophage, and hence of the gene into which it was inserted, may be determined.

Five cultures of strain EJ14 from five separate single colonies were grown in nutrient broth overnight at 30°C. Deletion mutants were selected by spreading 0.2ml of each culture on minimal medium supplemented with histidine, methionine, adenine, glycerol, xylose, lactose, and casamino acids, and incubating at 42°C for 24h. The xylose and lactose together select against the xylose-inducible lac operon of Mud(Ap^Rlac)I, as the strain is fda so any lactose entering the cell and being metabolised will lead to a build up of the toxic fructose-bisphosphate. The temperature of 42°C also selects against the prophage since the repressor protein of Mud(Ap^Rlac)I is temperature sensitive.

Since many colonies were obtained a preliminary screening was included before the final screening. The selection plates were replica-plated directly

as below.

Medium	Temp. °C	Phenotype	Genotype
nutrient, ampicillin	37	-	Ap^S
his, succ, mal	37	+	*mal*
his, G6P	37	+	*pgi*
his, ace	37	-	*ace*, *met*, *ade*, *ubi*
his, ade, gly	37	-	*met*
his, met, ade, succ	37	-	*ubi*
selection medium	42		

This gave many colonies which were *ace*, or *met*, several which were *pgi*, a few which were *mal* but none which were *ubi*. Colonies of these phenotypes were identified and patched out, then screened in full as below:

Medium	Phenotype	Genotype
nutrient, ampicillin	-	Ap^S
his, met, ade, succ, mal	+	*mal*
his, met, ade, G6P	+	*pgi*
his, met, ade, ace	-	*ace*
his, - , ade, gly	-	*met*
his, met, - , gly	-	*ade*
his, met, ade, succ	-	*ubi*
selection medium		

The actual numbers of colonies of each type of overall genotype merely reflect the time at which the deletion event occurred during growth. The number of times a particular genotype arose in separate cultures reflects more closely the frequency of the event. The overall genotypes found are listed in Table 3.11:

Table 3.11 Results of deletion mapping

ade	met	ace	pgi	mal	ubi	Number of Cultures
-	-	-	-	+	+	5
+	-	-	-	+	+	5
+	+	-	-	+	+	5
+	+	+	-	+	+	5
+	+	-	-	-	+	3
+	+	+	-	-	+	1

This indicates that the position of Mud(Ap^{R}lac)I, and hence xylE, is either between ace and pgi or between pgi and mal. This is compatible with and, therefore, supports the findings from the transduction mapping. It would be possible to find which of these positions was the correct one from deletion mapping if the right deletion was found. In this case, however, the necessary deletion (as indicated by the transduction mapping) would be pgi$^+$ malB. Since no ubi colonies were found, and it may not be possible to obtain deletions which are ubi in this way, one would have to rely on selecting malB. Therefore, the required strain would only have lost one marker which would not provide adequate evidence that a deletion had occurred. Such a strain was obtained by Dr. M.C. Jones-Mortimer from a Mud(Ap^{R}lac)II insertion in xylE indicating that xylE was between pgi and malB (Davis et al., 1984).

3.4 Selection of Mud(Ap^{R}lac)I Insertion Mutants in xylFG using a xylE Strain

To obtain mutants in the binding protein dependent xylose transport system a selection for xylose resistance in a strain deleted for xylE and sensitive to xylose was used. The dual function of the tricarboxylic acid cycle, to supply precursors for biosynthesis as well as to oxidise C_2-units from glycolysis, necessitates the occurrence of an anaplerotic reaction (Ashworth and Kornberg, 1966). This is normally achieved by the action of phosphoenolpyruvate carboxylase, coded for by the gene ppc. In ppc mutants the only anaplerotic pathway available is the glyoxylate cycle (Figure 3.7). However, the enzymes of this pathway are found in greatly reduced amounts during growth on glycolytic substrates (Kornberg, 1966). Therefore, such strains can only grow on substrates of glycolysis in the presence of utilisable intermediates of the tricarboxylic acid cycle (e.g. aspartate).

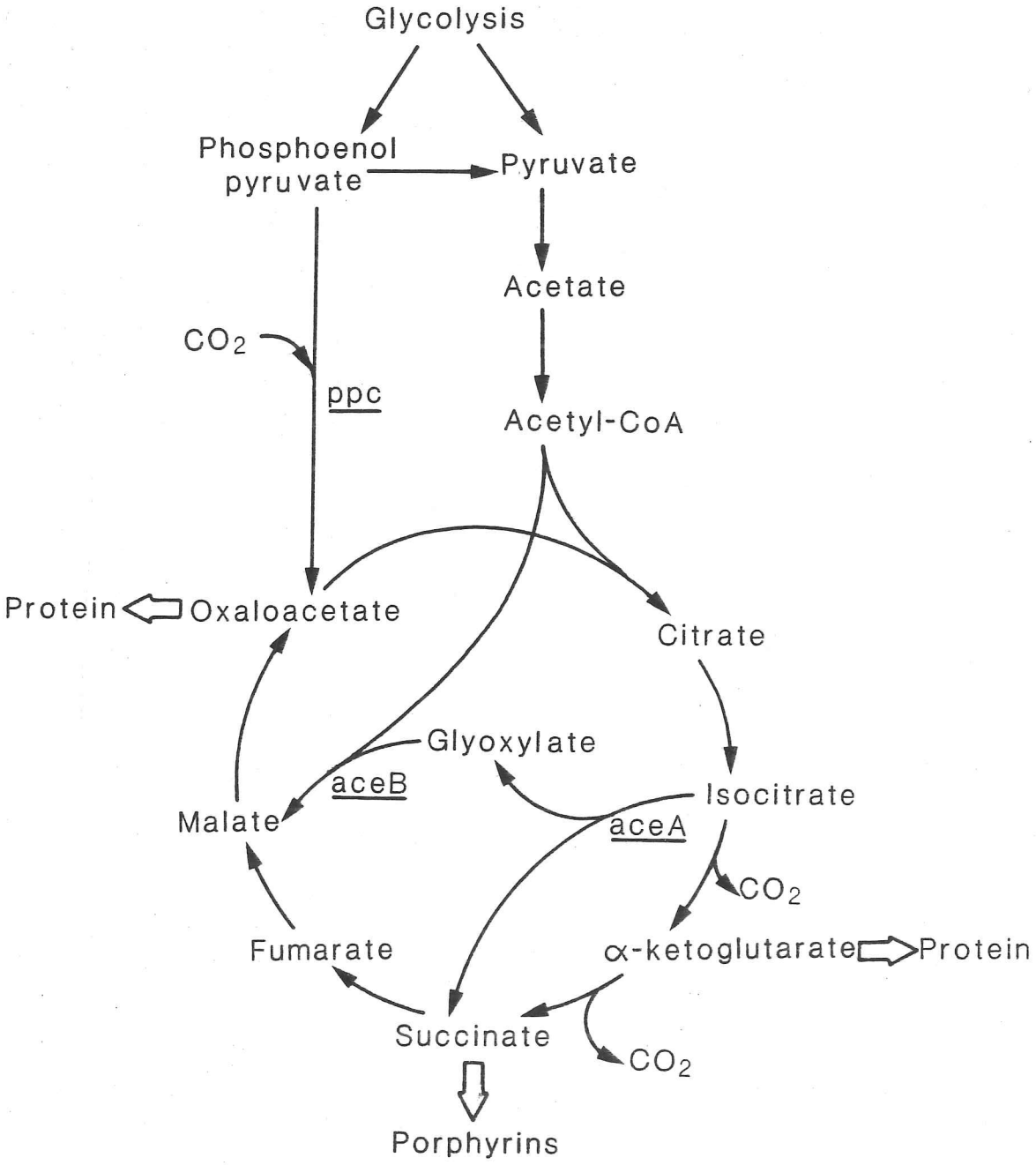


Figure 3.7 The glyoxylate cycle and associated metabolic pathways

Although *ppc* mutants can grow on acetate as sole carbon source, they are unable to grow on acetate in the presence of a glycolytic substrate, unless the metabolism of that substrate is blocked or a mutation (*iclR*^{c}) resulting in constitutive expression of isocitrate lyase (the first enzyme of the glyoxylate cycle) has occurred (Vinopal and Fraenkel, 1974). Thus the use of a *ppc* strain allows the selection of mutants blocked in metabolism or transport of a sugar (Kornberg and Smith, 1969).

Mud(Ap^{R}*lac*)I was used to lysogenise JM2390 (*ppc*, *xylE*$^{\Delta}$) as described in Chapter 2 (section 2.2.1). After allowing expression and segregation the cells were plated onto minimal agar supplemented with histidine (80μg ml^{-1}), arginine (80μg ml^{-1}), acetate (30mM), xylose (1mM) and ampicillin (25μg ml^{-1}). The selection was for both xylose resistance during growth on acetate and ampicillin resistance. Strains arising from this selection must be lysogens of Mud(Ap^{R}*lac*)I as they are ampicillin resistant, but the xylose resistance could be the result of insertion of Mud(Ap^{R}*lac*)I into one of the xylose genes or of a spontaneous mutation in one of the xylose genes with Mud(Ap^{R}*lac*)I inserted elsewhere in the chromosome. Thus the possible genotypes arising from the selection procedure are:

(a) Mu insertion into *xylR* (regulatory)

(b) Mu insertion into *xylA* (isomerase)

(c) Mu insertion into *xylB* (kinase)

(d) Mu insertion into *xylFG* (binding protein), *xylE* (proton-symport) being deleted in the starting strain

(e) A mutation in any of (a)-(d) occuring spontaneously and insertion of Mu elsewhere.

(f) A mutation (*iclR*^{c}) resulting in constitutive expression of isocitrate lyase and insertion of Mu elsewhere

As (f) can grow on xylose + aspartate, it may be distiguished from (a)-(e) where xylose metabolism or transport is blocked, the aspartate replacing the intermediates of the tricarboxylic acid cycle used in biosynthesis. Also, the cases where Mud(Ap^{R}*lac*)I is inserted into a xylose gene, (a)-(d), in the correct orientation for transcription of the *lac* structural genes from the *xyl* promoter may be distiguished from those where it is in the reverse orientation or elsewhere in the chromosome, (e)&(f), by screening for xylose inducible β-galactosidase.

Single colony isolates were, therefore, replica-plated to the following media (all minimal media containing histidine and arginine):

Medium	Required Phenotype	
xylose	-	
acetate, glucose	-	(*ppc*, *iclR*$^+$ not *iclR*^{c})
aspartate, xylose	-	
aspartate, glycerol, xylose	+	β-galactosidase +
aspartate, glycerol	+	β-galactosidase -
acetate, xylose	+	
aspartate, glucose	+	
nutrient, ampicillin	+	

The plate assays for β-galactosidase did not indicate clearly whether the activity was inducible or not. If the mutation was in *xylFG* induction might be a problem as there would be no transport system for the inducer to enter the cell. Therefore, 5 independent strains which appeared to have xylose-inducible β-galactosidase although there was some activity in the uninduced case, and which did not grow on xylose + aspartate were taken.

The remaining possiblities could be distinguished by making the strains *xylE*$^+$. For (d) XylE$^+$ XylFG$^-$ transductants would be expected to grow on xylose + aspartate, xylose entering the cell on the proton-symport system, whereas for (a)-(c) the XylE$^+$ transductants would be expected to be unable to grow on xylose + aspartate, being blocked in the metabolism of xylose. In addition, if the Mud(Ap^R*lac*)I is inserted into *xylFG* the β-galactosidase should become more clearly xylose-inducible in the XylE$^+$ strain.

The mutants obtained were transduced with P1 grown on strain EJ64 (*zjb*::Tn10) to tetracycline resistance, and screened by replica-plating to the following media (all minimal media containing histidine and arginine):

Medium	Required Phenotype	
xylose	-	
acetate, glucose	-	(*ppc*, *iclR*$^{+}$ not *iclR*^{c})
acetate, xylose	-	
acetate	+	
aspartate, xylose	+	
aspartate, glycerol, xylose	+	β-galactosidase +
aspartate, glycerol	+	β-galactosidase -
aspartate, glucose	+	
nutrient agar, 42°C	-	(*Muc*^{ts})
nutrient, ampicillin	+	
nutrient, tetracycline	+	

One of the mutant strains, EJ80, gave 71 out of 80 tetracycline resistant transductants which now grew on xylose + aspartate, and these showed clearly xylose-inducible β-galactosidase (one of which was taken, strain EJ81). Therefore, strain EJ80 must contain a Mud(Ap^{R}*lac*)I insertion in a gene, *xylFG*, that is only required for xylose utilisation in a *xylE* background. Since *xylE* specifies the proton linked transport system for xylose, this result indicates that *xylFG* specifies part of an alternative transport system.

3.5 Biochemical Characterisation of the Mutation in *xylFG*

Unequivocal evidence that the Mud(Ap^{R}*lac*)I insertion in strains EJ80 and EJ81 was in a gene coding for part of the xylose binding protein dependent transport system was obtained as below.

3.5.1 Transport of [^{14}C]-xylose and β-galactosidase activities

Uptake of [^{14}C]-xylose into strains JM2390 (XylE^{-}, XylFG^{+}), EJ80 (XylE^{-},XylFG^{-}) and EJ81 (XylE^{+},XylFG^{-}) was measured as described in Chapter 2 (section 2.6.1). The results (Table 3.12) clearly indicate that strain EJ80 is unable to transport xylose to any significant extent, the values obtained for induced EJ80 (0.14nmol mg^{-1} in 2min) being comparable with those for uninduced cultures of each of these strains (0.18nmol mg^{-1} in 2min for JM2390 and 0.13nmol mg^{-1} in 2min for EJ81), whereas the induced levels of transport activity for JM2390 (9.2nmol mg^{-1} in 2min) and EJ81 (3.1nmol mg^{-1} in 2min) are

Table 3.12 Xylose transport and β-galactosidase activities in strains JM2390, EJ80 and EJ81

The cultures were grown on glycerol plus aspartate in the presence or absence of 5mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.1 and 2.3.2). The values given are the means of duplicate measurements from each of two independent determinations.

Strain	Phenotype	Induced	Xylose Transport		β-galactosidase
			Initial Rate (nmol min^{-1} mg^{-1})	Extent at 2min (nmol mg^{-1})	Activity (units mg^{-1})
JM2390	$XylE^{-}XylFG^{+}$	+	6.4	9.2	-0.5
JM2390	$XylE^{-}XylFG^{+}$	-	0.2	0.2	-0.4
EJ80	$XylE^{-}XylFG^{-}$	+	0.2	0.1	29.1
EJ80	$XylE^{-}XylFG^{-}$	-	0.1	0.1	0.7
EJ81	$XylE^{+}XylFG^{-}$	+	2.3	3.1	48.2
EJ81	$XylE^{+}XylFG^{-}$	-	0.2	0.1	0.8

significantly higher. The level of activity of JM2390 ($XylE^-$,$XylFG^+$) is seen to be greater than that of EJ81 ($XylE^+$,$XylFG^-$) consistent with the XylFG system being of higher affinity. The β-galactosidase activity in strains EJ80 and EJ81 when assayed quantitatively was found to be inducible by factors of 40- and 60-fold respectively.

3.5.2 Xylose promoted pH changes

An alkaline pH change characteristic of proton symport was observed on the addition of xylose to a de-energised suspension of xylose-induced cells for strain EJ81 ($XylE^+$ $XylF^-$) but to a much lesser extent than had been observed for strain EJ15 ($XylE^+$ $XylF^+$). Cells of EJ81 grown on succinate exhibited a pH change to an extent of 1.1nmolH^+ mg^{-1} with an initial rate of 0.23nmolH^+ min^{-1} mg^{-1}, which may be compared with 2.6nmolH^+ mg^{-1} and 4.8nmolH^+ min^{-1} mg^{-1} for glycerol grown EJ15. The effect with EJ81 was only observable for succinate grown cells, the corresponding values being essentially zero for cells grown on glycerol plus aspartate. This low level of expression is probably due to poor induction of the symport system in EJ81 where there is no high affinity transport system which may be important for the entry of xylose to act as inducer. This suggestion is supported by the effect of the carbon source for growth as glycerol causes some catabolite repression of the proton symport system. Nevertheless, the presence of any alkaline pH change in EJ81 is indicative of the presence of a proton symport system for xylose, i.e. $XylE^+$ has been transduced into the $XylE^-$ $XylFG^-$ strain, EJ80.

3.5.3 Steady state kinetic analysis

The K_m for EJ81 ($XylE^+$,$XylFG^-$) was compared with the values obtained for JM2390 and EJ18, both ($XylE^-$,$XylFG^+$), for EJ15 ($XylE^+$,$XylFG^+$), and for vesicles of EJ15 (i.e. $XylE^+$). The results (Table 3.13 and Figure 3.8) clearly show that whereas JM2390 and EJ18 possess a high affinity transport system with low value for K_m (approx. 1μM) typical of a binding protein dependent system, EJ81 possesses a relatively low affinity system with a K_m of approx. 63μM, comparable with the value obtained for vesicles of EJ15 (approx. 69μM). In addition the double reciprocal plot for EJ81 is clearly linear, compared to that of EJ15 which is biphasic, so reintroduction of xylE has given a strain

Table 3.13 Steady state kinetic analysis of strains JM2390 and EJ81 and comparison with other strains

The cultures were grown on glycerol plus xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.1 and 2.6.2). Duplicate samples were taken for each point over the concentration range 2 to 400μM [^{14}C]-xylose, except for the determination on vesicles when the range was 4 to 400μM. The values for EJ18, EJ81, and EJ15 are the means from two independent determinations.

Strain	Phenotype	K_m (μM)	V_{max} (nmol min^{-1} mg^{-1})
JM2390	$XylE^-XylFG^+$	1.0	6.2
EJ18	$XylE^-XylFG^+$	0.9	5.8
EJ81	$XylE^+XylFG^-$	63.6	4.9
EJ15 vesicles	$XylE^+$	68.8	8.3
EJ15	$XylE^+XylFG^+$	111.4, 3.5	17.8, 7.0

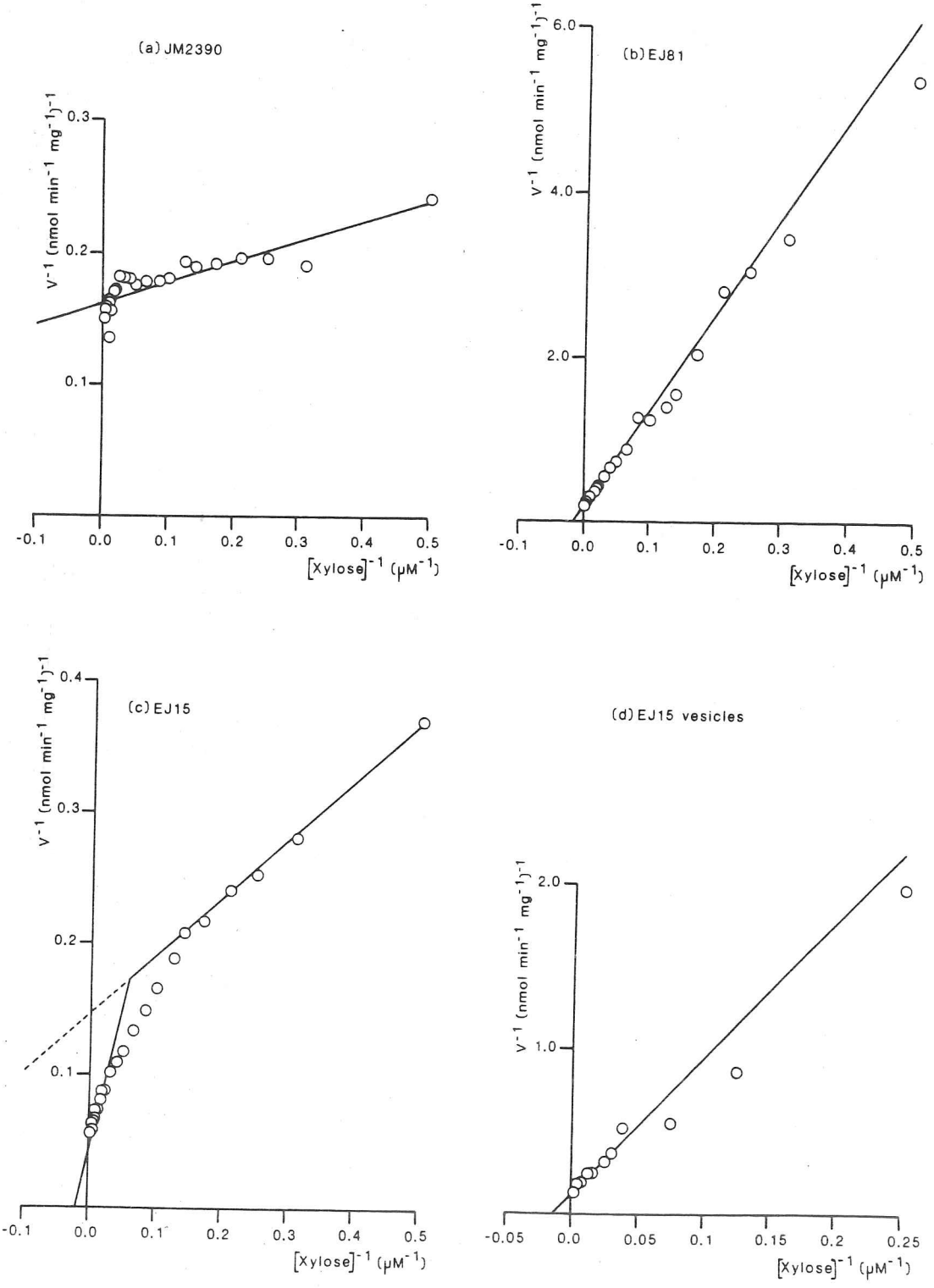


Figure 3.8 Steady state kinetic analysis of xylose transport in strains JM2390 (a), EJ81 (b), EJ15 (c), and EJ15 vesicles (d)

The cultures were grown on glycerol plus xylose as inducer, and prepared and transport measured for 15s time points as described in Chapter 2 (sections 2.6.1 and 2.6.2). The figure shows the means of duplicate measurements. The lines for (a), (b), and (d) are computer best fits by the least squares analysis of Cleland (1967). The K_m and V_{max} values are given in Table 3.13.

with only one transport system, a low affinity system.

Thus JM2390 possesses the high affinity system only, EJ80 has no transport system for xylose, and EJ81 possesses the low affinity system only. Therefore, the Mud(Ap^{R}lac)I insertion must be in xylFG, where xylFG are genes coding for the high affinity, binding protein dependent transport system.

3.5.4 Binding protein activities

Induced and uninduced cultures of JM2390 (xylFG$^+$) and induced cultures of EJ80 (xylFG::Mud(Ap^{R}lac)I) and of EJ70 (xylFG) were made into spheroplasts, and the shock fluids were concentrated and assayed for radio-labelled xylose binding as described in Chapter 2 (section 2.7). The results (Table 3.14) indicate the presence of a xylose-inducible xylose binding protein in JM2390. Strain EJ70 shows no binding protein activity, the values obtained in the assay being comparable with those for the uninduced culture of JM2390. Thus the mutation in EJ70 must lie within the gene coding for the binding protein, designated xylF.

However, strain EJ80 still possesses xylose binding protein activity comparable with that of the XylFG$^+$ strain JM2390. Thus, the Mud(Ap^{R}lac)I insertion in EJ80 must be into a gene coding for some other component of the high affinity transport system such as a membrane component, which shall be designated xylG. Also, since Mud(Ap^{R}lac)I insertions are polar mutations, either xylF and xylG are under the control of different promoters in separate operons, or if they are under the control of a single promoter in one operon xylF must be promoter proximal.

3.6 Mapping of the Mud(Ap^{R}lac)I Insertion in xylG

Preliminary P1 transductions with EJ70 had indicated that the genes coding for the binding protein dependent transport system mapped near to the xyl operon, at 80min (Bachmann, 1983) on the E. coli chromosome having a cotransduction frequency of 13-22% with mtl and of 35-43% with zhj::Tn10, to be compared with a cotransduction frequency of 46-51% for xylA and zhj::Tn10.

Table 3.14 Binding protein activities

The cultures were grown on glycerol plus 10mM xylose as inducer and prepared and assayed as described in Chapter 2 (section 2.7). The values given are the means of two independent determinations from samples taken in triplicate.

Strain	Induced	Binding Activity (pmol bound (mg protein)$^{-1}$)
JM2390	+	107
JM2390	-	-8
EJ70	+	-12
EJ80	+	113

The availability of a Mud(Ap^R*lac*)I insertion in *xylG* enabled the mapping of *xylG* relative to *xylA* by means of the ampicillin resistance determinant of the prophage. P1 generalised transduction was used in both directions and the markers were scored as follows: *xylA* by growth or not on xylose (5mM) (in the *xylE xylG* backgound poor growth occurred for *xylA*$^+$ and none at all for *xylA*); *xylG* by resistance or sensitivity to ampicillin. The Tn10 insertion used in the mapping was located between *mtl* and *xylA* (Campbell *et al.*, 1982).

Transduction V

Phage P1 grown on strain EJ39 was used to transduce strain EJ80 to tetracycline resistance:

P1.EJ39		*xylG*$^+$	(Ap^S)	*xylA*	*zhj*::Tn10	(Tc^R)	*mtl*
EJ80	*xylG*::Mud(Ap^R*lac*)I	(Ap^R)	*xylA*$^+$		(Tc^S)		*mtl*$^+$

The transductants were screened for *xylA* and ampicillin sensitivity, with the results given in Table 3.15:

Table 3.15 Results from transduction V, selection for Tc^R

	xylA	Ap	Observed	Expected
donor type	-	S	68	17.4
	+	S	15	65.6
	-	R	13	63.6
recipient type	+	R	290	239.4

These two markers were found to be linked (χ^2=237), indicating that *xylA* and *xylG* (Ap$^{R/S}$) lie close to each other on the linkage map on the same side of the Tn10 marker, but not giving their relative order.

Transduction VI

Phage P1 grown on strain EJ80 was used to transduce EJ82, a Muc$^+$ lysogen of EJ39, to ampicillin resistance:

P1.EJ80	*xylG*::Mud(Ap^R*lac*)I	(Ap^R)	*xylA*$^+$		(Tc^S)		*mtl*$^+$
EJ82		*xylG*$^+$	(Ap^S)	*xylA*	*zhj*::Tn10	(Tc^R)	*mtl*

Colonies lacking xylose-inducible β-galactosidase were disregarded as events where the Mu prophage had probably moved, then the remaining transductants

were screened for xylA and tetracycline sensitivity, with the results given in Table 3.16:

Table 3.16 Results from transduction VI, selection for Ap^R

	xylA	Tc	Observed	Expected
donor type	+	S	283	284.1
	-	S	38	35.9
	+	R	84	79.9
recipient type	-	R	7	10.1

In this case the results show no linkage between the unselected markers, there being no significant deviation from the expected values assuming no linkage for each class (χ^2=1.29). Therefore, xylA and zhj::Tn10 probably lie on opposite sides of the selected marker, xylG (Ap^R).

The largest class for transduction V was the recipient class, whereas for transduction VI it was the donor class. This is an effect of scoring markers close to a Mu insertion. For recombination to occur the prophage sequences must loop out when the homologous sequences align, and there will be a lower recombinational frequency as a result between the site of the Mu insertion and markers close by. Recombination will, therefore, occur most often further from the insertion site, resulting in the insertion and nearest markers often behaving as a group.

Taking these results in conjunction with the known position of the Tn10 marker used, between xylA and mtl, the xylG gene coding for a component of the high affinity xylose transport system, is thought to map between xylA and zhj::Tn10 with the probable gene order:

xylA xylG zhj::Tn10 mtl

3.7 Discussion

There was previous evidence for the existence of two transport systems for xylose in E. coli (Shamanna and Sanderson, 1979), a proton symport system (Lam et al.,1980) and a binding protein dependent system (Ahlem et al., 1982). As described above, it has been possible to construct strains lacking each or both of these systems. Strains with a single transport system for xylose are

useful for studying the properties of each system, and strains lacking both transport systems are suitable hosts for cloning experiments.

The mapping undertaken has shown the genes for the two transport systems to lie at distinct sites on the chromosome. The genes coding for the binding protein dependent system, xylFG, map close to the xylose metabolic genes, xylABR, at 80min, whereas the gene coding for the xylose proton symport system, xylE, maps separately at 91.4min, close to malB. This mapping is an essential preliminary to cloning the genes involved in xylose transport.

This genetic dissection of xylose transport confirms the existence of two distinct systems for xylose transport. Thus the transport of xylose in E. coli resembles that of galactose (Rotman et al., 1968; Wilson, 1974) and arabinose (Novotny and Englesberg, 1966; Schleif, 1969; Brown and Hogg, 1972) in being mediated by two independent systems.

The K_m for xylose transport by XylE (approx. 63μM) is lower than that for arabinose transport by AraE (145-168μM; Daruwalla et al., 1981) and that for galactose transport by GalP (220-450μM; Henderson and Giddens, 1977), although still of the same order of magnitude. The V_{max} is also lower for XylE (4-6nmol min^{-1} mg^{-1}) than for AraE (15-18nmol min^{-1} mg^{-1}) and GalP (approx. 19nmol min^{-1} mg^{-1}). Thus the xylose proton symport system has a slightly higher affinity but also a slightly lower maximum rate of transport than the arabinose or galactose proton symport systems.

Although the xylG mutation effectively abolishes xylose transport (from 1mM external concentration) in a strain which lacks the xylose proton symport system, the double mutant remains somewhat permeable to higher concentrations of xylose. This is shown both by the xylose-inducibility of β-galactosidase in the double mutant and by its poor but detectable growth on 5mM xylose. The transport system responsible for this residual transport has not been identified.

CHAPTER 4

INHIBITION AND LABELLING STUDIES WITH N-ETHYLMALEIMIDE

CHAPTER 4

INHIBITION AND LABELLING STUDIES WITH N-ETHYLMALEIMIDE

4. INHIBITION AND LABELLING STUDIES WITH N-ETHYLMALEIMIDE

NEM has been found to inhibit other sugar-proton symport systems (LacY (Fox and Kennedy, 1965; Jones and Kennedy, 1969), GalP (Kaethner and Horne, 1980; Macpherson et al., 1983), AraE (Macpherson et al., 1981)), and [^{3}H]-NEM has been used to label the transport proteins involved. Labelling has been improved by the prior protection of a sulphydryl group against reaction with NEM by a specific substrate analogue or competitive inhibitor of the relevant transport process, followed by removal of the protecting agent and then a labelling reaction with [^{3}H]-NEM. The measurement of the difference between the protected sample and an otherwise identical unprotected sample corrects for unspecific labelling of other residues (Macpherson et al., 1981), as only the sites which were protected in the first step should be available for the subsequent labelling step. The labelling reactions were carried out using subcellular vesicles because with intact cells it has not been possible to obtain an identifiable labelled peak for other symport systems (Henderson, personal communication). In order to use this technique for xylose-proton symport it was first necessary to investigate its interaction with NEM.

4.1 Inhibition of Xylose Transport into Vesicles by NEM and Protection by Sugars

4.1.1 Inhibition of xylose transport into vesicles by NEM

The inhibition of xylose transport into vesicles of strain EJ15 by NEM at 25°C and pH6.6 for 15min at concentrations varying from 0 to 4mM was measured as described in Chapter 2 (section 2.8.1). The results (Figure 4.1) indicate that xylose transport is susceptible to inhibition by NEM, with half-maximal inhibition occurring at 0.3mM NEM. NEM has been shown to have no effect on the generation of a pmf by ascorbate plus PMS (Kaback and Patel, 1978), so the inhibition of transport observed must be due to a direct effect on the transport protein.

The increase in the degree of inhibition was relatively gradual at concentrations of NEM above 1mM, which reduced the xylose transport activity to 10.5% and 8.3% of the control values for the 15s and 2min time points

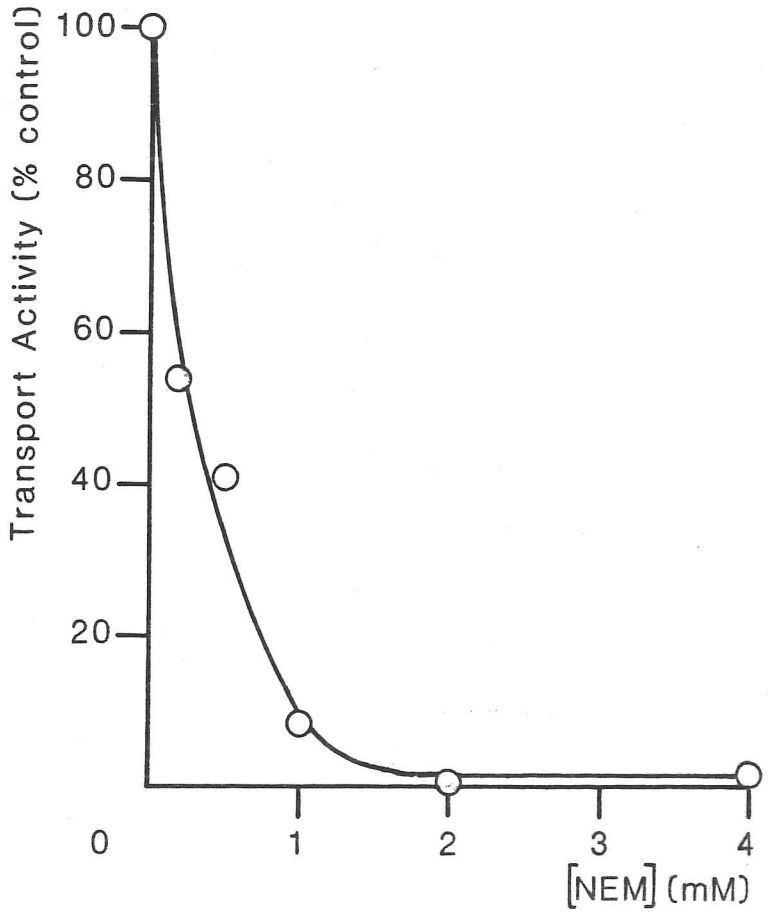


Figure 4.1 Inhibition of xylose transport by NEM in vesicles of strain EJ15

The vesicles were pre-incubated with 1mM NEM for 15min at 25°C and pH6.6, and energised transport assayed as described in Chapter 2 (section 2.8.1). The figure shows the means of duplicate measurements for samples taken 2min after the addition of xylose, expressed as a percentage of the control value, 1.01nmol mg^{-1}.

respectively. Therefore, it was decided to use 1mM NEM for the subsequent protection experiments, thus conforming to the work done on other symport systems where 1mM NEM was used (Macpherson et al., 1981; Macpherson et al., 1983). It was undesirable to use a higher concentration as this would probably require higher concentrations of protecting agent to achieve protection, which might in turn result in a decreased specificity of protection and so of labelling.

4.1.2 Protection by sugars against inhibition caused by NEM

Various sugars were screened for the ability to protect xylose transport against inhibition by NEM in vesicles of strain EJ15 as described in Chapter 2 (section 2.8.1). The sugars, each used at a concentration of 20mM, were xylose, lyxose, xylitol, D-ribose, D-arabinose, glucose, 2-deoxyglucose, 6-deoxyglucose, methyl-α-D-glucoside, methyl-α-D-xylopyranoside, methyl-β-D-xylopyranoside. Of these only xylose, 6-deoxyglucose and glucose gave unequivocal results indicating that they exhibited significant protection against inhibition by NEM. The other sugars either did not show protection against inhibition by NEM (xylitol, ribose, 2-deoxyglucose, methyl-β-D-xylopyranoside) or the results obtained were not sufficiently reproducible to ascertain the degrees of protection achieved. The degree of protection may be calculated based on the relevant differences in transport activity (A) according to the equation (Macpherson, 1982):

$$\% \text{ Protection} = \frac{A_{(\text{NEM + protecting agent})} - A_{(\text{NEM alone})}}{A_{(\text{control without NEM})} - A_{(\text{NEM alone})}} \times 100\%$$

The values obtained are given in Table 4.1.

Table 4.1 Degree of protection by sugars against inhibition by NEM of xylose transport in vesicles of strain EJ15

Sugar	% Protection	
	15s	2min
xylose	59.2	41.9
6-deoxyglucose	70.6	59.3
glucose	43.8	63.6

(The values given are the means of duplicate measurements and, in the case of 6-deoxyglucose, from three independent vesicle preparations.)

As 6-deoxyglucose appeared to be the most effective and it would be useful to use a non-metabolisable analogue it was decided to investigate its protecting ability in more detail.

4.1.3 Protection by 6-deoxyglucose against inhibition by NEM

The ability of 6-deoxyglucose to protect against inhibition by NEM was investigated over a concentration range of 0 to 40mM (Figure 4.2). The maximum protection obtained restored activity to approx. 80-85% of the control (in the absence of NEM) for the 15s points and to approx. 70-75% for the 2min points. Since there was no apparent increase in protection at concentrations greater than 10mM, these results indicated that 10mM 6-deoxyglucose should be sufficient to obtain labelling of the XylE protein. However, it was decided to use 20mM 6-deoxyglucose for the subsequent labelling experiments to conform with the conditions used for other symport systems (Macpherson et al., 1981; Macpherson et al., 1983).

4.2 Interaction of 6-deoxyglucose with the Xylose Proton Symport System

4.2.1 6-deoxyglucose as a substrate of XylE

The ability of 6-deoxyglucose to elicit an alkaline pH change in strain EJ15 was investigated. An alkaline pH change characteristic of proton symport was observed with 6-deoxyglucose but at a rate (0.8nmolH^+ min^{-1} mg^{-1}) lower than that observed with xylose (4.8nmolH^+ min^{-1} mg^{-1}). The extent of the alkaline pH change (5.8nmolH^+ mg^{-1}) was greater than that observed with

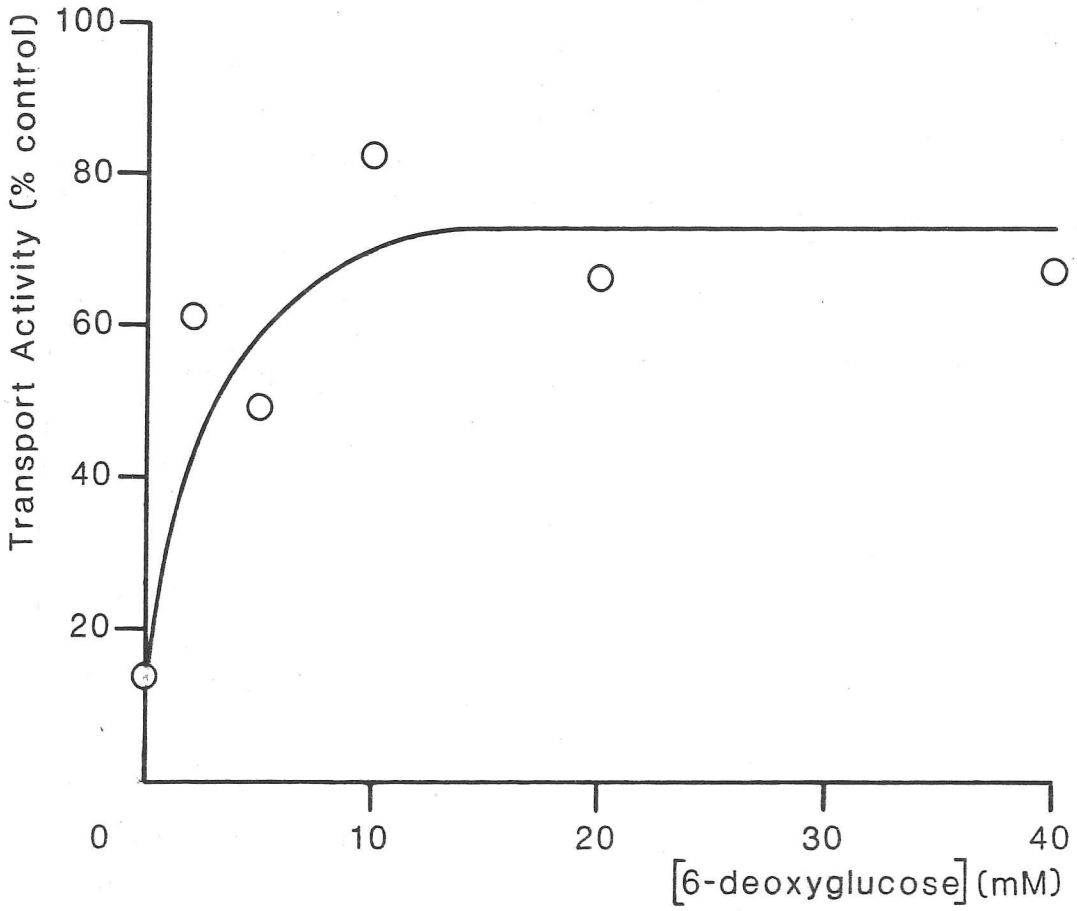


Figure 4.2 Protection by 6-deoxyglucose against inhibition by NEM of xylose transport in vesicles of strain EJ15

The vesicles were pre-incubated with NEM ± 6-deoxyglucose, washed, and energised xylose transport assayed as described in Chapter 2 (section 2.8.1). The figure shows the means of duplicate measurements from two independent determinations for samples taken 2min after the addition of xylose, expressed as a percentage of the control value, 0.66nmol mg^{-1}.

xylose for strain EJ15 (2.6nmolH^+ mg^{-1}) because the 6-deoxyglucose was not metabolised (Figure 4.3).

Measurement of radio-labelled sugar uptake in vesicles of strain EJ15 indicated that there was some transport of 6-deoxyglucose, but at a lower level than the transport of xylose itself (Figure 4.4). There was in addition a small amount of glucose uptake in these vesicles (Figure 4.4), but no transport of arabinose or galactose (data not shown).

Thus it would appear that 6-deoxyglucose is a poor substrate for transport by the xylose-proton symport system. 6-deoxyglucose is also a good substrate for GalP (Henderson, personal communication), so the absence of galactose transport was an important indication that the interaction being observed in this series of experiments was independent of GalP.

4.2.2 6-deoxyglucose as an inhibitor of XylE

The effect of 6-deoxyglucose on the transport of xylose by XylE was investigated.

Energised xylose transport was measured in strain EJ15 with or without the prior addition of an equal concentration of 6-deoxyglucose. The extent of the alkaline pH change when xylose was added 3min after 6-deoxyglucose (0.04nmolH^+ mg^{-1}) was reduced to about 1% of that in the absence of 6-deoxyglucose (2.6nmolH^+ mg^{-1}) (Figure 4.5).

The transport of [^{14}C]-xylose into vesicles of strain EJ15 was measured as in Chapter 2 (section 2.6.2), with the addition of 6-deoxyglucose 3min before that of xylose. The concentration of 6-deoxyglucose was varied from 0 to 20mM. The transport of xylose was strongly inhibited even at low concentrations of 6-deoxyglucose (Figure 4.6), with half-maximal inhibition occurring at a concentration of 0.2-0.3mM. The activity was reduced essentially to its minimum of 6.5% of the control activity by 5mM 6-deoxyglucose for the 15s points; the effect was more pronounced at lower concentrations for the 2min points when there was no significant increase in inhibition beyond 1mM 6-deoxyglucose, which retained only 5.7% activity of the control value.

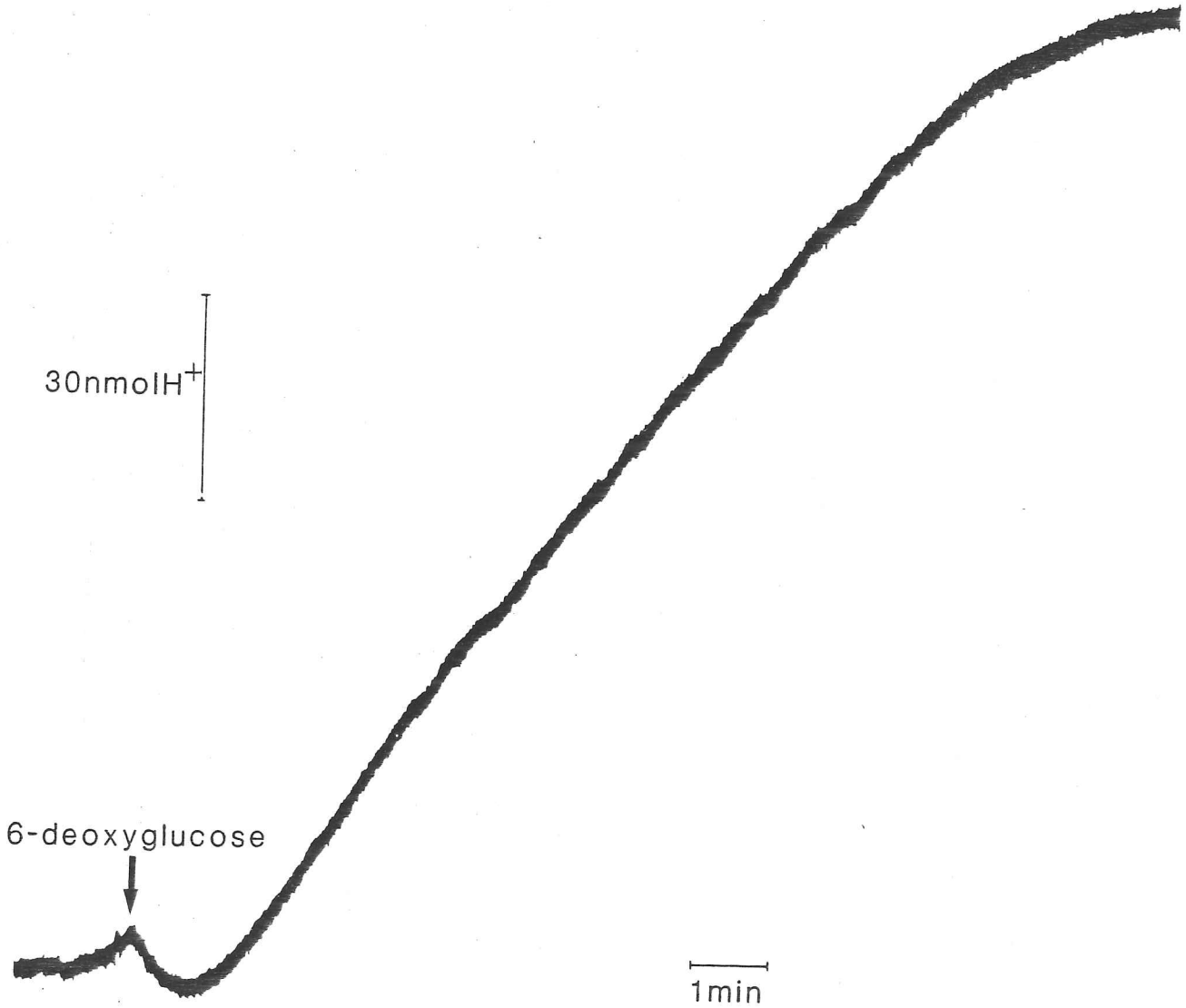


Figure 4.3 6-deoxyglucose-promoted alkaline pH change in EJ15

The cells were grown on glycerol plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3). The recording was calibrated by the addition of 3µl 0.01M NaOH immediately prior to the addition of substrate (20µl 0.5M).

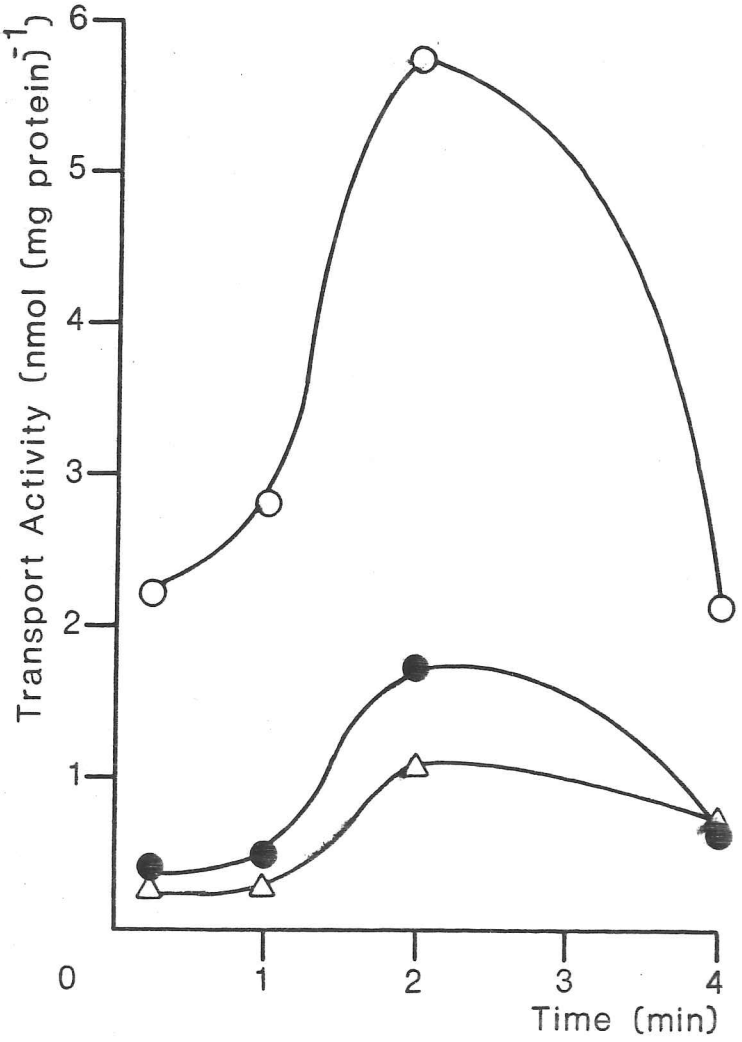


Figure 4.4 Transport of xylose, 6-deoxyglucose and glucose in vesicles of strain EJ15

PMS/ascorbate energised transport was measured as described in Chapter 2 (section 2.6.2), but taking 50µl samples at intervals. The figure shows the means of duplicate measurements: xylose, (○); 6-deoxyglucose, (●); glucose, (△).

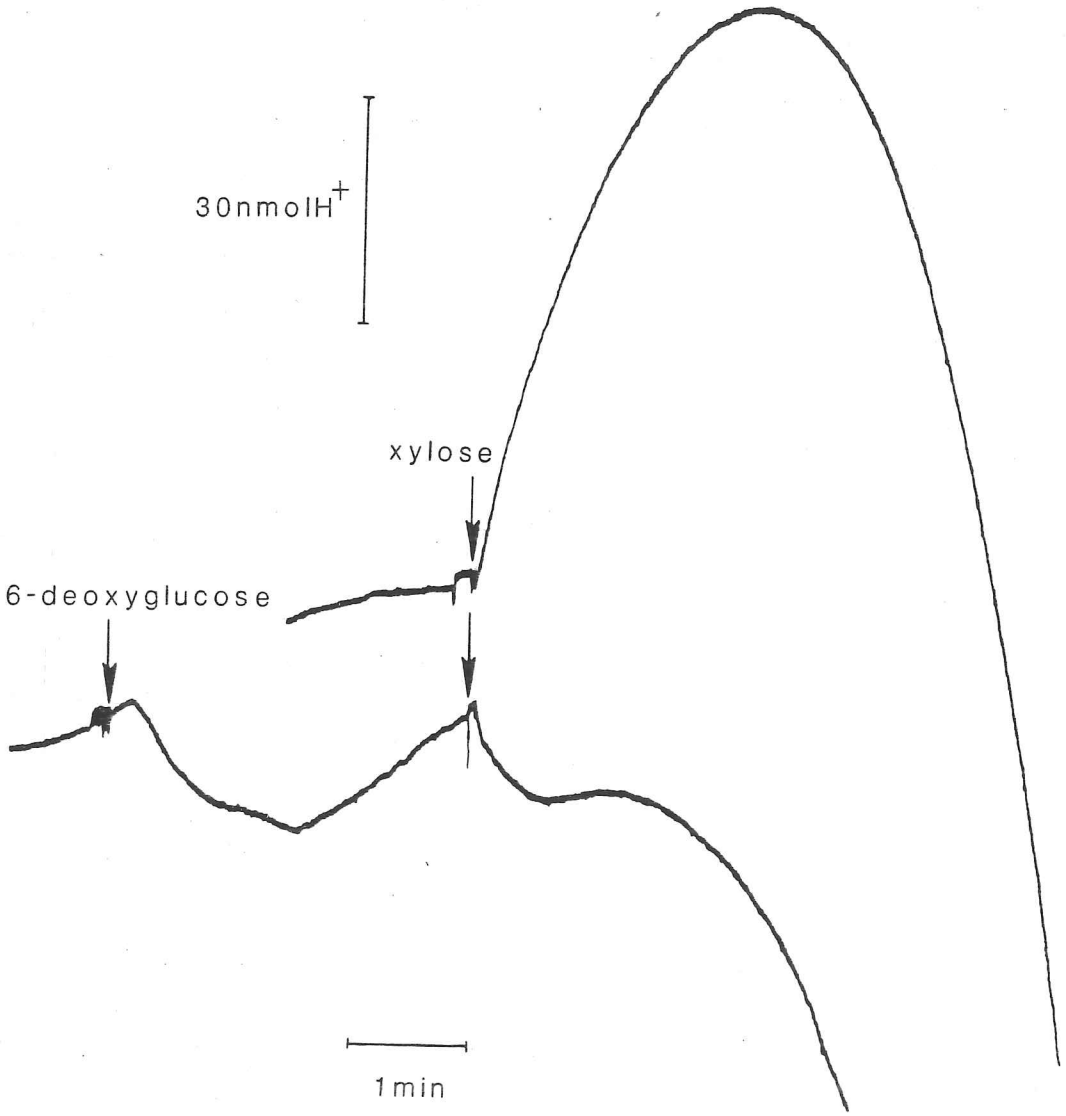


Figure 4.5 Inhibition of xylose-promoted alkaline pH change by 6-deoxyglucose in strain EJ15

The cells were grown on glycerol plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3). The recordings were calibrated by the addition of 3µl 0.01M NaOH immediately prior to the addition of sugar. Where included, an equal quantity of 6-deoxyglucose (20µl 0.5M) was added 3min before the addition of xylose.

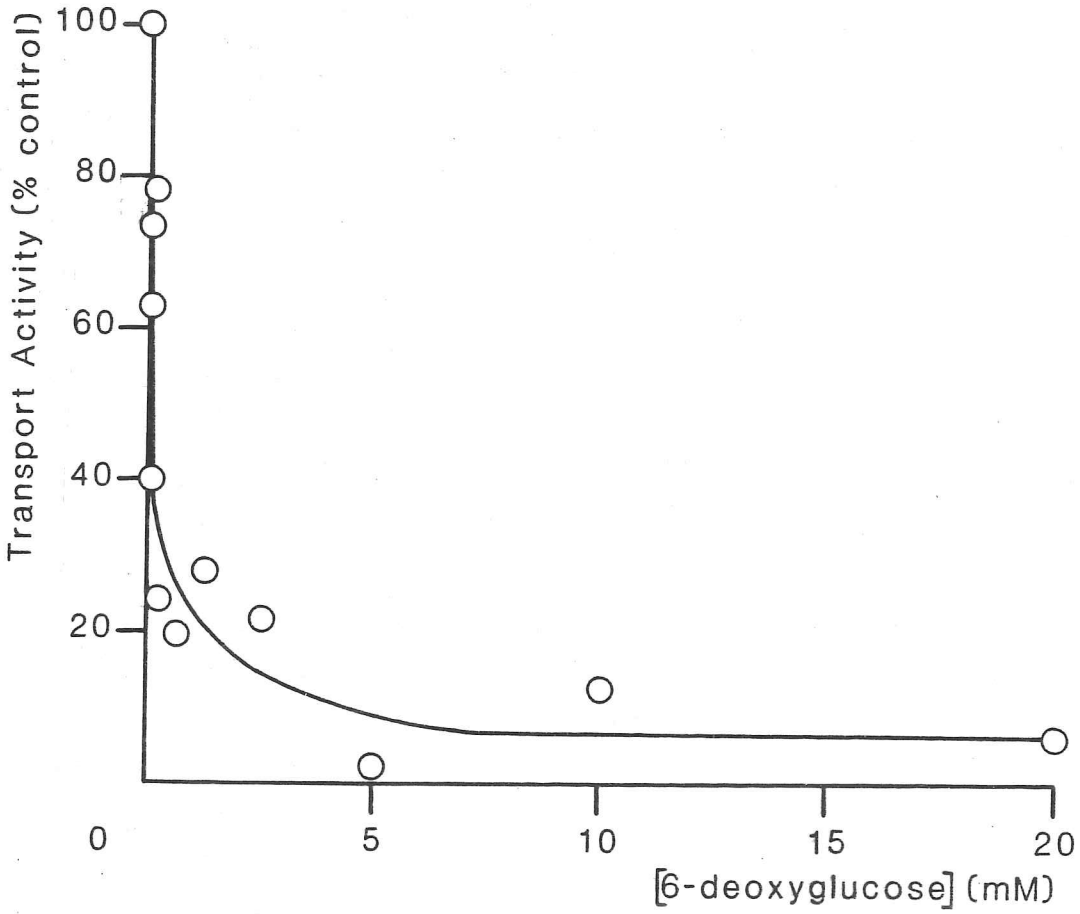


Figure 4.6 Inhibition of xylose transport by 6-deoxyglucose in vesicles of strain EJ15

PMS/ascorbate energised transport was measured as described in Chapter 2 (section 2.6.2), with varying concentrations of 6-deoxyglucose added 3min before the addition of [^{14}C]-xylose. The figure shows the means of duplicate measurements for 15s time points, expressed as a percentage of the control value, 0.37nmol mg^{-1} (mean of four measurements).

Thus it would appear that 6-deoxyglucose is a good inhibitor of xylose transport by the xylose-proton symport system and that the substrate binding site can accommodate the bulk of an extra methyl group on carbon-5 of xylose.

4.3 Specificity of 6-deoxyglucose Inhibition of Xylose Symport

If 6-deoxyglucose were to be useful in experiments to label selectively the xylose-proton symporter, then it should ideally bind only to this transport system and so leave only this free sulphydryl group for labelling by radioactive NEM. Accordingly, the effect of 6-deoxyglucose on the transport of xylose by XylFG and of glucose by PtsG was investigated. The interaction of 6-deoxyglucose with GalP under the conditions of the labelling experiments had been eliminated, for GalP was not expressed (section 4.2.1).

4.3.1 Inhibition of xylose binding protein transport

Inhibition of [^{14}C]-xylose transport in strain EJ18 (*xylE*$^-$ *xylFG*$^+$) by 6-deoxyglucose and unlabelled xylose, as a control, was investigated as in Chapter 2 (section 2.6.1). The unlabelled sugar was used at concentrations of 0.5mM and 5mM, 10-fold and 100-fold respectively the concentration of [^{14}C]-xylose used. Only slight inhibition of xylose transport was observed for 6-deoxyglucose at 0.5mM (approx. 75% transport activity remaining) despite the reduction in transport activity by unlabelled xylose at 0.5mM to approx. 10% of the control values (Table 4.2). At 5mM 6-deoxyglucose more substantial inhibition of xylose transport occurred, with a reduction in transport activity to approx. 40% of the control value. This may be compared with the essential abolition of xylose transport by 5mM unlabelled xylose, when only approx. 1% of the xylose transport activity remained (Table 4.2).

Thus 6-deoxyglucose is a much more powerful inhibitor of xylose transport by the proton symport system (section 4.2.2) than by the binding protein system. The inhibition of xylose transport in a XylE$^+$ XylF$^-$ strain would provide the ideal comparison with the experiment described in this section.

Table 4.2 Inhibition of XylFG transport activity in EJ18 by 6-deoxyglucose and xylose

The cells were grown on glycerol plus 10mM xylose as inducer and prepared and assayed as described in Chapter 2 (section 2.6.1). The unlabelled inhibitory sugar was added 3min before the [^{14}C]-xylose. The values given are the means of duplicate measurements.

Inhibitor	Concentration (mM)	Transport Activity (%)	
		Initial Rate	Extent at 2min
-	0	100	100
Xylose	0.5	13	9
Xylose	5	1	1
6-deoxyglucose	0.5	72	78
6-deoxyglucose	5	39	43

4.3.2 Inhibition of glucose transport by 6-deoxyglucose

It was possible that 6-deoxyglucose might interact with the glucose transport systems. Therefore, inhibition of transport of glucose and of methyl-α-D-glucoside in strain EJ15 by 6-deoxyglucose and by glucose as a control at a concentration of 5mM (100-fold the concentration of the substrate) was measured. Methyl-α-D-glucoside is a specific substrate of the ptsG system (Curtis and Epstein, 1975; Stock et al., 1982); the substrate specific for the ptsM system (2-deoxyglucose) was not used as the strain used was ptsM$^-$. No significant inhibition of the transport of either glucose or methyl-α-D-glucoside by 6-deoxyglucose was observed despite the drastic reduction in the transport activity by 5mM glucose (Table 4.3). It is possible that 6-deoxyglucose would inhibit glucose transport via ptsM, but by using strain EJ15, and related strains which are ptsM$^-$, it can be assumed that 6-deoxyglucose does not interact with the glucose transport systems.

4.4 Labelling of XylE with NEM

In order to improve the clarity of labelling it was decided to use a labelling method involving both [^{3}H]-NEM and [^{14}C]-NEM. The protected sample was labelled with one isotope and the unprotected sample was labelled with the other isotope, so that the two could be combined and run in one track on an SDS-polyacrylamide gel as described in Chapter 2 (sections 2.8.2 and 2.8.3). This enabled any variation in the slicing of two separate tracks to be eliminated. By running dual experiments with the isotopes the other way round in relation to the protected sample any differences in specific activity or impurities causing preferential labelling by one isotope of non-specific proteins would be detected, and averaging of the two sets of results could be used to remove this background effect.

4.4.1 Comparison of strain EJ15 (xylE$^+$) with strain EJ71 (xylE$^\Delta$)

Vesicles of strain EJ15 were labelled as in Chapter 2 (section 2.8.2) using both [^{3}H]-NEM and [^{14}C]-NEM. The results (Figure 4.7) indicated the presence of a 6-deoxyglucose protectable peak in strain EJ15 at an apparent molecular mass of 38 500-41 500Da. When the experiment was repeated with vesicles of strain EJ71 (xylE$^\Delta$) this peak was found to be absent (Figure 4.8), indicating

Table 4.3 Inhibition of glucose and methyl-α-glucoside transport activity in EJ15 by 6-deoxyglucose and glucose

The cells were grown on glucose and prepared and assayed as described in Chapter 2 (section 2.6.1). The inhibitory sugar was added to a final concentration of 5mM 3min before the addition of radioisotope labelled substrate. The values given are the means of duplicate measurements.

Inducer	Glucose Transport (%)		Methyl-α-glucoside Transport (%)	
	Initial Rate	Extent at 2min	Initial Rate	Extent at 2min
-	100	100	100	100
Glucose	6	3	0	1
6-deoxyglucose	92	89	110	114

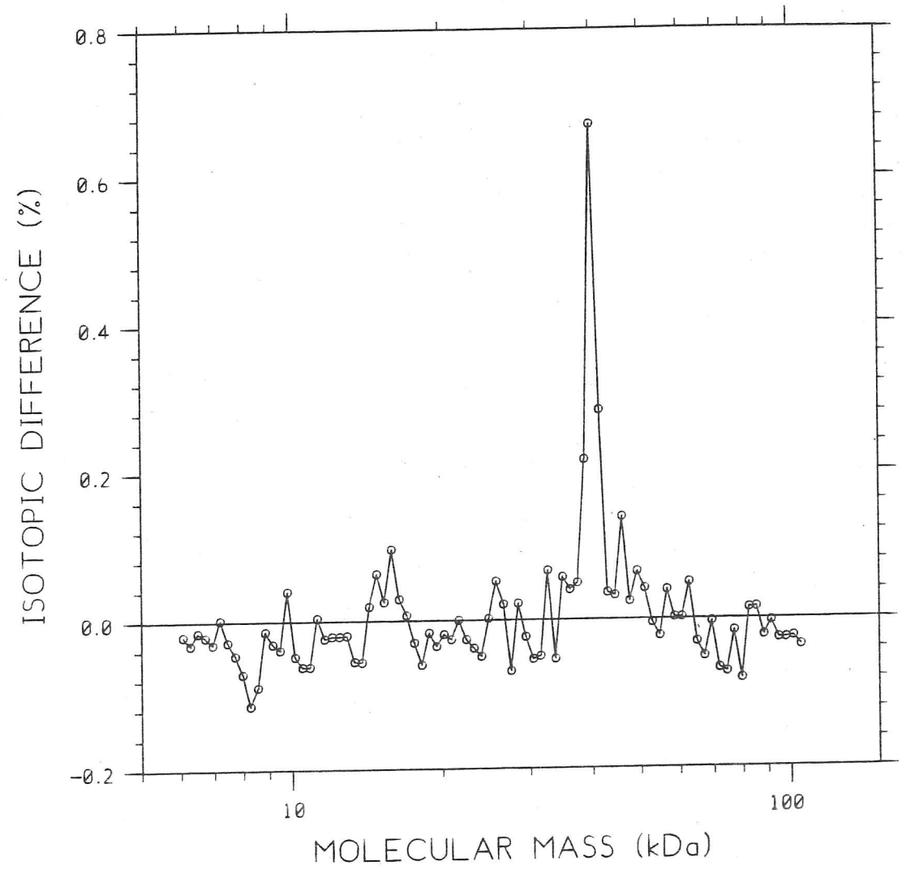


Figure 4.7 6-deoxyglucose protectable NEM-labelled proteins from vesicles of strain EJ15 ($xylE^{+}$)

The vesicles were labelled with [^{14}C]- or [^{3}H]-NEM after reaction with unlabelled NEM in the presence or absence of 6-deoxyglucose, as described in Chapter 2 (section 2.8.2). The proteins were separated by SDS-polyacrylamide gel electrophoresis, the tracks were sliced, and the readioactivity in each slice was determined as described in Chapter 2 (section 2.8.3). The figure shows the means of the protected minus the unprotected values from two independent experiments with the labels reversed.

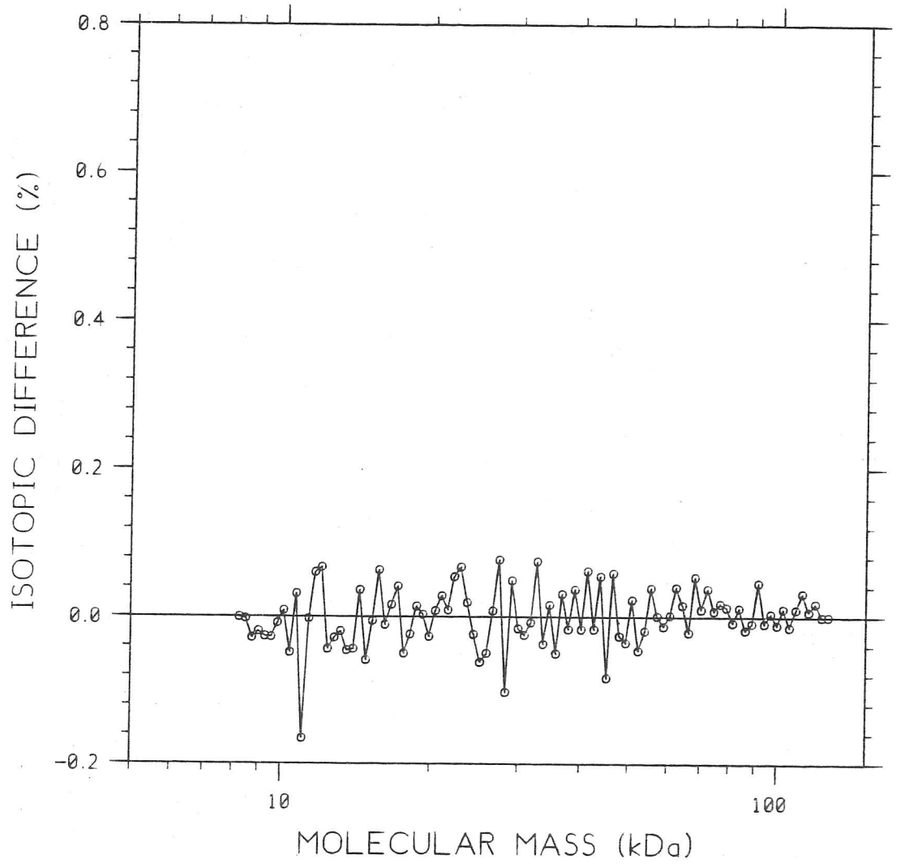


Figure 4.8 6-deoxyglucose protectable NEM-labelled proteins from vesicles of strain EJ71 ($xylE^{\Delta}$)

The conditions were as described for Figure 4.7. The figure shows the means of the protected minus the unprotected values from two independent experiments with the labels reversed.

that the peak occurring with strain EJ15 was indeed due to the XylE protein.

4.4.2 Comparison of strain EJ68 (*xylE*$^+$) with strain EJ54 (*xylE*::Mud(Ap^R*lac*)II)

Strains EJ68 and EJ54 are *xylA* (xylose isomerase negative) and EJ68 has higher levels of xylose-proton symport activity than EJ15 as a result. This is probably due to higher intracellular concentrations of inducer and the absence of any self-catabolite repression due to lack of metabolism. Such effects have previously been observed for the arabinose genes (Katz and Englesberg, 1971; Beverin *et al.*, 1971).

The labelling experiments with NEM were, therefore, repeated using vesicles of these strains. Again a 6-deoxyglucose protectable peak was found in the *xylE*$^+$ strain, EJ68, at an apparent molecular mass of 35 500-39 000Da, but in this case the degree of labelling was greater (1.4% compared with 0.7%, Figure 4.9). This peak was absent from EJ54 vesicles (Figure 4.10) confirming its identity as being due to XylE. Also, no high molecular mass peak was observed for EJ54, indicating that the hybrid XylE-LacZ protein produced does not possess a normal xylose binding site, although it includes a large part of the XylE protein (see Chapter 5). In this case, however, the insertion mutation *xylE*::Mud(Ap^R*lac*)II resulted in the expression of *lacY* under the control of the xylose promoter. Thus the vesicles of EJ54 could be checked for energised uptake of lactose as the EJ68 vesicles were for xylose. In addition, it was possible for Dr. P.J.F. Henderson to label LacY with NEM using TMG as a protecting agent (Henderson, personal communication), giving a peak of apparent molecular mass approx. 30 000Da, as expected for LacY (Jones and Kennedy, 1969) and thus demonstrating that a protectable NEM-labelled peak could be identified in the EJ54 vesicles.

4.5 Discussion

As with the other proton symport systems, LacY, AraE, and GalP, transport by XylE was susceptible to inhibition by the sulphydryl reagent N-ethylmaleimide. This inhibition was protected against by the substrate or a substrate analogue, indicating the presence of a sulphydryl group in the vicinity of the substrate binding site, and allowing relatively specific labelling of the transport protein with radioisotope-labelled NEM. In this

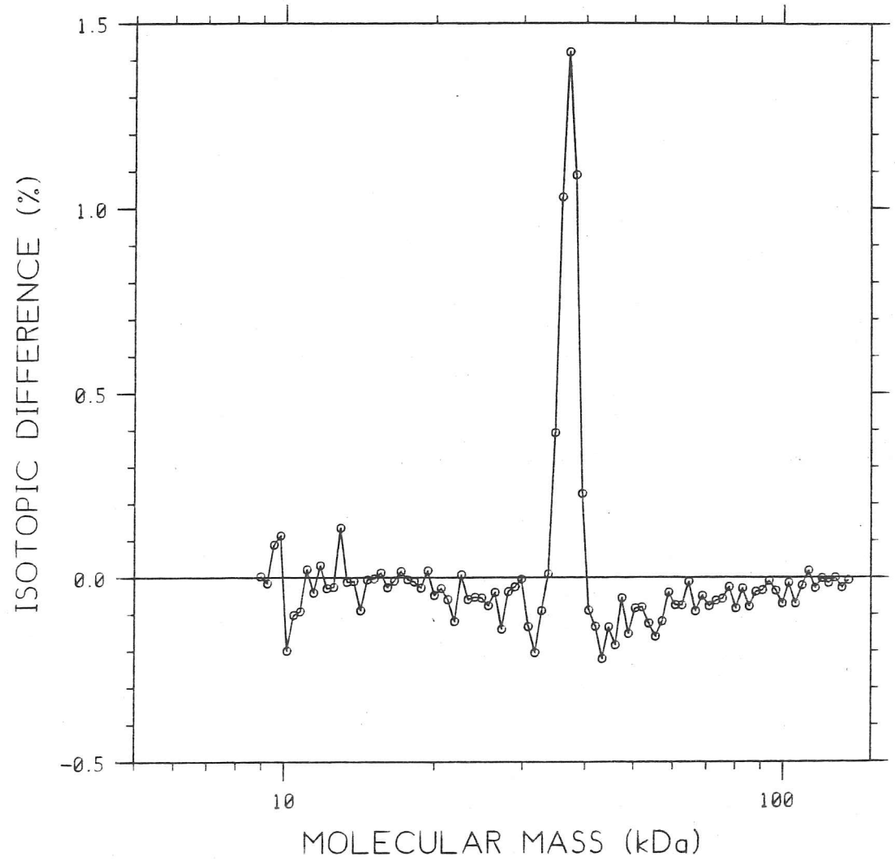


Figure 4.9 6-deoxyglucose protectable NEM-labelled proteins from vesicles of strain EJ68 (xylE$^+$)

The conditions were as described for Figure 4.7. The figure shows the means of the protected minus the unprotected values from two independent experiments with the labels reversed.

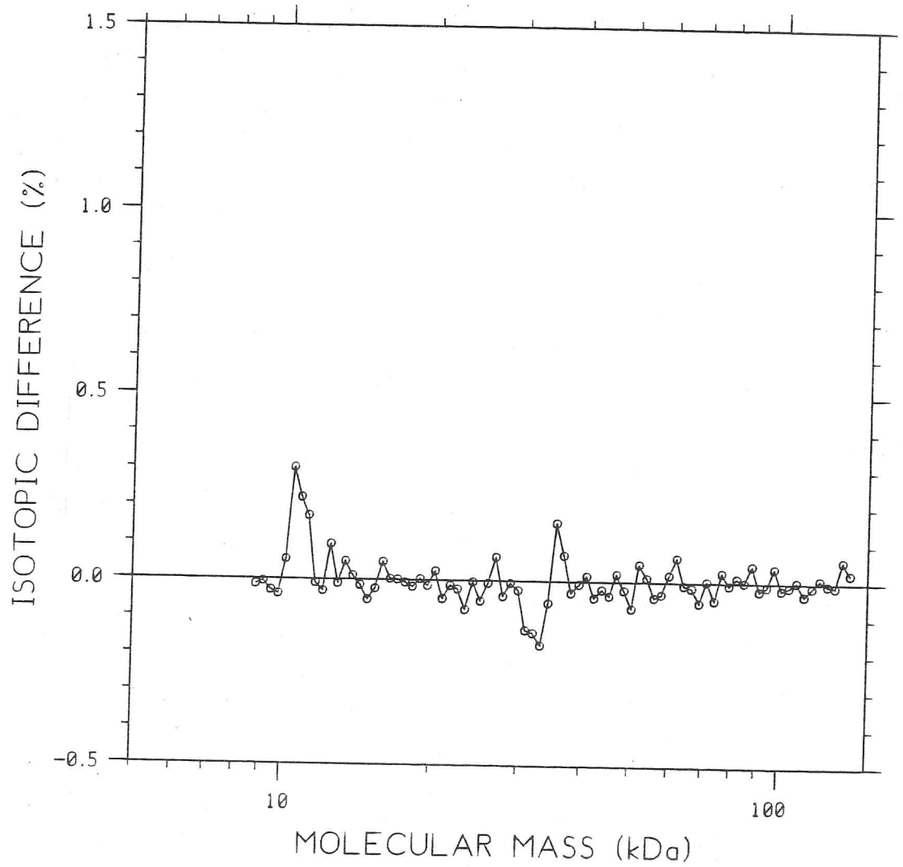


Figure 4.10 6-deoxyglucose protectable NEM-labelled proteins from vesicles of strain EJ54 (xylE::Mud(Ap^{R}lac)II)

The conditions were as described for Figure 4.7. The figure shows the means of the protected minus the unprotected values from two independent experiments with the labels reversed.

case 6-deoxyglucose was found to be a suitable protecting agent; although only a poor substrate for transport by XylE, 6-deoxyglucose proved to be a good inhibitor of xylose transport by XylE. It was not such a potent inhibitor of xylose transport by XylFG. Comparison of protected versus unprotected samples revealed a labelled peak of apparent molecular mass in the range 36 000-41 000Da for the $XylE^+$ strains, which was absent in the $XylE^-$ strains. This dual characteristic of protectability by a specific inhibitor of XylE and absence in $XylE^-$ strains identifies this peak as due to the XylE protein, and not a XylFG component. The value for the apparent molecular mass of XylE is similar to those of AraE (36 000-38 000Da, Macpherson et al., 1981) and of GalP (34 000-38 000Da, Macpherson et al., 1983) but different from that of LacY (approx. 30 000Da, Jones and Kennedy, 1969; Newman et al., 1981). However, the true molecular mass of LacY obtained from the DNA sequence is 46 502Da (Buchel et al., 1980) and the true molecular mass of XylE is also greater than that estimated from SDS-polyacrylamide gel electrophoresis (see Chapter 6). This discrepancy is thought to be due to an abnormally high binding of SDS to the relatively hydrophobic protein, resulting in an increased charge-to-mass ratio; at sufficiently high gel concentrations (20% and above) when the molecular sieving factor predominates over this increased charge-to-mass ratio the molecular mass found for LacY was approx. 46 000Da (Beyreuther et al., 1980). The molecular mass of LacY was also shown to be 47 500Da by gel filtration in the presence of hexamethylphosphoric triamide (Konig and Sandermann, 1982).

CHAPTER 5

PURIFICATION OF A XYLE-LACZ HYBRID PROTEIN

5. PURIFICATION OF A XYLE-LACZ HYBRID PROTEIN

A Mud(Ap^R*lac*)II insertion in the *xylE* gene was made by Dr. M.C. Jones-Mortimer (Davis *et al.*, 1984). Mud(Ap^R*lac*)II differs from Mud(Ap^R*lac*)I in that as well as there being no transcriptional termination signals between the end of Mu and the beginning of the *lac* structural genes, there are no translational termination sites (the *trp* sequences and the first eight codons of *lacZ* are deleted relative to Mud(Ap^R*lac*)I). Thus insertion of Mud(Ap^R*lac*)II into a gene in the correct orientation and reading frame gives rise to gene fusion and the production of a hybrid protein with the N-terminus of the gene into which the prophage is inserted fused to β-galactosidase.

The aim of the work described in this chapter was to purify the XylE-LacZ hybrid protein produced by this Mud(Ap^R*lac*)II insertion in *xylE*. By utilising the properties of the β-galactosidase part of the molecule for this purification, the N-terminal part of the XylE protein fused to it would be obtained. This would enable determination of the N-terminal sequence of the XylE protein for verification of the reading frame deduced from DNA sequencing of the cloned *xylE* gene.

5.1 Evidence that the XylE-LacZ Hybrid Protein is Membrane Bound

β-galactosidase assays on sonicated preparations of JM2336 (*xylE*::Mud(Ap^R*lac*)II) were performed as described in Chapter 2 (section 2.3.3). This confirmed the xylose-inducibility of the β-galactosidase activity in JM2336 and showed the majority of this activity to be membrane bound, compared to the mainly cytoplasmic xylose-inducible β-galactosidase of EJ18 resulting from the Mud(Ap^R*lac*)I insertion in *xylE* (Table 5.1).

The *xylA* derivative of JM2336, EJ54, showed much higher levels of β-galactosidase activity than JM2336 when induced with xylose (approx. 1 725 compared to approx. 316 units (mg dry mass)$^{-1}$), and was used for the purification work.

Table 5.1 Xylose inducibility and membrane association of β-galactosidase activity in JM2336

The cells were grown on glycerol in the presence or absence of 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.3.3). The values given are the means of duplicate measurements from each of two independent determinations.

Strain	Induced	β-galactosidase activity (units mg^{-1})	Cytoplasmic (%)	Membraneous (%)
JM2336	+	367.5	4	96
JM2336	-	2.2	-	-
EJ18	+	490.1	78	22
EJ18	-	2.5	-	-

5.2 Identification of the Hybrid Protein

The band corresponding to the hybrid protein on an SDS-polyacrylamide gel was identified among the proteins binding to a monoclonal antibody against β-galactosidase. The proteins solubilized from a membrane preparation were incubated with the antibody which was attached to a Sepharose resin. The resulting protein-antibody-resin complex was sedimented and washed, then the bound proteins were eluted with 6M urea, as described in Chapter 2 (section 2.9). The eluates were dialysed prior to analysis by SDS-polyacrylamide gel electrophoresis. Induced preparations of strains EJ54 and EJ18 and an uninduced preparation of strain JM2336 were compared.

The monomer of β-galactosidase has a large molecular mass (116 000Da) relative to most proteins in the cell. The hybrid protein must have a molecular mass geater than or equal to that of native β-galactosidase, since part of the XylE protein is fused to the β-galactosidase molecule. The results (Figure 5.1) show only one band of molecular mass greater than that of β-galactosidase (found in EJ18) which is present in the induced preparation of EJ54 but absent in the uninduced preparation of JM2336. This band, with an apparent molecular mass of approx. 145 000-150 000Da, must, therefore, correspond to the XylE-LacZ hybrid protein.

5.3 Solubilization of the XylE-LacZ Hybrid Protein

The conditions required to solubilize the hybrid protein were investigated in a series of small scale experiments prior to any attempt at purification.

5.3.1 Selection of a suitable detergent

The detergents Triton X-100, octyl-glucoside, Nonidet-P40 and Lubrol-PX were compared for their ability to solubilize the β-galactosidase activity of the hybrid protein. The protein concentration was 1mg ml^{-1} and the detergents were tested over a concentration range of 0-2% for Triton X-100 and octyl glucoside and of 0-4% for Nonidet-P40 and Lubrol-PX, as described in Chapter 2 (section 2.10.1). The activity of the supernatants after centrifuging in the airfuge at 100 000g for 1h was expressed as a percentage of the activity in

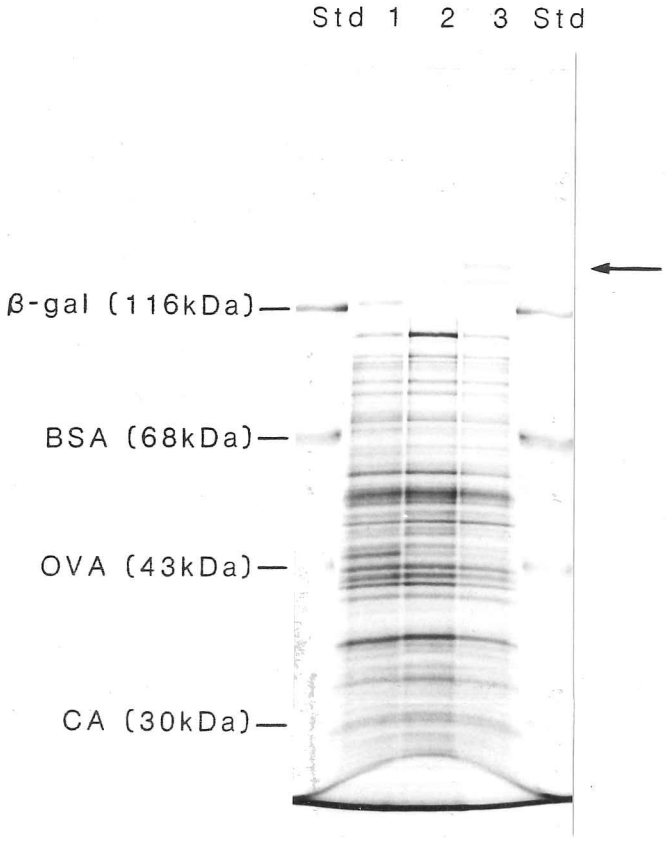


Figure 5.1 Identification of the XylE-LacZ hybrid protein

The band corresponding to the XylE-LacZ hybrid protein was identified as described in Chapter 2 (section 2.9). The figure shows a 10% SDS-polyacrylamide gel of the proteins binding to a monoclonal antibody against β-galactosidase: track 1, EJ54 induced; track 2, JM2336 uninduced; track 3, EJ18 induced. The arrow indicates the XylE-LacZ hybrid protein, present in EJ54. The positions of the standard molecular mass markers are indicated: β-gal, β-galactosidase; PhB, phosphorylase B; BSA, bovine serum albumin; OVA, ovalbumin; CA, carbonic anhydrase.

the total mixture, sampled immediately prior to centrifuging. In each case quite a large proportion (mean value of 41%) was 'solubilised' in the absence of detergent. This was probably due to the action of proteases splitting the hybrid protein into its two constituent parts during centrifugation as cooling was not possible using the airfuge. In other experiments this problem was much less severe. The results (Figure 5.2) indicate that Triton X-100 is the most effective of these detergents at solubilizing the hybrid protein at 1mg ml^{-1} protein concentration. Therefore, Triton X-100 was chosen for subsequent experiments, with the added consideration that to use a large column for gel filtration the cost of using octyl-glucoside, for example, would be prohibitive. However, Triton X-100 has the disadvantage of absorbing light of wavelength 280nm so preventing the continuous spectrophotometric monitoring of protein as it elutes from a column.

5.3.2 Determination of the conditions for solubilizing with Triton X-100

For a large scale preparation it would be necessary to solubilize much larger quantities of protein, but for chromatography on the gel filtration column it was important to keep the sample volume relatively small to maintain resolution. It would, therefore, be necessary to solubilize at higher protein concentrations.

The solubilization of the hybrid protein by 1% Triton at protein concentrations of 0.5-5mg ml^{-1} was compared using the airfuge method (Chapter 2, section 2.10.1). The results (not shown) indicated that the proportion of the protein solubilized decreased with increasing protein concentration.

The solubilization of the hybrid protein at 1mg ml^{-1} and at 7.5mg ml^{-1} was compared, varying the concentration of Triton from 1% to 5%, using the airfuge method. The results (Table 5.2) clearly indicate that whilst for a protein concentration of 1mg ml^{-1} raising the concentration of Triton from 1% to 5% does not increase the amount of the hybrid protein solubilized, for a protein concentration of 7.5mg ml^{-1} the higher concentration (5%) of Triton is required to achieve a reasonable degree of solubilization.

Subsequently, solubilization for the smaller scale experiments was done at 1mg ml^{-1} protein and 1% Triton X-100; and for the larger scale preparations at

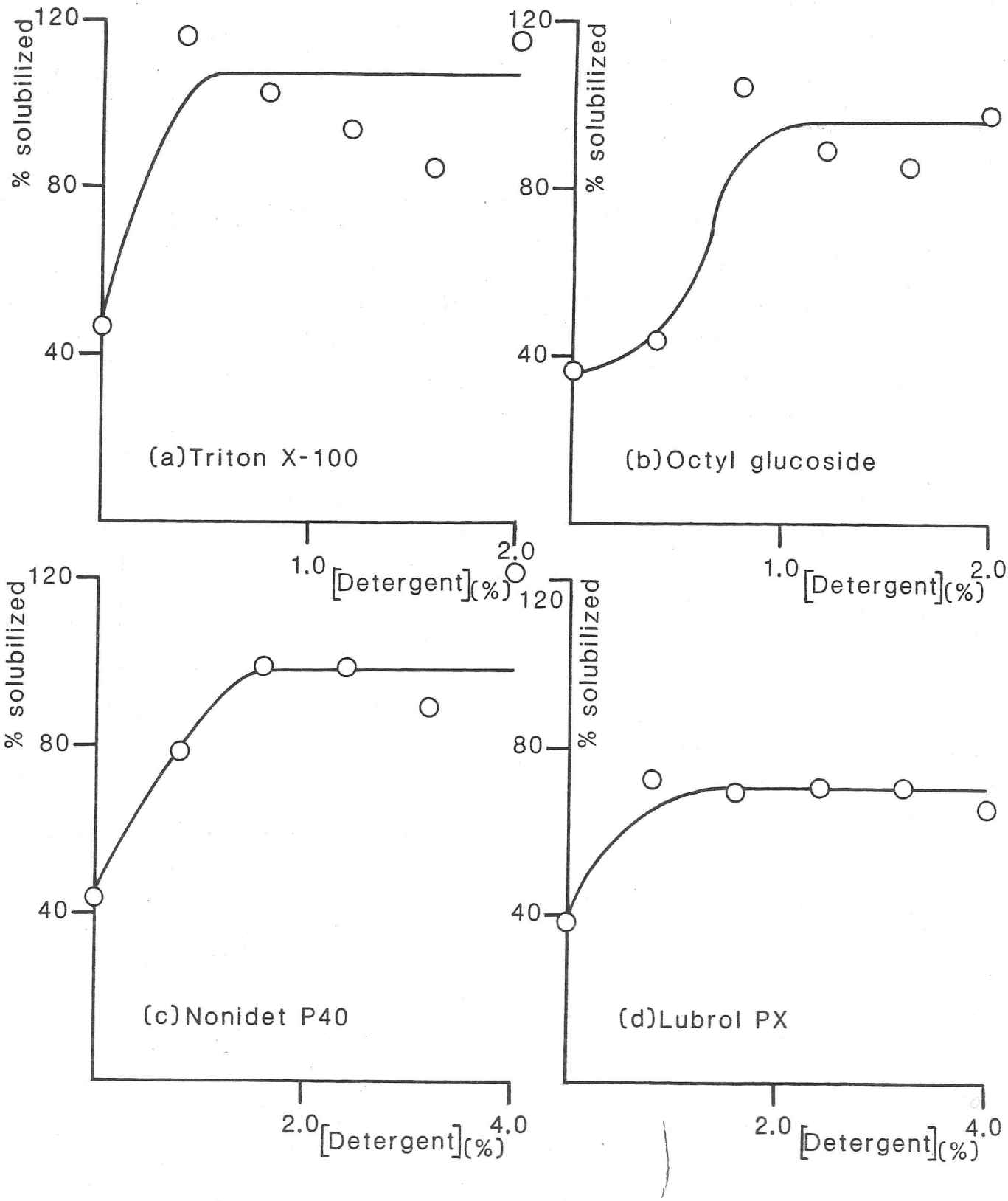


Figure 5.2 Solubilization of the XylE-LacZ hybrid protein by the detergents Triton-X100 (a), octyl glucoside (b), Nonidet P40 (c) and Lubrol PX (d)

The solubilization of the hybrid protein by each detergent was investigated as described in Chapter 2 (section 2.10.1). The figure shows the β-galactosidase activity solubilized as a percentage of the total β-galactosidase activity prior to centrifugation.

Table 5.2 Solubilization of XylE-LacZ hybrid protein by Triton X-100

Membranes made from xylose-induced EJ54 were prepared and solubilized as described in Chapter 2 (sections 2.1.4 and 2.10.1). The percentages solubilized were determined by assaying β-galactosidase activities as described.

Triton concentration	% solubilized	
(%)	1mg ml^{-1} protein	7.5mg ml^{-1} protein
1	46.2	29.9
2	43.1	33.5
5	49.5	55.5

7.5mg ml^{-1} protein and 5% Triton X-100, as in Chapter 2 (section 2.10.2).

5.4 Affinity Column Chromatography

The resin p-aminophenyl-β-D-thiogalactoside-agarose was shown by Steers *et al.* (1971) to be an affinity matrix for β-galactosidase. They described the purification of β-galactosidase on a column of this resin, loading the column in 50mM Tris-HCl/100mM NaCl/10mM $MgCl_2$, pH7.5 and eluting with 100mM sodium borate, pH10. They failed to obtain elution with 50mM lactose in the loading buffer, although it might have been expected that substrate would elute the protein from an affinity column.

It was hoped that the XylE-LacZ hybrid protein could be purified in a similar way to β-galactosidase on the affinity column. However, the hybrid protein was in fact very different from native β-galactosidase. β-galactosidase is a cytoplasmic enzyme whereas the hybrid protein was membrane bound, so the column would have to be run in the presence of detergent to keep the hybrid protein solubilized. Also, as only part of the hybrid protein was β-galactosidase, the XylE portion might interfere with the binding of the β-galactosidase moiety to the column. In addition, the set of proteins from which the hybrid protein was being purified was completely different to that from which β-galactosidase was purified. These factors meant that the conditions for chromatography on the column might need to be altered.

5.4.1 Effect of salt concentration

Initially chromatography on the column was performed as described by Steers *et al.* (1971). The hybrid protein was found to bind to the resin under these conditions and to elute with 100mM sodium borate, pH10. However, the column did not appear to have a very large capacity for binding the hybrid protein, and the eluted protein was not substantially freed of the other proteins. It appeared that the membrane proteins from which the hybrid protein needed to be purified might be binding to the column more strongly than the cytoplasmic proteins from which native β-galactosidase was purified.

Some experiments done with native β-galactosidase confirmed that a high concentration of lactose (200mM) failed to elute β-galactosidase from the affinity resin. Running a salt gradient in the loading buffer prior to elution with 100mM sodium borate, pH10 indicated that 1M NaCl did not elute β-galactosidase from the column, but the β-galactosidase eluted subsequently appeared to be more pure on an SDS-polyacrylamide gel than that obtained with no prior salt wash.

Therefore, a salt gradient was applied with a sample (2mg protein, approx. 12 000 units β-galactosidase activity) of the hybrid protein loaded onto the affinity column, after the initial wash through with loading buffer. The gradient was applied in two stages; first from 0.1M to 0.3M NaCl, then from 0.3M to 1M NaCl, each over 5 column volumes. Finally, the elution was with 100mM sodium borate, pH10.

The β-galactosidase activity, i.e. the hybrid protein, was found to elute mostly during the salt gradient, over a large number of fractions from approx. 0.3M to 0.5M NaCl, and coinciding with a major protein peak. This was, therefore, ineffective as a means of purification.

5.4.2 Effect of varying pH

After loading the sample (2mg protein, 14 000 units β-galactosidase activity) the column was washed through with 5 column volumes of the loading buffer, 50mM Tris-HCl/100mM NaCl/1% Triton, pH7.4. It was then eluted consecutively with 5 column volumes of the same buffer at pH8.5, pH9.0, and pH9.5, and finally with 100mM sodium borate, pH10.

The resulting profile (Figure 5.3) showed that a large proportion of the β-galactosidase acitivity, i.e. hybrid protein, was eluted from the column by the buffer at pH9.5, although a substantial amount remained bound to the resin until 100mM sodium borate, pH10 was passed through the column. Whereas the β-galactosidase activity peak from the elution with borate, pH10 coincided with a major protein peak, that from the elution with the Tris-HCl buffer at pH9.5 corresponded to a relatively low protein concentration. The specific activity of fractions 66 to 75 inclusive from this peak was 3.4-fold higher than that of the material applied to the column (23 600 compared with 7 000

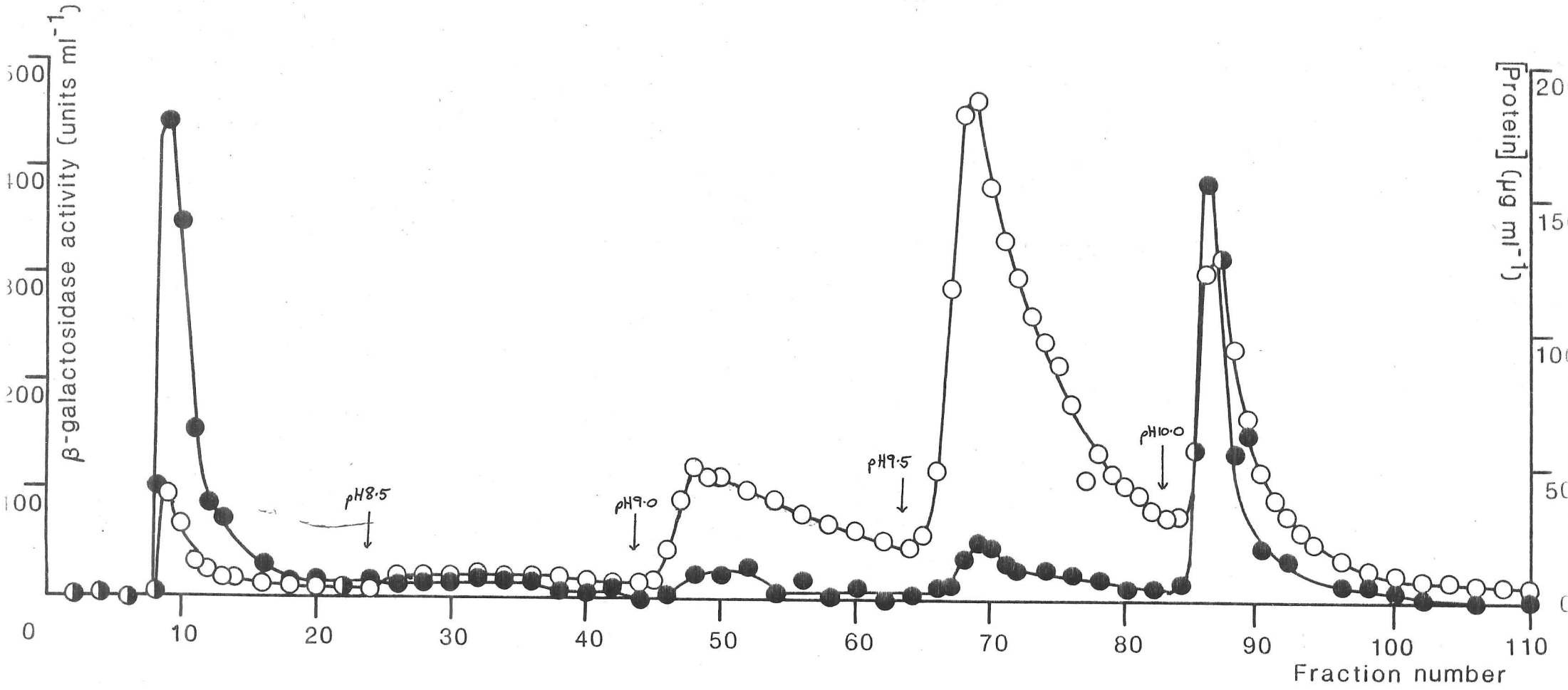


Figure 5.3 Elution profile from the p-aminophenyl-β-D-thiogalactoside affinity column

Chromatography was performed as described in Chapter 2 (section 2.11.1), with elution by 50mM Tris-HCl/100mM NaCl/1% Triton at pH's of 8.5, 9.0, 9.5 and by 100mM sodium borate/1% Triton, pH10.0 as indicated in the figure: β-galactosidase activity, (○); protein concentration, (●).

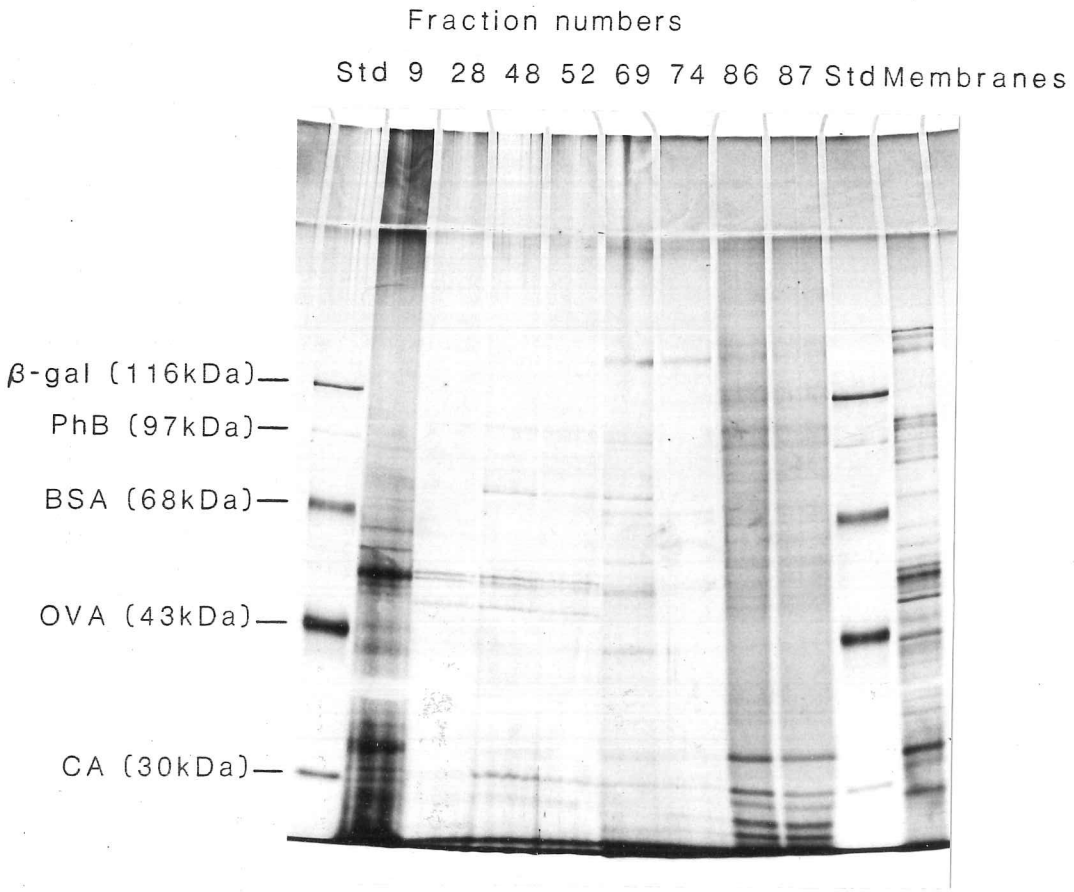


Figure 5.4 Analysis of fractions from the p-aminophenyl-β-D-thiogalactoside affinity column

The figure shows a 10% SDS-polyacrylamide gel of fractions eluted from the affinity column, silver stained by the method of Wray et al.(1981). The fractions were solubilized at 100°C for 2min; the fraction numbers are indicated for each track. The positions of the standard molecular mass markers are indicated (abbreviations as for Figure 5.1; PhB, phosphorylase B).

units mg^{-1}). Fractions from this peak were solubilized in SDS-dissolving buffer at 100°C for 2min and analysed on a 10% SDS-polyacrylamide gel, which was silver stained (Figure 5.4). Although several contaminating bands remained, the procedure yielded a considerable purification of the applied mixture.

5.5 DEAE Ion-exchange Column Chromatography

A published purification method for cytoplasmic β-galactosidase includes, as a major step, chromatography on a DEAE column (Miller, 1972). In this procedure the β-galactosidase is eluted with a salt gradient from 0.2M to 0.35M NaCl.

The salt concentrations at which the hybrid protein would bind and elute were checked in a test-tube experiment as described in Chapter 2 (section 2.11.2). Some elution occured at 0.2M NaCl so it was decided to apply a gradient from zero NaCl with the column.

The effect of processing a sample (2mg protein, 12 000 units β-galactosidase activity) of the hybrid protein on the DEAE column was investigated. After washing through with 3-4 column volumes of the loading buffer (10mM Tris-HCl/1% Triton, pH7.6) the column was eluted with a salt gradient from 0-0.3M NaCl as described in Chapter 2 (section 2.11.3).

The β-galactosidase activity, i.e. the hybrid protein, eluted in the middle of the gradient from approx. 0.15M to 0.2M NaCl, but the activity peak coincided with part of the major protein peak (Figure 5.5). The specific activity of fractions 64 to 73 inclusive from the activity peak was 1.7-fold higher than that of the applied material (10 370 compared with 6 170 units mg^{-1}).

Fractions were solubilized in SDS-dissolving buffer at 100°C for 2min and analysed on a 10% SDS-polyacrylamide gel, which was silver stained (Figure 5.6). This showed that many contaminating bands were still present. Nevertheless, this method using a small column would be suitable for concentration of dilute solutions of hybrid protein.

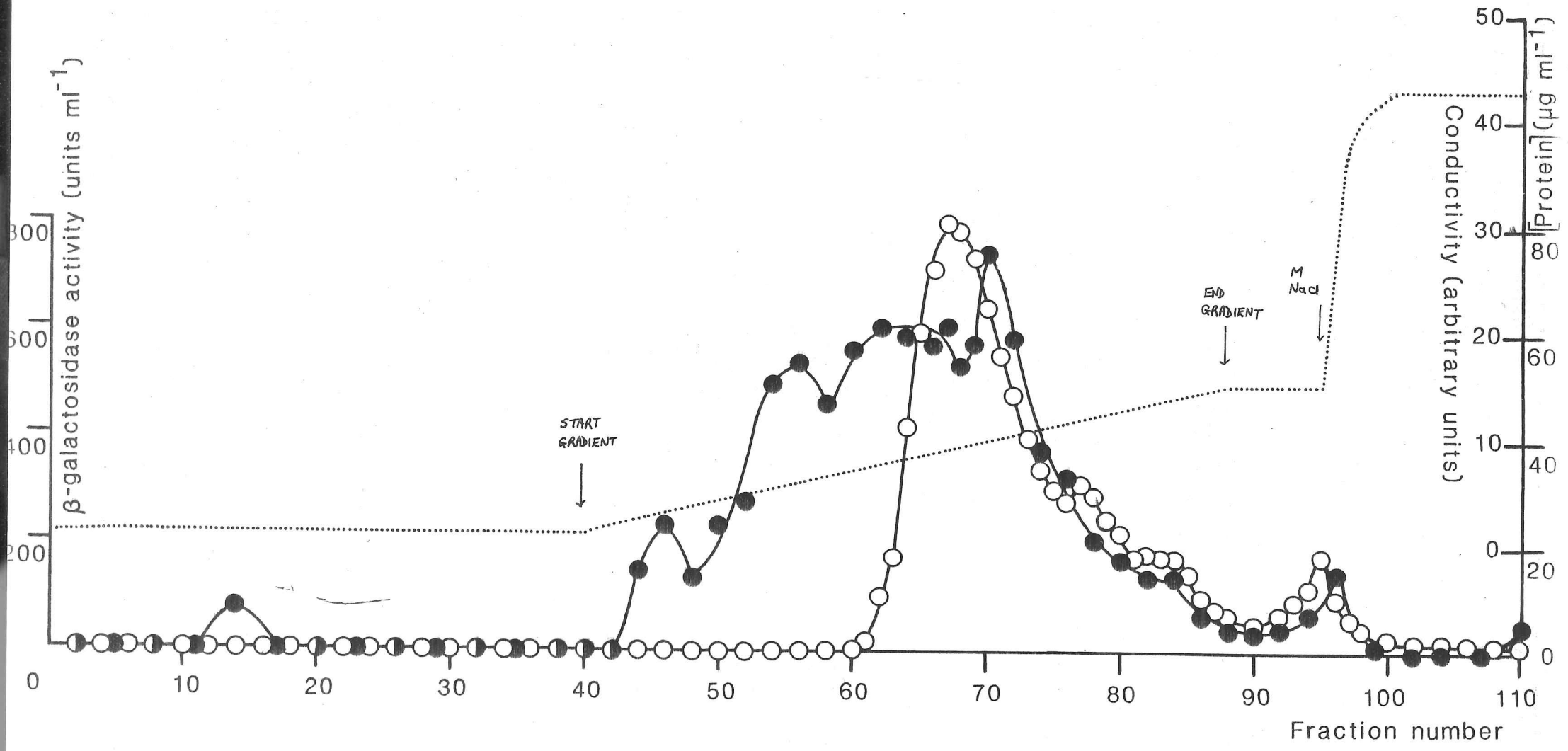


Figure 5.5 Elution profile from the DEAE column

Chromatography was performed as described in Chapter 2 (section 2.11.3), with elution by a salt gradient from 0-0.3M NaCl. The figure shows: β-galactosidase activity, (○); protein concentration, (●); conductivity, (·····).

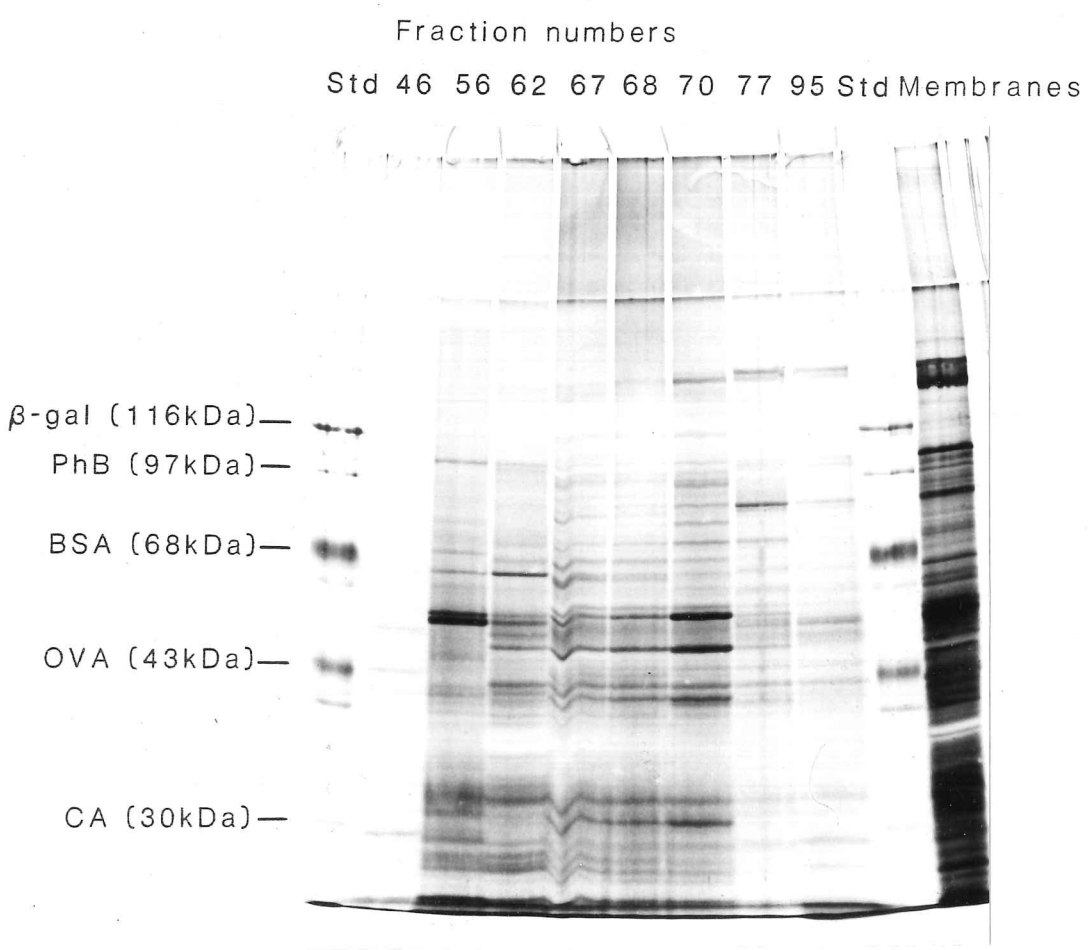


Figure 5.6 Analysis of fractions from the DEAE column

The figure shows a 10% SDS-polyacrylamide gel of fractions eluted from the DEAE column, silver stained by the method of Wray et al.(1981). The fractions were solubilized at 100°C for 2min; the fraction numbers are indicated for each track. The positions of the standard molecular mass markers are indicated (abbreviations as for Figure 5.4).

5.6 Bio-Gel A-5m Gel Filtration Column Chromatography

The purification of a MalF-LacZ hybrid protein has been reported by Shuman et al. (1980). This hybrid protein was purified without regard to its β-galactosidase activity by gel filtration under SDS-denaturing conditions, the fractions being assayed by electrophoresis of samples on gels.

The native β-galactosidase enzyme is a tetramer of the subunit (molecular mass 116 000Da) having an M_r of approx. 464 000Da. This molecular mass might be expected to be substantially larger than that of most membrane proteins, so it was hoped that some separation might be achieved using a non-denaturing gel filtration column, where the fractions could still be assayed for β-galactosidase activity.

A Bio-gel A-5m column was poured and calibrated with Blue Dextran, β-galactosidase, and cytochrome c as in Chapter 2 (section 2.11.4). A sample (approx. 20mg protein, approx. 100 000 units β-galactosidase activity) of the hybrid protein was applied to the column and eluted as in Chapter 2 (section 2.11.4). The hybrid protein eluted in fractions earlier than those corresponding to the position of native β-galactosidasein the calibration step, in accord with its larger molecular mass. In addition, the β-galactosidase activity peak eluted before the major protein peak, and a silver stained SDS-polyacrylamidegel of these fractions revealed relatively few contaminating bands.

Therefore, a larger sample (157mg protein, approx. 1 000 000 units β-galactosidase activity) of the impure hybrid protein in 20ml was applied to the column, giving the profile in Figure 5.7. The specific activity of fractions 51 to 58 of the eluted hybrid protein was 4.5-fold higher than that of the material applied to the column (70 300 compared with 15 600 units mg^{-1}).

Fractions were solubilized in SDS-dissolving buffer at 55°C for 8min and analysed on a 10% SDS-polyacrylamide gel. It was clear from this (Figure 5.8) that the hybrid protein eluted separately from most proteins and was contaminated with only a few other bands.

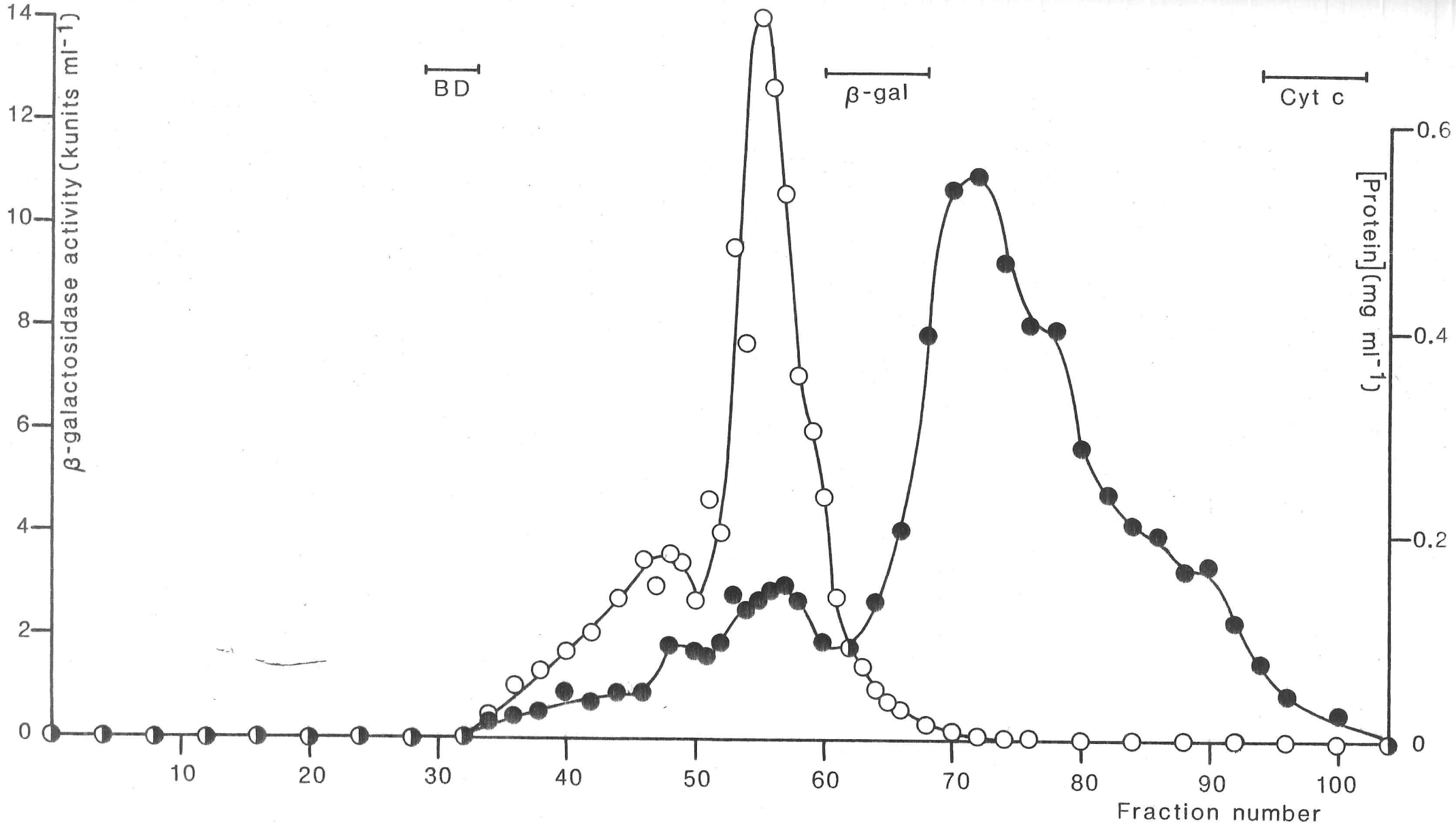


Figure 5.7 Elution profile from the Bio-gel column

Gel filtration was performed as described in Chapter 2 (section 2.11.4). The figure shows: β-galactosidase activity, (○); protein concentration, (●). The positions of the standards used to calibrate the column are indicated: BD, blue dextran; β-gal, β-galactosidase; Cyt c, cytochrome c.

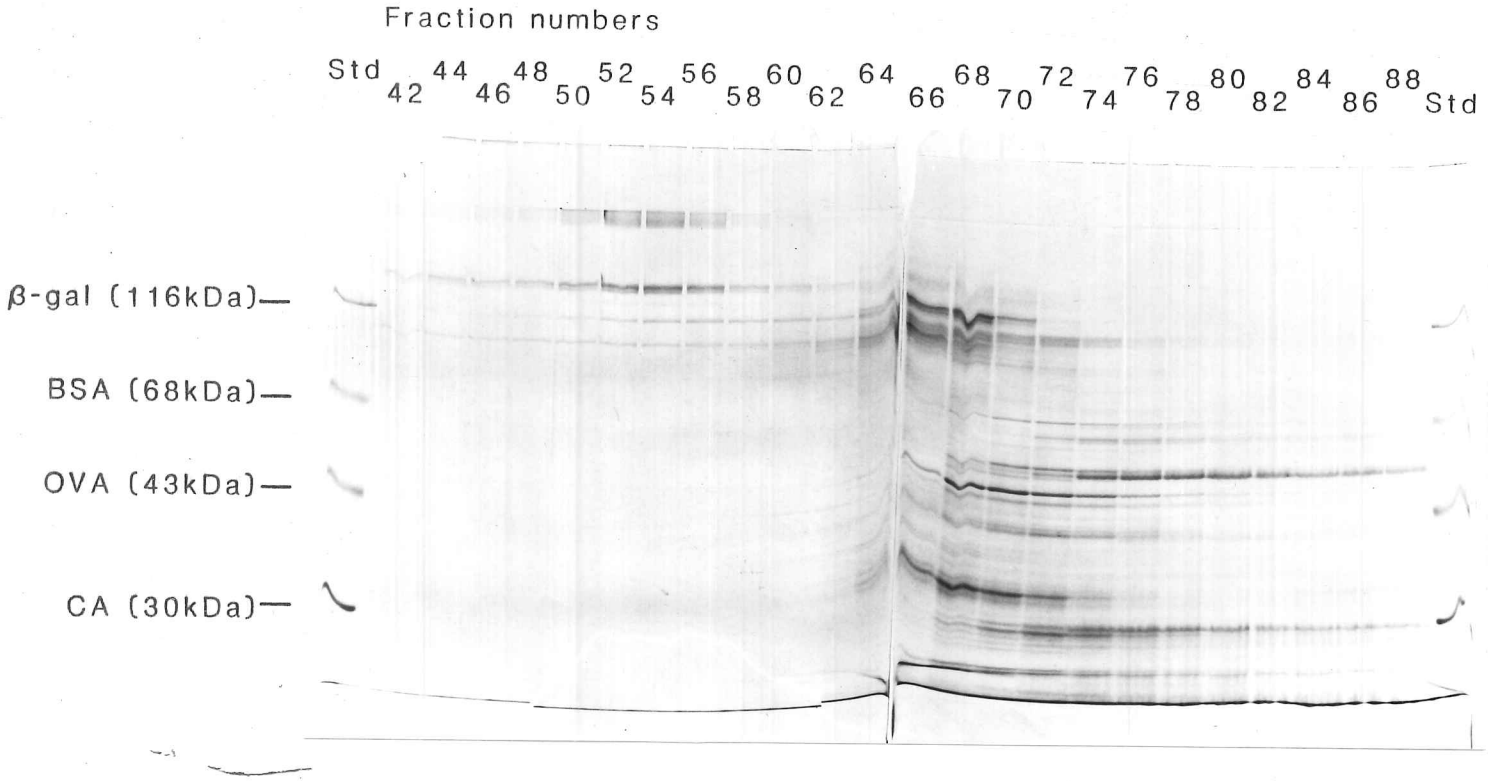


Figure 5.8 Analysis of fractions from the Bio-gel column

The figure shows 10% SDS-polyacrylamide gels of fractions eluted from the Bio-gel column, stained with Coomassie Blue. The fractions were solubilized at 55°C for 8min; the fraction numbers are indicated for each track. The positions of the standard molecular mass markers are indicated (abbreviations as for Figure 5.4).

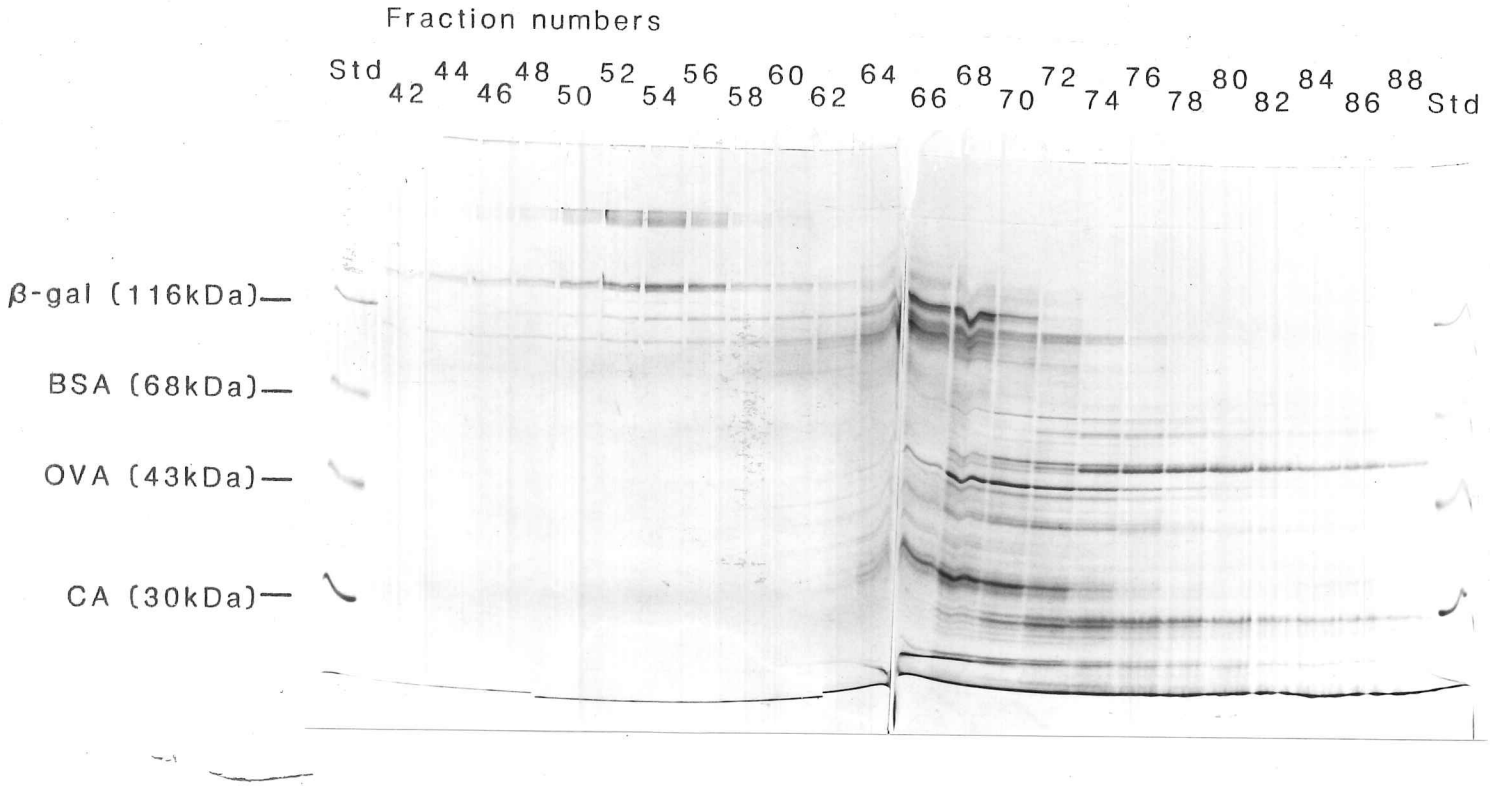


Figure 5.8 Analysis of fractions from the Bio-gel column

The figure shows 10% SDS-polyacrylamide gels of fractions eluted from the Bio-gel column, stained with Coomassie Blue. The fractions were solubilized at 55°C for 8min; the fraction numbers are indicated for each track. The positions of the standard molecular mass markers are indicated (abbreviations as for Figure 5.4).

Fractions 51 to 58 corresponding to the main part of the β-galactosidase activity peak were pooled and concentrated on a 2ml DEAE column as described in Chapter 2 (section 2.11.5).

5.7 Conditions for Solubilization of the Hybrid Protein in Sodium Dodecyl Sulphate

Samples of the pool of the β-galactosidase activity peak from chromatography of 157mg protein of the solubilizd membrane preparation of induced EJ54 on the Bio-gel column, and an analogous pool made from the equivalent fractions resulting from chromatography of 157mg protein of a solubilized membrane preparation of uninduced JM2336 on the same column, were used to compare conditions of solubilization in SDS-dissolving buffer for electrophoresis on SDS-polyacrylamide gels. By comparing induced and uninduced preparations confirmation of the identity of the band as that of the hybrid protein would be obtained.

Both an induced and an uninduced sample, with SDS-dissolving buffer added, were incubated at each of the following temperatures for the times indicated:

26°C for 35min
37°C for 20min
45°C for 15min
55°C for 8min
70°C for 5min
100°C for 2min

They were then analysed on a 5% SDS-polyacrylamide gel. The results (Figure 5.9) clearly show that optimal solubilization occurs in the middle of this temperature range, at 55°C. It is also apparent that the hybrid protein migrates at a different position on the gel when solubilized at 100°C compared to when solubilized at 37°C, and both bands occur at the intermediate temperatures.

This difficulty in solubilization of the hybrid protein is probably a result of the preferred conditions of solubilization for each part of the molecule, relating to the fact that the XylE and LacZ proteins are such different types of protein. β-galactosidase solubilizes optimally at 100°C,

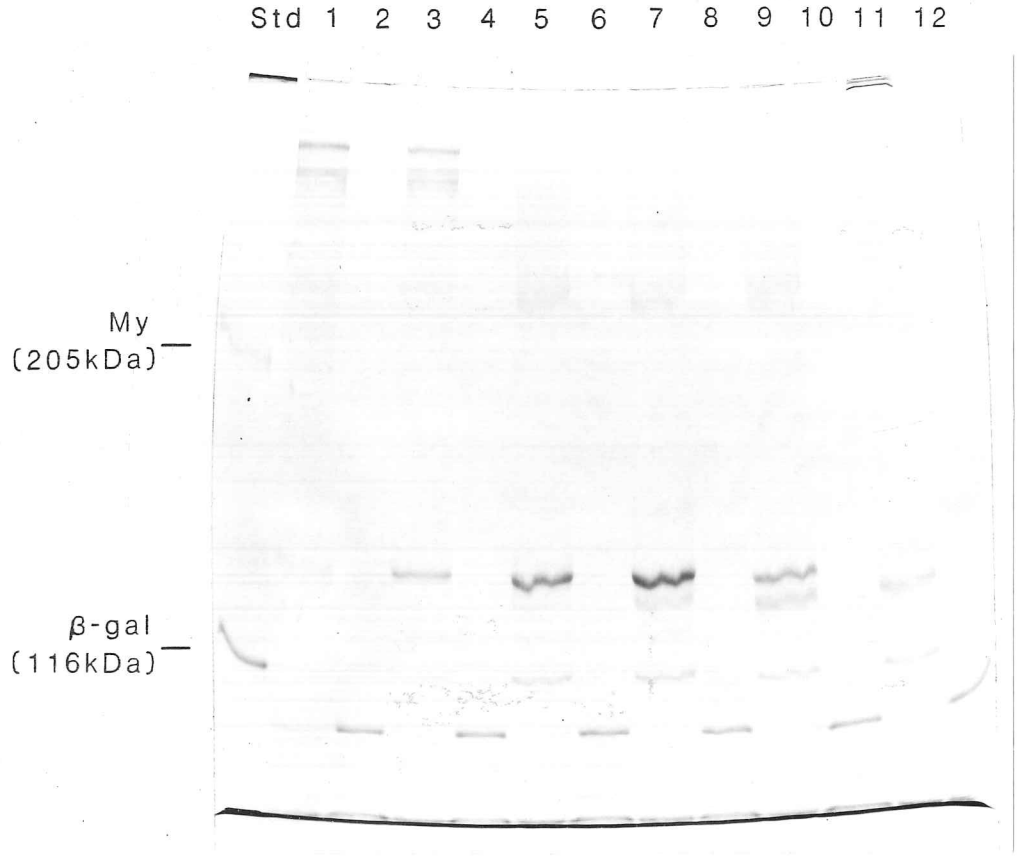


Figure 5.9 Effect of temperature on solubilization of the hybrid protein by SDS

Induced and uninduced samples of the equivalent fractions from the Bio-gel column were solubilized in SDS as indicated in the text (section 5.7). The figure shows a 5% SDS-polyacrylamide gel stained with Coomassie Blue: tracks 1, 3, 5, 7, 9, 11, induced; tracks 2, 4, 6, 8, 10, 12, uninduced; tracks 1 & 2, 26°C; tracks 3 & 4, 37°C; tracks 5 & 6, 45°C; tracks 7 & 8, 55°C; tracks 9 & 10, 70°C; tracks 11 & 12, 100°C. The positions of the standard molecular mass markers are indicated: My, myosin; β-gal, β-galactosidase.

whereas proton symport system membrane proteins are optimally solubilized at low temperatures (Teather et al., 1978; Macpherson et al., 1981). It may be that at high temperatures the XylE part of the molecule is still partly folded, whereas at low temperatures the β-galactosidase part is, resulting in the protein migrating in different positions on the gel, whilst at intermediate temperatures some of each form is found. Alternatively, the two positions on the gel may be the result of differential binding of SDS at the different temperatures, although if this is the case there must be two distinct states as opposed to a gradual alteration in the number of SDS molecules bound.

5.8 Further Purification After the Bio-gel Column Chromatography

In an attempt to improve the purification of the hybrid protein achieved by gel filtration, part of the pool from the Bio-gel column was processed by the following chromatographic methods.

5.8.1 Affinity column

A sample applied to the p-aminophenyl-β-D-thiogalactoside column was eluted consecutively with 5 column volumes of the loading buffer at pH9.0 and pH9.5, and finally with 100mM sodium borate, pH10. Fractions electrophoresed on a 10% SDS-polyacrylamide gel indicated that some improvement in the purification of the hybrid protein was achieved by this pass of the affinity column. However, the capacity of the resin was very small, and to process all the hybrid protein from the Bio-gel column would require several passes on a much larger column.

5.8.2 DEAE column

In an attempt to reduce the binding of contaminants to the DEAE column it was decided to run the column at a lower pH. A test-tube experiment (see Chapter 2, section 2.11.2) indicated that the hybrid protein still bound to the resin and retained its β-galactosidase activity at pH6.5, although a slight reduction of activity was observed at pH6.0. In a further experiment the binding of the hybrid protein to DEAE at pH6.5 was investigated over a range of salt concentrations, from 0 to 0.4M NaCl. The result clearly showed

that the hybrid protein eluted between 0.1M and 0.2M NaCl.

Therefore, a sample was applied to the DEAE column equilibrated with 10mM Tris-HCl/100mM NaCl/1% Triton, pH6.5, and eluted with a salt gradient from 0.1M to 0.2M NaCl in the same buffer over 5 column volumes. The pressure head was reduced to 25cmH_2O giving a flow rate of 11.5ml h^{-1}.

The β-galactosidase activity peak eluted over a large number of fractions from about 0.12M to 0.15M NaCl. Fractions analysed on a 10% SDS-polyacrylamide gel showed little improvement over the purification achieved by the Bio-gel column.

5.8.3 Repetition of the gel filtration on the Bio-gel column

The gel filtration step used to purify the MalF-LacZ hybrid protein by Shuman et al. (1980) was repeated loading a smaller sample volume to achieve greater resolution. Following their example, a concentrated pool of fractions from the first pass of the Bio-gel column (approx. 350 000 units β-galactosidase activity) was reapplied to the column but this time in a sample size of 1.5ml. The hybrid protein eluted largely in the same relative position although some activity eluted earlier. Fractions 53 to 57 of the activity peak from this second pass were again pooled and concentrated.

Fractions were solubilized in SDS-dissolving buffer at 55°C for 8min and electrophoresed on a 10% SDS-polyacrylamide gel (Figure 5.10). Some improvement in the purification was evident. A major contaminating band below the β-galactosidase standard was eliminated, although this was probably removed in the concentrating step on DEAE. A further band in the region of the ovalbumin standard was also removed. However, the second pass of the Bio-gel column did not appear to be necessary. The contaminating band at the same molecular mass as the β-galactosidase standard was thought to be the β-galactosidase moiety released from the hybrid protein due to protease action, and separation of the hybrid from this band was not achieved.

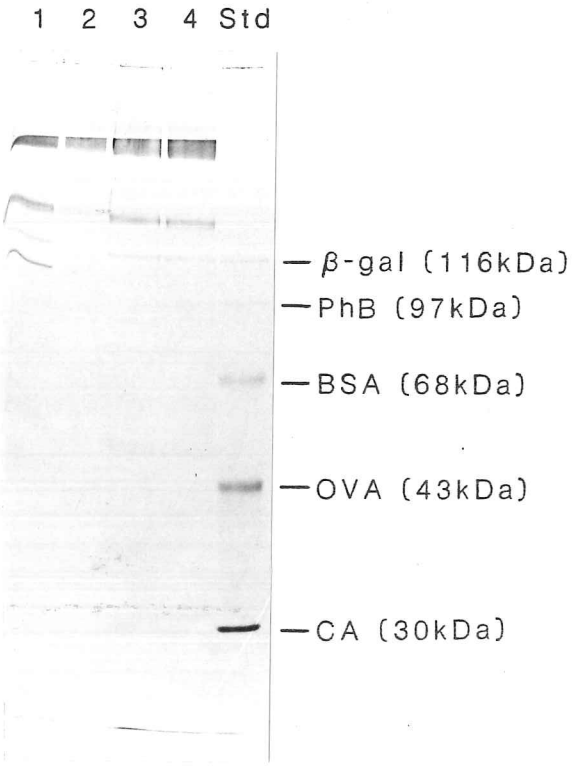


Figure 5.10 Analysis of fractions from first and second passes of the Bio-Gel column

The figure shows a 10% SDS-polyacrylamide gel of pools from the first and second passes of the Bio-gel column: track 1, pool of fractions 51-58 from first pass; track 2, pool of fractions 53-57 from second pass; track 3, concentrated pool from first pass; track 4, concentrated pool from second pass. The positions of the standard molecular mass markers are indicated (abbreviations as for Figure 5.4).

5.9 Overall Purification Scheme

The overall purification scheme resulted in a 32-fold increase in the specific activity of β-galactosidase over that of the lysed cells (90 400 compared with 2 800 units mg^{-1}, Table 5.3). However, the total β-galactosidase activity decreased during the purification procedure.

Table 5.3 Overall purification scheme of XylE-LacZ hybrid protein

Step	Volume (ml)	β-galactosidase activity (units ml^{-1})	(units mg^{-1})	Protein (mg ml^{-1})	Purification factor	Yield (%)
Crude lysate	23	146 900	2 800	52.3	-	100
Membranes	3.5	465 500	10 230	45.5	3.7	48
Solubilization	21	52 060	15 600	3.34	5.6	32
Bio-Gel column	40	8 830	67 900	0.13	24	10
Concentration	1.5	200 000	90 900	2.2	32	9
Bio-Gel column	25	4 080	72 900	0.06	26	3
Concentration	0.6	244 150	90 400	2.7	32	4

5.10 N-terminal Sequence Analysis

A sample (approx. 0.8mg protein) of the concentrated pool from the second pass of the Bio-gel column was prepared for N-terminal sequence analysis as described in Chapter 2 (section 2.12). The analysis was performed on a gas-phase sequenator by Dr. J.E. Walker of the MRC Laboratory of Molecular Biology. The following sequence of the first six residues was obtained:

Met-Asn-(Val/Thr)-(Gln/Thr)-Tyr-Asn-

Further sequence data were not obtained.

5.11 Discussion

The purification of the XylE-LacZ hybrid protein is not a straightforward repetition of the purification of β-galactosidase. This is primarily because the hybrid protein, being a membrane protein, has very different solubility

characteristics from native β-galactosidase. Detergent must be present throughout to keep the protein in solution. The XylE part of the molecule may interfere with its binding to, for example, the affinity column. In addition, the hybrid protein has to be purified from a different set of proteins to that from which native β-galactosidase is purified. Nevertheless, the β-galactosidase activity of the hybrid protein was a valuable means of assay. The hybrid protein was purified on the basis of its large size relative to the majority of membrane proteins by gel filtration. It would probably have been a better and more general purification scheme for the second gel filtration pass to be run under SDS-denaturing conditions, where the separation would be based on differences in monomer molecular mass. However, the hybrid could not then be assayed by its β-galactosidase activity.

Subsequent to these experiments, Ullmann (1984) described conditions for performing p-aminophenyl-β-D-thiogalactoside affinity column chromatography quite different from those originally used by Steers et al. (1971). The column was loaded and run in 1.6M NaCl which increased the capacity of the affinity column for β-galactosidase, by, it was suggested, preventing non-specific adsorption of foreign proteins. After washing with approx. 100 column volumes until no more material absorbing at 280nm appeared in the eluate, the protein was eluted with 100mM sodium borate, pH10. This yielded in one step a purification of 33- to 100-fold. However, it should be noted that these experiments were all done with cytoplasmic proteins. Also, the XylE-LacZ hybrid protein had been found to elute from the p-aminophenyl-β-D-thiogalactoside column at the much lower salt concentrations of 0.3-0.5M NaCl (section 5.4.1).

The N-terminal sequence was obtained from the material prepared, and although only the first six amino-acids were determined this was sufficient to confirm the reading frame and translation initiation point obtained from DNA sequencing of the cloned xylE gene (see Chapter 6).

A possible further use of the purified hybrid protein would be to raise antibodies which would be active against the XylE protein. This would provide an assay for the XylE protein without having to resort to reconstitution experiments, which could be useful for further work to purify the XylE protein itself.

CHAPTER 6

CLONING AND SEQUENCING OF THE XYLE GENE

6. CLONING AND SEQUENCING OF THE XYLE GENE

The mapping of xylE (Chapter 3) had indicated that it was close to malB (91.5min) on the E. coli chromosome. This proximity to malB was exploited in the cloning strategy. Induction of phage λ inserted in the malB region should yield phage carrying xylE. Comparison of the restriction map of the DNA from such a phage with that of a phage derived from a λplacMu insertion in xylE itself should indicate the position of xylE on the DNA carrying xylE. This would facilitate the identification of a suitable restriction fragment for subcloning and DNA sequencing.

6.1 Insertion of λplacMu3 into xylE

6.1.1 Selection of λplacMu insertion mutants in xylE

λplacMu was inserted into strain JM2087 as described in Chapter 2 (section 2.2.3). After allowing segregation as for the Mud(Ap^{R}lac) mutagenesis (Chapter 2, section 2.2.1), the cells were plated onto minimal medium supplemented with histidine (80μg ml^{-1}), glycerol (20mM) and xylose (5mM) and containing XGal (30μg ml^{-1}), and incubated at 30°C. The selection was for xylose resistance in an fda strain growing on glycerol, as for the Mud(Ap^{R}lac) insertion (Chapter 3, section 3.1), with a screening for β-galactosidase activity which hydrolyses XGal to release a blue dye.

This β-galactosidase activity might be xylose-inducible as a result of λplacMu being inserted in the correct orientation and reading frame in one of the xylose genes, or constitutive, due to λplacMu being inserted in some other expressed gene. Blue colonies from among many white colonies were streaked to single colonies, patched out, and replica-plated as for the Mud(Ap^{R}lac) insertions (Chapter 3, section 3.1). Of 10 independent selections with λplacMu3, two gave strains which had xylose-inducible β-galactosidase and failed to grow on histidine + xylose (i.e. were still fda): EJ61 and EJ62.

6.1.2 Identification of the λplacMu insertion mutation in xylE

As the position of xylE on the chromosome was known at this stage the simplest way to identify which, if any, of these insertions was in xylE was to map them by P1 transduction relative to the marker zjb::Tn10.

Strains EJ61 and EJ62 were transduced to tetracycline resistance (10μg ml^{-1}) with P1 grown on strain NB1 (zjb::Tn10). Colonies were patched out on the selection medium, replica-plated to:

histidine, xylose
histidine, glycerol, xylose
histidine, glycerol
nutrient, xylose
nutrient, tetracycline
nutrient

and incubated at 30°C.

For strain EJ61 all of 80 tetracycline resistant transductants retained xylose-inducible β-galactosidase activity, indicating that the λplacMu insertion in EJ61 was not in xylE.

For strain EJ62 nine out of 80 colonies had lost their β-galactosidase activity (plate ONPG assay on nutrient + xylose and on nutrient), and these nine failed to grow or grew only poorly on histidine, glycerol, xylose as compared with growth on histidine, glycerol (i.e. had become sensitive to xylose again). Thus the λplacMu insertion in EJ62 was in xylE.

6.1.3 Confirmation of the insertion of λplacMu into xylE

The λplacMu insertion into xylE in EJ62 was transduced into an fda^+ background, yielding EJ65.

When λplacMu is inserted into a gene resulting in β-galactosidase activity, a hybrid protein is formed consisting of the N-terminal region of the gene into which the phage is inserted and all but the first few amino-acids of β-galactosidase. In this case the insertion is in a membrane protein so it might be expected that, if the hybrid protein includes sufficient of the XylE protein, the β-galactosidase activity to become membrane bound.

Using a preparation as described in Chapter 2 (section 2.3.3) and quantitatively assaying for β-galactosidase, it was demonstrated that the β-galactosidase activity in EJ65 was both xylose-inducible and membrane associated. The distribution of the hybrid protein between the soluble and membrane fractions may be compared with that of native β-galactosidase (formed as a result of an operon fusion, as opposed to a gene fusion, in xylE) in EJ18 (Table 6.1).

The alkaline pH change indicative of proton symport observed on addition of xylose to EJ15 (prepared as in Chapter 2, section 2.6.3) was absent when xylose was added to EJ65 (Table 6.2 and Figure 6.1). However, the addition of IPTG, a substrate for the LacY symport system, to xylose-induced cells elicited an alkaline pH change in EJ65 but not in EJ15, indicating that the EJ65 cells were still capable of symport activity (Table 6.2 and Figure 6.1).

This provided conclusive evidence that the λplacMu insertion in EJ65, and thus in EJ62, was in xylE.

6.2 Insertion of λplacMu3 into malK-lamB

6.2.1 Selection of λplacMu insertion mutants in malK-lamB

λplacMu was inserted into strain JM2433 as described in Chapter 2 (section 2.2.3). The infection mixture (1ml) was diluted five-fold with nutrient broth containing 5mM maltose, split into aliquots and incubated overnight at 30°C to allow segregation.

Samples (0.1ml) of these cultures were superinfected with λvir at a multiplicity of infection of approx. 1, then plated out in soft agar lawns containing XGal and maltose and incubated at 37°C for 24 hours. The maltose was present to ensure that any unmutated lamB genes would be expressed well, and, therefore, that cells with intact lamB genes would be λ sensitive. The XGal would distinguish cells expressing β-galactosidase from those which did not.

From the λ resistant colonies resulting, several blue colonies (i.e. expressing either maltose-inducible or constitutive β-galactosidase) were

Table 6.1 Xylose inducibility and membrane association of β-galactosidase activity in EJ65

The cells were grown on glycerol in the presence or absence of 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.3.3). The values given are the means of duplicate measurements from each of two independent determinations.

Strain	Induced	β-galactosidase activity (units mg^{-1})	Cytoplasmic (%)	Membraneous (%)
EJ65	+	281.5	4	96
EJ65	-	1.6	-	-
EJ18	+	490.1	78	22
EJ18	-	1.6	-	-

Table 6.2 Xylose and IPTG promoted alkaline pH changes in EJ15 and EJ65

The cultures were grown on glycerol plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3).

Strain	Sugar	Extent ($nmolH^+ mg^{-1}$)	Number of measurements*
EJ15	Xylose	2.55	13(4)
EJ15	IPTG	0.01	4(2)
EJ65	Xylose	0.04	6(2)
EJ65	IPTG	1.51	4(2)

*: number of measurements (number of preparations)

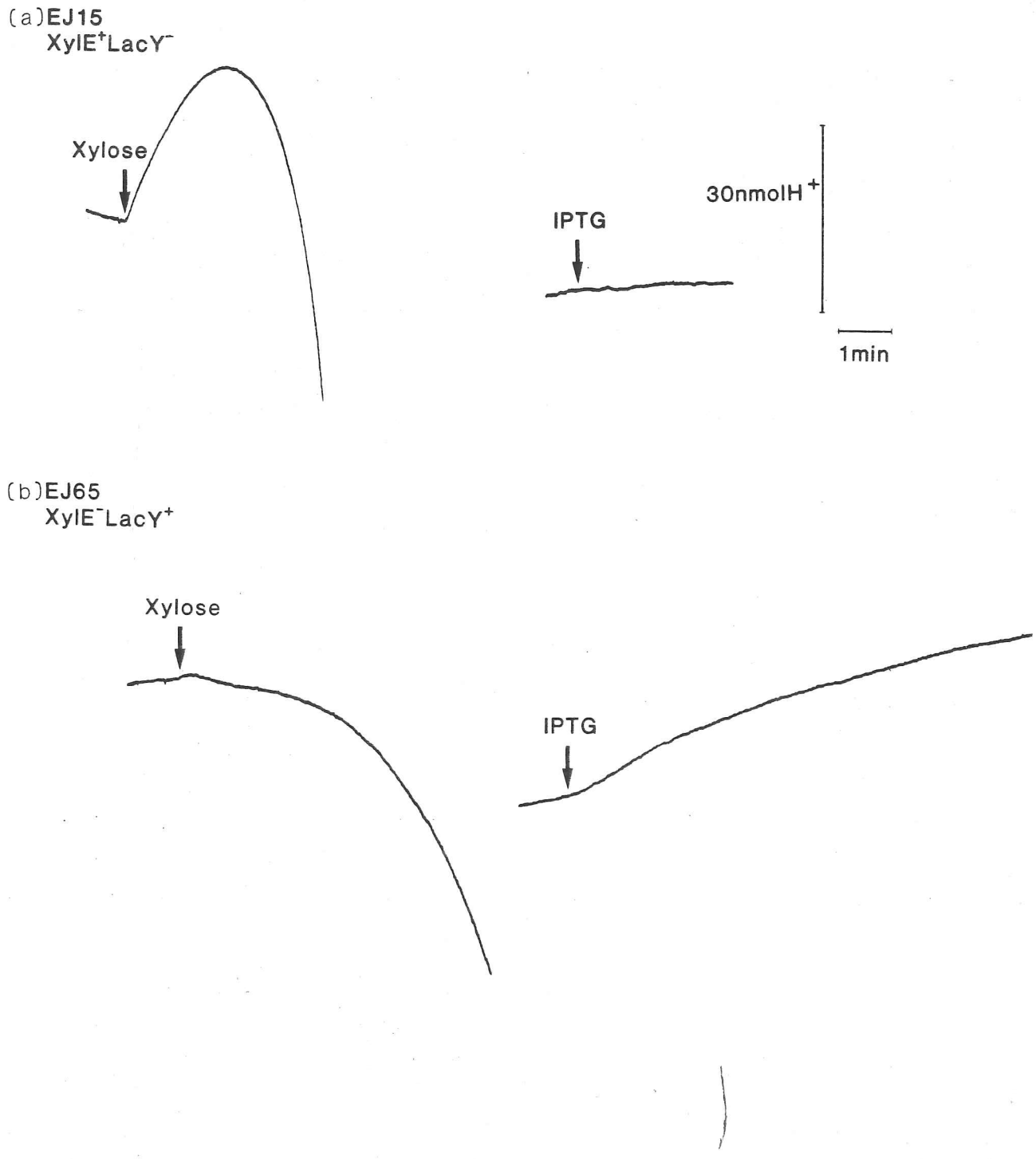


Figure 6.1 Xylose- and IPTG-promoted alkaline pH changes in strains EJ15 (a) and EJ65 (b)

The strains were grown on glycerol plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3). In each case the recording was calibrated by the addition of 3µl 0.01M NaOH immediately prior to the addition of substrate (20µl 0.5M).

streaked to single colonies. The ONPG plate assay (Chapter 2, section 2.3.1) on replica-plates with and without maltose did not indicate clearly whether the β-galactosidase was inducible or not. The strains taken were checked for λ resistance by spot tests.

6.2.2 Identification of the λplacMu insertion mutation in malK-lamB

To ascertain whether the λ resistance was due to λplacMu inserted in malK-lamB or to two events (spontaneous mutation to λ resistance with λplacMu inserted elsewhere in the chromosome), the strains were transduced to tetracycline resistance with P1 grown on EJ64 (zjb::Tn10) in order to map the λ resistance and the β-galactosidase activity.

In one case, 54 of 86 colonies became λ sensitive but all colonies retained β-galactosidase activity, becoming yellow in the ONPG plate assay relatively slowly. Thus the λ resistance and the β-galactosidase activity did not map in the same place and so the λ resistance could not be due to the λplacMu insertion.

In the case of strain EJ84, 46 of 85 colonies became λ sensitive and the same 46 colonies lost their β-galactosidase activity completely, whilst those colonies which retained their β-galactosidase activity went yellow in the ONPG plate assay very quickly. Thus in EJ84 the λ resistance and the β-galactosidase activity mapped together and so the λ resistance was due to the λplacMu insertion, i.e. λplacMu was inserted into malK-lamB. It could not be determined at this stage whether the insertion was in lamB itself or in malK and causing λ resistance by a polar effect.

6.3 Ultra-violet Induction of λplacMu Lysogens

The λplacMu lysogenic strains EJ65 (xylE::λplacMu) and EJ84 (malK-lamB::λplacMu) were u.v. induced and the phage lysates screened as in Chapter 2 (section 2.2.7).

6.3.1 Specialised transducing phage obtained

In the case of phage from EJ65 ten independent Spi$^-$ phage were selected, all of which possessed xylose-inducible β-galactosidase, and ten phage were picked from screening for xylose-inducible β-galactosidase, which on screening plaque purified phage were found to be Spi$^+$. These were screened to obtain an estimate of the size of the inserted host DNA by restriction digests of DNA prepared by the miniprep in Chapter 2 (section 2.13.1) with BamHI and with EcoRI. A Spi$^-$ phage with a large insert was used for the subsequent experiments, λΦ(*xylE'-lacZ*)7, in the hope that it would carry DNA extending into the *malB* region.

The phage from EJ84 were selected for being Spi$^-$ then screened for maltose-inducible β-galactosidase and the ability to complement *xylE* as described in Chapter 2 (section 2.2.7). Only one of 10 independent inductions yielded phage carrying *xylE*. This phage, λ(*xylE*)Φ(*malK'-lacZ*), was plaque purified twice before being used for DNA preparation and restriction mapping.

6.3.2 Confirmation of the presence of *xylE* on λ(*xylE*)Φ(*malK'-lacZ*)

The presence of *xylE* on the DNA carried by phage λ(*xylE*)Φ(*malK'-lacZ*) would be confirmed by the acquisition of the ability to transport xylose in a XylE$^-$ XylFG$^-$ strain lysogenic for λ(*xylE*)Φ(*malK'-lacZ*). Colonies from within a spot test for complementation of *xylE* on a lawn of EJ94 (*xylE*$^\Delta$ *xylG*::Mud(Ap^R*lac*)I) containing the helper phage λ*cI*^{857} were streaked out. This yielded colonies with maltose-inducible β-galactosidase (in addition to the xylose-inducible β-galactosidase due to *xylG*::Mud(Ap^R*lac*)I) and which had regained the ability to grow on xylose as sole carbon source. One such lysogenic strain, EJ97, was shown to have maltose-inducible β-galactosidase by quantitative assay (Table 6.3) and to have regained the ability to transport [^{14}C]-xylose to an extent comparable with the *xylE*$^+$ *xylG*::Mud(Ap^R*lac*)I strain EJ81 (Table 6.3). Xylose-promoted alkaline pH changes were not measured for the lysogen, but subsequent experiments with a fragment subcloned from λ(*xylE*)Φ(*malK'-lacZ*) into the plasmid pBR328 indicated the presence of symport activity by this criterion (section 6.6). Thus, the *xylE* gene must be carried on the phage λ(*xylE*)Φ(*malK'-lacZ*).

Table 6.3 Xylose transport and β-galactosidase activities in strains EJ94 and EJ97, a lysogen of λ(xylE)Φ(malK'-lacZ)

The cells were grown on succinate in the presence or absence of inducer and assayed as described in Chapter 2 (sections 2.6.1 and 2.3.2). The values given are the means of duplicate measurements from the number of independent determinations indicated in the parentheses.

Strain	Inducer	Xylose Transport		β-galactosidase
		Initial Rate	Extent at 2min	Activity
		(nmol min^{-1} mg^{-1})	(nmol mg^{-1})	(units mg^{-1})
EJ94	Xylose	0.4 (2)	0.1 (2)	173.8 (2)
EJ94	-	-	-	15.6 (2)
EJ94	Maltose	0.0 (1)	0.1 (1)	21.8 (2)
EJ97	Xylose	5.6 (2)	5.4 (2)	265.7 (3)
EJ97	-	0.3 (1)	0.2 (1)	18.9 (3)
EJ97	Maltose	0.6 (2)	0.7 (2)	95.6 (3)

6.4 Restriction Analysis of the λ Phage DNA

The Spi⁻ selection yields phage deleted for red and gam and which, therfore, have unidirectional deletions from this end of λ (see Chapter 1, section 1.3.4). Thus, for λplacMu insertions which are in the correct orientation for transcription this results in host DNA from the promoter side of the insertion being carried. Since the directions of transcription of the genes xylE and malK-lamB are divergent (Figure 6.2) Spi⁻ phage from insertions in these two genes should carry a region of overlapping host DNA.

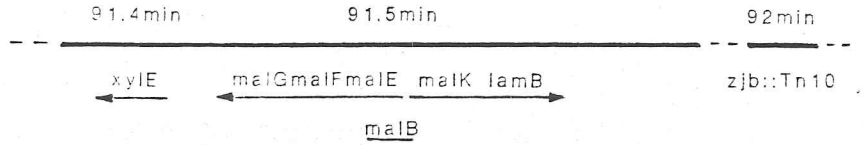


Figure 6.2 The xylE-malB region of the E. coli chromosome
The arrows indicate the directions of transcription of the genes.

DNA was prepared from the phages λΦ(xylE'-lacZ) and λ(xylE)Φ(malK'-lacZ) as in Chapter 2 (section 2.13.1) and used for restriction mapping. The λ(xylE)Φ(malK'-lacZ) DNA was analysed with the enzymes SacI, SmaI, SalI, BglII, BamHI, EcoRI, HindIII and PvuII. The single, double or triple digestions were separated on 0.7% or 1% agarose gels (e.g. Figure 6.3). Comparison of the restriction map deduced from the fragments obtained with published restriction data for the malB region (Marchal et al., 1978) indicated that the point of λplacMu insertion in EJ84 was malK.

The λΦ(xylE'-lacZ) DNA was similarly analysed and comparison of the restriction maps of the two phages (Figure 6.4) did indeed reveal a region of overlapping DNA between the two. This region is in opposite orientations relative to the phage λ DNA in the two phages. Alignment of the restriction sites in this region identified the fusion point of xylE'-lacZ within the DNA carring the intact xylE gene and so located xylE on this DNA.

6.5 Subcloning of the 8kb BamHI-SalI Fragment into pBR328

The 8kb BamHI-SalI fragment was cloned into the multicopy plasmid pBR328. λ(xylE)Φ(malK'-lacZ) DNA digested with BamHI and SalI was ligated into BamHI-SalI cut and phosphatased pBR328 at 15°C overnight as in Chapter 2 (section 2.14.2). The ligated plasmid was transformed into strain EJ96

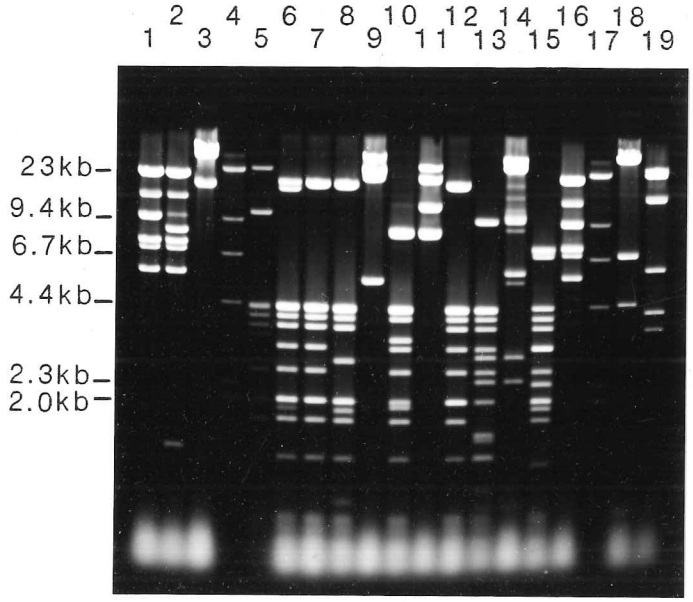


Figure 6.3 Analysis of restriction endonuclease digestions of the phage λ(xylE)Φ(malK'-lacZ)

Restriction digestions were carried out and the products were separated by electrophoresis on a 1% agarose gel as described in Chapter 2 (section (sections 2.14.1 and 2.5.2). The figure shows an example of the results obtained: track 1, BamHI; track 2, BamHI + SalI; track 3, SalI; track 4, λ HindIII standard; track 5, λplacMu PvuII standard; track 6, PvuII + SalI; track 7, PvuII; track 8, PvuII + SacI; track 9, SacI; track 10, PvuII + SmaI; track 11, SmaI; track 12, PvuII; track 13, PvuII + BglII; track 14, BglII; track 15, PvuII + BamHI; track 16, BamHI; track 17, λ HindIII standard; track 18, HindIII; track 19, EcoRI.

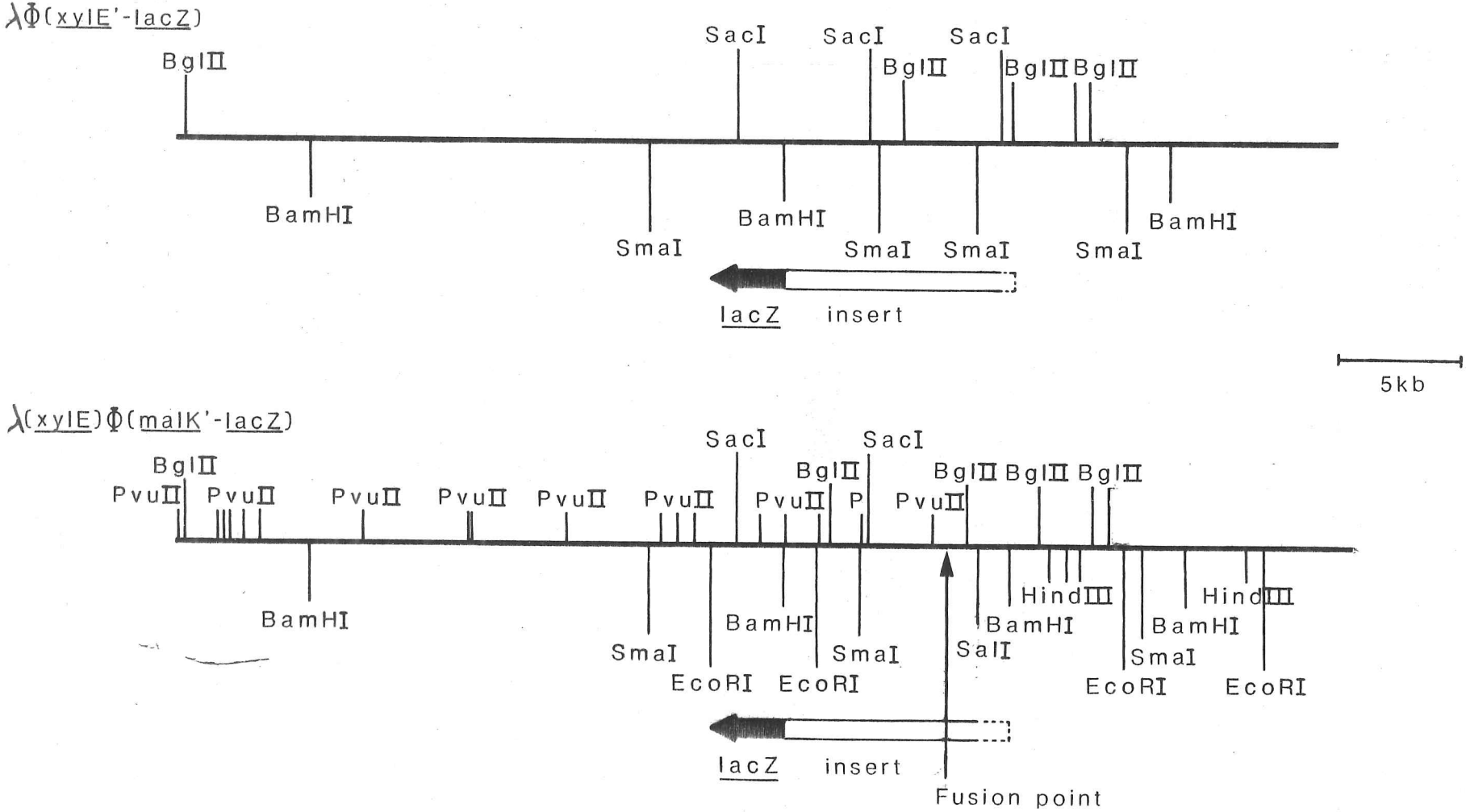


Figure 6.4 Restriction maps of phages λΦ(xylE'-lacZ)7 and λ(xylE)Φ(malK'-lacZ) and location of the xylE gene

The inserted chromosomal DNA is in the opposite orientation with respect to the phage DNA in the two phages. The arrow indicates the fusion point of xylE to lacZ in λΦ(xylE'-lacZ)7, i.e. a point within the xylE gene, on the DNA carried by λ(xylE)Φ(malK'-lacZ).

(*xylE*$^{\Delta}$ *xylG*::Mud(Ap^{R}*lac*)I *recA*) as described in Chapter 2 (section 2.15) selecting for chloramphenicol resistance (12.5μg ml^{-1}). The transformants were screened for tetracycline sensitivity, i.e. the presence of an insert in the tetracycline resistance gene.

The plasmid DNAs from 12 Cm^{R} Tc^{S} transformants were prepared by minipreps as in Chapter 2 (section 2.13.2) and digested with BamHI and SalI to identify those with the correct sized insert. Separation of the fragments on a 1% agarose gel indicated that 3 of these 12 transformants contained the desired 8kb fragment (Figure 6.5(a)). Further digestions with PvuII, HindIII or SacI, in addition to BamHI and SalI, on a plasmid containing the BamHI-SalI fragment and one containing the 7.2kb BamHI-BamHI fragment (cloned into incompletely digested vector DNA) from λ(*xylE*)Φ(*malK'-lacZ*) confirmed that these assignments were correct. The 8kb fragment was cut by SacI and by PvuII but not by HindIII, whereas the 7.2kb fragment was cut by HindIII but not by SacI or PvuII (Figure 6.5(b)), as expected from the restriction map of the λ(*xylE*)Φ(*malK'-lacZ*) phage DNA (Figure 6.4). This yielded the subclone pEJ1.

The presence of an intact *xylE* gene on the BamHI-SalI fragment carried by pEJ1 was shown by the restoration of xylose-inducible [^{14}C]-xylose uptake to EJ96 by the plasmid (Table 6.4) comparable to that observed for EJ81 (XylE^{+} XylFG^{-}).

Finer detail restriction mapping was possible using the subclone in pBR328. Plasmid pEJ1 DNA was digested with HincII which cuts more frequently than most 6-base recognition sequence enzymes owing to degeneracy in its recognition sequence (GTPyPuAC), in addition to the restriction enzymes known to have sites in the region of interest. This resulted in the restriction map of pEJ1 shown in Figure 6.6.

6.6 Subcloning of the 2.7kb HincII Fragment into pBR328

A 2.7kb HincII fragment was found to span the position of *xylE* with the fusion point near the centre of the fragment. This would be a suitable sized piece for DNA sequencing, provided the entire *xylE* gene was contained within it. To determine if this was the case the 2.7kb HincII fragment was purified as in Chapter 2 (section 2.16.1) and subcloned into PvuII cut and phosphatased

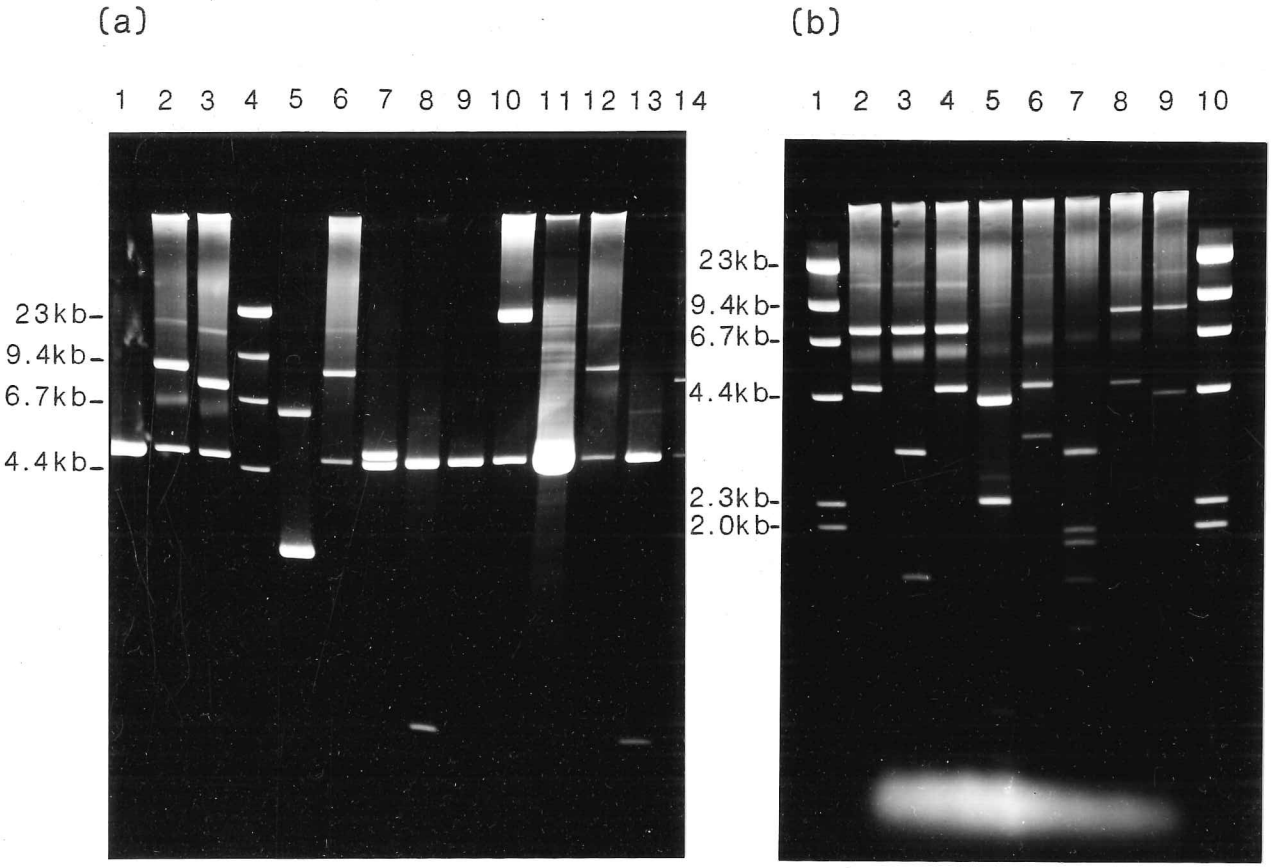


Figure 6.5 Subcloning of the BamHI-SalI fragment into pBR328

The figure shows 1% agarose gels of restriction digestions with BamHI and SalI, and where relevant the additional enzymes indicated, of plasmids derived from subcloning a BamHI-SalI digest of λ(xylE)Φ(malK'-lacZ) into pBR328.

(a) Plasmids obtained: track 4, λ HindIII standard; track 11, pBR328; tracks 2, 6 & 12, 8.0kb BamHI-SalI insert; tracks 3 & 14, 7.2kb BamHI-BamHI insert; track 7, 5.5kb BamHI-BamHI insert; tracks 8 & 13, 1.2kb BamHI-SalI insert; track 10, 19kb BamHI-BamHI insert; tracks 1 & 9, no insert; track 5, unidentified.

(b) Verification of assignment of 8.0kb and 7.2kb inserts: tracks 1 & 10, λ HindIII standard; tracks 2-5, 7.2kb BamHI-BamHI insert; tracks 6-9, 8.0kb BamHI-SalI insert; additional enzymes: tracks 2 & 6, SacI; tracks 3 & 7, PvuII; tracks 5 & 9, HindIII.

Table 6.4 Xylose transport activity in strain EJ96 with and without plasmid pEJ1

The cells were grown on succinate in the presence or absence of 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.1). The values given are the means of duplicate measurements from the number of independent determinations indicated in parentheses.

Strain/Plasmid	Induced	Xylose Transport	
		Initial Rate ($nmol\ min^{-1}\ mg^{-1}$)	Extent at 2min ($nmol\ mg^{-1}$)
EJ96	+	0.3 (4)	0.1 (4)
EJ96	-	0.1 (2)	0.1 (2)
EJ96/pEJ1	+	29.3 (2)	19.8 (2)
EJ96/pEJ1	-	1.8 (2)	2.4 (2)
EJ81	+	11.8 (5)	13.0 (5)
EJ81	-	0.3 (2)	0.2 (2)

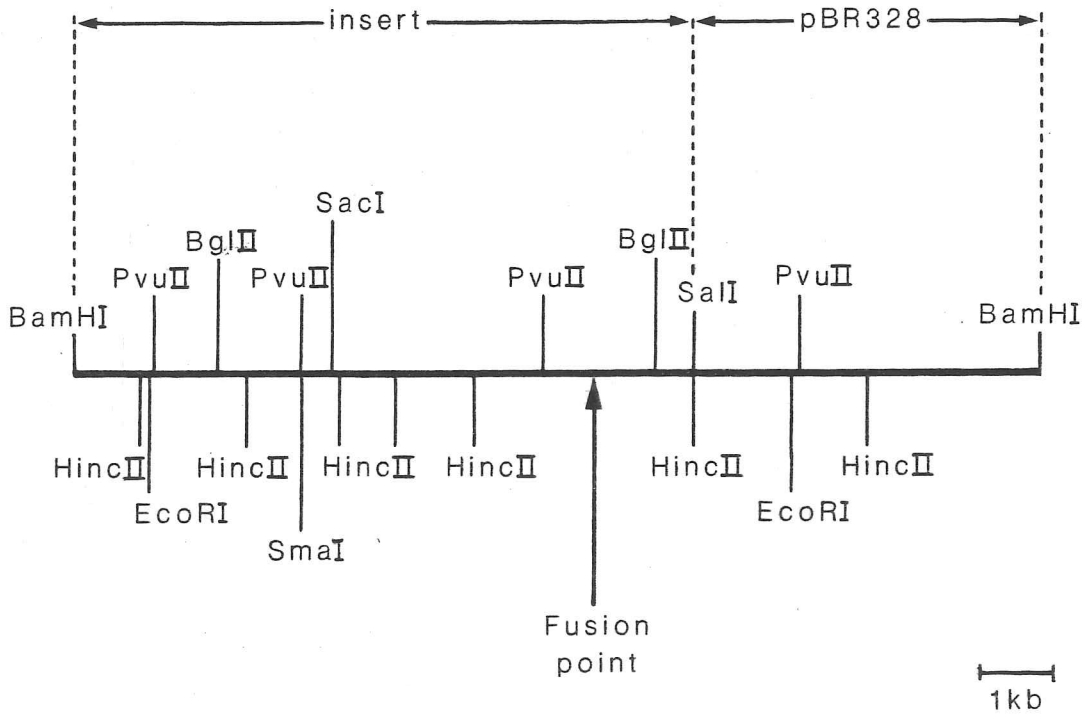


Figure 6.6 Restriction map of plasmid pEJ1

The figure shows plasmid pEJ1 linearized through the single BamHI site. The arrow indicates the fusion point of xylE to lacZ in λΦ(xylE'-lacZ)7, i.e. a point within the xylE gene. The total size of the plasmid is 12.8kb.

pBR328 (Figure 6.7(a)). The ligated plasmid was transformed into EJ96 selecting for tetracycline resistance. Transformants were screened for chloramphenicol sensitivity, i.e. the presence of an insert within the chloramphenicol resistance gene. Plasmid minipreps were performed on six Tc^R Cm^S transformants and the DNA was digested with BglII + PvuII to check for the correct insert, and with BglII + EcoRI to indicate the orientation of the insert (Figure 6.7).

A transformant harbouring a plasmid with the fragment in each orientation in relation to the direction of transcription of the chloramphenicol resistance gene into which it was inserted was assayed for [^{14}C]-xylose transport. Both plasmids, pEJ2 and pEJ3, restored xylose transport to strain EJ96 (Table 6.5). The plasmid with the insert in the opposite orientation to that of the cloramphenicol resistance gene, pEJ3, was used for subsequent experiments so that the transcription should be from the xylose promoter.

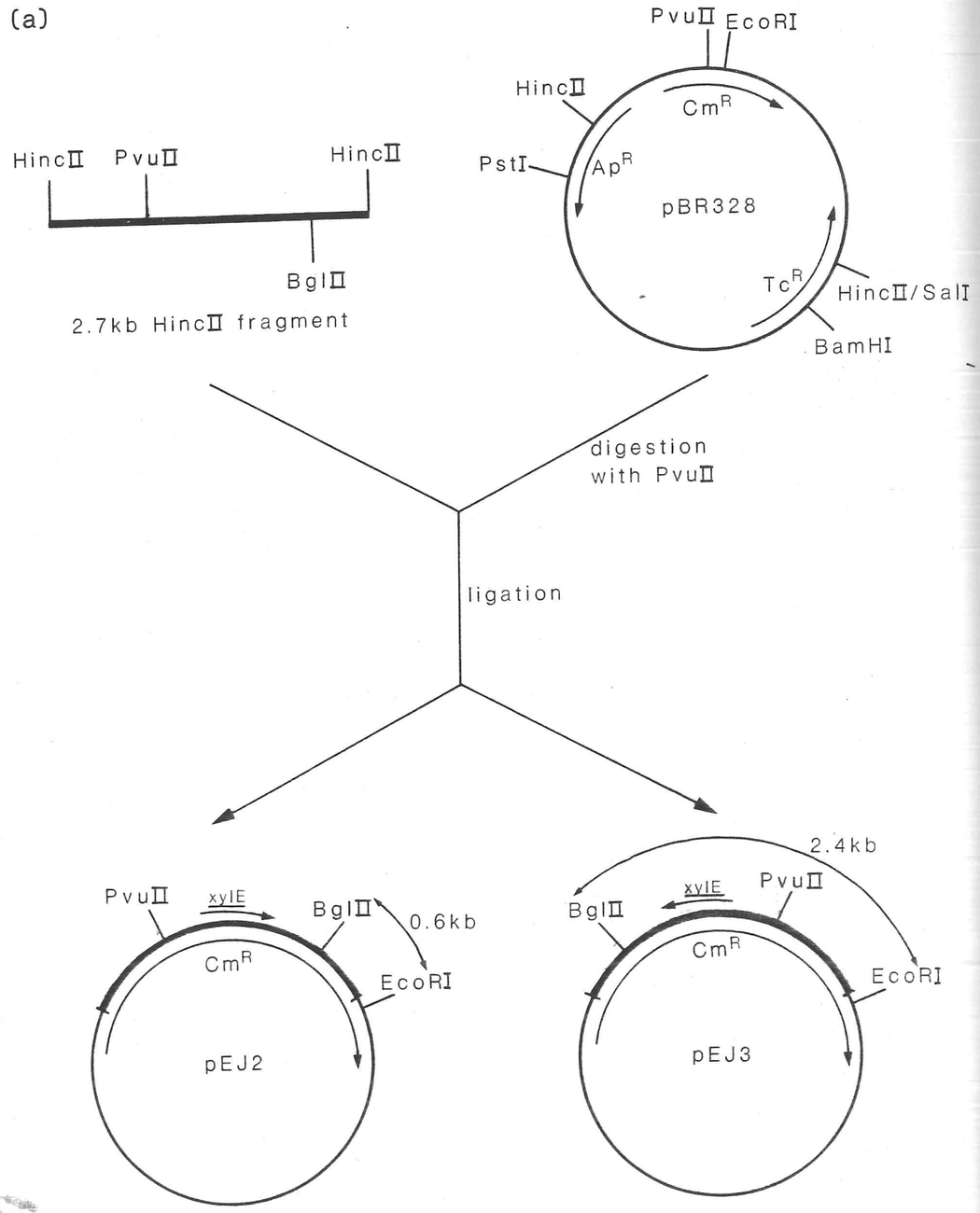
The addition of xylose to de-energised xylose-induced cells of strain EJ96 carrying pEJ3 elicited an alkaline pH change indicative of proton symport which did not occur for EJ96 itself (Figure 6.8).

Therefore, the intact xylE gene must be present on the 2.7kb HincII fragment.

6.7 DNA Sequencing of the 2.7kb HincII Fragment

The DNA sequence of the 2.7kb HincII fragment containing the xylE gene was determined by the Sanger dideoxy chain termination method (Sanger et al., 1977, 1980) using random clones. Purified HincII fragment was circularised and sonicated to produce random pieces; fragments in the size range 300-700bp were cloned into M13mp8 or M13mp10; DNA was prepared from these clones; and the sequencing reactions and gel electrophoresis were carried out all as described in Chapter 2 (sections 2.16.2 and 2.16.3). It was found that at higher voltages than those described for running the gels, the resolution of the sequencing 'ladder' was less satisfactory. An example of sequencing gels is shown in Figure 6.9.

(a)



(b)

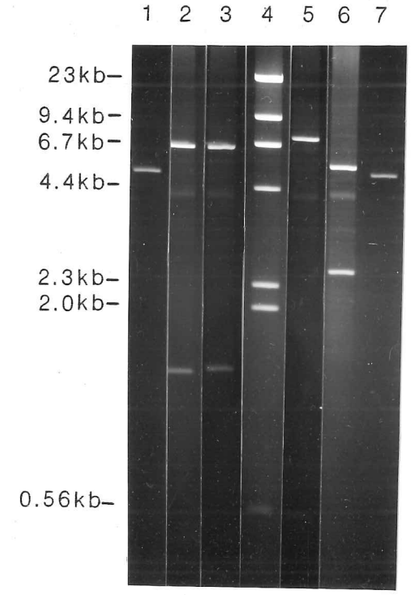


Figure 6.7 Subcloning of the HincII fragment into pBR328

(a) The HincII fragment was subcloned into the single PvuII site of pBR328 in the chloramphenicol resistance gene. The two possible orientations for the inserted DNA, the relation of the BglII site of the insert to the EcoRI site of the plasmid, and the directions of transcription of the genes are shown. (Not drawn to scale.)

(b) A 1% agarose gel of restriction digestions of subclones pEJ2 and pEJ3 to determine the orientation of the inserted DNA: track 4, λ HindIII standard; tracks 1 & 7, pBR328; tracks 2 & 5, pEJ2; tracks 3 & 6, pEJ3; tracks 1-3, BglII + PvuII; tracks 5-7, BglII + EcoRI.

Table 6.5 Xylose transport activities in strain EJ96 with and without plasmids pEJ2 and pEJ3

The cells were grown on succinate plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.1). The values given are the means of duplicate measurements from the number of independent determinations indicated in parentheses.

Strain/Plasmid	Xylose Transport	
	Initial Rate ($nmol\ min^{-1}\ mg^{-1}$)	Extent at 2min ($nmol\ mg^{-1}$)
EJ96	0.3 (4)	0.1 (4)
EJ96/pEJ2	6.5 (1)	7.4 (1)
EJ96/pEJ3	17.4 (2)	13.8 (2)
EJ81	11.8 (5)	13.0 (5)

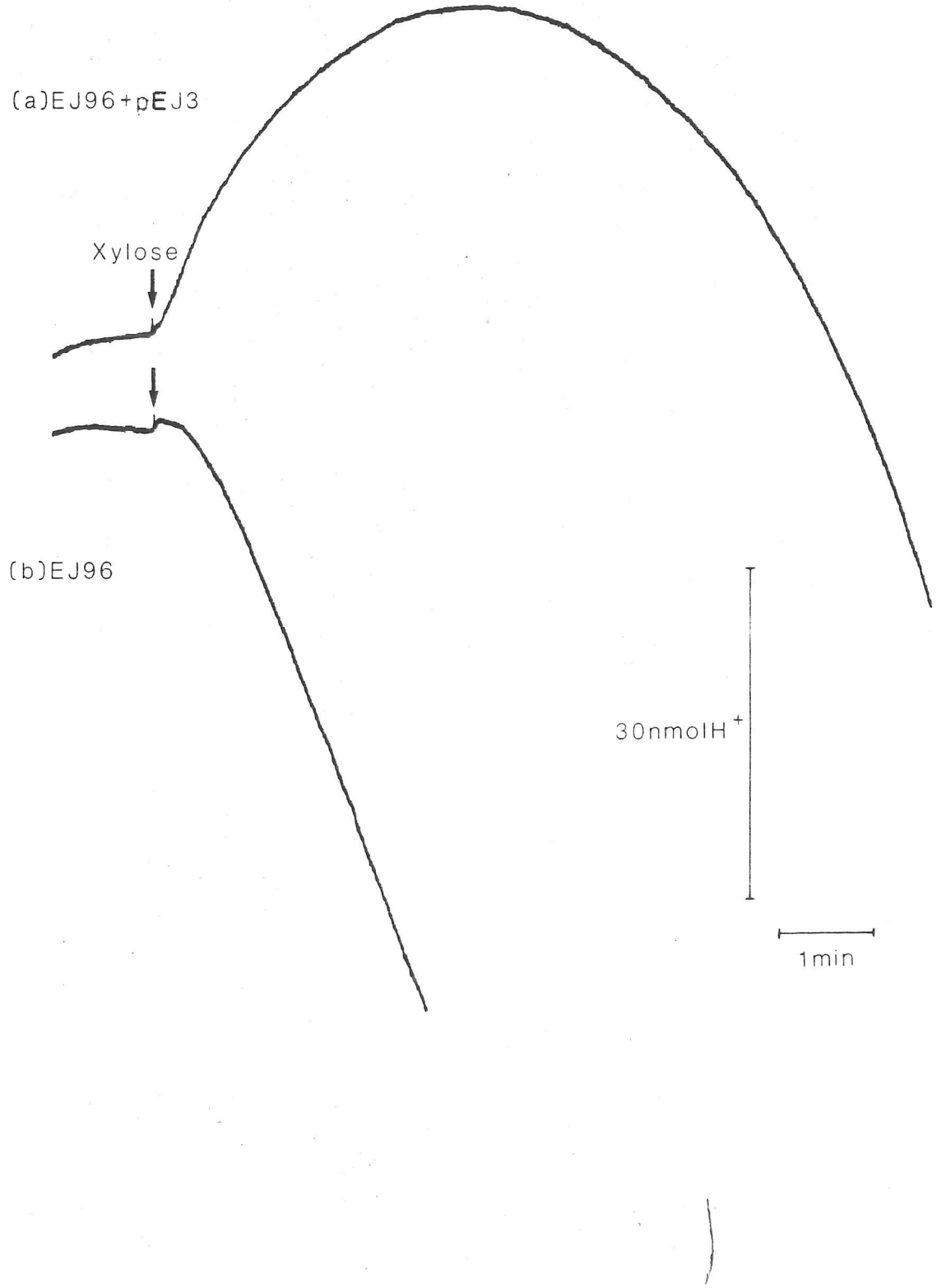


Figure 6.8 Xylose-promoted alkaline pH changes in strain EJ96 with (a) and without (b) plasmid pEJ3

The cells were grown on succinate plus 10mM xylose as inducer, and prepared and assayed as described in Chapter 2 (section 2.6.3). In each case the recording was calibrated by the addition of 3µl 0.01M NaOH immediately prior to the addition of xylose (20µl 0.5M).

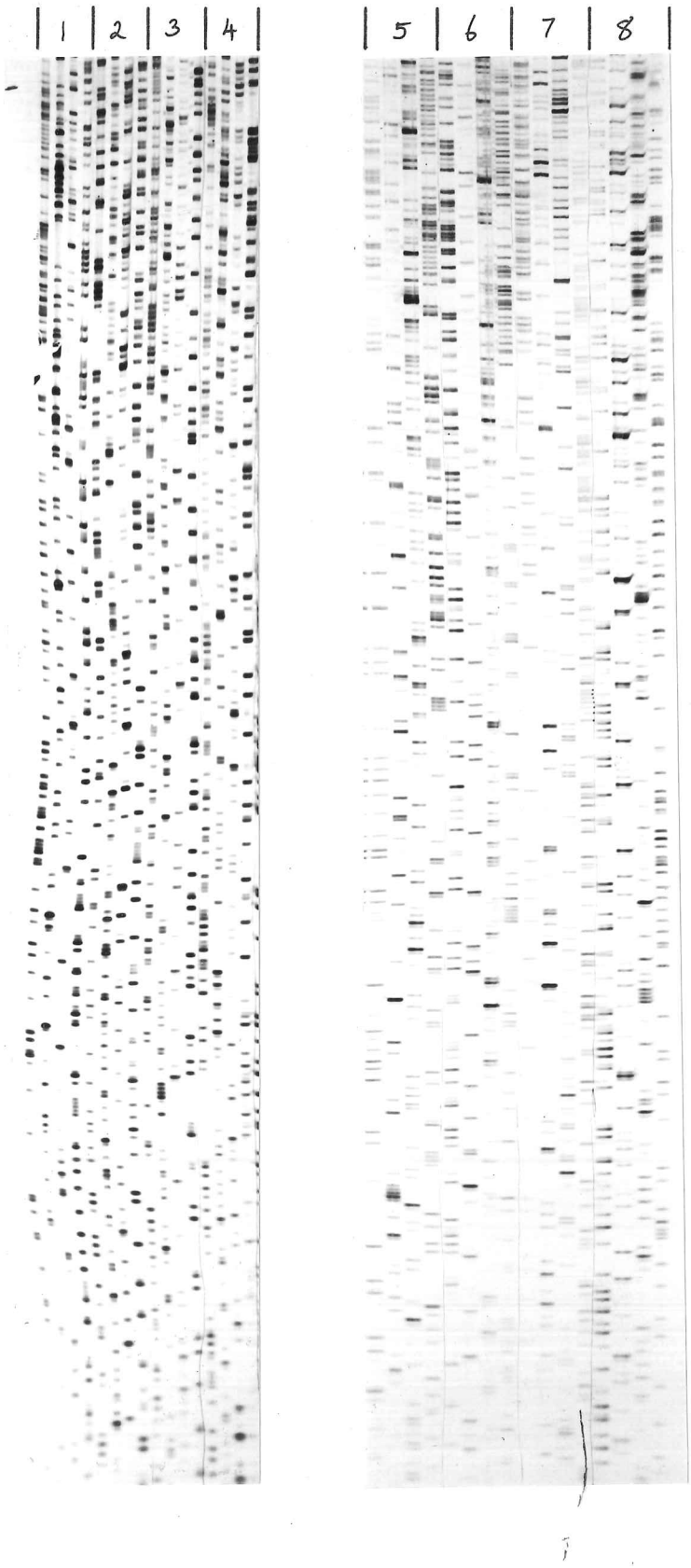


Figure 6.9 Examples of sequencing gels

Wedge gels (0.2-0.7mm) were run at 1500V (narrower tracks) or 2000V (wider tracks). For each clone the tracks from left to right read A, C, G, T. Numbers 1-8 represent different clones.

The sequence was determined on both strands an average of 10 times, giving a 'contig' of size 2 842 base pairs. The first 200 base pairs were only covered once on each strand but the sequence of the first 150 base pairs agreed exactly with that published by Dassa and Hofnung (1985) for malG whose sequence extended into this fragment by that amount. Analysis of the sequence revealed only one open reading frame of a suitable size for the xylE gene (Figure 6.10). In addition, the first six amino-acids of this reading frame (Met-Asn-Thr-Gln-Tyr-Asn-) agreed with the N-teminal sequence obtained from the XylE-LacZ hybrid protein (Chapter 5). Thus, this identifies the sequence of the xylE gene.

6.8 Discussion

The xylE gene is 1 473bp in length, corresponding to 491 amino-acids (Figure 6.11) and a protein molecular mass of 53 607Da. This is somewhat larger than the apparent molecular mass of the XylE protein on an SDS-polyacrylamide gel (36 000-41 000Da, Chapter 4) as has also been found for LacY (46 502Da compared with approx. 30 000Da, Buchel et al., 1980) and AraE (51 683Da compared with 36 000-38 000Da, M.C.J. Maiden, unpublished results). This is thought to be due to increased binding of SDS to these highly hydrophobic proteins (Beyreuther et al., 1980; see Chapter 4).

6.8.1 Promoter and Terminator Features

The start codon is preceded at a distance of 4 base pairs by the sequence TAAGGCAGGT, providing a ribosomal binding site; this can hybridise to the 3' end of the 16S ribosomal RNA, which has the sequence HO-AUUCCUCCA (Shine and Dalgarno, 1974). Also, there is a potential rho-independent transcriptional terminator sequence (a inverted repeat 10bp in length followed by a run of T residues) starting 10bp downstream of the translational stop codon (Figure 6.12). Two alternative possible Pribnow boxes (GATAAT and GACAAT) are marked on the DNA sequence. Although experiments such as S1 mapping and DNA footprinting are required to identify such regions with any degree of certainty, it is interesting to note that neither of these possibilities begin with a T although the remainder of the region is closely homologous with the consensus sequence (Pribnow, 1975; Rosenberg and Court, 1979). This is probably a reflection of the positive control of expression of the xylE

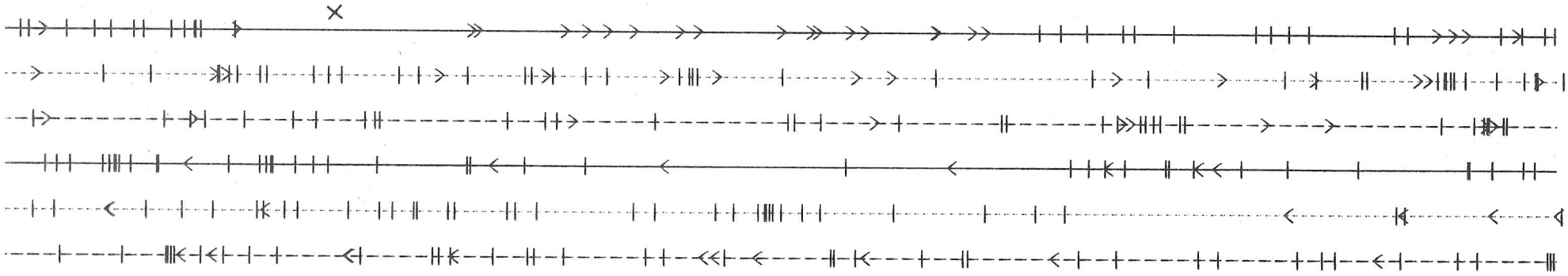


Figure 6.10 Identification of open reading frames in 2.8kb HincII fragment

The figure shows the six possible reading frames from the DNA sequence, with the positions of the initiation (>) and termination (|) codons marked. Only one open reading frame (marked x) is of sufficient length to encode the XylE protein.

```
                    5                  10                  15                  20
Met Asn Thr Gln Tyr Asn Ser Ser Tyr Ile Phe Ser Ile Thr Leu Val Ala Thr Leu Gly

                   25                  30                  35                  40
Gly Leu Leu Phe Gly Tyr Asp Thr Ala Val Ile Ser Gly Thr Val Glu Ser Leu Asn Thr

                   45                  50                  55                  60
Val Phe Val Ala Pro Gln Asn Leu Ser Glu Ser Ala Ala Asn Ser Leu Leu Gly Phe Cys

                   65                  70                  75                  80
Val Ala Ser Ala Leu Ile Gly Cys Ile Ile Gly Gly Ala Leu Gly Gly Tyr Cys Ser Asn

                   85                  90                  95                 100
Arg Phe Gly Arg Arg Asp Ser Leu Lys Ile Ala Ala Val Leu Phe Phe Ile Ser Gly Val

                  105                 110                 115                 120
Gly Ser Ala Trp Pro Glu Leu Gly Phe Thr Ser Ile Asn Pro Asp Asn Thr Val Pro Val

                  125                 130                 135                 140
Tyr Leu Ala Gly Tyr Val Pro Glu Phe Val Ile Tyr Arg Ile Ile Gly Gly Ile Gly Val

                  145                 150                 155                 160
Gly Leu Ala Ser Met Leu Ser Pro Met Tyr Ile Ala Glu Leu Ala Pro Ala His Ile Arg

                  165                 170                 175                 180
Gly Lys Leu Val Ser Phe Asn Gln Phe Ala Ile Ile Phe Gly Gln Leu Leu Val Tyr Cys

                  185                 190                 195                 200
Val Asn Tyr Phe Ile Ala Arg Ser Gly Asp Ala Ser Trp Leu Asn Thr Asp Gly Trp Arg

                  205                 210                 215                 220
Tyr Met Phe Ala Ser Glu Cys Ile Pro Ala Leu Leu Phe Leu Met Leu Leu Tyr Thr Val

                  225                 230                 235                 240
Pro Glu Ser Pro Arg Trp Leu Met Ser Arg Gly Lys Gln Glu Gln Ala Glu Gly Ile Leu

                  245                 250                 255                 260
Arg Lys Ile Met Gly Asn Thr Leu Ala Thr Gln Ala Val Gln Glu Ile Lys His Ser Leu

                  265                 270                 275                 280
Asp His Gly Arg Lys Thr Gly Gly Arg Leu Leu Met Phe Gly Val Gly Val Ile Val Ile

                  285                 290                 295                 300
Gly Val Met Leu Ser Ile Phe Gln Gln Phe Val Gly Ile Asn Val Val Leu Tyr Tyr Ala

                  305                 310                 315                 320
Pro Glu Val Phe Lys Thr Leu Gly Ala Ser Thr Asp Ile Ala Leu Leu Gln Thr Ile Ile

                  325                 330                 335                 340
Val Gly Val Ile Asn Leu Thr Phe Thr Val Leu Ala Ile Met Thr Val Asp Lys Phe Gly

                  345                 350                 355                 360
Arg Lys Pro Leu Gln Ile Ile Gly Ala Leu Gly Met Ala Ile Gly Met Phe Ser Leu Gly

                  365                 370                 375                 380
Thr Ala Phe Tyr Thr Gln Ala Pro Gly Ile Val Ala Leu Leu Ser Met Leu Phe Tyr Val

                  385                 390                 395                 400
Ala Ala Phe Ala Met Ser Trp Gly Pro Val Cys Trp Val Leu Leu Ser Glu Ile Phe Pro

                  405                 410                 415                 420
Asn Ala Ile Arg Gly Lys Ala Leu Ala Ile Ala Val Ala Ala Gln Trp Leu Ala Asn Tyr

                  425                 430                 435                 440
Phe Val Ser Trp Thr Phe Pro Met Met Asp Lys Asn Ser Trp Leu Val Ala His Phe His

                  445                 450                 455                 460
Asn Gly Phe Ser Tyr Trp Ile Tyr Gly Cys Met Gly Val Leu Ala Ala Leu Phe Met Trp

                  465                 470                 475                 480
Lys Phe Val Pro Glu Thr Lys Gly Lys Thr Leu Glu Glu Leu Glu Ala Leu Trp Glu Pro

                  485                 490
Glu Thr Lys Lys Thr Gln Gln Thr Ala Thr Leu ...
```

Figure 6.11 The amino acid sequence of the XylE protein

```
       260        270        280        290        300        310
TTGTTTTTAT CAATTTTGGA TAATTATCAC AATTAAGATC ACAGAAAAGA CATTACGTAA

       320        330        340        350        360        370
ACGCATTGTA AAAAATGATA ATTGCCTTAA CTGCCTGACA ATTCCAACAT CAATGCACTG

       380        390        400        410        420        430
ATAAAAGATC AGAATGGTCT AAGGCAGGTC TGAATGAATA CCCAGTATAA TTCCAGTTAT

       440        450        460        470        480        490
ATATTTTCGA TTACCTTAGT CGCTACATTA GGTGGTTTAT TATTTGGCTA CGACACCGCC

       500        510        520        530        540        550
GTTATTTCCG GTACTGTTGA GTCACTCAAT ACCGTCTTTG TTGCTCCACA AAACTTAAGT

       560        570        580        590        600        610
GAATCCGCTG CCAACTCCCT GTTAGGGTTT TGCGTGGCCA GCGCTCTGAT TGGTTGCATC

       620        630        640        650        660        670
ATCGGCGGTG CCCTCGGTGG TTATTGCAGT AACCGCTTCG GTCGTCGTGA TTCACTTAAG

       680        690        700        710        720        730
ATTGCTGCTG TCCTGTTTTT TATTTCTGGT GTAGGTTCTG CCTGGCCAGA ACTTGGTTTT

       740        750        760        770        780        790
ACCTCTATAA ACCCGGACAA CACTGTGCCT GTTTATCTGG CAGGTTATGT CCCGGAATTT

       800        810        820        830        840        850
GTTATTTATC GCATTATTGG CGGTATTGGC GTTGGTTTAG CCTCAATGCT CTCGCCAATG

       860        870        880        890        900        910
TATATTGCGG AACTGGCTCC AGCTCATATT CGCGGGAAAC TGGTCTCTTT TAACCAGTTT

       920        930        940        950        960        970
GCGATTATTT TCGGGCAACT TTTAGTTTAC TGCGTAAACT ATTTTATTGC CCGTTCCGGT

       980        990       1000       1010       1020       1030
GATGCCAGCT GGCTGAATAC TGACGGCTGG CGTTATATGT TTGCCTCGGA ATGTATCCCT

      1040       1050       1060       1070       1080       1090
GCACTGCTGT TCTTAATGCT GCTGTATACC GTGCCAGAAA GTCCTCGCTG GCTGATGTCG

      1100       1110       1120       1130       1140       1150
CGCGGCAAGC AAGAACAGGC GGAAGGTATC CTGCGCAAAA TTATGGGCAA CACGCTTGCA

      1160       1170       1180       1190       1200       1210
ACTCAGGCAG TACAGGAAAT TAAACACTCC CTGGATCATG GCCGCAAAAC CGGTGGTCGT

      1220       1230       1240       1250       1260       1270
CTGCTGATGT TTGGCGTGGG CGTGATTGTA ATCGGCGTAA TGCTCTCCAT CTTCCAGCAA
```

```
      1280       1290       1300       1310       1320       1330
TTTGTCGGCA TCAATGTGGT GCTGTACTAC GCGCCGGAAG TGTTCAAAAC GCTGGGGGCC

      1340       1350       1360       1370       1380       1390
AGCACGGATA TCGCGCTGTT GCAGACCATT ATTGTCGGAG TTATCAACCT CACCTTCACC

      1400       1410       1420       1430       1440       1450
GTTCTGGCAA TTATGACGGT GGATAAATTT GGTCGTAAGC CACTGCAAAT TATCGGCGCA

      1460       1470       1480       1490       1500       1510
CTCGGAATGG CAATCGGTAT GTTTAGCCTC GGTACCGCGT TTTACACTCA GGCACCGGGT

      1520       1530       1540       1550       1560       1570
ATTGTGGCGC TACTGTCGAT GCTGTTCTAT GTTGCCGCCT TTGCCATGTC CTGGGGTCCG

      1580       1590       1600       1610       1620       1630
GTATGCTGGG TACTGCTGTC GGAAATCTTC CCGAATGCTA TTCGTGGTAA AGCGCTGGCA

      1640       1650       1660       1670       1680       1690
ATCGCGGTGG CGGCCCAGTG GCTGGCGAAC TACTTCGTCT CCTGGACCTT CCCGATGATG

      1700       1710       1720       1730       1740       1750
GACAAAAACT CCTGGCTGGT GGCCCATTTC CACAACGGTT TCTCCTACTG GATTTACGGT

      1760       1770       1780       1790       1800       1810
TGTATGGGCG TTCTGGCAGC ACTGTTTATG TGGAAATTTG TCCCGGAAAC CAAAGGTAAA

      1820       1830       1840       1850       1860       1870
ACCCTTGAGG AGCTGGAAGC GCTCTGGGAA CCGGAAACGA AGAAAACACA ACAAACTGCT

      1880       1890       1900       1910       1920       1930
ACGCTGTAAT CTTCCTGTCC AGCACGCCGC GCCATTTCGG CGTGCTGACT TTTTACTCCC
```

Figure 6.12 The DNA sequence of the xylE gene and its flanking regions

The DNA sequence shown is numbered from the beginning of the HincII fragment (the sequence of the whole fragment is given in the appendix). A putative ribosomal binding site is boxed, the inverted repeat of a probable rho-independent transcriptional terminator is underlined, and two possible RNA polymerase binding sites are indicated by overbars. The beginning of the translational reading frame is indicated by the arrow. These features are discussed in more detail in the text (section 6.8.1).

gene, as although almost all genes not requiring an activator protein have Pribnow boxes beginning with a T, for almost all positively regulated genes the first position of the Pribnow box is occupied by other bases (Dairi et al., 1985). Also, poor sequence homology is found at the -35 region for positively regulated genes (Rosenberg and Court, 1979), where binding of the regulatory protein is required for DNA polymerase to bind.

6.8.2 Amino-acid composition and codon usage

The amino-acid composition (Table 6.6) reveals the large number of non-polar amino-acids present in the XylE protein. Using the same definitions of polar/non-polar amino-acids as Yazyu et al. (1984) XylE is 68% non-polar, comparable to the 71% non-polar found for LacY and 70% for MelB. This is a somewhat higher value than the occurrence of non-polar amino-acids among proteins generally, which average 53% (values taken from Klapper, 1977), reflecting the fact that these are intrinsic membrane proteins. Of the 491 amino-acids in XylE, 27 are negatively charged, and 35 are positively charged, giving a net positive charge of eight at neutral pH. Thus, the XylE protein is a basic protein as are LacY (Buchel et al., 1980) and MelB (Yazyu et al., 1984).

The codon usage (Table 6.7) shows no preferences such as those seen for strongly expressed genes (e.g. ribosomal proteins, RNA polymerase, omp products) and in this is similar to lacY and other weakly expressed genes (e.g. trp genes). For the strongly expressed genes the codons used optimise the codon-anticodon interaction energy and codons corresponding to minor tRNAs are avoided giving an optimal efficiency of translation (Grosjean and Fiers, 1982). The frequency of use of optimal codons (calculated according to Ikemura, 1981) in the xylE gene is 0.65 which is similar to the values of 0.62 for lacY and 0.57 for melB, and may be compared with values of 0.92 for ompA and 0.98 for lpp (lipoprotein) (Ikemura, 1981).

6.8.3 Hydropathic profiles

A 'hydropathic profile' may be obtained by assigning to each amino-acid a value representing its degree of non-polarity. The average of these values over a span of n amino-acids is plotted for this span, then the window is

Table 6.6 Amino acid composition* of XylE compared with LacY and MelB

Amino acid	XylE	LacY	MelB
Polar			
Arg	14	12	17
His	5	4	7
Lys	16	12	13
Asp	9	6	17
Glu	18	11	9
Asn	18	16	14
Gln	16	11	11
Ser	32	30	26
Thr	28	18	27
Nonpolar			
Ala	48	35	38
Cys	7	8	4
Gly	47	36	36
Ile	39	32	40
Leu	56	54	58
Met	18	14	16
Phe	32	56	34
Pro	18	12	17
Trp	12	6	8
Tyr	19	15	26
Val	39	29	51
% nonpolar	68	71	70

* Residues per molecule

Table 6.7 Codon usage in the xylE gene compared with the lacY gene

	xylE	lacY		xylE	lacY		xylE	lacY		xylE	lacY
TTT	20	32	TCT	4	7	TAT	11	7	TGT	2	5
TTC	12	24	TCC	11	3	TAC	8	8	TGC	5	3
TTA	9	9	TCA	3	5	TAA	-	-	TGA	-	-
TTG	1	1	TCG	6	8	TAG	-	-	TGG	12	6
CTT	5	5	CCT	3	0	CAT	3	4	CGT	7	5
CTC	8	4	CCC	0	2	CAC	2	0	CGC	7	4
CTA	1	3	CCA	6	4	CAA	7	4	CGA	0	0
CTG	32	32	CCG	9	6	CAG	9	7	CGG	0	2
ATT	24	23	ACT	6	2	AAT	6	8	AGT	4	2
ATC	13	9	ACC	14	7	AAC	12	8	AGC	4	6
ATA	2	0	ACA	2	2	AAA	12	10	AGA	0	1
ATG	18	14	ACG	6	7	AAG	4	2	AGG	0	0
GTT	11	9	GCT	10	6	GAT	5	3	GGT	27	13
GTC	9	4	GCC	15	12	GAC	4	3	GGC	14	19
GTA	7	8	GCA	11	5	GAA	15	9	GGA	2	2
GTG	12	8	GCG	12	12	GAG	3	2	GGG	4	2

moved along the sequence by one amino-acid and the process is repeated. Continuing in this way, a hydropathic plot for the whole protein is produced (Kyte and Doolittle, 1982). Averaging the values over such a window reduces the extreme fluctuations from one amino-acid to the next, and thus allows regions of overall high hydrophobicity to be detected.

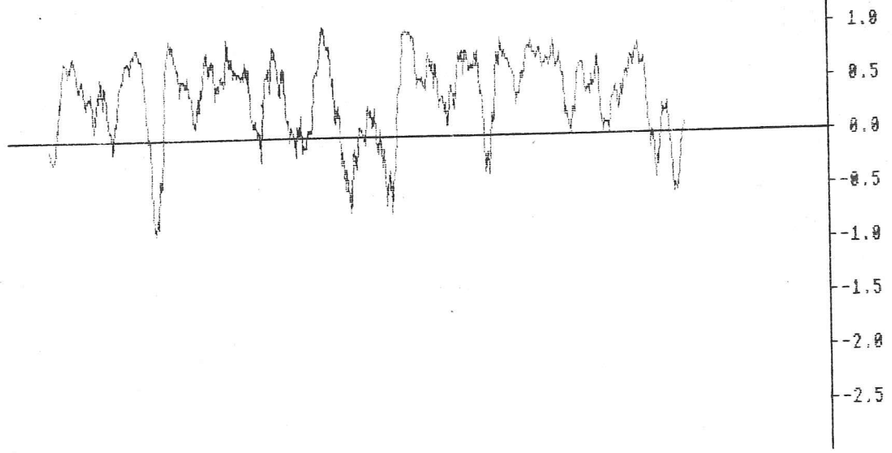
The hydropathic plots were calculated over spans of 9 and 21 residues by a computer program written by Dr. P.R. Alefounder. In an analogous way to the prediction that LacY contains 12 membrane spanning segments from the hydropathic profile, it could be suggested that XylE contains 12-13 membrane spanning segments from the hydropathic profile calculated for a span of 9 residues (Figure 6.13(a)). However, this can only be speculation until some structural information is obtained for the protein itself. The profiles (Figure 6.13(b)&(c)) for XylE and AraE (M.C.J. Maiden, unpublished results) calculated over a span of 21 residues are very similar. In both cases the protein is highly hydrophobic with a similar pattern of peaks, the most noticeable feature being a hydrophilic 'pocket' in the central region of the protein. It could be that this region represents an intra-membrane hydrophilic site. The most significant difference between the two pofiles is that XylE has an extra small hydrophobic peak within this hydrophilic 'pocket'.

When compared with LacY or MelB (Figure 6.13(d)&(e)) it can be seen that there is much less similarity in the detail of the hydropathic profile although the same general trend of hydrophobic peaks is apparent, as was found for LacY by Foster et al.(1983).

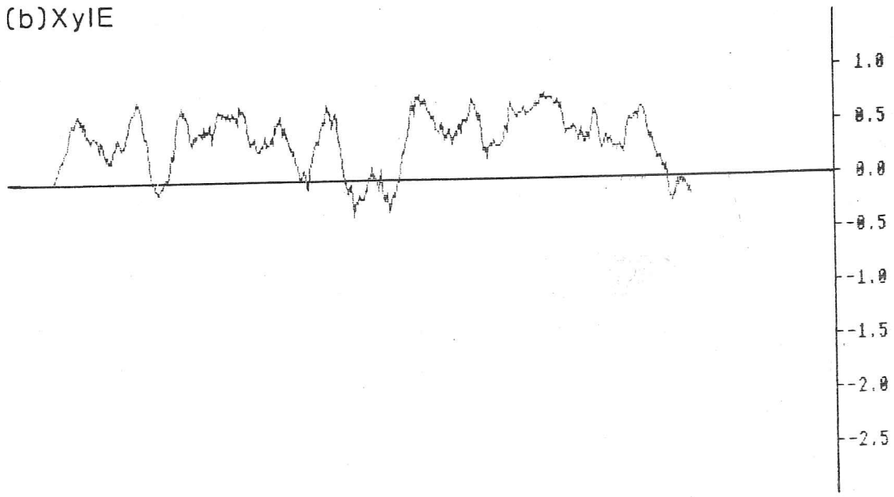
6.8.4 Diagon Analyses

A particularly effective way to look for similarities between two protein sequences is by use of the computer program DIAGON (Staden, 1982). This program produces a diagram showing all the matches between two sequences, where the x-axis represents one sequence and the y-axis the other. The program looks for sections where a proportion of the characters in the two sequences are similar by calculating a score for each pair of sections based on the degree of similarity between the pairs of amino-acids. This score is assigned to the central pair of amino-acids of the two sections being

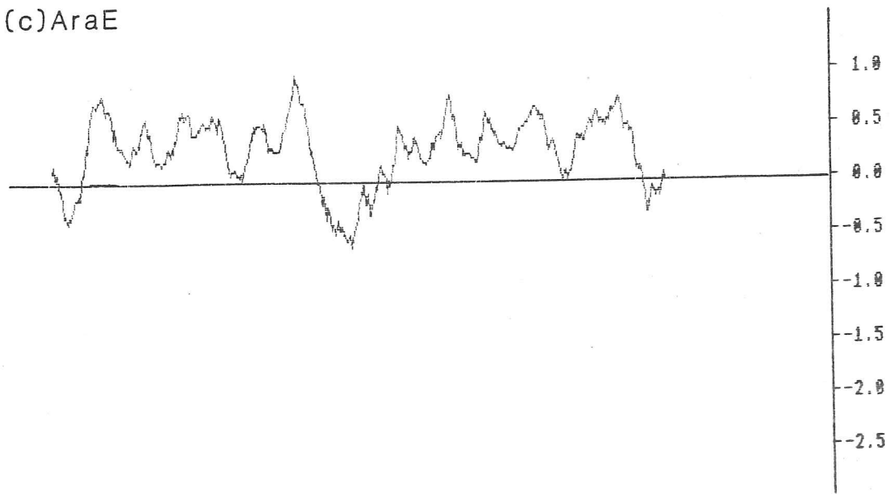
(a)XylE



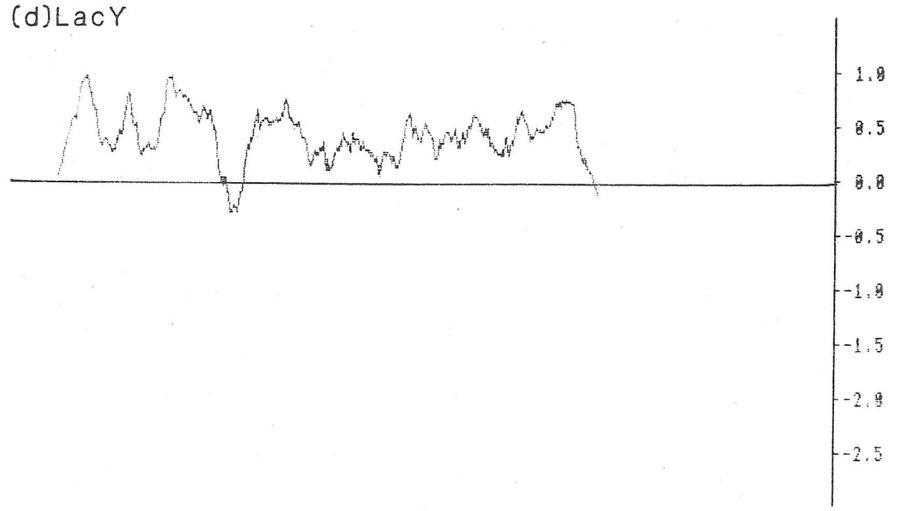
(b)XylE



(c)AraE



(d)LacY



(e)MelB

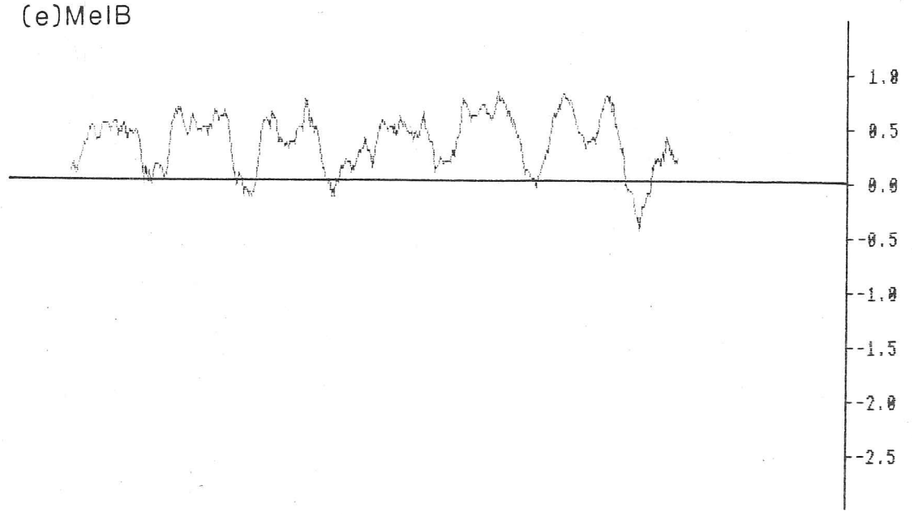


Figure 6.13 The hydropathic profiles of XylE (a) & (b), AraE (c), LacY and MelB (e)

The hydropathic profiles were calculated over a span of 9, (a), or 21, (b)-residues; they are discussed in the text (section 6.8.3).

compared, and if the value exceeds a certain minimum score a dot is plotted out at this point on the diagram. In this way a diagram is produced where each dot represents a certain degree of similarity over the span length between the two sequences being compared. If the comparison is between two identical sequences a central diagonal line should be found, and internal similarities should be seen as a symmetrical pattern of diagonal lines either side of this main diagonal. Thus, when comparing two different sequences the major point of interest is on the diagonal.

'Diagon' analyses of the protein sequences, comparing XylE with various other proteins, were performed with a span length of 21 residues and a minimum score of 240, using the program Microdiag written by Dr. P.R. Alefounder and based on the DIAGON program of Dr. R. Staden.

When XylE was compared with LacY (sequence from Buchel et al., 1980) or MelB (sequence from Yazyu et al., 1984), little similarity between the protein sequences was found (Figures 6.14 and 6.15). There was no correlation on the diagonal although some minor similarities of one region of one protein with other regions of the second protein were observed, reflecting generally hydrophobic regions. A similar pattern off the diagonal was found when comparing XylE with itself (Figure 6.16), indicating that any one part of the XylE protein was as similar to other parts of itself as to LacY or MelB.

When XylE was compared with AraE (sequence from M.C.J. Maiden, unpublished results), a strong similarity was found on the diagonal (Figure 6.17), even under quite stringent settings when the off-diagonal lines were removed. On alignment of the sequences XylE was found to consist of 28% identical residues to and 23% conservative changes from AraE. Although having similar generally hydrophobic regions there are some considerable differences between the two sequences so any amino acids that are conserved are likely to be significant. It is of interest to note that although XylE has 18 proline residues while AraE has only 11, 10 of these are conserved between the two proteins, and in most of these cases a short stretch of adjacent amino-acids to one side of the proline are either identical or of a similar nature. This is probably important in determining the structure of the protein as proline residues frequently cause an interruption in secondary structure. In view of the fact that Garcia et al., (1982) suggest histidine residues may be

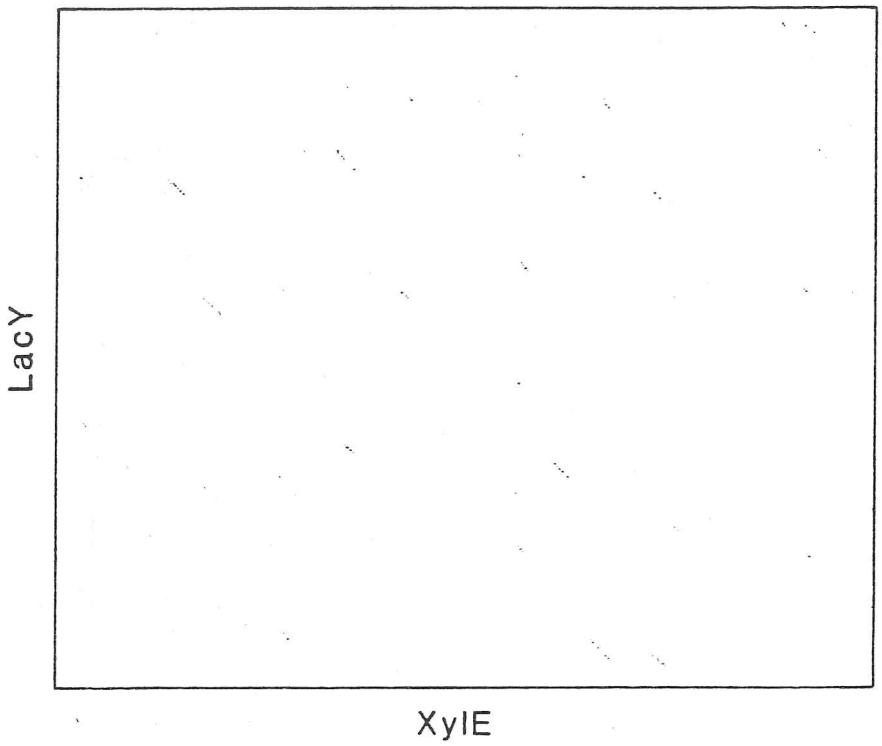


Figure 6.14 Comparison of the amino acid sequences of XylE and LacY by a Diagon analysis

The analysis was performed with a span length of 21 residues and the minimum score required for a point to be plotted was 240.

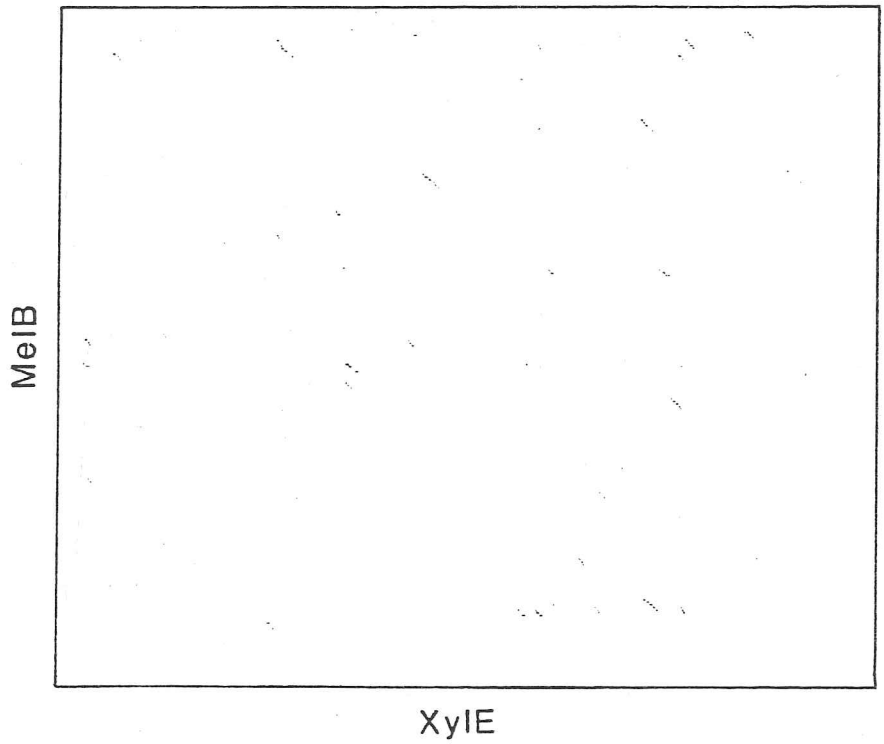


Figure 6.15 Comparison of the amino acid sequences of XylE and MelB by a Diagon analysis

The analysis was performed with a span length of 21 residues and the minimum score required for a point to be plotted was 240.

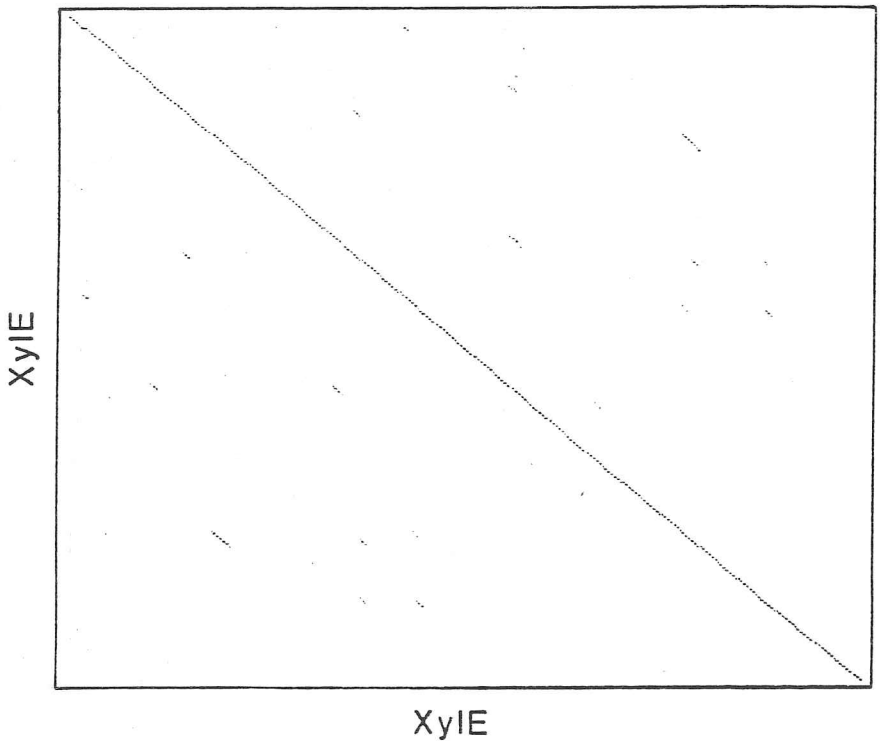


Figure 6.16 Comparison of the amino acid sequences of XylE with itself by a Diagon analysis

The analysis was performed with a span length of 21 residues and the minimum score required for a point to be plotted was 240.

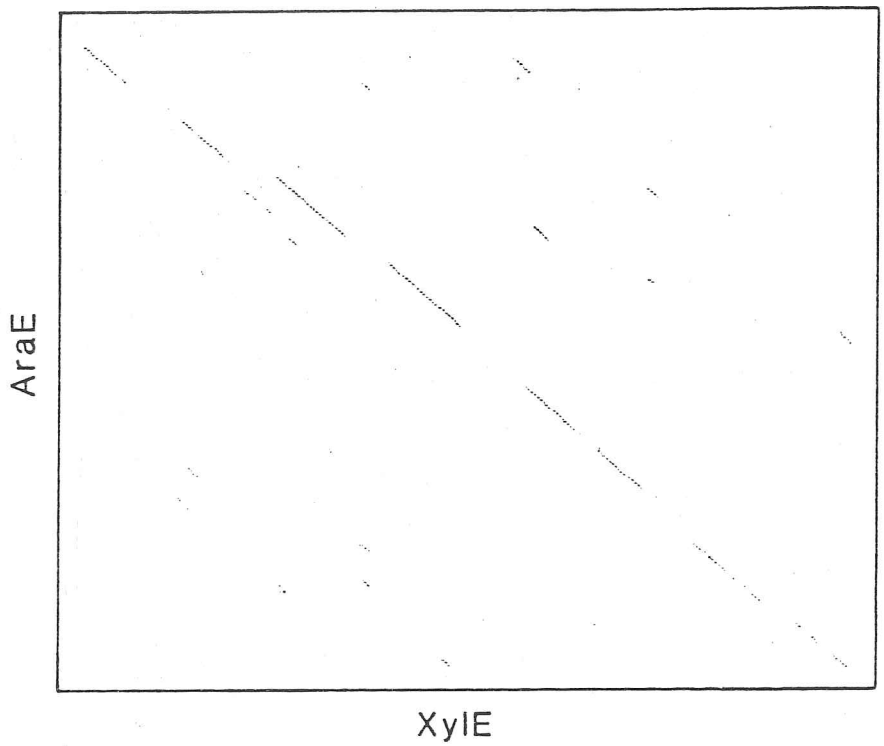


Figure 6.17 Comparison of the amino acid sequences of XylE and AraE by a Diagon analysis

The analysis was performed with a span length of 21 residues and the minimum score required for a point to be plotted was 240.

important in the mechanism of proton symport, it is worth noting that none of the five histidine residues in XylE are conserved in AraE. Also of interest is the distribution of cysteine residues. In AraE all five cysteine residues are concentrated in the C-terminal half of the protein, whereas in XylE only two of the seven cysteine residues are found in this region of the protein. It would thus seem likely that one of these two would be the residue labelled by N-ethylmaleimide. This proposal could be tested by site-directed mutagenesis, by isolation and amino-acid sequencing of the labelled peptide from purified protein, or by analysis of peptide mixtures containing the labelled peptide as has been done for LacY (Beyreuther et al., 1980).

The similarity between the XylE and AraE proteins, indicated in both the hydropathic profiles and the diagon analysis, suggests that there may be an evolutionary relationship between the genes coding for these two proton symport systems. Comparison of the two sequences provides the basis for future experiments to study the mechanism of action and structure-function relationships by, for example, site-directed mutagenesis.

CHAPTER 7

OVERVIEW

7. OVERVIEW

There are two transport systems for xylose in Escherichia coli (Shamanna and Sanderson, 1979): a proton symport system (Lam et al., 1980), coded for by the gene xylE, and a binding protein dependent system (Ahlem et al., 1982), coded for by the genes xylFG.

The work described here has concentrated on the proton symport system. Nevertheless, Mud(Ap^Rlac)I insertion mutations have been made in both transport systems, and the two systems have been shown to map at separate loci on the E. coli chromosome. At least two genes are involved in the binding protein dependent transport system, one coding for the binding protein (xylF), and the other, presumably, for a membrane component (xylG). Both these genes have been shown to map near to the xylose metabolic genes (xylAB) at about 80min on the E. coli chromosome. However, xylE is situated at a distinct locus, at 91.4min on the chromosome, near to malB. Thus xylose transport in E. coli resembles that of arabinose (araE and araFG; Novotny and Englesberg, 1966; Schleif, 1969; Brown and Hogg, 1972) and galactose (galP and mglP; Rotman et al., 1968; Wilson, 1974) in being mediated by two separate systems coded for by genes at distinct loci on the chromosome. Identification of the map position of xylE on the chromosome was a prerequisite to cloning the gene, and this location of xylE near to malB suggested a cloning strategy utilizing a phage λ insertion in malB.

In the course of this work to locate the genes coding for these two systems on the chromosome, strains lacking each or both of the transport systems were produced. Those having only a single system are useful for characterisation of that system, and strains lacking both systems were required for the identification of clones carrying xylE.

If the genes xylF and xylG form an operon xylF must be promoter proximal, as the Mud(Ap^Rlac)I insertion in xylG did not abolish xylose binding activity in a periplasmic shock fluid. Future work to determine the arrangement and operon structure of the genes xylABRFG, and any other transport component genes, in relation to each other could be undertaken. One way to approach this problem would be to analyse specialised transducing phages from λplacMu

insertions into these various genes by complementation and restriction mapping.

Also, by using a Mud(Ap^{R}lac)I insertion in *xylE* in a partial diploid *xylFG* strain containing in addition *xylE*$^+$, and comparing the induction of β-galactosidase and xylose transport activity it may be possible to gain an indication as to whether the XylE protein functions as a monomer or dimer (or higher oligomer) in transport.

The xylose proton symport system has been shown to be susceptible to inhibition by the sulphydryl reagent N-ethylmaleimide, as for the other sugar-proton symport sytems LacY (Fox and Kennedy, 1965; Jones and Kennedy, 1969), AraE (Macpherson, 1981), and GalP (Kaethner and Horne, 1980; Macpherson *et al.*, 1983). This inhibition may be alleviated by substrate and some substrate analogues, in particular by 6-deoxyglucose, indicating the sulphydryl group to be in the vicinity of the substrate binding site. Although only a poor substrate for transport by XylE, 6-deoxyglucose was found to be a good inhibitor of xylose transport by XylE, and a much less potent inhibitor of xylose transport by XylFG.

This has allowed the relatively specific labelling of the XylE protein with radioisotope-labelled NEM, after a protection step with 6-deoxyglucose and unlabelled NEM to block unprotected sites. Comparison of dual isotope labelled protected and unprotected samples identified a labelled peak in XylE$^+$ strains at an apparent molecular mass in the range 36 000-41 000Da, which was absent in XylE$^-$ strains. The absence of this peak in the XylE$^-$ strains provided strong evidence that it was due to the XylE protein, and not to a XylFG component. The value of the apparent molecular mass of XylE is similar to those of AraE (36 000-38 000Da; Macpherson *et al.*, 1981) and of GalP (34 000-38 000Da; Macpherson *et al.*, 1983), but different from that of LacY (approx. 30 000Da; Jones and Kennedy, 1969; Newman *et al.*, 1981). However, the true molecular mass of XylE calculated from the DNA sequence was found to be 53 607Da. A similar discrepancy has been reported for LacY, where the true molecular mass is 46 502Da (Buchel *et al.*, 1980). This has been attributed to abnormally high binding of SDS to the relatively hydrophobic protein (Beyreuther *et al.*, 1980) which results in an increased charge-to-mass ratio. A molecular mass of 47 500Da was also found for LacY by gel filtration in the

presence of hexamethylphosphoric triamide (Konig and Sandermann, 1982).

A XylE-LacZ hybrid protein formed as a result of a Mud(Ap^R*lac*)II insertion in *xylE* enabled the purification of part of the XylE protein by utilizing the β-galactosidase part of the molecule. The hybrid protein was membrane bound and had very different characteristics to native β-galactosidase. Detergent was required throughout the purification procedure to keep the protein in solution, and the XylE part of the molecule may have altered its binding characteristics to, for example, the p-aminophenyl-β-D-thiogalactoside affinity column. In addition, the protein was being purified from a different set of proteins to that from which native β-galactosidase is purified. Therefore, the purification of the hybrid protein was not a staightforward repetition of a purification procedure for β-galactosidase. The hybrid protein was purified on the basis of its large size, relative to the majority of membrane proteins, by gel filtration. This technique may prove useful for the purification of other membrane bound hybrid β-galactosidase proteins.

The sequence of the N-terminal six amino acids of this hybrid protein (determined by Dr. J.E. Walker) confirmed the reading frame and translation initiation point deduced from DNA sequencing of the cloned *xylE* gene.

For future work the hybrid protein may provide a means of raising antibodies which would be active against the XylE protein. Such antibodies would be useful to assay the XylE protein without having to resort to reconstitution experiments during procedures to purify the XylE protein itself.

The *xylE* gene has been cloned by using λ*plac*Mu and by exploiting the proximity of *xylE* to *malB*. Specialised transducing phages were obtained from λ*plac*Mu insertions in *xylE* and *malK* carrying an overlapping region of host chromosomal DNA. Comparison of the restriction maps of these phages located the *xylE* gene on the inserted DNA of λ(*xylE*)Φ(*malK'-lacZ*). Subcloning identified a 2.7kb HincII fragment containing the entire *xylE* gene, the DNA of which was sequenced by the Sanger dideoxy chain termination method.

Only one open reading frame of sufficient length to encode the XylE protein (apparent molecular mass 36 000-41 000Da) was present on this fragment, and the N-terminal sequence obtained from the XylE-LacZ hybrid

protein agreed with the first six amino acids of this reading frame. This revealed the true molecular mass of the XylE protein to be 53 607Da, corresponding to 491 amino acids. A putative ribosomal binding site was identified just upstream of the translational initiation codon, preceded by a possible RNA polymerase binding site, and a potential rho-independent transcriptional terminator was located downstream of the translational termination codon. The amino acid composition indicated the XylE protein to be a basic, highly hydrophobic protein, similar to LacY (Buchel et al., 1980) and MelB (Yazyu et al., 1984).

Comparison of the XylE sequence with those of LacY and MelB showed little similarity between them. However, comparison with the sequence of AraE (determined by M.C.J. Maiden) indicated that the two proteins were very similar, suggesting that there may be an evolutionary relationship between the genes xylE and araE.

Future work could include site-directed mutagenesis, based on a comparison of the XylE and AraE sequences, in an attempt to elucidate mechanistic features and structure function relationships. Also, it should be possible to overexpress the cloned xylE gene using a suitable expression vector, aiding the purification of the XylE protein. Purified protein would enable studies aimed at probing its secondary and tertiary structure.

REFERENCES

REFERENCES

Adler, J. (1975) Annu. Rev. Biochem. **44**, 341-356

Adler, J., Hazelbauer, G.L. & Dahl, M.M. (1973) J. Bacteriol. **115**, 824-847

Ahlem, C., Huisman, W., Neslund, G. & Dahms, A.S. (1982) J. Biol. Chem. **257**, 2926-2931

Alemohammad, M.M. & Knowles, C.J. (1974) J. Gen. Microbiol. **82**, 125-142

Ames, G.F.-L. & Higgins, C.F. (1983) Trends Biochem. Sci. **8**, 97-100

Ames, G.F.-L. & Spudich, E.N. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 1877-1881

Anderson, B., Weigel, N., Kundig, W. & Roseman, S. (1971) J. Biol. Chem. **246**, 7023-7033

Argos, P., Mahoney, W.C., Hermodson, M.A & Hanei, M. (1981) J. Biol. Chem. **256**, 4357-4361

Ashworth, J.M. & Kornberg, H.L. (1966) Proceedings of the Royal Society **B165**, 179-188

Bachmann, B. (1983) Microbiol. Rev. **47**, 180-230

Bankier, A.T. & Barrell, B.G. (1983) in 'Techniques in the Life Sciences' **B5**, 'Nucleic Acid Biochemistry' **B508**, (Flavell, R.A., ed.), pp1-34, Elsevier Scientific Publishers Ireland Ltd.

Bassilana, M., Damiano-Forano, E. & Leblanc, G. (1985) Biochem. Biophys. Res. Commun. **129**, 626-631

Bavoil, P., Hofnung, M. & Nikaido, H. (1980) J. Biol. Chem. **255**, 8366-8369

Bavoil, P. & Nikaido, H. (1981) J. Biol. Chem. **256**, 11385-11388

Berger, E.A. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 1514-1518

Berger, E.A. & Heppel, L.A. (1974) J. Biol. Chem. **249**, 7747-7755

Beverin, S., Sheppard, D.E. & Park, S.S. (1971) J. Bacteriol. **107**, 79-86

Beyreuther, K., Bieseler, B., Ehring, R., Griesser, H.-W., Mieschendahl, M., Muller-Hill, B. & Triesch, I. (1980) Biochem. Soc. Trans. **8**, 675-676

Beyreuther, K., Bieseler, B., Ehring, R. & Muller-Hill, B. (1981) in 'Methods in Protein Sequence Analysis' (Elzina, M., ed.), pp139-148, Humana Press, Inc., Clifton, N.J.

Bocklage, H. & Muller-Hill, B. (1983) Eur. J. Biochem. **137**, 561-565

Booth, I.R. & Hamilton, W.A. (1980) Biochem. J. **188**, 467-473

Booth, I.R., Mitchell, W.J. & Hamilton, W.A. (1979) Biochem. J. **182**, 687-696

Brass, J.M., Ehmann, U. & Bukau, B. (1983) J. Bacteriol. **155**, 97-106

Brass, J.M., Bauer, K., Ehmann, U. & Boos, W. (1985) J. Bacteriol. **161**, 720-726

Bremer, E., Silhavy, T.J., Weisemann, J.M. & Weinstock, G.M. (1984) J. Bacteriol. **158**, 1084-1093

Brickman, E., Silhavy, T.J., Bassford, P., Shuman, H. & Beckwith, J. (1979) J. Bacteriol. **139**, 13-18

Briggs, K.A., Lancashire, W.E. & Hartley, B.S. (1984) EMBO J. **3**, 611-616

Britton, P., Boronat, A., Hartley, D.A., Jones-Mortimer, M.C., Kornberg, H.L. & Parra, F. (1983) J. Gen. Microbiol. **129**, 349-358

Brooker, R.J. & Wilson, T.H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 3959-3963

Brown, C.E. & Hogg, R.W. (1972) J. Bacteriol. **111**, 603-613

Buchel, D.E., Gronenborn, B. & Muller-Hill, B. (1980) Nature (London) **283**, 541-545

Bukau, B., Brass, J.M. & Boos, W. (1985) J. Bacteriol. **163**, 61-68

Carrasco, N., Viitanen, P., Herzlinger, D. & Kaback, H.R. (1984a) Biochemistry **23**, 3681-3687

Carrasco, N., Herzlinger, D., Mitchell, R., DeChiara, S., Danho, W., Gabriel, T.F. & Kaback, H.R. (1984b) Proc. Natl. Acad. Sci. U.S.A. **81**, 4672-4676

Casadaban, M.J. & Chou, J. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 535-539

Casadaban, M.J. & Cohen, S.N. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4530-4533

Cecchini, G. & Kock, A.L. (1975) J. Bacteriol. **123**, 187-195

Celis, R.T.F. (1985) Eur. J. Biochem. **145**, 403-411

Chen, C.-C. & Wilson, T.H. (1984) J. Biol. Chem. **259**, 10150-10158

Cleland, W.W. (1967) Adv. Enzymol. **29**, 1-32

Clune, A., Lee, K.-S. & Ferenci, T. (1984) Biochem. Biophys. Res. Commun. **121**, 34-40

Cohn, D.E. & Kaback, H.R. (1980) Biochemistry **19**, 4237-4243

Cohn, D.E., Kaczorowski, G.J. & Kaback, H.R. (1981) Biochemistry **20**, 3308-3313

Curtis, S.J. & Epstein, W. (1975) J. Bacteriol. **122**, 1189-1199

Dairi, T., Inokuchi, K., Mizuno, T. & Mizushima, S. (1985) J. Mol. Biol. **184**, 1-6

Daruwalla, K.R., Paxton, A.T. & Henderson, P.J.F. (1981) Biochem. J. **200**, 611-627

Dassa, E. & Hofnung, M. (1985) EMBO J. **4**, 2287-2293

David, J.D. & Wiesmeyer, H. (1970) Biochim. Biophys. Acta **201**, 497-499

Davis, E.O., Jones-Mortimer, M.C. & Henderson, P.J.F. (1984) J. Biol. Chem. 259, 1520-1525

Davis, R.W., Botstein, D. & Roth, J.R. (1980) 'A Manual for Genetic Engineering: Advanced Bacterial Genetics', Cold Spring Harbour Lab., C.S.H., New York

Del Campillo-Campbell, A. & Campbell, A. (1982) J. Bacteriol. **149**, 469-478

Del Campo, F.F., Hernandez-Asensio, M. & Ramirez, J.M. (1975) Biochem. Biophys. Res. Commun. **63**, 1099-1105

Dills, S.S., Apperson, A., Schmidt, M.R. & Saier, M.H. (1980) Microbiol. Rev. **44**, 385-418

Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W. & Hofnung, M. (1984) J. Biol. Chem. **259**, 10606-10613

Ehring, R., Beyreuther, K., Wright, J.K. & Overath, P. (1980) Nature (London) **283**, 537-540

Flagg, J.L. & Wilson, T.H. (1977) J. Membr. Biol. **31**, 233-255

Foster, D.L., Boublik, M. & Kaback, H.R. (1983) J. Biol. Chem. **258**, 31-34

Fox, C.F. & Kennedy, E.P. (1965) Proc. Natl. Acad. Sci. U.S.A. **54**, 891-899

Fraenkel, D.G. & Vinopal, R.T. (1973) Annu. Rev. Microbiol. **27**, 69-100

Froshauer, S. & Beckwith, J. (1984) J. Biol. Chem. **259**, 10896-10903

Garcia, M.L., Patel, L., Padan, E. & Kaback, H.R. (1982) Biochemistry **21**, 5800-5805

Gilson, E., Nikaido, H. & Hofnung, M. (1982a) Nucleic Acids Res. **10**, 7449-7458

Gilson, E., Higgins, C.F., Hofnung, M., Ames, G.F.-L. & Nikaido, H. (1982b) J. Biol. Chem. **257**, 9915-9918

Goldkorn, T., Rimon, G. & Kaback, H.R. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3322-3326

Goldkorn, T., Rimon, G., Kempner, E.S. & Kaback, H.R. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1021-1025

Grenier, F.C., Reizer, J., Waygood, E.B. & Saier, M.H. (1985) J. Bacteriol. **163**, 243-247

Grenier, F.C., Waygood, E.B. & Saier, M.H. (1985) Biochemistry **24**, 4872-4876

Griesser, H.-W., Muller-Hill, B. & Overath, P. (1983) Eur. J. Biochem. **137**, 567-572

Grosjean, H. & Fiers, W. (1982) Gene **18**, 199-209

Hanahan, D. (1983) J. Mol. Biol. **166**, 557-580

Hanatani, M., Yazyu, H., Shiota-Niiya, S., Moriyama, Y., Kanazawa, H., Futai, M. & Tsuchiya, T. (1984) J. Biol. Chem. **259**, 1807-1812

Hayashi, S. & Lin, E.C.C. (1965) Biochim. Biophys. Acta **94**, 479-487

Heller, K.B., Lin, E.C.C. & Wilson, T.H. (1980) J. Bacteriol. **144**, 274-278

Henderson, P.J.F. & Giddens, R.A. (1977) Biochem. J. **168**, 15-22

Henderson, P.J.F., Giddens, R.A. & Jones-Mortimer, M.C. (1977) Biochem. J. **162**, 309-320

Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A. (eds.) (1983) 'Lambda II' Cold Spring Harbour Lab., C.S.H., New York

Hengge, R. & Boos, W. (1983) Biochim. Biophys. Acta **737**, 443-478

Hershey, A.D. (ed.) (1971) 'The Bacteriophage Lambda', Cold Spring Harbour Lab., C.S.H., New York

Herzlinger, D., Viitanen, P., Carrasco, N. & Kaback, H.R. (1984) Biochemistry **23**, 3688-3693

Herzlinger, D., Carrasco, N. & Kaback, H.R. (1985) Biochemistry **24**, 221-229

Higgins, C.F., Hiles, I.D., Whalley, K. & Jamieson, D.J. (1985) EMBO J. **4**, 1033-1039

Hobson, A.C., Gho, D. & Muller-Hill, B. (1977) J. Bacteriol. **131**, 830-838

Hobson, A.C., Weatherwax, R. & Ames, G.F.-L. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 7333-7337

Hogg, R.W. & Hermodson, M.A. (1977) J. Biol. Chem. **252**, 5135-5141

Hong, J.-S., Hunt, A.G., Masters, P.S. & Lieberman, M.A. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 1213-1217

Houssin, C., Le Maire, M., Aggerbeck, L.P. & Shechter, E. (1985) Arch. Biochem. Biophys. **240**, 593-606

Hoving, H., Koning, J.H. & Robillard, G.T. (1982) Biochemistry **21**, 3128-3136

Hoving, H., ten Hoeve-Duurkens, R. & Robillard, G.T. (1984) Biochemistry 23, 4335-4340

Howe, M.M. & Bade, E.G. (1975) Science **190**, 624-632

Ikemura, T. (1981) J. Mol. Biol. **151**, 389-409

Jablonski, E.G., Brand, L. & Roseman, S. (1983) J. Biol. Chem. **258**, 9690-9699

Jones, T.H.D. & Kennedy, E.P. (1969) J. Biol. Chem. **244**, 5981-5987

Kaback, H.R. (1968) J. Biol. Chem. **243**, 3711-3724

Kaback, H.R. (1971) Methods Enzymol. **22**, 99-120

Kaback, H.R. & Patel, L. (1978) Biochemistry **17**, 1640-1646

Kaethner, T.M. & Horne, P. (1980) FEBS Lett. **113**, 258-264

Katz, L. & Englesberg, E. (1971) J. Bacteriol. **107**, 34-52

Kaczorowski, G.J., LeBlanc, G. & Kaback, H.R. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 6319-6323

Kellermann, O. & Szmelcman, S. (1974) Eur. J. Biochem. **47**, 139-149

Kieser, T. (1984) Plasmid **12**, 19-36

Klapper, M.H. (1977) Biochem. Biophys. Res. Commun. **78**, 1018-1024

Klein, W.L. & Boyer, P.D. (1972) J. Biol. Chem. **247**, 7257-7265

Koch, A.L. (1985) Biochimie **67**, 137-140

Kolodrubetz, D. & Schleif, R. (1981) J. Mol. Biol. **151**, 215-227

Konig, B. & Sandermann, H. (1982) FEBS Lett. **147**, 31-34

Kornberg, H.L. (1966) in 'Essays in Biochemistry' **2**, (Campbell, P.N. & Greville, G.D., eds.), pp1-31, Academic Press Inc.

Kornberg, H.L. & Smith, J. (1969) Nature (London) **224**, 1261-1262

Kundig, W., Ghosh, S. & Roseman, S. (1964) Proc. Natl. Acad. Sci. U.S.A. **52**, 1067-1074

Kyte, J. & Doolittle, R.F. (1982) J. Mol. Biol. **157**, 105-132

Laemmli, U.K. (1970) Nature (London) **227**, 680-685

Lam, V.M.S., Daruwalla, K.R., Henderson, P.J.F. & Jones-Mortimer, M.C. (1980) J. Bacteriol. **143**, 396-402

Lancaster, J.R. & Hinkle, P.C. (1977) J. Biol. Chem. **252**, 7657-7661

Lombardi, F.J. & Kaback, H.R. (1972) J. Biol. Chem. **247**, 7844-7857

Lopilato, J., Tsuchiya, T. & Wilson, T.H. (1978) J. Bacteriol. **134**, 147-156

Macpherson, A.J.S. (1982) Ph.D. thesis (Cambridge)

Macpherson, A.J.S., Jones-Mortimer, M.C. & Henderson, P.J.F. (1981) Biochem. J. **196**, 269-283

Macpherson, A.J.S., Jones-Mortimer, M.C., Horne, P. & Henderson, P.J.F. (1983) J. Biol. Chem. **258**, 4390-4396

Maezawa, S., Hayashi, Y., Nakae, T., Ishii, J., Kameyama, K. & Takagi, T. (1983) Biochim. Biophys. Acta **747**, 291-297

Mahoney, W.C., Hogg, R.W. & Hermodson, M.A. (1981) J. Biol. Chem. **256**, 4350-4356

Maleszka, R., Wang, P.Y. & Schneider, H. (1982) Canadian J. Biochem. **60**, 144-151

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) 'Molecular Cloning: A Laboratory Manual', Cold Spring Harbour Lab., C.S.H., New York

Marchal, C., Greenblatt, J. & Hofnung, M. (1978) J. Bacteriol. **136**, 1109-1119

Markgraf, M., Bocklage, H. & Muller-Hill, B. (1985) Mol. Gen. Genet. **198**, 473-475

Meadow, N.D., Saffen, D.W., Dottin, R.P. & Roseman, S. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 2528-2532

Mieschendahl, M., Buchel, D., Bocklage, H. & Muller-Hill, B. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 7652-7656

Miller, J. (1972) 'Experiments in Molecular Genetics', Cold Spring Harbour Lab., C.S.H., New York

Misset, O., Brouwer, M. & Robillard, G.T. (1980) Biochemistry **19**, 883-890

Misset, O. & Robillard, G.T. (1982) Biochemistry **21**, 3136-3142

Mitaku, S., Wright, J.K., Best, L. & Jahnig, F. (1984) Biochim. Biophys. Acta **776**, 247-258

Mitchell, P. (1961) Nature (London) **191**, 144-148

Mitchell, P. (1970) Symp. Soc. Gen. Microbiol. **20**, 121-166

Mitchell, P. (1976) Biochem. Soc. Trans. **4**, 399-430

Nakae, T. (1976) Biochem. Biophys. Res. Commun. **71**, 877-884

Nelson, S.O. & Postma, P.W. (1984) Eur. J. Biochem. **139**, 29-34

Nelson, S.O., Wright, J.K. & Postma, P.W. (1983) EMBO J. **2**, 715-720

Nelson, S.O., Lengeler, J. & Postma, P.W. (1984) J. Bacteriol. **160**, 360-364

Neu, H.C. & Heppel, L.A. (1965) J. Biol. Chem. **240**, 3685-3692

Neuhaus, J.-M. & Wright, J.K. (1983) Eur. J. Biochem. **137**, 615-621

Neuhaus, J.-M., Schindler, H. & Rosenbusch, J.P. (1983) EMBO J. **2**, 1987-1991

Neuhaus, J.-M., Soppa, J., Wright, J.K., Riede, I., Blocker, H., Frank, R. & Overath, P. (1985) FEBS Lett. **185**, 83-88

Newman, M.J., Foster, D.L., Wilson, T.H. & Kaback, H.R. (1981) J. Biol. Chem. **256**, 11804-11808

Niiya, S., Yamasaki, K., Wilson, T.H. & Tsuchiya, T. (1982) J. Biol. Chem. **257**, 8902-8906

Novotny, C.P. & Englesberg, E. (1966) Biochim. Biophys. Acta **117**, 217-230

Novotny, M.J., Frederickson, W.L., Waygood, E.B. & Saier, M.H. (1985) J. Bacteriol. **162**, 810-816

Olsson, A., Moks, T., Uhlen, M. & Gaal, A.B. (1984) J. Biochem. Biophys. Methods **10**, 83-90

Osumi, T. & Saier, M.H. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 1457-1461

Ozeki, H. & Ikeda, H. (1968) Annu. Rev. Genetics **2**, 245-279

Padan, E., Patel, L. & Kaback. H.R. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 6221-6225

Parra, F., Jones-Mortimer, M.C. & Kornberg, H.L. (1983) J. Gen. Microbiol. **129**, 337-348

Patel, L., Garcia, M.L. & Kaback, H.R. (1982) Biochemistry **21**, 5805-5810

Peri, K.G., Kornberg, H.L. & Waygood, E.B. (1984) FEBS Lett. **170**, 55-58

Postma, P.W. & Roseman, S. (1976) Biochim. Biophys. Acta **457**, 213-257

Postma, P.W., Epstein, W., Schuitema, A.R.J. & Nelson, S.O. (1984) J. Bacteriol. **158**, 351-353

Prestidge, L.S. & Pardee, A.B. (1965) Biochim. Biophys. Acta **100**, 591-593

Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 784-788

Ramos, S. & Kaback, H.R. (1977a) Biochemistry **16**, 848-854

Ramos, S. & Kaback, H.R. (1977b) Biochemistry **16**, 854-859

Ramos, S. & Kaback, H.R. (1977c) Biochemistry **16**, 4271-4275

Ramos, S., Shuldiner, S. & Kaback, H.R. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 1892-1896

Reider, E., Wagner, E.F. & Schweiger, M. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 5529-5533

Richarme, G. (1985a) Biochim. Biophys. Acta **815**, 37-43

Richarme, G. (1985b) J. Bacteriol. **162**, 286-293

Richey, D.P. & Lin, E.C.C. (1972) J. Bacteriol. **112**, 784-790

Robertson, D.E., Kaczorowski, G.J., Garcia, M.L. & Kaback, H.R. (1980) Biochemistry **19**, 5692-5702

Robillard, G.T. & Konings, W.N. (1981) Biochemistry **20**, 5025-5032

Roseman, S. (1977) in 'Biochem. of Membrane Transport' FEBS Symp. 42, (Semenza, G. & Carafoli, E., eds.), pp582-597, Spring-Verlag, Berlin

Rosen, B.P. & Kashket, E.A. (1978) in 'Bacterial Transport' (Rosen, B.P., ed.), pp559-620, Marcel Dekker, Inc.

Rosenberg, M. & Court, D. (1979) Annu. Rev. Genetics **13**, 319-353

Rosenfeld, S.A., Stevis, P.E. & Ho, N.W.Y. (1984) Mol. Gen. Genet. **194**, 410-415

Rotman, B., Ganesan, A.K. & Guzman, R. (1968) J. Mol. Biol. **36**, 247-260

Rottenberg, H. (1976) FEBS Lett. **66**, 159-163

Saier, M.H. (1977) Bacteriol. Rev. **41**, 856-871

Saier, M.H. & Roseman, S. (1976a) J. Biol. Chem. **251**, 6598-6605

Saier, M.H. & Roseman, S. (1976b) J. Biol. Chem. **251**, 6606-6615

Saier, M.H., Simoni, R.D. & Roseman, S. (1976) J. Biol. Chem. **251**, 6584-6597

Sanger, F., Nicklen, S. & Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467

Sanger, F., Coulsen, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B.A. (1980) J. Mol. Biol. **143**, 161-178

Sanno, Y., Wilson, T.H. & Lin, E.C.C. (1968) Biochem. Biophys. Res. Commun. **32**, 344-349

Schaffner, W. & Weissmann, C. (1973) Anal. Biochem. **56**, 502-514

Schellenberg, G.D., Sarthy, A., Larson, A.E., Backer, M.P., Crabb, J.W., Lidstrom, M., Hall, B.D. & Furlong, C.E. (1984) J. Biol. Chem. **259**, 6826-6832

Schleif, R. (1969) J. Mol. Biol. **46**, 185-196

Seckler, R., Wright, J.K. & Overath, P. (1983) J. Biol. Chem. **258**, 10817-10820

Seckler, R. & Wright, J.K. (1984) Eur. J. Biochem. **142**, 269-279

Shamanna, D.K. & Sanderson, K.E. (1979a) J. Bacteriol. **139**, 64-70

Shamanna, D.K. & Sanderson, K.E. (1979b) J. Bacteriol. **139**, 71-79

Shine, J. & Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 1342-1346

Shiota, S., Yazyu, H. & Tsuchiya, T. (1984) J. Bacteriol. **160**, 445-447

Shiota, S., Yamane, Y., Futai, M. & Tsuchiya, T. (1985) J. Bacteriol. **162**, 106-109

Shuman, H.A. (1982) J. Biol. Chem. **257**, 5455-5461

Shuman, H.A. & Beckwith, J.R. (1979) J. Bacteriol. **137**, 365-373

Shuman, H.A. & Silhavy, T.J. (1981) J. Biol. Chem. **256**, 560-562

Shuman, H.A., Silhavy, T.J. & Beckwith, J.R. (1980) J. Biol. Chem. **255**, 168-174

Silhavy, T.J., Brickman, E., Bassford, P.J., Casadaban, M.J., Shuman, H.A., Schwartz, V., Guarente, L., Schwartz, M. & Beckwith, J.R. (1979) Mol. Gen. Genet. **174**, 249-259

Singh, A.P. & Bragg, P.D. (1976) FEBS Lett. **64**, 169-172

Staden, R. (1982) Nucleic Acids Research **10**, 2951-2961

Steers, E., Cuatrecasas, P. & Pollard, H.B. (1971) J. Biol. Chem. **246**, 196-200

Stock, J. & Roseman, S. (1971) Biochem. Biophys. Res. Commun. **44**, 132-138

Stock, J.B., Waygood, E.B., Meadow, N.D., Postma, P.W. & Roseman, S. (1982) J. Biol. Chem. **257**, 14543-14552

Stribling, D. & Perham, R.N. (1973) Biochem. J. **131**, 833-841

Stroobant, P. & Kaback, H.R. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 3970-3974

Szmelcman, S. & Hofnung. M. (1975) J. Bacteriol. **124**, 112-118

Szmelcman, S. & Schwartz, M. (1976) Eur. J. Biochem. **65**, 13-19

Tanaka, K., Niiya, S. & Tsuchiya, T. (1980) J. Bacetiol. **141**, 1031-1036

Teather, R.M., Hamelin, O., Schwarz, H. & Overath, P. (1977) Biochim. Biophys. Acta **467**, 386-395

Teather, R.M., Muller-Hill, B., Abrutsch, U., Aichele, G. & Overath, P. (1978) Mol. Gen. Genet. **159**, 239-248

Treptow, N.A. & Shuman, H.A. (1985) J. Bacteriol. **163**, 654-660

Trumble, W.R., Viitanen, P.V., Sarkar, H.K., Poonian, M.S. & Kaback. H.R. (1984) Biochem. Biophys. Res. Commun. **119**, 860-867

Tsuchiya, T. & Wilson, T.H. (1978) Membrane Biochem. **2**, 63-79

Tsuchiya, T., Raven, J. & Wilson, T.H. (1977) Biochem. Biophys. Res. Commun. **76**, 26-31

Tsuchiya, T., Lopilato, J. & Wilson, T.H. (1978) J. Membr. Biol. **42**, 45-59

Tsuchiya, T., Takeda, K. & Wilson, T.H. (1980) Membrane Biochem. **3**, 131-146

Tsuchiya, T., Ottina, K., Moriyama, Y., Newman, M.J. & Wilson, T.H. (1982) J. Biol. Chem. **257**, 5125-5128

Tsuchiya, T., Oho, M. & Shiota-Niiya, S. (1983) J. Biol. Chem. **258**, 12765-12767

Ullmann, A. (1984) Gene **29**, 27-31

Viitanen, P., Garcia, M.L., Foster, D.L., Kaczorowski, G.J. & Kaback, H.R. (1983) Biochemistry **22**, 2531-2536

Viitanen, P., Garcia, M.L. & Kaback, H.R. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1629-1633

Vinopal, R.T. & Fraenkel, D.G. (1974) J. Bacteriol. **118**, 1090-1100

von Meyenburg, K. & Nikaido, H. (1977) Biochem. Biophys. Res. Commun. **78**, 1100-1107

Wandersman, C., Schwartz, M. & Ferenci, T. (1979) J. Bacteriol. **140**, 1-13

Waygood, E.B. (1980) Canadian J. Biochem. **58**, 1144-1146

West, I.C. (1970) Biochem. Biophys. Res. Commun. **41**, 655-661

West, I.C. (1980) Biochim. Biophys. Acta **604**, 91-126

West, I.C. & Mitchell, P. (1972) J. Bioenergetics **3**, 445-462

West, I.C. & Mitchell, P. (1973) Biochem. J. **132**, 587-592

West, I.C. & Wilson, T.H. (1973) Biochem. Biophys. Res. Commun. **50**, 551-558

Wilson, D.B. (1974) J. Biol. Chem. **249**, 553-558

Wilson, D.M., Ottina, K., Newman, M.J., Tsuchiya, T., Ito, S. & Wilson, T.H. (1985) Membrane Biochem. **5**, 269-290

Wilson, T.H., Kusch, M. & Kashket, E.R. (1970) Biochem. Biophys. Res. Commun. **40**, 1409-1414

Winkler, H.H. & Wilson, T.H. (1966) J. Biol. Chem. **241**, 2200-2211

Witholt, B., Bockhout, M., Brock, M., Kingma, J., van Heerikhuizen, H. & de Leij, L. (1976) Anal. Biochem. **74**, 160-170

Wong, P.T.S., Kashket, E.R. & Wilson, T.H. (1970) Proc. Natl. Acad. Sci. U.S.A. **65**, 63-69

Wray, W., Boulikas, T., Wray, V.P. & Hancock, R. (1981) Anal. Biochem. **118**, 197-203

Wright, J.K. & Overath, P. (1984) Eur. J. Biochem. **138**, 497-508

Wright, J.K. & Seckler, R. (1985) Biochem. J. **227**, 287-297

Wright, J.K., Riede, I. & Overath, P. (1981) Biochemistry **20**, 6404-6415

Wright, J.K., Weigel, U., Lustig, A., Bocklage, H., Mieschendahl, M., Muller-Hill, B., & Overath, P. (1983) FEBS Lett. **162**, 11-15

Yazyu, H., Shiota-Niiya, S., Shimamoto, T., Kanazawa, H., Futai, M. & Tsuchiya, T. (1984) J. Biol. Chem. **259**, 4320-4326

Yazyu, H., Shiota, S., Futai, M. & Tsuchiya, T. (1985) J. Bacteriol. **162**, 933-937

Zilberstein, D., Shuldiner, S. & Padan, E. (1979) Biochemistry **18**, 669-673

APPENDIX

APPENDIX: DNA sequence gel readings

The consensus sequence is shown in heavy print. The numbers at the left refer to the individual gel readings, and the sign of these numbers indicates whether the sequence was determined on the positive or the negative strand.

```
            10         20         30         40         50         60
-69  AACGGCCTGACGGCAGGTGGTGTGAAAGGTTAAAGATGTTGTTCTGCCAATGTTATGCCG
 71  AACGGCCTGACGGCAGGTGGTGT
 86  *ACGGCCTGACGGCAGGTGGTGTGAAAGGTTAAAGATGTTGTTCTGCCAATGTTATGCCG
     AACGGCCTGACGGCAGGTGGTGTGAAAGGTTAAAGATGTTGTTCTGCCAATGTTATGCCG

            70         80         90        100        110        120
-69  CTGCACCCTCAACTTACGTTATCCCAACTTGTGACTGTTATTCGGCGCTCCACGGAGC
 86  CTGCACCCTCAACTTACGTTATCCCAACTTGTGACTGTTATTCGGCGCTCCACGGAGCCG
-52                                                  CCACGGAGCC*
     CTGCACCCTCAACTTACGTTATCCCAACTTGTGACTGTTATTCGGCGCTCCACGGAGCCG

           130        140        150        160        170        180
 86  TTTTTTTCTTTCGTCTGCAATCTGAATCGTTCGCCGGTTAATATTTCCATCATAGAGCTT
-52  *TTTTTTCTTTCGTCTGCAATCTGAATCGTTCGCCGGTTAATATTTCCATCATAGAGCTT
 48                                                 TATTTCCATCATAGAGCTT
-34                                                       AGCTT
     TTTTTTTCTTTCGTCTGCAATCTGAATCGTTCGCCGGTTAATATTTCCATCATAGAGCTT

           190        200        210        220        230        240
 86  ATTATTTTTACGTTATTTGTTTTCCCACTTACGATAATTCTCTTTCGTGCTCTGAGTCAC
-52  ATTATTTTTACGTTATTTGTTTTCCCACTTACGATAATTCTCTTTCGTGCTCTGAGTCAC
 48  ATTATTTTTACGTTATTTGTTTTCCCACTTACGATAATTCTCTTTCGTGCTCTGAGTCAC
-34  ATTATTTTTACGTTATTTGTTTTCCCACTTACGATAATTCTCTTTCGTGCTCTGAGTCAC
 91                                                       TCAC
     ATTATTTTTACGTTATTTGTTTTCCCACTTACGATAATTCTCTTTCGTGCTCTGAGTCAC

           250        260        270        280        290        300
 86  GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAG
-52  GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
 48  GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
-34  GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
 91  GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
 39        AGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
-40         GTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
     GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
```

```
            310        320        330        340        350        360
-52  CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
 48  CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
-34  CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
 91  CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
 39  CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
-40  CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
-110           *CGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
     CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT

            370        380        390        400        410        420
-52  CAATGCACTGATAAAAGATCAGAA
 48  CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
-34  CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
 91  CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
 39  CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
-40  CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
-110 CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
     CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA

            430        440        450        460        470        480
 48  TTCCAGTTATATATTT
-34  TTCCAGTTATATATTTTCGATTACCTTAGTCGCTACATTAGGTG
 91  TTCCAGTT
 39  T*CCAGTTATATATTTTCGATTACCTTAGTCGCTACATTAGGTGGTTTATTATTTGGCTA
-40  TTCCAGTTATATATTTTCGATTACCTTAGTCGCTACATTAGGTGGTTTATTATTTGGCTA
-110 TTCCAGTTATATATTTTCGATTACCTTAGTCGCTACATTAGGTGGTTTATTATTTGGCTA
-108  TCCAGTTATATATTT*CGATTACCTTAGTCGCTACATTAGGTGGTT*ATTATTTGGCTA
 21                              AGTCGCTACATTAGGTGGTTTATTATTTGGCTA
 77                                      ATTAGGTGGTTTATTATTTGGCTA
102                                     TTAGGTGGTTTATTATTTGGCTA
 37                                          GGTTTATTATTTGGCTA
     TTCCAGTTATATATTTTCGATTACCTTAGTCGCTACATTAGGTGGTTTATTATTTGGCTA

            490        500        510        520        530        540
 39  CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
-40  CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
-110 CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
-108 CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
 21  CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
 77  CG*CACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
102  CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
 37  CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
-17                  CCGGTACTGTTGAGTCACTCAATACCGTC*TTGTTGCTCCACA
 18                                              CGTCTTTGTTGCTCCACA
     CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
```

```
                550         560         570         580         590         600
  39  AA*CTTAAGTGAATCCGCTGCCAA*TCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
 -40  AAACTTAAGTGAATCCGCTGCCAACTC
-110  AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTT*CGTGGCCAGCGCTCTGAT
-108  AAA*TTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
  21  AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
  77  AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
 102  AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
  37  AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
 -17  AAACTTAAGTGAATCCGCTGCCAACT*CCTGTTAGGGTTTT*CGTGGCCAGCGCTCTGAT
  18  AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
 -94                    CTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT
  -5                     TGCCAACTCCCTGTTAGGGTTTT*C*TGGCCAGCGCTCTGAT
      AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGCTCTGAT

                610         620         630         640         650         660
  39  TGG
-110  TGGTTGCATCATCGGC*GTGCCCTCGGTGGT
-108  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
  21  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
  77  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
 102  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
  37  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
 -17  TGGTTGCATCATCG*C*GTGCCCTC*GTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
  18  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
 -94  TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAA
  -5  TGGTTGCATCATCG*C*GTGCCCTC*GTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
 -88    GTT*CATCATCGGCGGTGCCCTCG*TGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
      TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA

                670         680         690         700         710         720
-108  TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
  21  TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
  77  TTCACTTAAGATTGCTGCTGTCCTGTTTTTT
 102  TTCACTTAAGATTGCTGCTGTCCTG
  37  TTCACTTAAGATT-C-GCTGTCCTGTTTTTTATTTCTGGTGTAGG
 -17  TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCT*GTGTAGGTTCTGCCTGGCCAGA
  18  TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
  -5  TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
 -88  TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
 -32                    CTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
  45                       TTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
 -30                          TTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
   9                             *TGGTGTAGGTTCT*CCTGGCCAGA
 -66                                          -GTGCC*GGCCAGA
      TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
```

```
          730        740        750        760        770        780
-108  ACTTGGTTT*ACCTCTATAAACCCGGACAACACTGT*CCTG
  21  A*TTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
 -17  ACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
  18  ACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
  -5  ACTT*GTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
 -88  ACTTGGTTT-*CCTCTATAAACCCGGACAACACTGTGCCTGT-TATCTGGCAGGTTATGT
 -32  ACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
  45  ACTTGGTTTTACCTCTATA--CCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
 -30  ACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATG*
   9  ACT-GGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
 -66  AC--GG---TACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
 100    TTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
      ACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT

          790        800        810        820        830        840
  21  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
 -17  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
  18  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
  -5  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
 -88  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATG*T
 -32  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
  45  CCCGGA*TTTGTTATTTATCGCATTA
 -30  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
   9  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
 -66  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
 100  CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
  11      GAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
  -8                                CGGTATTGGCGTTGGTTTAGCCTCAATG*T
      CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT

          850        860        870        880        890        900
  21  CTCGCCAATGTATATTGCGGAACTGGCTC*AGCTCATATT*GCGG
 -17  CTCGCCA-TGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGG
  18  CTCGCCAATGTATATTGCGGAACTGGCTC*AGCT*ATATTCGCGGGAAACTGGTCTCTTT
  -5  CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
 -88  CTCGCCA-TGTATATTGCGG
 -32  CTCGCCA-TGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAA
 -30  CTCGCC*ATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
   9  CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
 -66  CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
 100  CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
  11  CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
  -8  CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
  96                                                          TTT
      CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
```

```
            910        920        930        940        950        960
 18 TAACCAGTTTGCGATTATTT
 -5 TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGC
-30 TAACCAGTTTGCGATT
  9 TAACCAGTTTGCGATTATTT
-66 TAACCAGTTTGCGATTATTTT
100 TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTG*
 11 TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGC
 -8 TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGC
 96 TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGC
 29                                  TAGTTTACTGCGTAAACTATTTTATTGC
 20                                                 ATTTTATTGC
    TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGC

            970        980        990       1000       1010       1020
 -5 C**TTCCGGTGATGCCAGCTGGCTGAATA*TGA
 11 CCGTTCCGGTGATGCCAGCTGGCTGAATACTGACGGCTGGCGTTATATGTTTGC
 -8 C*GTTCCGGTGATGCCAGCTGGCTGAATACTGACGGCTGGCGTTATATGTTTGCCTCGGA
 96 CCGTTCCGGTGATGCCAGCTGGCTG
 29 CCGTTCCGGTGATGCCAGCTGGCTGAATACTGACGGCTGGCGTTATATGTTTGCCTCGGA
 20 CCGTTCCGGTGATGCCAGCTGGCTGAATACTGACGGCTGGCGTTATATGTTTGCCTCGGA
104       C*GTGATGCCAGCTGGCTGAATACTGACGGCTGGCGTTATATGTTTGCCTCGGA
 97                                                TATATGTTTGCCTCGGA
    CCGTTCCGGTGATGCCAGCTGGCTGAATACTGACGGCTGGCGTTATATGTTTGCCTCGGA

            1030       1040       1050       1060       1070       1080
 -8 ATGTATCCCTGCACTGCTGTTCTTAATGCTGCTGTATACC*TGCCAGAAAGTCCTCGCTG
 29 ATGTATCCCT**ACTGCTGTTCTTAATGCTGCTGTATACCGTGCCAGAAAGTCCTCGCTG
 20 ATGTATCCCTGCACTGCTGTTCTTAATGCTGCTGTATACCGTGCCAGAAAGTCCTCGCTG
104 ATGTATCCCTGCACTGCTGTTCTTAATGCTGCTGTATACCGTGCCAGAAAGTCCTCGCTG
 97 ATGTATCCCTGCACTGCTGTTCTTAATGCTGCTGTATACCGTGCCAGAAAGTCCTCGCTG
-111    TATCCCTGCACTGCTGT*CTTAATGCTGCTGTATACC*TGCCAGAAA*TCCTCGCTG
 51                      TCTTAATGCTGCTGTATACCGTGCCAGAAAGTCCTCGCTG
 -2                                                CAGAA**TCCTCGCTG
-13                                                      CT*
    ATGTATCCCTGCACTGCTGTTCTTAATGCTGCTGTATACCGTGCCAGAAAGTCCTCGCTG

            1090       1100       1110       1120       1130       1140
 -8 G
 29 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAA*-ATGGGCAA
 20 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
104 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
 97 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
-111 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
 51 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
 -2 GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
-13 GCT-AT-TC-C-CGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTAT*GGCAA
    GCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATTATGGGCAA
```

```
          1150        1160        1170        1180        1190        1200
  29 CACGCTTGCAACTCAGGCAGTACAGGAAATTAAACACTCCCTGGATCATGGCCGCAAAAC
  20 CACGCTTGCAACTCAGGCAGTACAGGAAATTAAACA*TCCCTGGATCATGGCCGCAAAAC
 104 CACGCTTGCAACTCAGGCAGTACAGG
  97 CACG*TTGCAACT-AGGCAGTACAGGAAATTAAACA*TCCCTGGATCATGGCCGCAAAA*
-111 CACGCTTGCAACTCAGGCAGTACAGGAAATTAAACACTCCCTGGATCATGGCCGCAAAAC
  51 CACGCTTGCAACTCAGGCAGTACAGGAAATTAAACACTCCCTGGATCATGGCCGCAAAAC
  -2 CACGCTTGCAACTCAGGCAGTACAGGAAATTAAACACTCCCTGGATCATGGCCGCAAAAC
 -13 CACGCTTGCAACTCAGGCA*TACA*GAAATTAAACACTCCCTGGATCATGGCCGCAAAAC
     CACGCTTGCAACTCAGGCAGTACAGGAAATTAAACACTCCCTGGATCATGGCCGCAAAAC

          1210        1220        1230        1240        1250        1260
  29 CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATGCTCTCCAT
  20 CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATG
  97 CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTG
-111 CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATGCTCTCCAT
  51 CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATGCTCTCCAT
  -2 CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATGCTCTCCAT
 -13 CG*TGGTCGTCTGCTGATGTTTGGC-TGGGCGTGATTGTAATCGGCGTAATGCTCTCCAT
  85                          TTTGGCGTGGGCGTGATTGTA*TCGGCGT*ATGCTCTCCAT
 -25                                T*GGC*TGATTGTAATCG*CGTAATGCTCTCCAT
 -27                                     TGATTGTAATCGGCGTAATGCTCTCCAT
  42                                       ATTGTAATCGGCGTAATGCTCTCCAT
  15                                                GGCG-AATGCTCTCCAT
  19                                                  CGTAATGCTCTCCAT
     CGGTGGTCGTCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATGCTCTCCAT

          1270        1280        1290        1300        1310        1320
  29 CTTCCAGCAA
-111 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAA*TGTTCAAAAC
  51 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAA*C
  -2 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
 -13 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCA-AAC
  85 CTTCCAGCAATTTGTCGGCATCAATGTG
 -25 CTTCCAGCAATTTGTCGGCATCAAT-TGGTGCTGTACTAC-CGCC*GAAGTGTTCAAAAC
 -27 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAA*TGTTCAAAAC
  42 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
  15 CT*CCAGCAA*TTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
  19 CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
  46                TCGGCATCA*TGTGGTGCTGTACTACGCGCCGGA-GTGTTCAAAAC
  75                TCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
  31                          TGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
     CTTCCAGCAATTTGTCGGCATCAATGTGGTGCTGTACTACGCGCCGGAAGTGTTCAAAAC
```

```
          1330        1340        1350        1360        1370        1380
-111 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
  51 GCTGGGGGCCAGCACGGATATC
  -2 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
 -13 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
 -25 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCG*AGTTATCAACCT
 -27 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
  42 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
  15 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
  19 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
  46 GCTGGGGGCCAGCACGG-TATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCA*CCT
  75 GCTGGGGGCCAGCACGGATATCGCGCTCTTGCAGACCAT-ATTGTCGGAGT-ATCAACCT
  31 GCTGGGGGCCAGCACGGATATCGCGCTGTTGCAGACCATTATTGTCGGAGTTATCAACCT
 -26                                                 TCG*AGTTATCAACCT
 -65                                                   G-AGTTATCAACCT
 -53                                                     GTTATCAACCT
    GCTGGGGGCCAGCACGGATATCGCGCT-TTGCAGACCATTATTGTCGGAGTTATCAACCT

          1390        1400        1410        1420        1430        1440
-111 CACCTTCACCGTTCTGG
  -2 CACCTTCACCGT*CTGGCAAT*AT
 -13 CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGC
 -25 CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
 -27 CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGC
  42 CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
  15 CACCT*CACCGTTCTGGCAA*TATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
  19 CACCTTCAC*GTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
  46 CACCTTCACCGTTCTGGCA*TTATGACGGTG*AT*AATTTGGTCGTAA*CCACTGCAA*T
  75 CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
  31 CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
 -26 CACCTTCACC*TTCTGGCAATTATGACGGT*GATAAATTTG*TC*TAAGCCACTGCAAAT
 -65 CACCTTCACC-TTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT
 -53 CACCTTCACCGTTCTGGCAATTATGACGGT*GATAAATTT*GTC*TAAGCCACTGCAAAT
    CACCTTCACCGTTCTGGCAATTATGACGGTGGATAAATTTGGTCGTAAGCCACTGCAAAT

          1450        1460        1470        1480        1490        1500
 -25 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
  42 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
  15 *ATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
  19 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
  46 TATCGGCGCACTC
  75 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
  31 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
 -26 TATCG*CGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGC-TTTTACACTCA
 -65 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCC-CGGTACCGCGTTTTACACTCA
 -53 TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTAC-CGTTTTACA8TCA
    TATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGGTACCGCGTTTTACACTCA
```

```
          1510        1520        1530        1540        1550        1560
-25  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
 42  GGCA*CGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGC*TTTGCCATGTC
 15  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
 19  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
 75  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
 31  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
-26  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
-65  GGCACCGGGTATTGTGGCGCTACTGTC
-53  GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC
 90                                          TATGTTGCCGCCTTTGCCATGTC
 -3                                            TTGCCGCCTTTGCCATGTC
     GGCACCGGGTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCTTTGCCATGTC

          1570        1580        1590        1600        1610        1620
-25  CTGGGGTCCGGTATGCTGGGTACTGCTGTCG
 42  CTGGGGT*CGGTATGCTGG
 15  CTGGG*TCCGGTATG*TGGGTA*TGCTGTCGGAAATCTTCCCGAATGCTATTCGTGGT*A
 19  CTGGGGTCCGGTATGCTGGGTACTGCTGTCGGAAATCTT*CCGAATGCTATTCGTGGTAA
 75  CTGGGGTCCGGTATGCTGGGTACTGCTGTCGGAAA
 31  CTGGGGTCCGGTATGCTGGGTACTGCTGTCG
-26  CTGGGGTCCGGTATGCTGGGTACTGCTGTCGGAAATCTTCCCGAATGCTATTCGTGGTAA
-53  CTGGGGTCCGGTATGCTGGGTACTGCTGTCGGAAATCTTCCCGAATGCTATTCGTGGTAA
 90  CTGGGGTCCGGTATGCTGGGTACTGCTGTCG
 -3  CTGGG*TCCG-TATGCTGGGTACTGCTGTCGGAAATCTTCCCGAATGCTATTC-TGGTAA
-74              ATGCTGGGTACTGCTGTCGGAAATCTTCCCGAATGCTATTCGTGGTAA
     CTGGGGTCCGGTATGCTGGGTACTGCTGTCGGAAATCTTCCCGAATGCTATTCGTGGTAA

          1630        1640        1650        1660        1670        1680
 15  AGCGCTGGCAATC
 19  AGCGCTGGCAATCGCGGTGGCG**CCAG
-26  AGCGCTGGCAATCGCGGTGGCGGCCCAGTGGCTGGCGAACTACTTCGTCTCCTGGACCTT
-53  AGCGCTGGCAATCGCGGTGGCGGCCCAGTGGCTGGCGAACTACTTCGTCTCCTGGACC
 -3  AGCGCTGGCAATCGC*GTGGCGGCCCAGTGGCTGGCGAACTACTTCGTCTCCTGGACCTT
-74  A-CGCTGGCAATCGCGGTGGCGGCCCAGTGGCTGGCGAACTAC*
-10                    GTGGCGGCCCA-TGGCTGGCGAACTACTTCGTCTCCTGGACCTT
 54                   GGTGGCGGCCCAGTGGCTGGCGAACTACTTCGTCTCCTGGACCTT
 36                     TGGC*GCCCAGTGGCTGGCGAACTACTTCGTCTCCTGGACCTT
-89                              CCA*TGGCTGGCGAACTACTTCGTCTCCTGGACCTT
     AGCGCTGGCAATCGCGGTGGCGGCCCAGTGGCTGGCGAACTACTTCGTCTCCTGGACCTT

          1690        1700        1710        1720        1730        1740
-26  CCCGATGATGGACAAAAACTCCTGGCTGGTGGCCCATTTCCACAACGG
 -3  CCCGATGATGGACAAAAACTCCTGGCTGGTGGCCCATTTCCACAACGGTTTCTCCTACTG
-10  CCC*ATGATGGACAAAAACTCCTGGCT*GTGGCCCATTTCCACAACGGTTTCTCCTACTG
 54  CCCGATGATGGACAAAAA*TCCTGGCTGGTGGCCCATTTCCACAA-GGTTTCTCCTACTG
 36  CCCGATGATGGACAAAAACTCCTGGCTGGTGGCCCATTTCCACAACGGTTTCTCCTACTG
-89  CCCGATGATGGACAAAAACTCCTGGCT*GTGGCCCATTTCCACAACGGTTTCTCCTACTG
-35                                                  CTCCTACTG
     CCCGATGATGGACAAAAACTCCTGGCTGGTGGCCCATTTCCACAACGGTTTCTCCTACTG
```

```
          1750        1760        1770        1780        1790        1800
 -3 GATTTACGGTTGTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTGTCCCGGAAAC
-10 GATTTACGGTTGTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTGTCCCGGAAAC
 54 GATTTACGGTTGTATGGGCGTTCTGGCAGCACT-TTTATGTGGA
 36 GATTTACGGTTGTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTGTCCCGGAAAC
-89 GATTTACGGT*GTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTG*CCCGGAAAC
-35 GATTTACGGTTGTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTGTCCCGGAAAC
 22   TTT*CGGTTGTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTGTCCCGGAAAC
    GATTTACGGTTGTATGGGCGTTCTGGCAGCACTGTTTATGTGGAAATTTGTCCCGGAAAC

          1810        1820        1830        1840        1850        1860
 -3 CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGG
-10 CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
 36 CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
-89 CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
-35 CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
 22 CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
 76       GTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
 24                                AAGCGCTCTGGGAACCGGAAACGAAGAAAACACA
    CAAAGGTAAAACCCTTGAGGAGCTGGAAGCGCTCTGGGAACCGGAAACGAAGAAAACACA

          1870        1880        1890        1900        1910        1920
-10 ACAAACTG*TACGCTGTAATCTTCCTGTCCAGCACGCCGCGCCATTTCGGCGTG*TGACT
 36 ACAAA
-89 ACAAACT-*TACGCTGTAA
-35 ACAAACT*CTACGCTGTAATCTTCCTGTCCAGCACGC
 22 ACAAACTGCTACGCTGTAATCTTCCTGTCCAGCACGCCGCGCCATTTCGGCGTGCTGACT
 76 ACAAACTGCTACGCTGTAATCTTCCTGTCCAGCACGCCGCGCCATTTCGGCGTGCTGACT
 24 ACAAACTGCTACGCTGTAATCTTCCTGTCCAGCACGCCGCGCCATTTCGGCGTGCTGACT
-112                               TCCAGCACGCC**GCCATT*CG*C*TG*T*A*T
    ACAAACTGCTACGCTGTAATCTTCCTGTCCAGCACGCCGCGCCATTTCGGCGTGCTGACT

          1930        1940        1950        1960        1970        1980
-10 TTTTACTCCCGCTTCAGCC*
 22 TTTTACTCCCGCTTCAGCCGTTTCGAATTACACAGCCACAGGGTGATCACCAGTAACAGG
 76 TTTTACTCCCGCTTCAGCCGTTTCGAATTACACAGCCACAGGGTGATCACCAGTAACAGG
 24 TTTTACTCCCGCTTCAGCCGTTTCGAATTACACAGCCACAGGGTGATCACCAGTAACAGG
-112 TT*TACTCCCGCTTCAGCC*TT*C*AATTACACAGCCACAGG*TGATCACCAGTAACAGG
 -41          GCTTCAGCCG*TTCGAATTACACAGCCACA*GGTGATCACCAGTAACAGG
  -6             TCAGCCGTTTCGAATTACACAGCCACA*GGTGATCACCAGTAACAGG
  12                                        AGGGTGATCACCAGTAACAGG
 -28                                       GTGATCACCAGTAACAGG
    TTTTACTCCCGCTTCAGCCGTTTCGAATTACACAGCCACAGGGTGATCACCAGTAACAGG
```

```
          1990       2000       2010       2020       2030       2040
  22 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
  76 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
  24 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
-112 ATCGCAGCCGA*TAGATCAACACATCCA*TGGCGATTTATGATCGACGATGATCAAGCGC
 -41 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
  -6 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
  12 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
 -28 ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC
     ATCGCAGCCGAGTAGATCAACACATCCAGTGGCGATTTATGATCGACGATGATCAAGCGC

          2050       2060       2070       2080       2090       2100
  22 ACA*TGGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
  76 ACAATGGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
  24 ACAATGGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
-112 ACAATGGCG*TGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
 -41 ACAATGGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
  -6 ACAATGGCG*TGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
  12 ACAATGGCGGTGATCCCAATATAGACAAAGTAACG
 -28 ACAATGGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
  33      GGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA
  67                                          GGAAGTGAAAACCGGACTGA
 101                                               AAACCGGACTGA
     ACAATGGCGGTGATCCCAATATAGACAAAGTAACGTAGCGGGAAGTGAAAACCGGACTGA

          2110       2120       2130       2140       2150       2160
  22 AAGTACTTCACAATCAGCGCGAT*AATTCG
  76 AAGTACTTCACAATCAGCGCGATAA
  24 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCT
-112 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
 -41 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
  -6 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
 -28 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
  33 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
  67 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
 101 AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT
     AAGTACTTCACAATCAGCGCGATAAATTCGAAATAGAGAAAGTAAACCACCAGTCCTTCT

          2170       2180       2190       2200       2210       2220
-112 ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCACCGTT
 -41 ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCACCGTT
  -6 ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCACCGTT
 -28 ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCA
  33 ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCACCGTT
  67 ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCACCGTT
 101 ACCAGCTCATATTTGCTGGTTTGTTCTGGC
  23                                                CCAGATGCACCGTT
     ACCAGCTCATATTTGCTGGTTTGTTCTGGCGCGAACAGCACATCAGCCAGATGCACCGTT
```

```
          2230       2240       2250       2260       2270       2280
-112  TCTTTGCCGAGGAAGACAACCAAAATCAGGCCGAGGCACAGCAGGCCAAGATTGAGTACG
 -41  TCTTTGCCGAGGAAGACAACCAAAATCAGGCCGAGGCACAGCAGGCCAA
  -6  TCTTTGCCGAGGAAGACAACCAAAATCAGGCCGAGGCACAGCAGGCCAAGATTGAGTACG
  33  TCTTTGCCGAGGAAGACAACCAAAATCA*GCCGAGGCACAGCAGGCCAAGATTGAGTACG
  67  TCTTTGCCGAGGAAGACAACCAAAATCA*GCCGAGGCACAGCAGGCCAAGATTGAGTACG
  23  TCTTTGCCGAGGAAGACAACCAAAATCAGGCCGAGGCACAGCAGGCCAAGATTGAGTACG
 103        GCCGAGGAAGACAACCAAAATCAGGCCGAGGCACAGCAGGCCAAGATTGAGTACG
      TCTTTGCCGAGGAAGACAACCAAAATCAGGCCGAGGCACAGCAGGCCAAGATTGAGTACG

          2290       2300       2310       2320       2330       2340
-112  GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGA
  -6  GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
  33  GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
  67  GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
  23  GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
 103  GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
  93                              TCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
  98                                CACGCGCGGACGAGAGAGTGACGTCATAGCAGTA
      GTCTGCAAAATGGTGGAGATAAACTCCACGCGCGGACGAGAGAGTGACGTCATAGCAGTA

          2350       2360       2370       2380       2390       2400
  -6  CCTCCTTCACACGTGGCGAG
  33  CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
  67  CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
  23  CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
 103  CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
  93  CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
  98  CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
 106                  GCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT
      CCTCCTTCACACGTGGCGAGCTTCCTGTATAACGCAAAAATGTGACGGAGATCTATATTTT

          2410       2420       2430       2440       2450       2460
  33  TTGTGTAT
  67  TTGTGTATGTTTTGTTCAACCTGGATTTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
  23  TTGTGTATGTTTTGTTCAACCTGGATTTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
 103  TTGTGTATGTTTTGTTCAACCTGGATTTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
  93  TTGTGTATGTTTTGTTCAACCTGGATTTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
  98  TTGT
 106  TTGTGTATGTTTTGTTCAACCTGGATTTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
 -92   TGTGTATG--TTGT*CAACCTGGA*TTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
 -63                                                    ATATACCT
      TTGTGTATGTTTTGTTCAACCTGGATTTGGAGCGCGCCTCCACCAAAGCCGGATATACCT
```

```
          2470       2480       2490       2500       2510       2520
 67 ACCCGTAGGCCGGATAAGCCGCGCCT
 23 ACCCGTAGGCCGGATAAGCCGCGCCTGGCGGCGCATCCGGCAAGACAA
103 ACCCGTAGGCCGGATAAGCCGCGCCTGGCGGCGCATCCGGCAAGACAAACCGCACCAAAG
 93 ACCCGTAGGCCGGATAAGCCGCGCCTGGCGGCGCATCCGGCAAGACAAACCGCACCAAAG
106 ACCCGTAGGCCGGATAAGCCGCGCCTGGCGGCGCATCCGGCAAGACAAACCGCACCAAAG
-92 ACCCGTAGGCCGGAT
-63 ACCCGTAGGCCGGATAAGCCGCGCCTGGCGGCGCATCCGGCAAGACAAACCGCACCAAAG
107                                        GCATCCGGCAAGACAAACCGCACCAAAG
    ACCCGTAGGCCGGATAAGCCGCGCCTGGCGGCGCATCCGGCAAGACAAACCGCACCAAAG

          2530       2540       2550       2560       2570       2580
103 CCGGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAA
 93 CCGGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAATGCGCGAAAAAT
106 CCGGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAATGCGCGAAAAAT
-63 CCGGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAATGCGCGAAAAAT
107 CCGGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAATGCGCGAAAAAT
105   GGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAATGCGCGAAAAAT
-59                                        CCCAGCCACCAGCAATGCGCGAAAAAT
    CCGGTAATTCAAAAATCACCCATTCAGGTGATTCCCAGCCACCAGCAATGCGCGAAAAAT

          2590       2600       2610       2620       2630       2640
103
 93 CGACAATGTTCCTCTGGAACAAATGAAACGTAACCT
106 CGACAATGTTCCTCTGGAACAAATGAAACGT
-63 CGACAATGTTCCTCTGGAACAAATGAAACGTAACCTGATAAAGGTGGTAAAAATGAAGCC
107 CGACAATGTTCCTCTGGAACAAATGAAACGTAACCTGATAAAGGTGGTAAAAATGAAGCC
105 CGACAATGTTCCTCTGGAACAAATGAAACGTAA
-59 CGACAATGTTCCTCTGGAACAAATGAAACGTAACCTGATAAAGGTGGTAAAAATGAAGCC
 73                                        GTAACCTGATAAAGGTGGTAAAAATGAAGCC
 95                                           TGATAAAGGTGGTAAAAATGAAGCC
    CGACAATGTTCCTCTGGAACAAATGAAACGTAACCTGATAAAGGTGGTAAAAATGAAGCC

          2650       2660       2670       2680       2690       2700
-63 TTATTTTGCTGCTTTGATGTTATCAGTCTCTGAGCCTCAAGACGATCCTGAATGTAATAA
107 TTATTTTGCTGCTTTGATGTTATCAGTCTCTGAGCCTCAAGACGATCCTGAATGTAATAA
-59 TTATTTTGCTGCTTTGATGTTATCAGTCTCTGAGCCTCAAGACGATCCTGAATGTAATAA
 73 TTATTTTGCTGCTTTGATGTTATCAGTCTCTGAGCCTCAAGACGATCCTGAATGTAATAA
 95 TTATTTTGCTGCTTTGATGTTATCAGTCTCTGAGCCTCAAGACGATCCTGAATGTAATAA
    TTATTTTGCTGCTTTGATGTTATCAGTCTCTGAGCCTCAAGACGATCCTGAATGTAATAA

          2710       2720       2730       2740       2750       2760
-63 GCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGCCCGCGAAATGCCGGGCTG
107 GCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGCCCGCGAAATGCCGGGCTG
-59 GCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGCCCGCGAAATGC
 73 GCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGCCCGCGAAATGCCGGGCTG
 95 GCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGCCCGCGAAATGCCGGGCTG
-60                                        CGC4AAATGCCGGGCTG
    GCGTTCATGGCTGAACTCCTGAAATAGCTGTGAAAATATCGCCCGCGAAATGCCGGGCTG
```

```
          2770       2780       2790       2800       2810       2820
-63   ATTAGGAAA
107   ATTAGGAAAACAGGAAAGGGGGTTAGTGAATGCTTTTGCTTGATCTCAGTTTCAGTATTA
 73   ATTAGGAAAACAGGAAAGGGGGTTAGTGAATGCTTTTGCTTGATCTCAGTTTCAGTATTA
 95   ATTAGGAAAACAGGAAAGGGGGTTAGTGAATGCTTTTGCTTGATCTCAGTTTCAGTATTA
-60   ATTAGGAAAACAGGAAAGGGGGTTAGTGAATGCTTTTGCTTGATCTCAGTTTCAGTATTA
      ATTAGGAAAACAGGAAAGGGGGTTAGTGAATGCTTTTGCTTGATCTCAGTTTCAGTATTA

          2830       2840
107   ATATCCATTTTTTATAAGCGTC
 73   ATATCCATTTTTTATAAGCGTC
 95   ATATCCATTTTTTATAAGCGTC
-60   ATATCCATTTTTTATAAGCGTC
      ATATCCATTTTTTATAAGCGTC
```