STUDIES ON THE STRUCTURE AND FUNCTION OF A MAMMALIAN SUGAR TRANSPORT PROTEIN

by Richard A. J. Preston B.Sc. A.R.C.S. (London)

Department of Biochemistry and Chemistry
The Royal Free Hospital School of Medicine
Rowland Hill Street
London, NW3 2PF

A dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the University of London.

February 1995

ABSTRACT

The GLUT1 isoform of the mammalian passive glucose transporter family is of paramount importance in satisfying the intracellular demand for sugar by many tissues. In the absence of a three-dimensional structure at atomic resolution, the central aim of this project was to utilise a combination of methodologies to examine the structure/function relationship of GLUT1.

From an amino acid sequence alignment of the sugar transporter family, the two-dimensional model of GLUT1 was refined using several predictive algorithms. Furthermore, a three-dimensional model of the protein was constructed in accord with the hypothesis that the current transporter phenotype arose from the gene duplication of an ancestral six-helix protein. Experimental evidence in favour of this arrangement was obtained by the co-expression of the *N*- and *C*-terminal halves of GLUT1 in *Xenopus* oocytes.

Additional evidence in support of the predicted model for GLUT1 was obtained using a membrane-impermeant biotin derivative, which identified an exofacial lysine residue within the *C*-terminal half of GLUT1. In order to examine further the topography of the transporter, mutants of GLUT1 possessing additional sites for biotinylation were prepared for expression in CHO cells. However, analysis of a CHO cell line over-expressing wild-type GLUT1 revealed that, although high expression of functionally active GLUT1 could be achieved, the majority of the expressed protein was not targeted to the plasma membrane. These findings therefore precluded its use in topographical analyses. Nevertheless, the mutagenesis strategy also enabled the modification of proteolytic cleavage sites to aid the identification of the binding sites for transport inhibitors, such as ATB-BMPA. These mutants were shown to be functionally active in *Xenopus* oocytes, and one of them was characterised further by expression in the *Sf*9/baculovirus system. This mutant retained the ability to be labelled with ATB-BMPA and, furthermore, a labelled fragment was successfully immunopurified providing definitive evidence for the *C*-terminal location of the ATB-BMPA-binding site. It should now be possible to identify precisely the site of labelling by *N*-terminal sequencing.

ACKNOWLEDGEMENTS

First and foremost, I should like to express my gratitude to my supervisor, Dr. Stephen A. Baldwin. It has been a privilege to work under his expert guidance, and I hope that I have acquired something of his enthusiasm for science.

I would like also to thank Dr. B.M. Charalambous for many helpful discussions.

Many people have had various influences at different times throughout this project and a full description could make entertaining reading (!), however, some have played particularly significant roles.

I am extremely grateful to Dr. Dan Donnelly and Adam Nealis for all their assistance in using the programs to obtain structural data. I would also like to acknowledge Jocelyn Baldwin, not only for her generosity when I arrived in Leeds, but also for her supervision and advice with the *Xenopus* oocytes. In addition, the contribution of Dr. Alan Cox in establishing the *Sf*9/baculovirus expression system has proven invaluable. I also very much appreciate the advice of my good friend Kyriacos with many of the molecular biological aspects of the project - 'Here's to the next Chapter!'

Finally, I would like to thank everyone in the Baldwin/Henderson labs for their support and friendship and also for making the last year such a special one. I would particularly like to mention Gary, Giles, Angela, Arle, Tim and Katie who have been superb. Special thanks must go also to George, whose friendship, organisational skills and proofreading ability will always be appreciated.

TABLE OF CONTENTS

ABSTRACT 2

ACKNOWLEDGEMENTS 3

TABLE OF CONTENTS 4

LIST OF FIGURES 10

LIST OF TABLES 13

ABBREVIATIONS 14

CHAPTER 1. GENERAL INTRODUCTION 16

1.1 Introduction 16

1.2 Facilitative glucose transporters 17

1.2.1 GLUT1 19

1.2.2 GLUT2 20

1.2.3 GLUT3 21

1.2.4 GLUT4 21

1.2.5 GLUT5 22

1.2.6 GLUT7 22

1.3 Membrane topology and structure of GLUT1 23

1.3.1 Oligomeric state 27

1.4 Locations of substrate-binding site(s) 28

1.5 Heterologous expression systems 29

1.6 Site-directed mutagenesis 33

1.7 Aims of the study 37

CHAPTER 2. MATERIALS AND METHODS 38

2.1 Materials and suppliers 38

2.2 Techniques for protein chemistry 39

2.2.1 Purification of GLUT1 from erythrocyte membranes 39

2.2.2 Protein determination 40

2.2.3 Cytochalasin B ligand binding assay 41
2.2.4 SDS-polyacrylamide gel electrophoresis 42
2.2.5 Immunoblot (Western blot) analysis 43
2.2.5.1 Electrotransfer 43
2.2.5.2 Immunostaining 44
2.2.6 Preparation and use of an immunoaffinity column to purify antibodies raised aginst residues 460-477 of human GLUT1 45
2.2.7 Assessment of purified IgG by ELISA 47
2.2.8 Immobilisation of antibody raised against residues 460-477 of GLUT1 on protein A-Sepharose 47
2.2.9 Immunoprecipitation of GLUT1 48
2.3 Techniques for DNA manipulations 48
2.3.1 Phenol extraction and precipitation of nucleic acids 49
2.3.2 Plasmid DNA preparations 49
2.3.3 Removal of RNA from preparations of plasmid DNA 50
2.3.4 Digestion of DNA with restriction endonucleases 50
2.3.5 Agarose gel electrophoresis 51
2.3.6 Quantification of DNA 51
2.3.7 Purification of DNA fragments from agarose gels 52
2.3.8 Ligation of DNA fragments 52
2.3.9 Preparation of competent cells and transformation 53
2.3.10 Manual sequencing 54
2.4 Techniques for the baculovirus expression system 56
2.4.1 Methods for insect cell culture 56
2.4.1.1 Monolayer culture 56
2.4.1.2 Freezing and storage of insect cells 57
2.4.1.3 Thawing cells 57
2.4.2 Plaque assay 57
2.4.3 Virus amplification 58
2.4.4 Preparation of infectious AcNPV DNA 59

CHAPTER 3. SEQUENCE ALIGNMENT AND ANALYSIS OF THE SUGAR TRANSPORTER FAMILY 60

3.1 Introduction 60
3.2 The sugar transporter family 64
3.2.1 Multiple sequence alignment of the sugar transporter family 66
3.3 Phylogenetic relationships of the sugar transporters 70
3.4 Identification and location of secondary structural elements 73
3.4.1 Consensus secondary structure prediction 73
3.4.2 Hydropathic analysis of the sugar transporter family 77
3.5 Refinement of helix boundaries 81
3.6 Periodicity analysis of the sugar transporter family 87
3.7 Three-dimensional modelling of GLUT1 96
3.8 Discussion 99

CHAPTER 4. PROBING THE TOPOLOGY OF GLUT1 USING A MEMBRANE-IMPERMEANT REAGENT 100

4.1 Introduction 100
4.2 Optimisation of biotinylation procedure 105
4.3 Confirmation of the membrane-impermeant character of NHS-LC-Biotin 107
4.4 Proteolytic digestion of the biotinylated transporter 113
4.5 Discussion 116

CHAPTER 5. GENERATION OF GLUT1 MUTANTS FOR STRUCTURAL STUDIES AND THE IDENTIFICATION OF BINDING-SITES FOR TRANSPORT INHIBITORS 118

5.1 Introduction 118

5.2 Mutagenesis strategy 122
5.2.1 Mutants to aid identification of inhibitor binding-sites 122
5.2.2 Production of the *N*- and *C*-terminal halves of GLUT1 125
5.3 PCR mutagenesis 127
5.4 Generation of mutants for inhibitor binding-site identification 129
5.4.1 Construction of KK451/456RR mutant GLUT1 cDNA 133
5.4.2 Preparation and subcloning of GLUT1 mutant cassette 137
5.4.3 Construction of GLUT1 mutants KK.Q360K and KK.Q427K 137
5.4.4 Restriction analysis of mutant plasmid constructs 141
5.5 Partial sequencing of GLUT1 mutants 143
5.6 *Xenopus* oocyte expression system 146
5.6.1 Isolation of oocytes for expression studies 146
5.6.2 Injection of mRNA 147
5.6.3 Preparation of 5'-capped mRNA by *in vitro* transcription of plasmid DNA 148
5.6.4 Analysis of functional expression 151
5.6.5 Western blot analysis of *Xenopus* oocytes injected with wild-type and mutant GLUT1 mRNA 151
5.6.6 Time course of [^{3}H]-2DG uptake in *Xenopus* oocytes injected with mRNA encoding rat GLUT1 152
5.7 Functional expression of GLUT1 mutants 154
5.7.1 Expression of point mutants designed for inhibitor binding-site studies 154
5.7.2 Co-injection of mRNA encoding the *N*- and *C*-terminal halves of GLUT1 156
5.8 Discussion 160

CHAPTER 6. APPLICATION OF THE BACULOVIRUS EXPRESSION VECTOR SYSTEM TO THE IDENTIFICATION OF TRANSPORT INHIBITOR BINDING SITES 164

6.1 Introduction 164
6.2 Generation of recombinant pAcYM1 transfer vectors containing GLUT1 cDNA mutants 169
6.3 Generation of recombinant baculovirus 174
6.3.1 Preparation of transfer vector DNA 174
6.3.2 Co-transfection of insect cells 176
6.4 Production of GLUT1 via infection of *Sf*9 cells with wild-type and recombinant baculovirus 178
6.4.1 Immunoblot analysis of wild-type and double-lysine mutant GLUT1 over-expression in insect cells 178
6.5 Purification of polyclonal antibodies to the peptide corresponding to residues 460-477 of human GLUT1 179
6.6 Photolabelling of wild-type and the double-lysine mutant of GLUT1 expressed in insect cells with ATB-BMPA 181
6.7 Immunopurification of photolabelled GLUT1 expressed in insect cells 185
6.8 Discussion 189

CHAPTER 7. CHARACTERISATION OF GLUT1 EXPRESSED IN CHINESE HAMSTER OVARY (CHO) CELLS 191

7.1 Introduction 191
7.2 Generation of CHO cell line over-expressing GLUT1 193
7.3 CHO cell culture 195
7.3.1 Long-term storage of CHO cells 195
7.4 2-Deoxy-D-glucose uptake in CHO cells 196

7.5 Quantification of GLUT1 expressed in CHO cells 199
7.6 Photoaffinity labelling of membranes from CHO cells with [^{3}H]-cytochalasin B 201
7.7 Subcellular location of GLUT1 in wild-type and over-expressing CHO cell clones 204
7.8 Discussion 209

CHAPTER 8. GENERAL DISCUSSION 212

REFERENCES 219

APPENDIX 240

LIST OF FIGURES

Figure 1.2	Current topological model of GLUT1	25
Figure 3.1	Percent sequence identities for the sugar transporter family	68
Figure 3.2	Phenogram of the sugar transporter family	72
Figure 3.3	Consensus α-helix and β-turn prediction for the sugar transporter family	76
Figure 3.4	Consensus hydropathic analysis of the sugar transporter family	80
Figure 3.5	Refined topological model of GLUT1 based upon several consensus predictive schemes	86
Figure 3.6	Patterns in periodicity of Family I	92
Figure 3.7	Patterns in periodicity of the sugar transporter family	93
Figure 3.8	Possible arrangement of helices of GLUT1 in the lipid bilayer	95
Figure 3.9	Three-dimensional model of GLUT1	98
Figure 4.1	Reaction of NHS-LC-Biotin with a protein	103
Figure 4.2	Potential sites of exofacial biotinylation	104
Figure 4.3	Demonstration of linear relationship between amount of biotinylated protein applied to nitrocellulose and the intensity of the stained band	106
Figure 4.4	Time course for biotinylation of intact erythrocytes	108
Figure 4.5	Western blot analysis of biotinylated erythrocytes	109
Figure 4.6	Western blot demonstrating biotinylation of exofacial amino groups	111
Figure 4.7	Demonstration of biotinylation of purified GLUT1	112
Figure 4.8	Trypsin digest of the biotinylated, membrane-bound transporter	114
Figure 5.1	Diagram illustrating the mutations designed to facilitate the identification of inhibitor binding-sites	123
Figure 5.2	Production of the *N*- and *C*-terminal halves of GLUT1	126
Figure 5.3	Recombinant PCR	128

Figure 5.4 Design of the primers to generate the mutants for subsequent inhibitor binding-site identification 130
Figure 5.5 Design of the primers used to produce the *N*- and *C*-terminal halves of GLUT1 131
Figure 5.6 Diagram of the pGEM.GT construct 134
Figure 5.7 PCR of the double-lysine mutant of GLUT1 cDNA 136
Figure 5.8 The subcloning strategy used to prepare the GLUT1 mutant constructs 138
Figure 5.9 Production of Q360K and Q427K triple mutants of GLUT1 140
Figure 5.10 Partial restriction map and restriction analysis of GLUT1 mutant constructs 142
Figure 5.11 Sequencing of GLUT1 mutant constructs 145
Figure 5.12 *In vitro* transcription 149
Figure 5.13 Time course of sugar uptake in oocytes expressing rat GLUT1 153
Figure 5.14 Functional expression of GLUT1 mutants in *Xenopus* oocytes 155
Figure 5.15 Functional expression of GLUT1 half molecules in *Xenopus* oocytes 158
Figure 5.16 Western blot analysis of membranes prepared from *Xenopus* oocytes injected with GLUT1 mutant mRNA 159
Figure 6.1 Schematic representation of the baculovirus life cycle 165
Figure 6.2 Partial restriction enzyme map of transfer vector pAcYM1 170
Figure 6.3 Strategy used for the subcloning of GLUT1 mutants into the transfer vector pAcYM1 171
Figure 6.4 Restriction digests of recombinant transfer vectors containing wild-type and mutant GLUT1 cDNA fragments 173
Figure 6.5 Allelic replacement reaction between recombinant transfer vector and viral DNA 175
Figure 6.6 Western blot demonstrating expression of wild-type GLUT1 and the double-lysine mutant of GLUT1 in *Sf*9 cells 180
Figure 6.7 Analysis of purified antibody by ELISA 182

Figure 6.8	Western blot analysis with purified antibody	183
Figure 6.9	Structure of ATB-BMPA	184
Figure 6.10	D-glucose inhibition of ATB-BMPA photolabelling	186
Figure 6.11	Electrophoretic profile of ATB-BMPA-photolabelled insect cell membranes	188
Figure 7.1	Diagram of vector for stable transfection of CHO cells with GLUT1 cDNA	194
FIgure 7.2	Sugar uptake by wild-type and OE-CHO cells	197
Figure 7.3	Quantification of GLUT1 in CHO and OE-CHO cells	200
Figure 7.4	Electrophoretic profile of photoaffinity-labelled GLUT1 expressed in CHO cells	203
Figure 7.5	Subcellular location of GLUT1 within wild-type CHO cells	206
Figure 7.6	Subcellular location of GLUT1 within OE-CHO cells	207

LIST OF TABLES

Table 1.1	Properties of the mammalian passive glucose transporters	18
Table 3.1	Properties of transporters related to the mammalian GLUT family	65
Table 3.2	Predicted membrane spanning helix boundaries	83
Table 3.3	Predicted number of residues per turn for each putative transmembrane helix	90
Table 5.1	Optimum free Mg^{2+} ion concentration required for efficient PCR of primary products, plus predicted sizes	139

ABBREVIATIONS

AcNPV	*Autographa californica* nuclear polyhedrosis virus
ATB-BMPA	2-*N*-[4-(1-azi-2,2,2-trifluoroethyl)benzoyl]-1,3-bis-(D-mannos-4-yloxyl)-2-propylamine
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pairs
BSA	bovine serum albumin
$C_{12}E_8$	octaethylene glycol dodecyl ether
CD	circular dichroism
CHO	Chinese hamster ovary
Ci	curie (3.7×10^{10} Bq)
CIP	calf intestinal phosphatase
2DG	2-deoxy-D-glucose
DEPC	diethyl pyrocarbonate
dsDNA	double-stranded DNA
DMF	dimethyl formamide
DMSO	dimethyl sulphoxide
dpm	disintegrations per minute (60Bq)
DTT	dithiothreitol
ECL	enhanced chemiluminescence
ECV	extracellular virus particles
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
FTIR	Fourier transform infrared
GET	glucose, EDTA and Tris-HCl buffer
GMEM-S	Glasgow modified eagle's medium
HEPES	*N*-hydroxyethylpiperazine-*N'*-2-ethanesulphonic acid
IAPS-forskolin	3-iodo-4-azidophenethylamido-7-*O*-succinyldeacetyl-forskolin
IPTG	isopropyl β-D-thiogalactoside

kb	kilobase pairs
kDa	kilodaltons
KRH	Krebs Ringer HEPES
LB	Luria-Bertani broth
MOI	multiplicity of infection
MOPS	3-(*N*-morpholino)propanesulfonic acid
MSX	methionine sulphoximine
NBT	nitro blue tetrazolium
NHS-LC-Biotin	sulphosuccinimidyl-6-(biotinamido)hexanoate
NTP	(unspecified) nucleoside triphosphate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
pCMBS	*p*-chloromercuribenzenesulphonate
PCR	polymerase chain reaction
p.i.	post-infection
PMSF	phenylmethanesulphonyl fluoride
RC	reaction centre
SDS	sodium dodecyl sulphate
Sf	*Spodoptera frugiperda*
SGHMS	St. George's Hospital Medical School
TAE	Tris-acetate\EDTA buffer
TBE	Tris-borate\EDTA buffer
TBS	Tris-buffered saline
TE	Tris-EDTA buffer
TEMED	*N*,*N*,*N*',*N*'-tetramethylethylenediamine
Tris	tris(hydroxymethyl)-aminomethane
TTBS	Tris-buffered saline, 0.2% Tween-20

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

Membranes play a vital role in all living cells in that they control both the organisation and the function of many cellular processes. The plasma membrane defines the external boundary of the cell and regulates intra- and inter-cellular communication by controlling the passage of ions and solutes into and out of the cell. With respect to the internal environment of the cell, the intricate arrangement of membranes allows individual metabolic pathways to be separated into compartments which therefore enables stricter control. Internal membranes are also vitally important for ensuring that newly synthesised proteins reach their final intracellular or extracellular locations.

Glucose is one of the most abundant natural organic compounds and, as a consequence, represents a major source of metabolic energy for most mammalian cells. For example, the human brain is almost entirely dependent upon a sustained presence of glucose as the energy supply for oxidative metabolism (Lund-Andersen, 1979). However, glucose cannot enter cells by simple diffusion; the presence of carrier proteins is required to ensure the passage of the hydrophilic glucose molecules across the hydrophobic lipid bilayer of the membrane. Such proteins are termed the glucose transporters.

The process of solute transport across a cell membrane may be passive or active. Passive transport, often termed facilitative diffusion, requires no metabolic energy and is a type of diffusion in which an ion or molecule crossing the membrane moves down its electrochemical or concentration gradient with the aid of a transport protein. In contrast, active transport uses metabolic energy to move an ion or molecule against its electrochemical gradient. Such import or export of a small molecule against a concentration gradient may be achieved by coupling its flux to that of another molecule or ion, usually H^+ or Na^+, either in the same or opposite direction to that of the first molecule. These

processes are termed symport or antiport respectively. In mammals, active, sodium-linked symport systems for glucose occur in only a few tissues, such as the small intestine and kidney, where transepithelial transport against a concentration gradient is necessary (Hediger *et al.*, 1987). However, most mammalian cells take up glucose by passive facilitated diffusion (Elbrink and Bihler, 1975).

1.2 Facilitative glucose transporters

The purification and characterisation of the human erythrocyte glucose transporter (Kasahara and Hinkle, 1977, Baldwin *et al.*, 1982) provided the basis for the isolation of human and rat glucose transporter cDNA clones from HepG2 hepatoma cell and rat brain cDNA libraries, respectively (Mueckler *et al.*, 1985, Birnbaum *et al.*, 1986). These cDNA clones were subsequently used as probes to screen cDNA libraries from other mammalian tissues at low hybridisation stringency. With these procedures, five homologous facilitative glucose transporter isoforms have now been identified and characterised. Table 1.1 describes some of the properties of the facilitative glucose transporters but they will not be discussed in detail here because several extensive reviews of this area have been published recently (Gould and Bell, 1990, Lienhard *et al.*, 1992, Baldwin, 1993). Each glucose transporter (GLUT, the gene symbol for facilitative glucose transporter) isoform is numbered in the order of its cDNA discovery, i.e. GLUT1-GLUT7, the terminology of Fukumoto *et al.* (1989). GLUT6 represents a glucose transporter pseudogene-like sequence which is part of a mRNA that appears to be expressed in all human tissues (Kayano *et al.*, 1990). However, this sequence cannot encode a functional glucose transport protein due to the presence of multiple stop codons and frame shifts.

The facilitative glucose transporter isoforms display a marked tissue-specific pattern of expression and their kinetic and biochemical properties differ. This diversity of the transport proteins allows precise control of blood glucose concentration to be maintained in mammals over a range of physiological

Table 1.1 Properties of the mammalian passive glucose transporters.

Type	M_r (residues)	Tissue/cell expression	Chromosomal Location
GLUT1 erythrocyte/brain, HepG2	54,117 (492)	erythrocyte (human), blood-brain barrier, placenta, fetal tissues in general	1
GLUT2 liver	57,000 (524)	liver, pancreatic β-cell, kidney, small intestine	3
GLUT3 brain	53,933 (496)	brain (neurones)	12
GLUT4 adipocyte/muscle, insulin-regulatable	54,797 (509)	brown and white adipocytes, heart and skeletal muscle	17
GLUT5 small intestine	54,983 (501)	small intestine	1
GLUT7* hepatic microsomal	53,000 (528)	endoplasmic reticulum of hepatocytes	N.D.

* All isoform properties relate to the human glucose transporters, except for GLUT7 where the size of the rat protein is indicated.

conditions. The transporter isoforms, which vary in size from 492 to 528 residues, exhibit considerable similarity in their amino acid sequences with ≥40% of the residues being identical (Mueckler *et al.*, 1985, Fukumoto *et al.*, 1988, Kayano *et al.*, 1988, James *et al.*, 1989, Kayano *et al.*, 1990, Waddell *et al.*, 1992). This degree of sequence similarity suggests that the membrane topologies of these isoforms are essentially identical to that proposed for GLUT1 (Mueckler *et al.*, 1985) in which the protein spans the plasma membrane twelve times, with its *N*- and *C*-termini cytoplasmically oriented (see Section 1.3). The greatest sequence divergences are found in the cytoplasmic domains at the *N*- and *C*-termini, the hydrophilic central loop, and the extracellular loop between the first two putative transmembrane regions, implying that these regions may contribute to the unique features of these proteins including their intrinsic activities and subcellular localisation. With the exception of human GLUT5, which has been shown to transport fructose (Burant *et al.*, 1992), each of the transporter isoforms has been shown to be capable of transporting D-glucose following their expression in heterologous systems (discussed in Section 1.5). Furthermore, in each case glucose transport has been found to be inhibited by the fungal metabolite cytochalasin B. It is noteworthy that in addition to functioning as sugar carriers, the facilitative glucose transporters may play a role in transport of nicotinamide (Sofue *et al.*, 1992) and vitamin C (Vera *et al.*, 1993), and may also serve as membrane water channels (Fischbarg *et al.*, 1990).

1.2.1 GLUT1

GLUT1 is also known as the HepG2 or erythrocyte/brain type transporter, because the isolation of cDNAs encoding human and rat GLUT1 was achieved by screening human hepatoma carcinoma cell line (HepG2) and rat brain libraries using antibodies raised against the purified erythrocyte glucose transporter (Büchel *et al.*, 1980, Mueckler *et al.*, 1985). cDNAs encoding the transporter isoform have also been isolated from other species (Baldwin, 1993).

GLUT1 has a high degree of stereospecificity for pentose and hexose monosaccharides in the pyranose ring form (LeFevre, 1961, Barnett *et al.*, 1973). For example, the K_m for D-glucose of zero-trans uptake at 37°C is approximately 7 mM (Lowe and Walmsley, 1986), whereas for L-glucose the K_m is greater than 3 M (Carruthers, 1990). However, GLUT1 has a very low affinity for fructose (K_m= about 1.5 M, LeFevre and Marshall, 1958). The highest levels of polypeptide expression of GLUT1 are found in foetal tissues, including placenta and brain (Asano *et al.*, 1988, Santalucia *et al.*, 1992). In adults, the GLUT1 isoform is particularly abundant in cells that contribute to blood-tissue barriers (Takata *et al.*, 1990). Since low levels of GLUT1 mRNA or protein can be detected in most tissues, this transporter isoform may be responsible, at least in part, for constitutive glucose uptake. Interestingly, studies of the developmental regulation of GLUT1 mRNA in rat tissues have shown that the transporter is also highly expressed in tissues such as heart, liver, skeletal muscle and brown fat during foetal life, but its expression in these tissues diminishes rapidly after birth, implying the existence of a circulating factor responsible for the enhanced expression of GLUT1 during foetal life (Santalucia *et al.*, 1992). However, the nature of this signal involved in the developmental regulation of GLUT1 remains unknown.

1.2.2 GLUT2

The presence of GLUT2 in the liver, kidney and small intestine (Thorens *et al.*, 1988) suggests that it mediates the uptake and release of glucose by the liver, and that it participates in the transepithelial transport of absorbed and reabsorbed glucose by the small intestine and kidney, respectively. GLUT2 is also expressed in the insulin-producing β-cell of the endocrine pancreas which suggests that it may play a role in the regulation of glucose-stimulated insulin secretion. This possibility is supported by the finding that a decrease in the level of the GLUT2 isoform in the pancreatic β-cell precedes the development of type II non-insulin-dependent diabetes mellitus (Johnson *et al.*, 1990).

1.2.3 GLUT3

RNA blotting studies have shown that GLUT3 mRNA is found at various levels in most adult human tissues but is most abundant in brain, kidney and placenta (Kayano *et al.*, 1988). Similar studies in monkeys, rabbits, rats and mice have indicated that the GLUT3 mRNA has a more restricted tissue distribution in these species compared to humans, and that it is expressed at high levels only in the brain (Yano *et al.*, 1991, Maher *et al.*, 1992, Nagamatsu *et al.*, 1992). Furthermore, *in situ* hybridisation studies have demonstrated that GLUT3 mRNA is widely distributed in mouse brain and is present at the highest levels in the hippocampus, cerebellum and cerebral cortex (Nagamatsu *et al.*, 1992). The GLUT3 isoform is also expressed in primary cultured rat cerebellar neurones and neuronal cell lines (Maher and Harrison, 1991). It is therefore possible that the physiological role of GLUT3 is the transport of glucose into neuronal cells, whereas GLUT1 contributes to the uptake of glucose across the blood-brain barrier (Takata *et al.*, 1990). Further support for this hypothesis has been provided by Gould *et al.* (1991) who have shown that GLUT3 possesses a higher affinity for sugar than GLUT1 which would ensure efficient uptake of glucose by neuronal cells at the extracellular glucose concentrations which prevail in the brain. The latter are known to be lower than the glucose concentrations found at the blood side of the blood-brain barrier.

1.2.4 GLUT4

The existence of a novel insulin-regulatable glucose transporter was suggested by the poor cross-reactivity between antibodies specific for GLUT1 and rat adipocyte glucose transporter (Oka *et al.*, 1988). The insulin-regulatable glucose transporter, termed GLUT4 or muscle/fat type transporter, was first identified using a monoclonal antibody raised against a low-density microsomal fraction of rat adipocytes (James *et al.*, 1988). RNA blotting studies have shown that the highest levels of GLUT4 mRNA are found in insulin-responsive tissues such as brown and white fat, and cardiac and skeletal muscle (Birnbaum, 1989,

Charron *et al.*, 1989, Fukumoto *et al.*, 1989, James *et al.*, 1989, Kaestner *et al.*, 1989). The major mechanism by which insulin regulates transport in these tissues is by stimulating the translocation of glucose transporters from an intracellular membranous pool (the so-called low-density microsomal pool) to the plasma membrane of the cell (Cushman and Wardzala, 1980).

1.2.5 GLUT5

cDNA clones encoding the GLUT5 isoform, the small intestine type transporter, have been isolated from a human small intestine cDNA library by low stringency screening with GLUT1 cDNA (Kayano *et al.*, 1990). Immunohistochemical studies have shown that it is located mainly at the luminal surface of mature small intestinal enterocytes, where the active Na^+/glucose cotransporter (SGLT1) is also found (Davidson *et al.*, 1992). Thus, the precise role of this protein in the small intestine remains unclear and is, indeed, complicated by the results of a recent study based on the substrate specificity of GLUT5 which suggested that this isoform functions as a fructose transporter (Burant and Bell, 1992b). When expressed in *Xenopus* oocytes it mediates only very low levels of glucose uptake, but transports fructose with a high affinity (K_m= 6 mM). It is therefore possible that GLUT5 may subserve the role of a fructose transporter in tissues, such as intestine, kidney and sperm that are known to use fructose (Kayano *et al.*, 1990, Harris *et al.*, 1992, Burant *et al.*, 1992). Interestingly, Mantych *et al.* (1993) have very recently demonstrated that the GLUT5 isoform is also expressed in the endothelial cells of human brain microvasculature. The presence of the GLUT5 isoform in the brain, where fructose is not used as a substrate, leaves the physiological role of GLUT5 unclear.

1.2.6 GLUT7

Glucose release from the liver plays a crucial role in the maintenance of constant glucose concentrations in blood. The final step in glucose production

by both glycogenolysis and gluconeogenesis occurs within the lumen of the endoplasmic reticulum (ER), catalysed by the enzyme glucose 6-phosphatase (reviewed in Burchell and Waddell, 1991). Recently, the existence of a novel glucose carrier for the transport of glucose across the ER membrane has been demonstrated (Waddell *et al.*, 1991). cDNA clones encoding the liver microsomal glucose transporter, termed GLUT7, have been isolated from a rat liver cDNA library (Waddell *et al.*, 1992). Sequence analysis of the 528-residue protein shows that GLUT7 shares 68% identity with the sequence of rat GLUT2, which catalyses glucose flux across the hepatocyte plasma membrane. The major difference between the two transporter isoforms is the presence of an extra six residues at the *C*-terminus of GLUT7, which include a consensus motif for retention of membrane-spanning proteins in the ER (Waddell *et al.*, 1992). Expression of GLUT7 in COS 7 cells has shown that the protein is indeed a glucose transporter, and that it is localised to the ER membrane rather than the plasma membrane. The latter finding has been confirmed by Western blotting studies on subcellular membrane fractions of liver, which show that the transporter isoform is present in the microsomes (plasma-membrane free) and not in the plasma membrane (Waddell *et al.*, 1991).

1.3 Membrane topology and structure of GLUT1

Using hydropathic and secondary structure predictions as a guide, Mueckler *et al.* (1985) proposed a model for the two-dimensional arrangement of the GLUT1 protein in which the protein spans the plasma membrane twelve times in the form of α-helices of 21 amino acid residues each, accounting for about 50% of the 492 amino acid residues of the polypeptide (Figure 1.2). The *N*-terminus (residues 1-12), *C*-terminus (residues 451-492), and a very hydrophilic region (residues 207-271) in the centre of the protein are all predicted to lie on the cytoplasmic side of the membrane. Putative transmembrane domains 1 and 2 are connected by an extracellular segment of 33 amino acids (residues 34-66) and the remaining membrane-spanning domains by short regions of 7-14 amino acids. There are two potential *N*-linked glycosylation sites in GLUT1 at Asn_{45}

and Asn_{411}. Although, the latter site is predicted to be in transmembrane domain 11 and is therefore unlikely to be modified. In addition, the model predicts that several of the putative transmembrane domains (3, 5, 7, 8 and 11) may form amphipathic α-helices and contain abundant hydroxyl and amide side chains that could participate in glucose binding or line a transmembrane pore through which the sugar moves.

Several features of the model for the orientation of GLUT1 have been confirmed. The glycosylation of Asn_{45} has been shown by expressing GLUT1 *in vitro* in the presence of pancreatic microsomes (Mueckler and Lodish, 1986). This finding is in good agreement with proteolytic digestion studies of the human erythrocyte glucose transporter, which have localised the site of glycosylation to the amino terminal half of the protein (Cairns *et al.*, 1987, Davies *et al.*, 1990). Cleavage of the transport protein with trypsin yields two large membrane-bound fragments in addition to a number of small, water-soluble peptide fragments (Cairns *et al.*, 1984, Deziel and Rothstein, 1984). The *N*-terminal fragment (residues 1-212) migrates as a broad band of apparent M_r 23,000-42,000 on SDS-polyacrylamide gels, which is indicative of glycosylation. The other fragment (residues 270-456) containing the potential site of glycosylation at Asn_{411} is not glycosylated and migrates as a sharp band of apparent M_r 18,000 (Cairns *et al.*, 1984, 1987, Davies *et al.*, 1987, 1990). Direct evidence for the cytoplasmic orientation of the hydrophilic central and *C*-terminal regions of the GLUT1 has also been obtained by using vectorial proteolytic digestion and site-specific anti-peptide antibodies (Cairns *et al.*, 1987, Davies *et al.*, 1987, 1990, Haspel *et al.*, 1988b).

Biophysical studies of the human erythrocyte transporter provide direct evidence in support of the proposed model. On the basis of circular dichroism (CD) spectroscopy, Chin *et al.* (1987) have estimated the secondary structural composition of purified GLUT1 in reconstituted liposomes. The results show the transporter is composed of approximately 82% α-helices, 10% β-turns, and 8% random coil structure. The fact that the erythrocyte transporter is predominantly

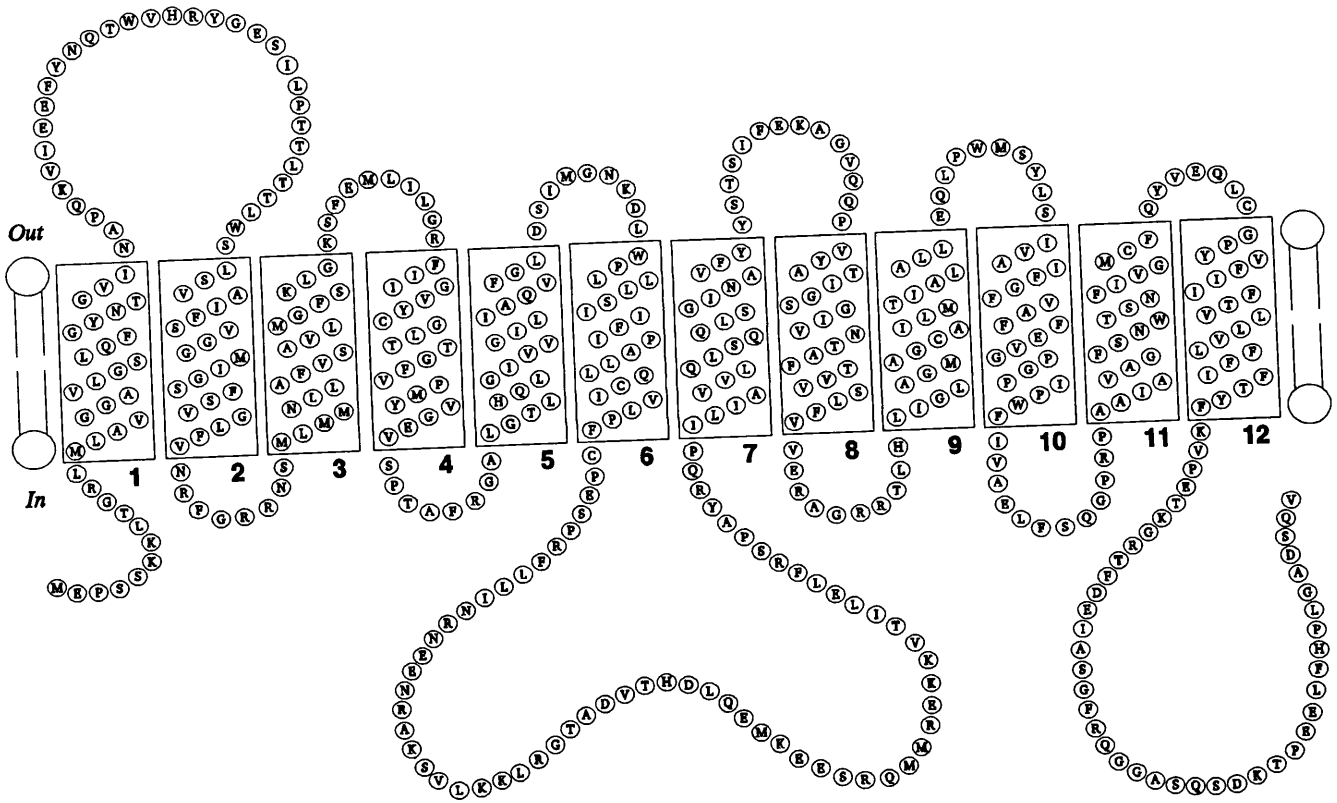


Figure 1.2 Current topological model of GLUT1.

composed of α-helical structure is supported by the Fourier transform infrared (FTIR) spectroscopic studies of (Alvarez *et al.*, 1987). However, the detection of putative β-strands in the protein in the latter study conflicts with the failure to detect such structures in the CD spectroscopic studies (Alvarez *et al.*, 1987, Chin *et al.*, 1987). The predominant α-helical structure may be a common feature of many membrane proteins with transport function. For example, CD spectroscopy of the lactose permease of *Escherichia coli*, a protein of similar molecular weight and related function to the glucose transporter, has shown that the permease is also largely composed of α-helices which account for about 85% of the total mass (Foster *et al.*, 1983). The presence of a significant amount of α-helical structure in the extramembranous domain of the erythrocyte transporter has also been suggested (Mueckler *et al.*, 1985, Chin *et al.*, 1986, 1987, Alvarez *et al.*, 1987). Direct evidence for this was obtained by FTIR spectroscopy of the transporter before and after trypsin digestion (Cairns *et al.*, 1987). The analysis showed that both the membrane-embedded domains and the central, hydrophilic and *C*-terminal domains of the protein contain α-helical structure.

Evidence obtained from deuterium and tritium exchange studies suggests the existence of an intraprotein aqueous channel which penetrates the membrane (Jung *et al.*, 1986, Alvarez *et al.*, 1987). The hydrogen exchange experiments reveal that a large proportion of the transport protein is readily accessible to solvent, and that about 20-25% of the peptide amide hydrogens appear to be free and in contact with water. Furthermore, the high deuterium exchange rate of the glucose transporter contrasts with the behaviour of other membrane proteins like rhodopsin or bacteriorhodopsin which exchange much more slowly (Osborne and Nabedryk-Viala, 1978, Downer *et al.*, 1986). Considering the fact that the glucose transporter is a transmembrane protein with most of its mass embedded in a hydrophobic environment (Mueckler *et al.*, 1985), the extent and rapidity of hydrogen exchange strongly support the existence of a water-filled channel in the transport protein. These findings are compatible with the proposed model of Mueckler *et al.* (1985) which predicted the presence of

amphipathic helices. Additional support for the presence of an aqueous channel is provided by the polarised FTIR spectroscopic studies of Chin *et al.* (1986). These studies suggest that the α-helices in the glucose transporter are orientated preferentially in a plane perpendicular to the membrane. It is likely that some of these helices, which are predicted to be amphipathic, participate in forming a hydrophilic channel across the lipid bilayer.

1.3.1 Oligomeric state

Despite many investigations, it remains unclear whether GLUT1 exists as oligomers and whether the functional properties of the transporter are determined by its monomeric or oligomeric structure. Several studies on the size of GLUT1, measured by irradiation inactivation of either D-glucose flux (Cuppoletti *et al.*, 1981) or cytochalasin B binding (Jung *et al.*, 1980, Jarvis *et al.*, 1986), and by freeze-fracture electron microscopy (Sogin and Hinkle, 1978), have yielded estimates of molecular size that are compatible with either a dimer (Jarvis *et al.*, 1986, Sogin and Hinkle, 1978) or a tetramer (Cuppoletti *et al.*, 1981, Jung *et al.*,1980). Recent studies by hydrodynamic techniques have also shown that the transporter exists as an oligomeric structure in a cholate solution (Hebert and Carruthers, 1991, 1992). In contrast, analyses of octyl glucoside-solubilised GLUT1 by high performance molecular-sieve chromatography have shown that the transporter exists as a monomer in this detergent immediately after ion-exchange chromatography (Lundahl *et al.*, 1991).

A number of kinetic studies have also suggested that glucose transporters may function as oligomers (Rampal *et al.*, 1986, Hebert and Carruthers, 1991, 1992). Additional evidence for this has come from the studies of Pessino *et al.* (1991) who have expressed chimeric GLUT1 molecules bearing the *C*-terminal region of GLUT4 in Chinese hamster ovary (CHO) cells. Immunoprecipitation of these recombinant transporters from detergent-solubilised membranes with anti-GLUT4 *C*-terminal peptide antibody also precipitated endogenous CHO cell

GLUT1, indicating that functional GLUT1 exists as oligomers. Interestingly, Burant and Bell (1992a) have recently studied the transport properties of oocytes co-expressing two glucose transporter isoforms (GLUT2 and GLUT3) with distinct kinetic parameters and found that these isoforms do not form heteromultimers with altered kinetic properties. They have also demonstrated that expression in oocytes of varying amounts of a functionally-inactive form of GLUT3 has no effect on the transport activity of co-expressed wild-type GLUT3. This indicates that when expressed in *Xenopus* oocytes, the monomer is a sufficient unit for functional activity of the glucose transporter.

1.4 Locations of substrate-binding site(s)

The availability of potent inhibitors that bind to the transporter more tightly than D-glucose itself has aided the identification of the substrate-binding site(s) of GLUT1. Greater detail concerning the locations of substrate binding-sites is found within Chapter 4, however, the significance of these molecules to studies of glucose transport warrants a brief description. Although the binding of transport inhibitors is normally reversible, some of them are photoactivable (or react covalently with the transporter upon UV illumination) and so can be used to label the binding sites irreversibly. Examples of such inhibitors are the fungal metabolite cytochalasin B (Carter-Su *et al.*, 1982), the diterpene forskolin and its derivatives (Wadzinski *et al.*, 1987) and a series of bis-mannose derivatives (Holman *et al.*, 1986). Kinetic studies have revealed that forskolin (Devés and Krupka, 1978) and cytochalasin B (Sergeant and Kim, 1985) bind to the transporter in competition with glucose solely at the cytoplasmic surface of the membrane. It is possible that these inhibitors bind to an inward-facing conformation of the substrate-binding site in a manner comparable with the binding of glucose (Joost *et al.*, 1988). In contrast, the bis-mannose derivatives are exofacial inhibitors (Holman *et al.*, 1986, Clark and Holman, 1990).

These studies suggest also that a structural separation of the internal and external binding sites exists, although both are believed to reside within the *C*-

terminal half of the protein. Chemical cleavage of the cytochalasin B-labelled protein at cysteine and tryptophan residues has led to the conclusion that the site of labelling lies within the region containing residues 389-412 (Holman and Rees, 1987). Because cytochalasin B competes with glucose for binding to the cytoplasmic face of the transporter, this region may represent part of the internal substrate binding site. In contrast, 2-*N*-[4-(1-azi-2,2,2-trifluoroethyl)benzoyl]-1,3-bis-(D-mannos-4-yloxyl)-2-propylamine(ATB-BMPA) a non-transported, membrane-impermeant sugar analogue (Clark and Holman, 1990) has been shown to label the transporter within putative transmembrane helix 8, between residues Ala_{301} and Arg_{330} (Davies, 1991). Because this analogue inhibits transport only when present at the extracellular surface of the membrane, helix 8 may contribute to the external substrate binding site of the transporter.

Extensive digestion of the membrane-bound transporter with trypsin destroys the ability to transport sugar, yet the cytochalasin B is still able to bind although with a reduced affinity (Baldwin *et al.*, 1980). Moreover, this binding can be inhibited by D-glucose and since trypsin digestion removes much of the cytoplasmic regions of the protein, it has been concluded that the site(s) of glucose and inhibitor binding must be located within the membrane-embedded portions of the transporter (Cairns *et al.*, 1987). The trypsinised transporter is still also able to bind ATB-BMPA, although the affinity is reduced by 12-fold (Clark and Holman, 1990), probably because the proteolysed transporter exists predominantly in an inward-facing conformation.

1.5 Heterologous expression systems

To understand the mechanism of action of a particular protein at the molecular level requires a knowledge of its three-dimensional structure at high resolution. As a consequence, the molecular details of many membrane-mediated events such as insulin-dependent glucose transport in diabetes and chloride transport in cystic fibrosis, are largely unknown due to a lack of structural information.

One of the limiting factors of structure-function studies of hydrophobic membrane proteins is the difficulty of obtaining sufficient quantities of the chosen protein in its native state. Such studies have progressed slowly since most of the biophysical and biochemical techniques used have been developed for water soluble proteins. However, structure determinations of bacterial porin, bacterial photosynthetic reaction centres and bacteriorhodopsin have been achieved mainly because all these proteins are highly abundant in their native biological membranes, often forming semi-crystalline arrays, enabling 10-100 mg quantities to be obtained with relative ease. They are also extremely robust as they exhibit long term stability in many detergents. In contrast, eukaryotic integral membrane proteins have not proved either as easy to purify or maintain in an active form in sufficiently large amounts for biophysical characterisation. Consequently, the application of recombinant DNA methodologies for obtaining high-level expression of membrane proteins in heterologous systems has been widespread recently.

Expression of foreign genes in a prokaryotic host system has been attempted most commonly in *E. coli* principally due to the ease of growing bacteria on a large scale. High level expression of many prokaryotic and eukaryotic proteins has been achieved and the wealth of genetic information on *E. coli* also renders it a good host for genetic manipulation experiments. Despite the recent development of a wide variety of prokaryotic expression vectors, there are still numerous difficulties associated with the production of protein from eukaryotic genes cloned into *E. coli*. For example, bacteria do not possess the enzymatic machinery required to post-translationally process eukaryotic proteins and thus glycosylation, proteolysis, phosphorylation, accurate disulphide bond formation and oligomerisation, which are often necessary for correct function, do not occur properly in prokaryotic systems. The expression of the GLUT1 and GLUT2 isoforms of the mammalian passive glucose transporter family in *E. coli* was reported several years ago (Sarkar *et al.*, 1988, Thorens *et al.*, 1988). However, use of this expression system yielded only very small amounts of functional protein, and these results appear to have not been reproducible by

other research groups.

As a consequence of these problems, considerable efforts have been made to develop systems to express eukaryotic proteins in a eukaryotic environment. Examples of such expression systems are yeast (Gunge, 1983), insect (Summers and Smith, 1987), and cultured mammalian cells (Asano *et al.*, 1993), as well as in whole organisms by the generation of transgenics (Palmiter *et al.*, 1982). There are, however, additional problems attached to heterologous expression systems. For example, the intracellular environment is not always favourable for the correct folding of membrane proteins. In addition, the expressed proteins may exert toxic effects since they may disrupt cellular activities by altering the physical nature or permeability properties of the bilayer. Recently, however, successful expression has been reported for several mammalian glucose transporter isoforms in a variety of eukaryotic expression systems. These systems have included *Xenopus* oocytes [GLUT1-5 (Birnbaum, 1989, Gould and Lienhard, 1989, Keller *et al.*, 1989, Vera and Rosen, 1989, Permutt *et al.*, 1989, Kayano *et al.*, 1990, Gould *et al.*, 1991)], mammalian cells [GLUT1 (Gould *et al.*, 1989, Asano *et al.*, 1989, Harrison *et al.*, 1990), GLUT3 (Asano *et al.*, 1992)], insect cells [GLUT1 (Yi *et al.*, 1992)] and transgenic mice [GLUT4 (Liu *et al.*, 1992)]. The *Xenopus* oocyte system is particularly well suited for the study of facilitative glucose transporters, since uninjected oocytes exhibit only low levels of endogenous glucose transport activity. The transporter proteins expressed in this system show similar kinetic properties to those of their native counterparts and can be distinguished on the basis of their affinity for nonmetabolised glucose analogues, such as 2-deoxy-D-glucose and/or 3-*0*-methyl-D-glucose. Moreover, sugar transport in this system is inhibited by cytochalasin B. However, the oocyte system cannot be used to express heterologous DNAs on a scale large enough for biochemical and biophysical analyses of the expressed proteins.

For detailed studies of structure and function relationships in these proteins, and in particular for the investigation of their structures by crystallisation, high

levels of expression are a prerequisite. The ability to introduce DNA into cultured cells has provided a powerful means for studying the function and control of mammalian genes. Transient transfection of cell lines is suitable for functional studies that require small amounts of protein and cells are usually harvested 48 to 72 hours post-transfection for analysis of the expressed product. The optimal time interval depends upon the cell type, the doubling time of the cells and the specific characteristics of expression for the transfected gene. COS cells (an African green monkey kidney cell line, CV-1, transformed with an origin-defective SV40 virus) have proved popular for transient transfection studies because of their capacity to replicate circular plasmids containing an SV40 origin of replication to very high copy number. Although this eventually leads to cell death, COS cells have the advantage of being capable of generating micrograms of recombinant protein within a few days.

However, there are limitations of the transient transfection procedure. For example, from the studies of Schürmann *et al.* (1993) a 3-5 fold increase was noted in the glucose transporter immunoreactivity of plasma membranes and in the transport activity reconstituted from these membranes. The glucose transport activity in intact cells, however, was increased by less than 2-fold. As a consequence, the differences in activity of transporters and mutants can only be assayed via the reconstitution of transport activity from isolated membranes. These authors also noticed that the differences in apparent molecular weight of constitutive and expressed GLUT1 suggests a different mode of glycosylation. That is, when subjected to low glucose conditions, an immunoreactive band of apparent M_r 45,000 was detected which was possibly a glucose transporter isoform possessing incomplete glycosylation. Thus, the transfection procedure appears to exhaust completely the capacity of the cells to glycosylate newly synthesised transporter.

In contrast, the goal of stable, long-term transfection is to isolate and propagate individual clones containing transfected DNA. Therefore, it is necessary to distinguish and select for those cells which have taken up exogenous DNA from

the bulk of non-transfected cells. This screening can be accomplished by drug selection when an appropriate drug resistance marker is included in the transfected DNA. Alternatively, morphological transformation can be used as a selectable trait in certain cases. Typically, cells are maintained in non-selective medium for 1 to 2 days post-transfection, then trypsinised and replated in selective medium containing the drug. The use of the selective medium is continued for several weeks, with frequent changes of medium to eliminate the non-viable cells and debris, until distinct colonies can be visualised. Individual colonies are then trypsinised and propagated in the presence of selective medium. Although several different drug selection markers are commonly used for long-term transfection studies, an alternative strategy is to use a vector carrying an essential gene that is defective in a given cell line. For example, CHO cells deficient in expression of the dihydrofolate reductase (DHFR) gene survive only in the presence of added nucleosides. However, these cells, when stably transfected with DNA expressing the DHFR gene, will synthesise the required nucleosides (Stark and Wahl, 1984).

1.6 Site-directed mutagenesis

One of the main objectives of expressing membrane proteins in mammalian cells is to determine the functional role of a protein by analysing the effects of its expression. In addition, it is possible to elucidate a functional role of specific domains or amino acid residues in the protein by examining the outcome of introducing mutations into certain regions of the protein.

In the absence of a three-dimensional structure for any sugar transport protein, many experimenters have applied the technique of site-directed mutagenesis in order to obtain information governing structural and/or functional roles of individual amino acids or domains of integral proteins. Several transporter proteins have been used in mutagenesis strategies, the progress of which have been extensively reviewed, for example the lactose permease of *E.coli* (LacY)

(Kaback, 1992a), GLUT1 (Baldwin, 1993), the lactose-H^+ symporter (LacS) of *Streptococcus thermophilus* (Poolman and Konings, 1993) and the MelB protein (Leblanc *et al.*, 1993). Mutants of these proteins have provided insights into the structural features, translocation events and the binding of substrates and inhibitors have been determined. Of all the transport proteins subjected to mutagenesis studies, the LacY protein has proven to be the most popular as over 300 different mutants have now been isolated and characterised (Kaback *et al.*, 1994).

Site-directed mutagenesis has been applied to structural and functional studies of GLUT1. For example (Mori *et al.*, 1994) have shown that GLUT1 possessing a Y293I mutation does not exhibit either an altered affinity for D-glucose or an appreciable change in affinity for the exofacial ligands ATB-BMPA and 4,6-O-ethylidene-D-glucose, but that it does exhibit a dramatic 300-fold decrease in affinity for cytochalasin B. As a consequence, this mutation is envisaged to lock the transporter in an outward facing conformation which thereby might implicate Tyr_{293} in closing the exofacial site around C4 and C6 of D-glucose in the transport process. Another study that has utilised CHO cells for the investigation of conformational changes during glucose transport has involved the mutation of a proline residue at position 385 (Tamori *et al.*, 1994). The substitution of Pro_{385} by isoleucine resulted in a marked decrease in transport activity together with a loss of labelling by ATB-BMPA, but no effect upon the binding of cytochalasin B was detected. This finding was suggestive of an inability to adopt the outward facing conformation.

The utility of expression studies also extends to the analysis of particular domains of proteins. For example, Baldwin and colleagues have suggested that the *C*-terminus of GLUT1 is not directly involved in substrate or inhibitor binding, but is essential for glucose transport itself (Cairns *et al.*, 1987). Indeed, Oka *et al.* (1990) have exploited the techniques of site-directed mutagenesis to confirm this suggestion by showing that deletion of most of the *C*-terminal domain (37 out of 42 amino acids) of GLUT1 abolishes sugar transport activity,

possibly by locking the glucose binding site into an inward-facing form. Truncation of the protein in this way does not, however, affect the ability of the protein to be photoaffinity labelled with cytochalasin B. These findings imply that although the *C*-terminal domain probably plays an essential role in the conformational changes that accompany sugar transport, it is not directly required for ligand binding. Additional support for this conclusion comes from recent studies of Katagiri *et al.* (1992). These authors expressed in CHO cells a mutant GLUT1 whose *C*-terminal domain had been replaced with the corresponding domain of GLUT2, a transporter isoform that has a higher K_m for glucose transport and a much lower affinity for cytochalasin B than GLUT1. The mutant was found to have an affinity for cytochalasin B similar to that of GLUT1, but had a higher K_m value for glucose transport than that of GLUT1. These results indicate that the cytoplasmic *C*-terminal domain of GLUT1 may play an important role in determination of the affinity for glucose. It is noteworthy that the *C*-terminal region is most diverse in amino acid sequence and size among the GLUT family.

It has been also suggested that putative transmembrane helices 10 and 11, which include two tryptophan residues (Trp_{388} and Trp_{412}) possibly involved in ligand binding, play an important role in the transport activity of GLUT1 (Cairns *et al.*, 1987, Holman and Rees, 1987). The substitution of Leu for Trp_{412}, which is located in helix 11 and is a highly conserved residue among the GLUT family, dramatically decreases the transport activity of the GLUT1 but the effect on cytochalasin B binding remains unclear (Katagiri *et al.*, 1991, Garcia *et al.*, 1992). In contrast, mutation of the helix 10 residue Trp_{388} to Leu results in a minor reduction in intrinsic activity of the transporter but significantly decreases the affinity of the transporter for cytochalasin B. In addition, substitution of the highly conserved residue Asn_{415}, located in helix 11 of GLUT1, by Asp gives similar results to those obtained in the Trp_{412} mutant (Katagiri *et al.*, 1991). Furthermore, Hashiramoto *et al.* (1992) have shown that transmembrane helix 7 constitutes part of the outward-facing binding site of GLUT1. They mutated residues Gln_{282} (in helix 7) to Leu and Asn_{288} (in helix 7) and Asn_{317} (in helix 8)

to Ile, respectively. The results showed that mutations at Asn_{288} and Asn_{317} had little effect upon transport activity or upon labelling by the exofacial ligand ATB-BMPA or the endofacial ligand cytochalasin B. However, substitution at Gln_{282} strongly decreased the affinity for exofacial ligands such as ATB-BMPA and 4,6-0-ethylidene-D-glucose, while having little effect upon transport activity or cytochalasin B binding. Consequently, through the prudent choice of potential sites for mutagenesis, it is possible to gain a substantial amount of information regarding the structure and/or function of a protein for which the crystal structure is unknown.

1.7 Aims of the study

The aim of the research described in this study is part of a continuing interest towards the elucidation of the structures and molecular mechanisms of proteins that effect the transport of sugars into eukaryotic cells. However, there is a great deal of evidence that suggests a common ancestor for the proteins of a variety of organisms which are involved in the transport of sugars. As a consequence, it is probable that the structure determination, and hence mechanism of action, of one transport protein such as GLUT1 will reveal the structural and mechanistic features inherent to transport proteins from many species.

In order to understand the mechanism of glucose transport at the molecular level, a high resolution structure of GLUT1 is required. In the absence of such information, however, the central aim of this project was to obtain structural and functional data about GLUT1 from a variety of methodologies. An initial objective was to analyse a multiple sequence alignment of the sugar transporter family, with a view to obtaining consensus information regarding the secondary structural elements of GLUT1 and thus refine the current two-dimensional model. It was intended that certain features of this two-dimensional model, such as the location of lysine residues on the exofacial face of the transporter would be investigated using amino group-specific membrane-impermeant probes.

In addition, it was hoped that by extending the theoretical analyses, a three-dimensional model of GLUT1 would be assembled for the purpose of rationalising the design of mutagenesis experiments to probe the structure/function relationship of the transporter. In particular, the intention was to devise a strategy for mutagenesis that would possess the capacity to generate information regarding the membrane assembly of the transporter, and also provide hard evidence regarding the precise locations of transport inhibitor binding-sites. To this end, it would also prove necessary to assess the suitability of various expression systems for the purpose of mutant analysis.

CHAPTER 2. MATERIALS AND METHODS

2.1 Materials and suppliers

Analytical grade chemicals were used in this study, and all solutions were prepared using deionised water. Materials were obtained from the following suppliers:-

General laboratory reagents and equipment

British Drug House Chemicals Limited, Poole, Dorset
Bethesda Research Laboratories, Cambridge
Fisons Scientific Apparatus Laboratory Suppliers, Loughborough, Leics.
May and Baker Ltd., Dagenham, Kent
Sigma Chemicals Co., Poole, Dorset
Whatman International Ltd., Maidstone, Kent

Culture media

Difco Labs., Detroit, Michigan, USA
Oxoid Ltd., Basingstoke, Hampshire
Gibco BRL (Life Technologies), Paisley, Renfrewshire

Radiolabelled compounds

Amersham Radiochemicals Ltd., Amersham, Buckinghamshire
NEN Research Products, Du Pont (UK) Ltd., Stevenage, Hertfordshire

Recombinant DNA technology products

Amersham International plc., Amersham, Buckinghamshire
Boehringer-Mannheim, Bell Lane, Lewis, East Sussex
Promega UK, Southampton
Pharmingen (Cambridge BioScience) Cambridge

2.2 Techniques for protein chemistry

2.2.1 Purification of GLUT1 from erythrocyte membranes

All operations were carried out on ice or at 4°C. Eythrocytes were washed with ice-cold phosphate-buffered saline (PBS, 5 mM sodium phosphate, 150 mM sodium chloride, pH 8.0) by centrifugation at 4,500 x *g* for 10 minutes and aspiration of the supernatant and layer of white cells on top of the pellet. This was repeated until the supernatant was clear and all plasma components removed. Ghost membranes were prepared by lysis of erythrocytes in ice-cold lysis buffer (5P8) consisting of 5 mM NaH_2PO4, 1 mM ethylenediaminetetra-acetic acid (EDTA), pH 8.0, 0.11 mM phenylmethanesulphonyl fluoride (PMSF). The erythrocyte membranes were harvested and concentrated with repeated washing steps by centrifugation at 11,500 x *g* for 20 minutes. The ghosts were homogenised by three passes of a hand-held homogeniser and assayed for protein (Section 2.2.2).

Erythrocyte membranes were stripped of their peripheral proteins by treatment under alkali conditions. Ghosts (4 mg/ml) were mixed with 15.4 mM sodium hydroxide, 2 mM EDTA, 0.2 mM dithiotreitol (DTT) which had been purged with nitrogen for 5 minutes immediately before use. After 10 minutes in the resulting pH 12 conditions, the membranes were collected by centrifugation at 18,000 x *g* for 20 minutes. The supernatant was discarded and the pellet washed twice with 50 mM Tris-HCl, pH 6.8 and resuspended in 10 ml of buffer. After homogenisation by three passes of a hand-held homogeniser, the homogeneous preparation was assayed for protein (Section 2.2.2).

Alkali-stripped membranes were solubilised with octyl-β-D-glucopyranoside such that final concentrations were 2 mg/ml protein and 1.35% (w/v) octyl-β-D-glucopyranoside in 46.5 mM Tris-HCl, 2 mM DTT, pH 7.4. After shaking on ice for 20 minutes, the non-solubilised material was removed by centrifugation at 45,000 x *g* for 1 hour at 4°C. The supernatant was removed immediately to ice

and made to 25 mM in NaCl in readiness for anion exchange chromatography on DEAE-cellulose. The anion exchange matrix (DE-52, Whatman) was equilibrated with 47.5 mM Tris-HCl, 1% octyl-β-D-glucopyranoside, 2 mM DTT, 25 mM NaCl, pH 7.4. Fractions were collected on ice and assayed for protein by their absorbance at 280nm, and the peak fractions pooled. After the addition of NaCl and EDTA to final concentrations of 100 mM and 1 mM, respectively, the resulting preparation was reconstituted by dialysis against 50 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA, pH 7.4 for 48 hours using four changes of buffer (2l). The reconstituted protein was then assayed for protein (Section 2.2.2) and functionality by its ability to bind cytochalasin B (Section 2.2.3).

2.2.2 Protein determination

Protein concentrations were determined by the method of Lowry *et al.* (1951). This assay relied on the formation of a protein-copper complex (Biuret reaction) and the reduction of Folin reagent by the tyrosine and tryptophan residues of the protein.

The assay required a standard curve, which was established with duplicate samples of 0, 12.5, 25, 37.5 and 50 μg of bovine serum albumin (BSA). Solutions used in the assay were as follows; 1) Solution A: 2% (w/v) sodium carbonate in 0.1 M NaOH, 2) Solution B: 1% (w/v) copper sulphate (pentahydrate) in water, 3) Solution B*: 2% (w/v) sodium/potassium tartrate in water, 4) Solution C: 0.4 ml of solution B and 0.4 ml of solution B* mixed with 39 ml of solution A and 1 ml of 20% SDS, and 5) Solution D: 2 ml of Folin's reagent plus 2 ml of water.

Protein standards (0-50 μg of BSA) and membrane samples (5-20 μl) were made up to 0.2 ml with distilled water, mixed with 1 ml of solution C, vortexed, and then incubated for 20 minutes at room temperature. Following incubation, 0.1 ml of solution D was added to each sample with immediate vortexing. After

incubating for 30 minutes at room temperature, the absorbances of the samples at 750 nm were measured using the solution of standard (0 µg of BSA) as blank. The protein concentration of the membrane samples was then determined from comparison with the standard curve.

2.2.3 Cytochalasin B ligand binding assay

The capacity of a glucose transporter preparation to bind cytochalasin B is a good indication of the protein's activity and was determined by equilibrium dialysis using the methodology of Zoccoli *et al.* (1978). The apparatus used consisted of chambers of 50 µl drilled in perspex, pairs of which are separated by dialysis membrane with a cut-off of 12,000 daltons (previously boiled once in 20 mM Na_2CO_3, 1 mM EDTA and three times in water to remove heavy metal ions). When clamped together, one of each pair of chambers was filled with 40 µl of 8 x 10^{-8} M [^{3}H]-cytochalasin B and the other filled with the glucose transporter sample. The [^{3}H]-cytochalasin B (New England Nuclear) was kept as a stock of 8 x 10^{-6} M in ethanol and was diluted before use into the same buffer as that containing the protein sample. Each sample was assayed in triplicate and incubated on a shaker for 18 hours at room temperature, with the tops of the chambers sealed to prevent evaporation. To correct for non-specific binding to membrane lipids, samples previously boiled at 100°C for 5 minutes were assayed in parallel, again in triplicate. Samples (25 µl) were then removed from each chamber, mixed with 2 ml of scintillation fluid and assayed for radioactivity by liquid scintillation counting using a Beckman LS5000 liquid scintillation counter. Ratios of bound to free cytochalasin B were calculated in the following manner:-

The radioactivity in the chamber containing the protein sample represented bound plus free [^{3}H]-cytochalasin B, whereas the radioactivity present in the chamber containing no protein represented free [^{3}H]-cytochalasin B. Thus the specific bound/free value was obtained by subtracting the non-specific bound/free value from the total bound/free value. Finally, by dividing the specific

bound/free value by the glucose transporter concentration, the binding activity of the preparation was determined.

2.2.4 SDS-polyacrylamide gel electrophoresis

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) is a powerful method for the separation of proteins according to size and was carried out using a discontinuous buffer system essentially as described by Laemmli (1970). Protein samples were routinely run on 10 or 12% polyacrylamide (acrylamide:bis-acrylamide = 37:1, w/w) slab gels. The slab gel comprised a 2 cm stacking gel of high porosity and a 10 cm separating gel of low porosity. The latter contained 10 or 12% polyacrylamide in 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.016% (v/v) TEMED. The stacking gel contained 3% polyacrylamide in 125 mM Tris-HCl (pH 6.8), 0.1% SDS and was polymerised by addition of 0.1% ammonium persulphate and 0.05% TEMED. Prior to loading, the proteins were solubilised in a loading buffer [40 mM Tris-HCl, pH 6.8, 0.8 mM EDTA, 0.8% SDS, 4 mM dithiothreitol DTT, 10% (v/v) glycerol, and 0.12% (w/v) pyronin Y]. Between 5 and 50μg of each of the solubilised samples was then loaded into each track of a 1.5 or 3 mm-thick gel. Low molecular weight range markers (M_r 14,400-97,000, Bio-Rad) were run alongside the sample tracks. Coomassie blue-prestained molecular markers (M_r 18,500-106,000) from Bio-Rad were also used when the samples were to be blotted onto nitrocellulose or nylon membranes. Electrophoresis was carried out using a Bio-Rad protein MK I electrophoresis cell and Pharmacia EPS 500/400 power supply. A constant current of 30 mA was applied for two 1.5 mm thick gels during migration of the tracking dye through the stacking gel, and then the current was increased to 60 mA during passage of the proteins through the separating gel. These values were doubled when 3 mm gels were employed. The gels were run until the pyronin Y marker had migrated about 9 cm from the top of the separating gel. The gel running buffer used was 25 mM Tris, 190 mM glycine and 0.1% SDS, pH 8.3.

Following electrophoresis, the gels were either stained with coomassie blue or subjected to electrotransfer for Western blotting. For the coomassie staining, the gels were fixed overnight in 10% (v/v) acetic acid, 25% (v/v) isopropanol, and then soaked for 8 hours in staining solution 1 [10% acetic acid, 25% isopropanol, 0.025% (w/v) coomassie blue R 250]. The gels were then stained for a further 16 hours in staining solution 2 (10% acetic acid, 10% isopropanol, 0.0025% (w/v) coomassie blue R250) followed by destaining in 10% acetic acid.

2.2.5 Immunoblot (Western blot) analysis

The immunoblot analysis of proteins comprises four main stages:-

a) Separation of the protein samples by SDS-PAGE (Section 2.2.1)
b) Electrophoretic transfer of the separated proteins from the gel to a membrane
c) Immunoreaction of primary antibody with the proteins bound to the membrane
d) Detection of the specifically bound primary antibody

2.2.5.1 Electrotransfer

Following SDS-PAGE, the proteins were electrophoretically transferred from the gels onto nitrocellulose membrane essentially according to the methods described by Towbin *et al.* (1979). After removal of the stacking gels, the separating gels were equilibrated in transfer buffer [39 mM glycine, 48 mM Tris, 0.0375% SDS, 20% (v/v) methanol] at room temperature for 20 min with gentle shaking. A piece of the membrane, pre-wet with transfer buffer, was placed on 3 stacked filter papers, also soaked in the buffer. The gel was then laid on top of the membrane and sandwiched by another 3 wet filter papers. The blotting apparatus (LKB Multiphor II semi-dry blotter) was run at a constant current of 1.6 mA per cm^2 of gel for 60 to 90 min, using a LKB Macrodrive I power supply. Following electrotransfer, the efficiency of the blotting was checked by staining

the gel, as described in section 2.2.4. To visualise the molecular weight markers, the piece of the membrane containing the markers was briefly stained in 0.1% (w/v) amido black, 25% isopropanol, 10% acetic acid and then destained in 10% acetic acid.

2.2.5.2 Immunostaining

After electrotransfer, the blot was washed in Tris-buffered saline (TBS, 20 mM Tris-HCl, 500 mM NaCl, pH 7.5) for 10 minutes. The membrane was then placed in a 150 ml Sterilin bottle containing blocking buffer [5% (w/v) dried skimmed milk powder in TTBS (TBS containing 0.2% Tween-20)] and incubated for 2 hours on a roller mixer to block non-specific binding sites on the membrane. The nitrocellulose was washed twice for 5 minutes with 100 ml TTBS with rolling and then incubated overnight at room temperature with 15 to 20 µg of relevant affinity-purified antibody or 20 µl of antiserum in 10 ml of antibody buffer (1% dried milk powder in TTBS) with rolling. Following incubation, the membrane was washed three times for 15 minutes in TTBS and immunoreactive bands were then detected in one of the following ways:-

1) *Alkaline phosphatase method:* The membrane was transferred into a 150 ml roller bottle (Sterilin) containing 10 ml antibody buffer and 3.3 µl (1:3,000 dilution) of an alkaline phosphatase conjugated goat anti-rabbit IgG or goat anti-mouse IgG (Bio-Rad) and incubated for 1 to 2 hours at room temperature with rolling. The nitrocellulose was washed three times for 15 minutes in TTBS and twice for 10 minutes in TBS to remove Tween-20, which might otherwise have formed a precipitate with the colour development reagents. The membrane was then incubated in colour development solution [0.37 mM nitro blue tetrazolium (NBT), 0.35 mM 5-bromo-4-chloro-3-indolyl phosphate (BCIP) in 100 mM sodium bicarbonate, 1 mM $MgCl_2$, pH 9.8) until strong purple bands were present. The reaction was stopped by washing the membrane several times in distilled water.

2) *Enhanced Chemiluminescence (ECL) method:* Bound primary antibody was detected according to the protocol supplied with the ECL detection Kit (Amersham). The nitrocellulose was incubated for 1 hour at room temperature on the roller mixer with 1 µl of donkey anti-rabbit IgG peroxidase-conjugate in 10 ml antibody buffer (1:10,000 dilution). The membrane was washed three times for 30 minutes in TTBS and then incubated for 1 minute with a mixture of equal volumes of detection solution 1 and 2 (0.125 ml/cm^2). The membrane was blot-dried with paper towelling, mounted on a 3MM filter paper and covered with Saran wrap, prior to exposure to a sheet of X-ray film.

2.2.6 Preparation and use of an immunoaffinity column to purify antibodies raised aginst residues 460-477 of human GLUT1

Antiserum from a rabbit immunised with the synthetic peptide corresponding to residues 460-477 was available in the laboratory (from the work of A. Davies) and it was decided to affinity purify antibodies to this peptide using a method involving the attachment of cysteine-containing peptides to iodoacetyl-agarose (Sulfolink™ from Pierce).

A 1 x 20cm Sephadex G10 (exclusion limit 700) column was prepared to purify the peptide corresponding to residues 460-477 of GLUT1. The Sephadex G10 was incubated overnight in 50 mM Tris-HCl, 5 mM EDTA, pH 8.5 and then used to pour a 1 x 20cm column. The column was equilibrated at 4°C by pumping through several column volumes of the de-gassed buffer, at a flow rate of 24 ml/hour. The peptide was prepared for chromatography by dissolution of 3 mg synthetic peptide in 500 µl of 50 mM Tris-HCl, 5 mM EDTA, pH 8.5 in a ground-glass stoppered tube. To ensure full reduction of the peptide -SH groups, 25 µl of 1M DTT (final concentration 47.6 mM) was added to the tube which was then flushed with nitrogen, sealed and incubated for 1 hour at room temperature. The peptide solution was then pumped onto the column in the cold room at a flow rate of 24 ml/hour and eluted with the Tris buffer to separate the reduced peptide from the excess DTT. Fractions (1 ml) were

collected and monitored spectrophotometrically at 230nm, with the peak void-volume fractions containing the peptide (3-4 ml) being pooled.

The Sulfolink™ coupling gel (3 ml) was washed at room temperature in a sinter funnel with at least 6 volumes of 50 mM Tris-HCl, 5 mM EDTA, pH 8.5 to remove the storage buffer (10 mM EDTA, 0.05% sodium azide, 50% glycerol). It was then transfered to a 10 ml screw-capped glass tube, centrifuged briefly and the supernatant removed with a pasteur pipette. The purified peptide solution from the G10 column was then added to the Sulfolink gel and incubated at room temperature, with rotation, for 25 minutes in the dark and then left to stand for a further 30 minutes. The mixture was centrifuged as briefly as possible and the supernatant removed with a pasteur pipette. Excess peptide was washed out of the gel by five repeated washing steps with 50 mM Tris-HCl, 5 mM EDTA, pH 8.5. In order to block excess iodoacetyl groups on the gel, 5 ml 50 mM cysteine in 50 mM Tris-HCl, 5 mM EDTA, pH 8.5 was added and incubated at room temperature, slowly rotating the tube, for 25 minutes in the dark, followed by a further 30 minute incubation without rotation. The non-covalently bound peptide was then removed by transferring the gel back to a sinter funnel and washing with 50 ml 1M NaCl. The gel was equilibrated in 10 mM sodium phosphate, 145 mM NaCl, pH 7.2 (PBS) by washing with 50 ml of this buffer then poured into a 5 ml econo-column and washed with several column volumes of PBS at a flow rate of 10 ml/hour.

Serum (8 ml) obtained from a rabbit immunised with the synthetic peptide was then loaded onto the PBS-equilibrated peptide column at a flow rate of 10 ml per hour. Non-specifically bound IgG was eluted from the column with 10 mM sodium phosphate, 800 mM NaCl, pH 7.2, until A_{280} values dropped to near zero. Finally, the antibody remaining on the column was eluted with 0.2 M glycine-HCl, pH 2.4. The fractions possessing the highest A_{280} values were pooled, immediately neutralised by the addition of 2M Tris, then dialysed against PBS. The purified IgG was then assayed by enzyme-linked immunosorbent assay (ELISA) using peptide-coated plates (Section 2.2.7).

2.2.7 Assessment of purified IgG by ELISA

A 96-well plate (GIBCO Life Technologies) was coated with either purified glucose transporter or the peptide corresponding to residues 460-477 of GLUT1. Purified glucose transporter (400 ng in 50 mM sodium carbonate buffer, pH 9.6) was placed in each well and left at room temperature overnight in the dark. When the original peptide was used as the antigen, a 1 mg/ml solution was prepared in dimethyl sulphoxide (DMSO), diluted to 0.25 µg/ml with 50 mM sodium carbonate buffer, pH 9.6, prior to drying-down 20 ng in each well. Once the antigens had been dried down, the plates were washed five times with PBS containing 0.02% sodium azide and 0.05% Tween-20 (PBSA-T) to remove unbound peptide. Protein binding sites on the plastic were then blocked by adding 200 µl blocking buffer (PBSA-T containing 5% milk powder) to each well and incubated at 37°C for 2 hours. The plates were then washed five times with PBSA-T prior to incubation with the test antibody. Samples (100 µl) of serial 2-fold dilutions (1/100 to 1/12800) of the affinity purified antibody in PBSA-T containing 1% milk powder were added to triplicate wells and incubated overnight at 37°C. After the plates had been washed five times with PBSA-T, 100 µl of a 1:3000 dilution of goat anti-rabbit IgG conjugated to alkaline phosphatase (BioRad) was added to each well and incubated for 2 hours at 37°C. The plates were then washed five times with PBSA-T and 100 µl of colour development reagent (1 mg/ml *p*-nitrophenyl phosphate (disodium salt) dissolved in 10 mM diethanolamine-HCl, 1 mM $MgCl_2$, pH 9.8) was added to each well. After 45 minutes, the absorbance at 405 nm was measured using a plate reader (TiterTek Plus, ICN Flow).

2.2.8 Immobilisation of antibody raised against residues 460-477 of GLUT1 on protein A-Sepharose.

Antibodies, affinity-purified as described in Section 2.2.6, were cross-linked to protein A-Sepharose CL-4B using the bifunctional imidoester, dimethylpimelimidate, in the following way. Protein A-Sepharose CL-4B was

washed by centrifugation at 1,000 x *g* for 5 minutes in PBS five times. Affinity-purified antibody (4 mg) was added to the resin and incubated for 2 hours at room temperature with inversion. The sepharose was retrieved by centrifugation at 1,000 x *g* for 5 minutes and equilibrated in two changes of 0.2 M ethanolamine, pH 8.2 for 5 minutes at room temperature. Following centrifugation at 1,000 x *g* for 5 minutes, 40 ml 0.2 M triethanolamine/HCl, pH 8.2, 40 mM NaOH was added to the Sepharose and then left at room temperature for 45 minutes. The Sepharose was pelleted (1,000 x *g* for 5 minutes) and incubated in 2 ml 20 mM ethanolamine, pH 8.2 for 5 minutes at room temperature. The immobilised antibody was then stored as a 10% slurry at 4°C.

2.2.9 Immunoprecipitation of GLUT1

Photolabelled transporter (100 µg) in a phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA, pH 7.4) containing proteinase inhibitors pepstatin A (1 µg/ml) and PMSF (0.2 mM) was solubilised in a final volume of 1 ml with 0.25% SDS and 2.5% Triton X-100. Samples (0.5 ml) were then incubated with 100 µl of the 10% slurry of cross-linked antibody (Section 2.2.8) overnight at 4°C with rolling. The Sepharose beads were washed three times by centrifugation at 12,000 x *g* for 5 minutes with 0.5 ml phosphate buffer containing 1% Triton X-100 and 0.1% SDS, then once with 0.5 ml phosphate buffer containing 0.1% Triton X-100 and 0.01% SDS. Adsorbed polypeptides were then eluted with gel loading buffer (50 mM Tris-HCl, pH 6.8, 1 mM EDTA, 2% SDS, 6 M urea and 10% glycerol prior to SDS-PAGE (Section 2.2.4).

2.3 Techniques for DNA manipulations

The methods used for DNA manipulations were essentially as described by Sambrook *et al.* (1989) with some modifications. Plasticware, glassware, and media were sterilised by autoclaving for 15-20 min at 121°C (15 pounds/inch^2). Chemicals were either autoclaved or filter sterilised and the highest grades

available were used.

2.3.1 Phenol extraction and precipitation of nucleic acids

Phenol extraction is a rapid method for purification of nucleic acids from cellular extracts and inactivating enzymes. An equal volume of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0)-saturated phenol:chloroform:isoamyl alcohol (25:24:1) was added to a DNA preparation and mixed thoroughly by vortexing. The mixture was then centrifuged at 12,000 x *g* in a microfuge for 2 minutes to separate the phases. The top (aqueous) layer containing DNA was carefully removed and re-extracted with chloroform as above. The recovery of nucleic acids was then achieved by ethanol precipitation.

The ethanol precipitation was carried out by the addition of 0.1 volume of 3 M sodium acetate (pH 4.8) and 2 volumes of cold, absolute ethanol. The mixture was left at -20°C for 30 minutes and the precipitate was collected by centrifugation at 12,000 x *g* for 10 minutes in a microfuge. The pellet was then washed with 70% (v/v) ethanol, and dried either by inversion for 20 minutes at room temperature or under vacuum for a few minutes. Finally, the nucleic acids were dissolved in either TE or sterile distilled water.

2.3.2 Plasmid DNA preparations

A tube containing 5 ml of LB [1% (w/v) tryptone (Oxoid), 0.5% yeast extract (Oxoid), 0.5% NaCl, pH 7.0] or 2TY (1% tryptone, 1% yeast extract, 0.5% NaCl, pH 7.0) medium was inoculated with a single bacterial colony from a freshly streaked LB plate and grown overnight at 37°C in LB plus 50 µg/ml of ampicillin with constant shaking. Samples (1.5 ml) of each overnight culture were then transferred to 1.5 ml eppendorf tubes and centrifuged at 6,000 x *g* in a microfuge for 30 seconds. The pellets were each resuspended in 250 µl of GET buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0) and left on ice 5 minutes. The cells were lysed by the addition of 250 µl of a solution of 0.2 M

NaOH, 1% SDS and incubated for 5 minutes on ice. Precipitation of chromosomal DNA was achieved by neutralising the mixture with 200 µl of 3 M sodium acetate (pH 4.8) and incubating on ice for 10 to 60 minutes. The precipitate of cellular DNA and debris was removed by centrifugation at 12,000 x *g* for 10 minutes in a microfuge. The supernatants containing plasmid DNA were transferred to fresh 1.5 ml eppendorf tubes. To each supernatant was then added 0.9 ml of cold ethanol, followed by incubation at -20°C for 30 minutes to precipitate nucleic acids. The precipitate was collected by centrifugation at 12,000 x *g* for 10 minutes, resuspended in 200 µl of TE and phenol-extracted (Section 2.3.1).

2.3.3 Removal of RNA from preparations of plasmid DNA

DNA obtained by the alkaline lysis method also contained a large amount of RNA. To remove this, the DNA was treated with the enzyme RNase A (Sigma) that had been heated for 10 min at 100°C to inactivate any DNase activity. The RNase A was added to a DNA preparation at a concentration of 50 µg/ml and incubated for 30 to 60 min at 37°C. The reaction was terminated by phenol/chloroform extraction and the DNA was recovered by ethanol precipitation (Section 2.3.1).

2.3.4 Digestion of DNA with restriction endonucleases

Restriction endonucleases type II are DNases that recognise specific oligonucleotide sequences. Restriction enzymes were purchased from NBL or Promega. The restriction reaction was typically composed of the substrate DNA incubated for at least 1 hour at 37°C in a solution buffered near pH 7.5, containing Mg^{2+}, frequently Na^{+}, and the desired restriction enzyme. Digestions were carried out where possible in their appropriate `Reaction buffer' (supplied with the enzymes). Enzymes were used at a concentration of 2 to 5 units per µg of DNA. The enzyme reactions were monitored by agarose gel electrophoresis (Section 2.3.5) and the digested DNA was recovered by

phenol/chloroform extraction and ethanol precipitation (Section 2.3.1).

2.3.5 Agarose gel electrophoresis

Separation of DNA fragments generated after restriction digestion was achieved by agarose gel electrophoresis, using a horizontal submarine gel apparatus (Bio-Rad). Gel concentrations used were dependent on the size of the DNA fragments to be analysed and the degree of separation required. Most of the DNA fragments used in this study for further manipulations were of such a size that they were usually resolved in agarose gels with concentrations between 0.7 and 1.0%. Gels were prepared by microwaving the required amount of agarose (Sigma) in 1 x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) or 0.5 x TBE (45 mM Tris-borate buffer, 1 mM EDTA, pH 8.0), cooling the solution to 45°C and adding ethidium bromide to a final concentration of 0.5 µg/ml. Ethidium bromide intercalates between the bases of DNA and emits fluorescent radiation on UV illumination (302 nm recommended). The agarose was then poured into a gel casting tray and allowed to set for about 1 hour at room temperature. DNA samples were mixed with 0.1 volume of 10 x loading buffer [0.4% bromophenol blue and 67% (w/v) sucrose in water] prior to loading. Electrophoresis was carried out in 1 x TAE buffer at a constant voltage of 60 to 100V for 20 to 60 min. Molecular weight markers were provided by lambda DNA restricted with *Hin*d III (0.5 to 23 kb, NBL). DNA bands were visualised using a LKB UV transilluminator (2011 Macrovue) and photographed using a Polaroid land camera with Polaroid type 667 film.

2.3.6 Quantification of DNA

DNA concentrations were accurately measured by spectrophotometric absorbance readings of diluted DNA solutions at 260 nm. An absorbance of 1.0 at 260 nm was taken to be equivalent to a concentration of 50 µg/ml for double-stranded DNA. The purity of DNA preparations could be estimated by the ratio between the absorbances at 260 nm and 280 nm. Pure DNA preparation

should have an A_{260}/A_{280} ratio of ≥ 1.8. Amounts of DNA fragments in agarose gels were estimated by visual comparison of their fluorescence intensities with those of known amounts of lambda molecular markers.

2.3.7 Purification of DNA fragments from agarose gels

Purification of DNA by preparative electrophoresis is greatly facilitated by the use of chemically modified forms of agarose that gel and melt at low temperature without loss in the strength of the hardened gel. Such properties provide a simple way for the recovery and purification of DNA fragments before and after enzymic modification, ligation with other fragments, or sequencing. For such purifications, samples of DNA were electrophoresed at 4°C, typically at 30-40V, following which the DNA bands were visualised using UV transillumination and excised. The gel piece of interest was transferred to a 50 ml polypropylene tube containing 5 volumes of TE and then melted at 65°C in a water bath. The solution was then extracted with phenol/chloroform twice followed by an extraction with chloroform (Section 2.3.1). DNA was recovered by ethanol precipitation at -20°C and resuspended in an appropriate volume of TE (Section 2.3.1).

2.3.8 Ligation of DNA fragments

Both insert and vector DNA should be digested with appropriate restriction enzymes to generate compatible ends for cloning. If a single restriction enzyme is used to prepare the vector, the DNA should be treated with calf intestinal alkaline phosphatase (CIP) to remove 5' phosphate groups and thus prevent recircularisation of the vector during ligation. The missing 5' phosphate residues required in the ligation reaction can be provided only by the insert DNA, thus favouring the intermolecular joining event. To the linearised vector DNA (5 µg) was added 0.01 u/pmol ends of CIP in a total volume of 30 µl of 10 mM Tris, 1 mM EDTA, pH 8.0, and the mixture was incubated at 37°C for 60 minutes. The enzyme was then inactivated by the addition of 2 µl 0.5M EDTA followed

by extraction with phenol and chloroform as described in Section 2.3.1. To obtain the optimal ratio of vector to insert DNA, 1:1, 1:3 and 3:1 molar ratios of vector:insert DNA were tested routinely for the sub-cloning of inserts. A typical ligation reaction consisted of vector DNA and insert DNA at the appropriate ratios, 1 u T4 DNA ligase, 1 µl ligase 10 x buffer (300 mM Tris-HCl, pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP) in a final volume of 10 µl. The ligation mixtures were then left overnight at 16°C then transformed into competent JM109 cells (as described in Section 2.3.9).

2.3.9 Preparation of competent cells and transformation

In order to introduce DNA into cells, competence was artificially induced in *E. coli* cells (e.g. JM109 and DH1) by treating them with calcium chloride prior to adding DNA. This was performed by the method of Cohen *et al.* (1972) with some modifications.

A single colony of JM109 from a fresh streak plate was picked into 2 ml of LB medium and grown overnight at 37°C with constant shaking. The overnight culture (0.5 ml) was then used to inoculate 50 ml of LB in a 250 ml flask. The cells were grown at 37°C with agitation (200 rpm) for 2 to 3 hours until the absorbance reading at 600 nm (A_{600}) reached 0.4 to 0.5. Then, the cells were chilled on ice for 10 to 60 minutes and collected by centrifugation at 800 x *g* for 10 minutes at 4°C. The pellet was resuspended in 10 ml of ice-cold, 100 mM $CaCl_2$. After 30 minutes on ice, the cells were pelleted again as before and then resuspended gently in 2.5 ml of ice-cold 100 mM $CaCl_2$. At this point, the cells were either dispensed into 0.2 ml aliquots including 15% (v/v) glycerol, frozen on dry-ice and stored at -70°C or used immediately for transformation.

For transformation, up to 100 ng of DNA in a volume of less than 10 µl was added to a 200 µl aliquot of competent cells. After incubating on ice for 45 minutes, with occasional mixing every 15 min, the cells were heat-shocked at 42°C for exactly 90 sec and then immediately returned to ice for 1 minute. A

small volume (0.8 ml) of LB medium prewarmed at 37°C was added to each tube and then incubated at 37°C for 30 minutes to allow the expression of the antibiotic resistance gene. Following incubation, 100 µl of the transformed cells were plated out onto prewarmed LB agar plates containing ampicillin (50 µg/ml) and grown overnight at 37°C. The remaining cells were microfuged briefly and resuspended in 100 µl of LB prior to plating.

2.3.10 Manual sequencing

Although the chain-termination method is known to work best when using single-stranded templates, many laboratories have begun using double-stranded DNA (dsDNA) directly for sequencing because of the simplicity and convenience of the method. In order to be able to read sequence very close to the primer, low concentrations of nucleotide were used in the labelling reactions and manganese, which affects the termination step, was included in the reaction buffer, as described by the manufacturers and detailed below.

The DNA of all plasmid constructs to be sequenced was extracted by the alkaline lysis method, RNase-treated, and purified by phenol and chloroform extraction as detailed in Section 2.3.1. Before sequencing, the dsDNA was alkali-denatured by adding 6 µl of 1 M NaOH, 1 mM EDTA to 5 µg of plasmid DNA dissolved in 25 µl of TE and incubating at room temperature for 10 minutes. The DNA was then separated from alkali by passage through a "Spin-Column." Such columns were prepared by piercing the bottom of a 0.5 ml eppendorf tube with the tip of a 21 gauge needle and then adding 20 µl of glass beads (Sigma G-1145, 150-212µm), followed by 500 µl of Sepharose CL-6B equilibrated in TE. The 0.5 ml tube was placed inside a 1.5 ml eppendorf tube, also pierced with the needle. The eppendorf tubes, supported in a glass test tube, were centrifuged in a bench centrifuge at 3,000 x *g* for 5 minutes to remove the Sepharose buffer. Following alkali denaturation, the DNA was transferred to a column prepared as above and centrifuged at 3,000 x *g* for 3 minutes.

In order to anneal the template DNA to the sequencing primer, 6 µl of the denatured template DNA was mixed with 2 µl of 5 x annealing buffer (100 mM $MgCl_2$, 200 mM Tris-HCl, pH 7.5, 250 mM NaCl) and 2 µl (3 ng/µl) of sequencing primer. The mixture was heated to 65°C for 2 minutes and then allowed to cool slowly to room temperature over a period of about 30 minutes. The mixture was then placed on ice and used within 4 hours. The labelling reaction involved adding 1 µl of 100 mM DTT, 2 µl of diluted labelling mix (0.75 µM dGTP, dCTP and dTTP), 1 µl of [α-^{35}S]dATP at 10 µCi per µl and 10 µM (1000 Ci/mmol), 1 µl of Mn buffer (0.15 M sodium isocitrate, 0.1 M $MnCl_2$) and 2 µl of Sequenase Version 2.0 enzyme (diluted 1:8 in 10 mM Tris-HCl, pH 7.5, 5 mM DTT, 0.5 mg/ml BSA) to the annealed primer-template mixture. The labelling reaction was incubated at room temperature for 2 minutes and then immediately placed on ice.

A sample (3.5 µl) of the above labelling mix was transferred to each of 4 eppendorf tubes (labelled G, A, T and C), and 2.5 µl of the appropriate, prewarmed termination buffer was spotted on to the side of each tube. The extension reactions were then started simultaneously by brief centrifugation in a microfuge, followed by incubation at 37°C for 5 minutes. Reactions were stopped by the addition of 4 µl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) and stored on ice until loading. Sequencing mixtures were denatured at 90°C for 5 minutes before loading onto 6% polyacrylamide, 8 M urea gels (0.4 mm thick), prepared and pre-run in 1 x Tris-borate buffer (TBE, 0.08 M Tris-borate, 0.02 M EDTA, pH 8.0). Gels were electrophoresed at a constant power (70 watts) until the bromophenol blue dye had run about 80% of the length of the gel. Following electrophoresis, the gel was fixed in 10% (v/v) acetic acid for 30 minutes, dried onto Whatman 3MM paper and autoradiographed at -70°C overnight.

2.4 Techniques for the baculovirus expression system

2.4.1 Methods for insect cell culture

2.4.1.1 Monolayer culture

Spodoptera frugiperda (*Sf*) cells are commonly employed as the host permissive cell line to support AcNPV replication and protein synthesis. The cell line used in this study was the *Sf*9 line and derives from the pupal ovarian tissue of the alfalfa looper (Vaughn *et al.*, 1977). The cells were maintained according to the method described by Summers and Smith (1987), with some modifications.

*Sf*9 cells exhibited a doubling time of approximately 18 to 24 hours in complete TC100 medium [TC100 medium (Gibco-BRL), 10% (v/v) foetal calf serum (Flow), 1% of antibiotics (penicillin 5,000 units/ml + streptomycin 5,000 µg/ml, Gibco-BRL)] at 28 ± 1.0°C. The *Sf*9 cells did not require CO_2, and neither trypsin nor other enzymes were used during subculturing. Subculturing in tissue culture flasks was performed by discarding the old culture medium and resuspending the cells by tapping the side of the flask 3 to 5 times and rapidly pipetting fresh medium across the monolayer, avoiding excessive foaming. The cell suspension was split (typically 1:4) and transferred into new flasks containing a suitable quantity of complete TC100 medium (10 ml for a 75 cm^2 or 5 ml for a 25 cm^2 flask) and incubated at 28°C. Cell viability was checked by adding 0.1 ml of trypan blue (0.4% stock, pH 7.3) to 1 ml of cells and examining under a microscope. The cells that took up trypan blue were considered non-viable. The viability should be more than 97% for healthy log-phase cells. The morphology of the healthy cells appeared to be rounded but not granular. When the *Sf*9 cells were overgrown or in older culture, they began to float in the medium.

2.4.1.2 Freezing and storage of insect cells

Cells used for freezing were healthy log-phase cultures (>97% viability). They were pelleted by centrifugation at 1 000 x *g* for 5 minutes and resuspended in fresh complete medium at a density of 4-5 x 10^6 cells/ml. The cell suspension was diluted with an equal volume of fresh freezing medium [20% (v/v) DMSO in complete TC100] to yield a final DMSO concentration of 10% and maintained on ice. The diluted cell suspension was then dispensed into 1 ml aliquots. The cells were frozen slowly by placing freezing vials in an insulated container at -20°C for 1 hour and then at -70°C overnight. Finally, the cells were stored in liquid nitrogen.

2.4.1.3 Thawing cells

A vial was removed from liquid nitrogen and thawed rapidly with gentle agitation in a 37°C waterbath. The outside of the vial was decontaminated quickly with 70% ethanol. Cells were placed into a 25 cm^2 flask containing 5 ml of fresh complete medium and incubated at 28°C. The old medium was discarded and replaced with fresh complete TC100 the next day.

2.4.2 Plaque assay

The plaque assay was used either to purify a virus stock or to determine the number of infectious virus particles (i.e. the titre) in a stock. The assay was carried out essentially according to the methods described in Summers and Smith (1987) and Emery (1991a).

*Sf*9 cells were seeded into 60 mm tissue culture dishes (2 x 10^6 cells/dish) and allowed to attach for 1 hour at room temperature. Serial (10-fold) dilutions of virus inoculum were prepared ranging from 10^{-1} to 10^{-5}. Following incubation, the medium was removed from the attached cells and then 100 µl of the diluted virus was added to each plate, ensuring an even distribution of the inoculum.

The dishes were then incubated for a further 1 hour at room temperature or 28°C. While incubating, equal volumes of 2% (w/v) "Seaplaque (FMC)", a low melting point agarose, and complete TC100 medium were mixed to yield a final concentration of 1.0% agarose, which was kept at 45°C in a water bath until required. Following incubation, the inoculum was removed and 4 ml of the agarose/TC100 mixture was added slowly to the edge of each dish, ensuring even spreading. Once the agarose had hardened the dishes were incubated for 5 to 7 days at 28°C in a humidified container. Plaques appeared as clear, circular areas approximately 1 to 3 mm in diameter.

Since each plaque derives from a single infectious virus particle, the concentration of infectious units (i.e. the titre) in a virus stock can be determined by counting the number of plaques formed by different dilutions of the virus on the plates, and is expressed in plaque forming units per ml (pfu/ml). The titre (pfu/ml) Is calculated as follows (Summers and Smith, 1987): pfu/ml = 1/dilution x number of plaques x 1/(ml inoculum/dish).

2.4.3 Virus amplification

*Sf*9 cells ($2x10^7$) were seeded into a 75 cm^2 tissue culture flask and allowed to attach at room temperature for 15 minutes. The low titre recombinant stock (co-transfection supernatant) is used to infect the cells at a multiplicity of infection (MOI) of less than 1 to prevent repetitive infections selecting for deletion mutants. The cells were incubated at 28°C for 3 days and the supernatant from the plate was harvested after spinning down the cells at 2,500 x *g* for 5 minutes. The virus titre was then determined using the plaque assay procedure (Section 2.4.1) before the amplification was repeated until a high virus titre was obtained.

2.4.4 Preparation of infectious AcNPV DNA

Baculovirus particle-containing supernatants may be stored either for up to 6 months at 4°C, however, the best way to preserve a recombinant virus is to isolate its DNA and store it at -70°C. Confluent monolayers of *Sf*9 cells were infected with recombinant baculovirus at a MOI of about 1. Three to five days after infection, the cells were removed by centrifugation at 1,000 x *g* for 10 minutes. The supernatants containing extracellular virus particles were then subjected to ultracentrifugation at 100,000 x *g* for 1 hour at 4°C. The viral pellet was resuspended in 1 ml of TE and then laid onto a gradient consisting of equal volumes of 10 and 50% sucrose in TE (w/w). Following centrifugation in a swing-out rotor for 90 minutes at 100,000 x *g*, the band of virus at the 10-50% sucrose interface was carefully removed, diluted with TE to 50 ml, and pelleted at 100,000 x *g* in a fixed angle rotor for 1 hour at 4°C. The viral pellet was resuspended in 1 ml of TE, to which was added 0.6 ml of lysis buffer [10% (w/v) sodium *N*-lauryl sarcosinate, 10 mM EDTA], and incubated at 60°C for 20 minutes. The mixture was laid onto a 54% (w/v) cesium chloride/TE gradient containing 200 µl of ethidium bromide (10 mg/ml) and subjected to ultracentrifugation at 200,000 x g for 18 hours at 20°C. Using UV light, the viral DNA bands (supercoiled and open circular DNA) were collected, extracted with water-saturated butan-1-ol to remove the ethidium bromide, and dialysed overnight against sterile TE. The viral DNA was stored at 4°C.

CHAPTER 3. SEQUENCE ALIGNMENT AND ANALYSIS OF THE SUGAR TRANSPORTER FAMILY

3.1 Introduction

The ability to predict the three-dimensional structure of a protein from its primary structure is one of the ultimate aims of molecular biology. Although the prediction of soluble protein structure is becoming more accurate, the prediction of integral membrane protein structure may be further from realisation because there are still very few known structures for this class of proteins. However, membranes are essentially two-dimensional and consequently provide a powerful constraint upon the allowed arrangement of the secondary structural elements that cross them. However, there is a requirement for effective prediction techniques to be developed for membrane proteins is an important factor with respect to the design of experiments aimed at providing a deeper understanding of how these proteins function and maintain a cell's contact with the external environment.

A landmark achievement towards understanding the structure of membrane proteins was the crystallisation and structure determination of the photosynthetic reaction centre (RC) from the purple bacterium *Rhodopseudomonas viridis* (Deisenhofer *et al.*, 1985), followed by that of the RC from *Rhodobacter sphaeroides* (Allen *et al.*, 1988). The RCs from these bacteria are integral membrane protein-pigment complexes which carry out the initial steps of photosynthesis, and are composed of three membrane-associated protein subunits and several cofactors. A central feature of the structural organisation of the RC is the presence of eleven hydrophobic α-helices, approximately 20-30 residues in length, which represent the membrane-spanning portion of the RC.

The formation of apolar α-helices is one of the ways in which a protein can adapt its structure to a lipid environment (Jennings, 1989) since it allows a

favourable interaction between the apolar amino acid side chains and the lipid environment, whilst fully satisfying the hydrogen bonding potential of the peptide units in a regular secondary structure. From electron diffraction studies on the structure of bacteriorhodopsin (Henderson *et al.*, 1990), and X-ray crystallographic data regarding the RC (Deisenhofer *et al.*, 1985), it appears that transmembrane segments are typically apolar helices, 17-25 residues in length. The alternative arrangement, which has been noted in the crystallographically-determined structure of a bacterial outer membrane porin (Weiss and Schulz, 1992), shows the transmembrane segments to be anti-parallel β-strands arranged in a 16-stranded barrel. This structure is characterised by the presence of alternating hydrophobic and hydrophilic residues, with the former interacting with the lipid and the latter facing inwards towards the aqueous centre of the barrel. These transmembrane segments can be as short as 6-7 residues in length because of the extended conformation.

The general consensus of opinion is that most transmembrane segments adopt the structure of an α-helix, and this has led to the development of integral membrane protein structure prediction methods with this assumption at their core. The accuracy of this assumption will, of course, only become apparent upon the solution of more integral membrane protein structures. However, biophysical analyses of those proteins currently available in sufficient quantity have revealed that the majority of integral membrane proteins techniques so far investigated have a very high helix content. The evidence in favour of this with respect to GLUT1 has been described in Section 1.3, and although the location of the helices relative to the membrane has not been determined, it appears from infrared spectra of the protein, before and after proteolytic removal of the cytoplasmic regions, that the membrane embedded portions of the protein are predominantly α-helical (Cairns *et al.*, 1987).

The intrinsic properties of membrane proteins that make them distinct from globular proteins centre upon the regions of the protein which interact directly with the lipid bilayer. These regions are necessarily hydrophobic in nature in

order to allow the interactions between the hydrophobic amino acid side chains and the acyl chains of membrane phospholipids. As the apolar surfaces of the transmembrane helices generate an interface capable of interacting with the lipid environment, so the more polar character of the inward facing residues provides for the formation of helical bundles which protect the polar surfaces from the lipid. As a consequence, the folding of a multi-spanning α-helical protein appears to rely upon the membrane insertion of correctly oriented transmembrane segments. Recent studies have indicated the presence of topogenic signals within integral membrane proteins, that is, sequence patterns which correlate with the topology of the membrane spanning segments. The most evident of these signals is the prevalence of a membrane spanning hydrophobic core bounded on the cytoplasmic side by positively charged amino acid residues [the 'positive inside rule' (von Heijne, 1992)]. Such topogenic signals can be used to evaluate the plausibility of predicted integral membrane structures, as has been demonstrated by von Heijne (Sipos and von Heijne, 1993).

Unfortunately, the size of the three-dimensional structure dataset for membrane proteins is only a fraction of that of its one-dimensional counterpart. However, analysis of the primary structure database has revealed some general patterns and some rules have been deduced from searches and comparisons of the sequences. Examination of the sequence and structure of the RC has demonstrated a number of important similarities between soluble and integral membrane proteins (Rees *et al.*, 1989). For example, although the total surface area of the RC exposed to lipid is similar to the solvent exposed surface area of a similarly sized soluble protein, the most striking difference between soluble and membrane proteins is the chemical nature of the exposed groups. In order to minimise surface energies, soluble proteins fold to generate a polar face that contacts the aqueous environment whilst the character of the surface amino acids in the RC is largely apolar. The average hydrophobicity of non-exposed residues and the internal atomic packing density are similar, however, which is indicative of internal van der Waals contacts being a primary feature of

structure stabilisation within both protein classes. Although it is not advisable to generalise about membrane protein structure from such a small set of molecules, it is likely that the atomic interactions in the interior of α-helical integral membrane proteins are similar to those observed in soluble proteins and that the same packing principles apply.

The task of predicting the structure of α-helical integral membrane proteins can be broken down to a set of discrete problems which range from the definition of the regions of sequence spanning the lipid bilayer to the orientation and arrangement of the individual helices that constitute a three-dimensional model of the protein. Within these conceptual extremes reside many other facets of prediction which possess the inherent ability to alter dramatically the outcome of a three-dimensional modelling problem. Obviously, each step requires critical analysis in order to assess the plausibility of any conclusions.

This chapter is devoted to the generation of a three-dimensional model for GLUT1 via the implementation of a variety of sequence analysis tools upon the sugar transporter family. In essence, the general strategy was to generate an amino acid sequence alignment of as many proteins homologous to the mammalian sugar transporters as possible, with the intention of gaining structural information from various predictive algorithms. It was hoped that hydrophobicity and periodicity analyses would suggest possible secondary and tertiary structural features of GLUT1. The principal objective underlying this approach was to rationalise experimental ideas, that is, structural features emerging from such analyses would form the basis of future site-directed mutagenesis strategies. Naturally, each of the these tools is comprised of implicit assumptions which will be elaborated upon through the progression of sequence to 'structure'. However, due to the existence of a large number of homologous proteins believed to possess the twelve membrane spanning α-helix architecture of the sugar transporter family, it was hoped that a consensus result would be obtained from each type of analysis which would be more significant than if only the GLUT1 sequence had been used.

3.2 The sugar transporter family

One of the main outcomes from primary structure database search and comparison studies is that the vast majority of membrane proteins so far examined belong to the α-helical transmembrane class of proteins with hydrophobic stretches long enough to span the membrane. The sugar transporter family contains the passive sugar transporters of mammalian tissues, as well as both passive transporters and active sugar/H^+ symporters of higher plants, green algae, protozoans, yeasts, cyanobacteria and eubacteria.

The principal criterion for selection of proteins to be included within the sugar transporter family was the presence of regions of sequence similarity. Table 3.1 lists some fundamental properties of the proteins found to be homologous to the mammalian sugar transporters, and it is evident that there is a marked variation in sequence length and activity. However, a twelve membrane spanning α-helix architecture is predicted for each protein that implies a fundamentally similar mechanism of transport at the molecular level. It was intended that information supporting this notion might be obtained by a thorough analysis of the family at the level of amino acid sequence.

Having amassed such a number of related sequences, the problem is to try and unlock the intrinsic structural information in the absence of a crystal structure for any member of the superfamily. Clearly, the secondary and tertiary structural features of the transporter family must, in some way, be reflected in the similarity of the amino acid sequences. The initial aim, therefore, was to apply numerous sequence analysis methods to all of the sequences in the hope that certain common structural features would become apparent, and would perhaps illuminate some functional aspects of sugar transport. It is important to realise from the outset that sequence analysis methods are approximate and possess many inherent limitations. The fundamental question centres upon whether two or more sequences exhibit similar three dimensional structure and/or function based upon the similarities observed in their amino acid sequences.

Table 3.1 Properties of transporters related to the mammalian GLUT family.

Designation	Organism	Substrate	Size (res)	Reference
SNF3	*Saccharomyces cerevisiae*	Glucose	884	Celenza *et al.*, 1988
GAL2	*S.cerevisiae*	Galactose	574	Nehlin *et al.*, 1989
HXT1	*S.cerevisiae*	Glucose	569	Lewis and Bisson, 1991
HXT2	*S.cerevisiae*	Glucose	541	Kruckeberg and Bisson, 1990
RAG1	*Klebsiella lactis*	Glucose	567	Goffrini *et al.*, 1990
MAL61	*S.carlsbergensis*	Maltose	614	Yao *et al.*, 1989
LAC12	*K.lactis*	Lactose	587	Chang and Dickson, 1988
IRT1	*S.cerevisiae*	Myo-inositol	584	Nikawa *et al.*, 1991
IRT2	*S.cerevisiae*	Myo-inositol	612	Nikawa *et al.*, 1991
HUP1	*Chlorella kessleri*	Glucose	533	Sauer and Tanner, 1989
STP1	*Arabidopsis thaliana*	Glucose	522	Sauer *et al.*, 1990
glcP	*Synechocystis*	Glucose	468	Schmetterer, 1990
Pro-1	*Leishmania enrietti*	Glucose	567	Langford *et al.*, 1994
qa-y	*Neurospora crassa*	Quinate	537	Geever *et al.*, 1989
qutD	*Aspergillus nidulans*	Quinate	533	Hawkins *et al.*, 1988
AraE	*Escherichia coli*	Arabinose	472	Maiden *et al.*, 1988
GalP	*E.coli*	Galactose,	464	Henderson *et al.*, 1992
XylE	*E.coli*	Xylose	491	Davis and Henderson, 1987
glf	*Zymomonas mobilis*	Glucose	473	Barnell et al., 1990
cit+	Klebsiella pneumoniae	Citrate	444	van der Rest *et al.*, 1990
citA	*E.coli* pWR61	Citrate	431	Sasatsu *et al.*, 1985
kgtP	*E.coli*	α-ketoglutarate	432	Seol and Shatkin, 1991
bap3	*Strep. hygroscopicus*	Bialaphos	448	Raibaud *et al.*, 1991
CmlA	*Pseudomonas aeruginosa*	Chloramphenico	419	Bissonnette *et al.*, 1991
TetC	E.coli pBR322	Tetracyclines	396	Henderson and Maiden, 1990
TetA	E.coli pRP1	Tetracyclines	399	Waters *et al.*, 1983
TetB	*E.coli* Tn10	Tetracyclines	401	Hillen and Schollmeier, 1983
norA	*Staphylococcus aureus*	Quinolones	388	Yoshida *et al.*, 1990
Bmr	*Bacillus subtilis*	Multidrug	468	Neyfakh, 1992
TetL	*B.subtilis* pTHT15	Tetracyclines	458	Hoshino *et al.*, 1985
pNS1981	*B.subtilis*	Tetracyclines	458	Sakaguchi *et al.*, 1988
BNS908	*B.subtilis*	Tetracyclines	458	Sakaguchi *et al.*, 1988
TetK	S.aureus pT181	Tetracyclines	295	Khan and Novick, 1983
mmr	*Streptomyces*	Methylenomycin	475	Neal and Chater, 1987
qacA	*S.aureus*	Antiseptic	514	Rouch *et al.*, 1990
ACTII	*Streptomyces coelicolor*	Actinorhodin	578	Fernandez-Moreno *et al.*, 1991
ATR1	*S.cerevisiae*	Aminotriazole	547	Kanazawa *et al.*, 1988
LacY	*E.coli*	Lactose	417	Büchel *et al.*, 1980
LacY	*K.pneumoniae*	Lactose	416	McMorrow *et al.*, 1988
RafB	*E.coli*	Raffinose	425	Aslanidis *et al.*, 1989
MelB	*E.coli*	Melibiose	469	Yazyu *et al.*, 1984

3.2.1 Multiple sequence alignment of the sugar transporter family

At the core of sequence comparison is the concept of an alignment which defines the relationship between sequences on a residue-by-residue basis. Aligned sequences are presumed to be related in an evolutionary and/or functional sense; residues occupying equivalent positions are believed to share common ancestors and/or to have equivalent biological roles. Thus, the alignment of two or more protein sequences can provide a wealth of information to guide further experimentation, particularly if one of the aligned proteins has been chemically or crystallographically well characterised. Clearly, a multiple alignment of the whole sugar transporter family was required so that common sequence features could be highlighted and used as a basis for predictions regarding the consensus structure of the family.

An attempt to generate the alignment using MULTAL software (Taylor, 1990) was made (data not shown) which represents a compromise between obtaining a good alignment and the time taken to do so. Although the program was capable of aligning the most homologous sequences, little success was achieved with subsequent alignment of distantly related proteins. Therefore, since any inference from the alignment is crucially dependent on its accuracy, it was felt necessary to produce the alignment manually.

There is a wealth of genetic evidence supporting the roles of insertions and deletions in the evolution of macromolecules, and it is customary to allow for the presence of unrelated segments reflecting these events in sequence alignments via the introduction of 'gaps'. Unfortunately, the inclusion of gaps can greatly increase the number of identities or similarities seen between sequences that may not be related. Although the introduction of gaps is usually obvious, it is the positioning which requires the greatest consideration lest the significance of identities be over-interpreted. However, since the ultimate aim of this alignment was to investigate possible structural similarities in the sugar transporter family, the guiding influence was residue size and hydrophobicity.

The alignment (Appendix) will not be discussed in great detail here because general features have been reviewed recently elsewhere (Griffith *et al.*, 1992, Henderson *et al.*, 1992). However, it is important to emphasise certain common characteristics. The sequence similarities of the sugar transporter family are probably indicative of a similarity of three-dimensional structure and that, at the molecular level, these passive transporters, active symporters and antiporters must share many features. Perhaps the most interesting property is the sequence similarity between the passive, mammalian sugar transporters and the active, bacterial transporters where sugar transport is driven by the proton gradient existing across the cytoplasmic membrane. This finding indicates that, mechanistically, the processes of passive and active transport are likely to be comparable, although examination of the sequence alone is not sufficient to be able to place a protein into either of these categories.

It is evident that within the sugar transporter family are sub-families of proteins that exhibit sequence motifs which are not always present in the other sub-families. The transporter superfamily appears to comprise four sub-families (designated I-IV) of homologous transport proteins. The first sub-family (I) includes sugar transporters from organisms as diverse as mammals, plants, algae, fungi and bacteria, the substrate specificities of which range from pentoses, hexoses and dissacharides to a carboxylate compound, quinate. In most cases the degree of sequence identity lies between 45% and 70% (Figure 3.1) for the members of Family I. The inclusion of the rat SV2 sequence in Family I might appear to be incongruous from its extraordinarily low sequence identities (<10% to all members of the family) but the justification for its positioning is partly due to its possession of the characteristic internal sequence duplications. That is, it contains PESPR and PETRG motifs at the *C*-terminal ends of helices 6 and 12, plus the distinctive (N/D)(R/K)XGR(K/R) and (N/D)(R/K)XGR(K/R) motifs which occur between helices 2 and 3, and helices 8 and 9. Thus, a low sequence identity does not imply that a similarity is not biologically important. Indeed, even if the similarity of two complete sequences is so low as not to reach statistical significance, the presence of short regions

Figure 3.1 Percent sequence identities for the sugar transporter family.

HUMAN GLUT1																																																							
RABBIT GLUT1	97																																																						
RAT GLUT1	98	97																																																				
PIG GLUT1	97	97	98																																																		
MOUSE GLUT1	96	96	98	96																																																
HUMAN GLUT2	55	55	56	55	56																																														
RAT GLUT2	54	54	54	53	54	82																																													
MOUSE GLUT2	52	52	52	52	#	82	94																																											
RAT GLUT7	48	48	48	48	48	59	68	66																																										
HUMAN GLUT3	63	63	63	65	63	51	49	49	45																																								
MOUSE GLUT3	63	63	63	66	63	50	48	47	44	82																																						
HUMAN GLUT4	65	65	66	65	66	52	52	51	44	57	56																																					
RAT GLUT4	65	64	65	64	65	52	52	51	44	57	55	95																																				
MOUSE GLUT4	65	64	66	64	65	52	52	51	44	56	55	94	97																																			
HUMAN GLUT5	40	41	41	42	41	39	37	37	37	38	40	41	40	40																																		
RAT SV2	6	6	6	6	7	8	7	7	6	7	7	6	6	6	7																																	
YEAST SNF3	25	25	25	24	25	25	24	24	20	24	24	24	23	24	25	9																																
YEAST GAL2	25	26	25	26	25	24	24	24	22	24	24	24	23	23	23	8	28																															
YEAST HXT1	23	23	24	24	24	23	22	22	20	23	24	22	21	21	21	8	27	68																														
YEAST HXT2	25	25	25	26	25	25	23	23	22	25	25	23	22	22	22	8	29	68	66																													
YEAST RAG1	24	25	25	25	24	23	22	22	21	23	24	22	21	21	23	9	27	70	74	69																												
YEAST MAL61	19	19	19	20	19	17	17	18	17	19	19	17	17	17	17	7	19	19	18	20	19																											
YEAST LAC12	19	20	19	20	19	17	18	18	18	19	20	18	18	18	18	6	19	21	20	22	21	17																										
YEAST ITR1	25	24	25	26	25	23	24	24	20	25	26	25	24	24	22	7	24	23	23	25	24	17	18																									
YEAST ITR2	25	24	25	25	24	24	24	24	20	24	25	24	23	23	21	7	24	24	24	25	25	17	17	87																								
CHLORELLA HUP1	24	23	23	24	23	25	24	25	21	24	25	27	26	25	24	7	25	23	23	25	24	16	20	24	23																							
ARABIDOPSIS STP1	26	26	26	26	26	25	25	25	22	24	25	26	26	26	25	7	23	23	24	24	25	15	18	24	24	45																						
SYNECHOCYSTIS glcP	26	26	27	25	26	28	28	28	27	26	27	28	28	27	25	8	26	25	28	28	27	18	22	25	26	27	29																					
LEISHMANIA Pro-1	19	19	19	19	19	17	18	17	16	19	18	19	18	18	17	7	14	16	16	19	16	11	14	15	14	17	17	18																				
NEUROSPORA qa-y	22	23	23	23	23	23	23	23	19	21	23	22	22	23	23	7	21	24	22	23	23	18	19	23	23	25	24	26	14																			
ASPERGILLUS qutD	22	22	22	22	22	22	22	22	19	22	24	22	22	22	20	6	24	24	23	26	24	17	20	24	24	24	23	27	13	61																		
E.coli AraE	25	25	25	24	26	25	26	26	24	25	25	26	25	25	25	7	24	22	23	24	23	17	20	28	27	30	28	29	17	25	24																	
E.coli GalP	26	25	26	25	26	27	26	26	24	25	24	27	27	27	26	8	26	25	24	25	25	18	19	29	28	32	30	31	18	26	25	63																
E.coli XylE	27	26	27	27	26	26	25	25	22	26	27	24	24	24	23	7	24	26	26	27	27	18	18	24	25	24	26	39	16	25	25	29	32															
Z.mobilis glf	24	24	24	24	24	24	23	23	21	25	26	25	26	25	22	8	22	25	25	25	26	19	18	22	21	24	26	30	16	22	23	26	30	42														
K.pneumoniae cit+	12	12	12	11	12	15	15	15	15	13	12	13	12	13	13	9	15	13	14	15	14	11	13	16	15	15	15	14	13	15	15	15	15	15	17													
E.coli citA	14	14	14	14	14	17	17	17	15	15	14	14	14	14	14	7	14	14	15	17	14	13	14	18	19	17	15	16	13	15	15	15	17	17	19	65												
E.coli kgtP	11	11	11	11	11	12	13	13	11	12	13	13	13	14	13	7	12	13	12	13	12	10	13	16	15	13	12	13	10	13	13	15	16	13	13	30	30											
S.hygroscopicus bap3	13	13	13	13	13	11	13	13	11	13	12	12	12	12	11	7	11	11	12	11	11	12	13	15	14	12	13	11	12	13	12	15	13	13	13	28	28	36										
Tn1696 cmlA	10	10	10	9	10	9	8	8	9	11	10	10	11	10	10	7	10	10	10	12	10	10	11	12	12	10	8	12	10	9	10	11	10	10	15	12	12	13	12									
E.coli TetC	15	16	15	14	15	13	15	14	12	15	14	14	15	15	12	7	13	12	13	12	13	15	13	12	13	11	12	15	13	15	16	15	14	14	14	15	13	9	12	15								
E.coli Tet A	14	14	13	12	13	13	14	13	12	13	13	14	14	15	13	8	13	12	15	12	13	14	12	12	13	12	12	13	12	15	15	14	13	14	14	14	13	12	12	15	78							
E.coli TetB	12	12	12	11	12	11	10	10	9	12	14	12	12	12	11	8	12	10	9	11	11	11	13	13	13	11	11	13	11	11	13	14	12	12	13	13	11	12	10	13	45	44						
S.aureus norA	5	5	5	4	5	6	5	6	7	5	6	6	5	5	4	6	5	6	5	5	6	7	6	6	8	8	6	7	4	6	6	6	6	8	5	6	5	5	7	8	5	5	7					
B.subtilis Bmr	6	5	6	5	6	7	7	7	7	5	6	8	7	7	7	6	5	6	4	7	6	6	6	5	4	5	3	5	5	5	5	5	4	7	7	5	6	7	7	8	4	6	8	8				
TetL	12	12	12	11	12	12	11	11	12	11	11	13	13	13	11	6	11	11	11	10	10	8	13	11	10	9	10	10	8	11	11	14	11	10	11	12	10	13	10	13	15	15	15	7	6			
pNS1981	12	12	12	11	12	12	11	11	12	11	11	13	13	13	11	6	11	11	11	10	10	8	13	11	10	9	10	10	8	11	11	14	11	10	11	12	10	13	10	13	15	15	15	7	6			
BS908	12	12	12	10	12	10	9	10	10	11	10	12	12	12	12	7	11	12	10	9	10	7	13	10	10	10	9	11	8	11	10	14	11	10	11	11	11	11	10	13	14	13	15	7	6	81	81	
S.aureus TetK	10	9	10	8	10	10	11	12	12	10	10	11	11	11	10	9	11	11	11	9	10	8	12	10	10	9	10	11	7	10	10	13	12	10	10	12	11	12	11	10	13	12	14	8	5	61	61	60
mmr	10	10	10	9	10	9	10	10	10	11	12	11	11	11	11	8	10	9	9	10	8	8	8	10	10	13	10	13	10	8	8	12	12	10	11	10	9	12	14	15	15	16	13	7	7	16	16	16	15							
qacA	10	10	10	10	9	12	11	11	12	12	14	12	12	11	11	7	13	11	10	12	12	11	9	11	12	13	12	13	9	10	11	14	16	11	12	13	12	10	10	13	14	15	15	8	6	14	14	15	13	22						
ActII	10	10	10	9	10	12	11	11	11	11	11	11	11	11	10	9	9	11	12	12	11	9	9	11	11	10	11	12	9	10	11	13	12	10	11	9	9	10	10	8	11	11	12	5	8	13	13	12	12	18	18					
ATR1	7	7	7	7	7	9	8	9	9	8	8	6	7	7	8	7	7	9	9	9	8	8	8	6	8	9	10	9	6	10	8	9	10	10	9	8	9	8	9	11	9	9	10	8	4	12	12	12	13	10	13	16				
LacY	12	11	12	11	11	13	13	13	12	12	12	13	13	13	12	6	12	13	12	12	12	8	12	14	15	12	13	10	10	8	9	10	12	11	12	12	11	9	9	11	14	13	13	9	7	12	12	11	11	9	12	8	10			
K.pneu	11	11	11	12	11	12	13	13	11	11	12	13	14	13	10	6	12	10	10	10	12	9	11	13	12	10	11	13	9	10	10	11	12	11	11	13	12	10	8	11	13	13	12	8	6	11	11	11	10	7	11	6	8	59		
RafB	11	12	12	11	11	11	12	12	10	11	11	13	13	12	11	6	13	10	8	8	9	8	10	12	12	10	11	12	9	9	9	9	10	11	12	12	11	10	11	11	13	13	13	9	7	9	9	8	10	9	13	7	7	56	58	
MelB	10	10	10	9	9	11	11	10	10	10	10	10	10	10	13	5	8	9	9	9	9	9	7	8	8	9	10	9	8	8	8	8	8	9	8	11	10	8	7	7	9	8	10	8	6	6	6	5	8	7	9	7	6	9	10	10

corresponding to well-conserved motifs, at similar locations in the two proteins, is strongly suggestive of a common evolutionary origin.

The second family (II) contains transporters for citrate, α-ketoglutarate and bialaphos. According to the nomenclature of Griffith *et al.* (1992), this family is designated as family IV. However, with regard to the presence of internal sequence duplications and conserved motifs this family does possess greater similarity to Family I than Family III.

The third family (III) contains the TetA, B and C tetracycline transporters, the NorA quinolone resistance protein and the Bmr multidrug resistance protein. This latter protein confers resistance to several structurally dissimilar compounds including puromycin, chloramphenicol and ethidium bromide.

The fourth family (IV) contains the TetL and TetK tetracycline transporters, the QacA quaternary ammonium compound transporter, the Mmr methylenomycin resistance protein and a protein, ActII, which is thought to be involved in the export of the the polycyclic antibiotic actinorhodin.

In contrast to the transporters described above, whose sequences clearly demonstrate a common evolutionary origin, it is apparent from Figure 3.1 that the disaccharide transporters of bacteria show little sequence identity with the rest of the sugar transporter family, the *lac* permease of *E.coli* is the best understood of the bacterial proton-linked sugar transporters. However, although only about 12% of the residues in GLUT1 are identical with those in either of the *lac* permeases from *E.coli* or *K.pneumoniae*, the sequences can be aligned to reveal identical residues or conservative substitutions at numerous positions characteristic of the sugar transporter family (see Appendix). For example, a DKLGLR motif is found between putative helices 2 and 3 that resembles the mammalian (D/N)(R/K)XGR(K/R) motif. In addition, the distinctive PESP motif of the sugar transporters at the *C*-terminal end of helix 6 is found in the forms of TDAP and PESS in the respective lactose transporters. It is likely, therefore,

that the bacterial disaccharide transporters do represent a distinctly related branch of the sugar transporter family. In the present context, such a relationship is important because the *lac* permease of *E.coli* is, without doubt, the most intensively studied of the bacterial proton-linked sugar transporters (Kaback, 1992b, 1994). Consequently, information about its structure and function is likely to illuminate the understanding of the mechanism of action of the mammalian sugar transporters.

One of the most striking traits of the sugar transporter family of sequences is that the regions corresponding to the putative transmembrane helices can be aligned readily with very few gaps. In contrast, the intervening sequences contain insertions and deletions, which are strongly suggestive of a location exposed to solvent. It is these characteristics that provide strong evidence for the unifying hypothesis which envisages a common ancestor for the sugar transporters and homologues.

Evidence to substantiate this comes from a closer examination of the aligned sequences which reveals the presence of internal sequence duplications and conserved residues that must play important roles in the structure and/or function of these proteins. The most prominent of these is the (N/D)(R/K)XGR(K/R) motif that occurs between helices 2 and 3 and the (E/D)(R/K)XGR(K/R) motif which recurs between helices 8 and 9. Similar motif twins exist at the *C*-terminal ends of helices 6 and 12 in the forms of PESPR and PETKG, respectively. Although the functions of these motifs are unknown, their presence does reveal that the proteins probably evolved as a result of an internal gene duplication event (Maiden *et al.*, 1987). Assuming that the ancestral protein did possess only six helices, then the logical progression is to assume that the more recent transporters are composed of two bundles of six helices disposed about a pseudo two-fold axis of symmetry (Baldwin, 1992).

3.3 Phylogenetic relationships of the sugar transporters

An alternative way in which the relationship of a set of related protein sequences can be expressed quantitatively is in the form of a phylogenetic tree. The topology of the tree gives an indication of how the sequences should be grouped; the branch lengths provide some sense of the true evolutionary distances. However, the accuracy of the tree depends naturally upon the alignment of the sequences. Clearly, the most critical feature of the alignment is the positions of the gaps which are necessarily different in a multiple sequence alignment than in a dual sequence alignment. A useful method of visualising the pairwise comparison data is to apply the technique of single linkage cluster analysis. This provides a convenient representation in the form of a phenogram (Saitou and Nei, 1987) that can illustrate some of the inter-relationships between the members of a sequence group. For, not only are sequences grouped by similarity, but the maximum level of similarity between the groups is readily apparent. Further, given that the relationship between significance score and alignment accuracy is known, the phenogram in conjunction with pairwise scores can help to identify quickly which pairs of sequences may align to high accuracy.

Figure 3.2 is a phenogram of the sugar transporter family and was constructed using the single-linkage, or nearest neighbour agglomerative method (Everitt, 1980). The distances between each pair of sequences were calculated as the number of mismatched amino acids and expressed as a percentage of sequence length. Gaps in the aligned sequences were treated as other residues, with aligned gaps being counted as matches. The trees were produced directly from the University of Wisconsin Genetics Computer Group (GCG) Package (Devereux *et al.*, 1984), using routines developed by A.B. Heath (Div. of Informatics, NIBSC) to incorporate SAS (SAS Institute Inc.) software statistical clustering algorithms and graphics language.

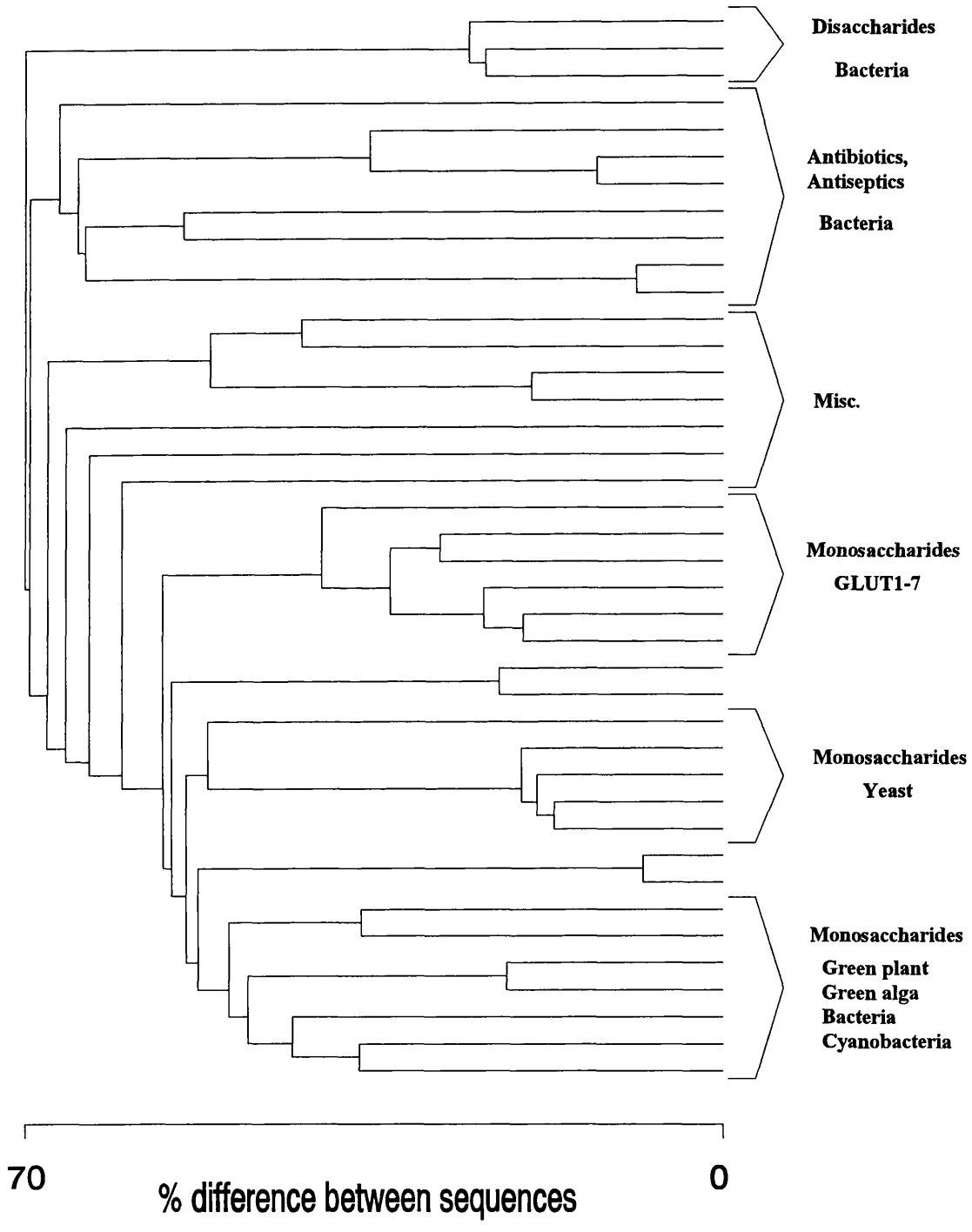


Figure 3.2 Phenogram of the sugar transporter family.

The branches of the phylogenetic tree are not intended to imply an evolutionary path or distance, as the scale represents observed differences rather than an expected number of nucleotide substitutions. The purpose of the tree is to illustrate the differences in the relationship between the individual sequences of the transporter family in a manner independent of hydropathy or secondary structural terms. It is, therefore, interesting to note that the clustering of the sequences appears to be substrate-specific.

The accuracy of any scheme of phylogenetic tree construction is crucially dependent upon proper sequence alignment, since a 'correct' alignment may not necessarily be the mathematically optimal alignment. Thus, the strategy adopted was not to omit or move a gap that occurred between two similar sequences just because an additional match might have been made with some very distantly related sequence, which is often the case when mathematically optimised alignments are generated. Thus, a greater weighting was applied to more recently diverged sequences than to distant relationships in an attempt to generate a more accurate alignment. This was deemed necessary in order to minimise the formation of untenable conclusions regarding the consensus structure obtained from subsequent secondary structure and hydropathic analyses.

3.4 Identification and location of secondary structural elements

3.4.1 Consensus secondary structure prediction

The amino acid sequences of a great variety of proteins have been derived from gene sequencing, rather than physical characterisation of the proteins themselves. Consequently, the development and application of various predictive schemes to infer secondary structure has been widespread. The Chou-Fasman (Chou and Fasman, 1974) and Robson (Garnier *et al.*, 1978) algorithms are probably the methods that have been applied most commonly to estimate the local structural elements present, namely α-helices, β-sheets,

random coils and turns. To each amino acid, a potential for α-helix, β-strand and β-turn conformation is assigned that has been determined from proteins of known three-dimensional structure. For a protein of unknown structure, the profiles of the three potentials along the amino acid sequence are compared to predict the secondary structure. The predictive power of such an analysis is limited, but in many cases it is the only piece of structural information available.

This moderate credibility is unfortunately reduced when the procedure is applied to membrane proteins (Wallace *et al.*, 1986). In general, all lipid embedded portions of membrane proteins are predicted to be in β-strand conformation which appears, in the main, not to be the case. The reason for this shortcoming is that the Chou-Fasman potential for β-strand conformation is highest for Val, Ile, Tyr, Phe, Trp and Leu. These are the most hydrophobic residues and, therefore, are found predominantly within the interior of soluble proteins as a hydrophobic core of β-strand conformation. Therefore, since the integral parts of membrane proteins are always hydrophobic, they are predicted to form β-strands.

Despite these shortcomings, the rationale for attempting to generate a consensus secondary structure prediction in the present study, was to try and form structural ideas concerning the extramembranous regions of GLUT1, namely the large central cytoplasmic loop, the *C*-terminal region and the turn/loop regions connecting the putative transmembrane helices. Turns have a very marked tendency to occur at local maxima of hydrophilicity because, by their nature, they are almost always exposed at the surface and rely heavily upon side-chain to main-chain H-bonding, which is indicative of hydrophilic side chains. Since the connecting loops between α-helices or between β-strands must contain at least one β-turn, the search for β-turns seems especially useful. An additional aid in predicting α-helices and β-strands is the condition that membrane spanning α-helices must comprise about 20 residues, whereas membrane spanning β-strands are about 10 residues long.

Consensus secondary structure predictions were based on the alignment of the sugar transporter family (Appendix) and utilised the Robson method (Garnier *et al.*, 1978) comprising the original and updated parameter sets (Garnier and Robson, 1989). The predictions were averaged using FORTRAN software (Perkins *et al.*, 1988) that produced the percentage of the sequences predicted to be in a particular conformation at each residue position. The averaging of scores with this procedure produces a superior prediction than with the analysis of single sequences. This is because hydrophilic residues appear to be located sometimes in transmembrane segments. These residues, which may be involved in substrate binding, might be expected to differ between transporters with different substrates, and so their contribution to the overall prediction would be reduced in a consensus prediction. Figure 3.3, for the sake of clarity, only demonstrates the consensus structure prediction in terms of the helix and turn conformations.

From Figure 3.3 it can be seen that there are distinctive α-helical predictions between residues 212-270 and 467-480, but elsewhere the helical predictions are erratic and inconclusive, even though biophysical techniques provide strong evidence for a highly α-helical nature of GLUT1 (Alvarez *et al.*, 1987). Indeed, although not shown in Figure 3.3, strong β-sheet predictions were observed throughout the alignment, except within the central cytoplasmic loop and the *C*-terminus. Two strong helical predictions are observed, separated by a β-turn prediction, within the central cytoplasmic loop between residues 232-236. Perhaps the most interesting features of Figure 3.3 are the predictions for the locations of turns within the family, particularly within the *N*-terminal half. Prominent predictions are seen at positions thought to be between membrane-spanning regions. These turn predictions are located at residue numbers 52, 72, 94, 154, 184 and 209 within the *N*-terminal half, whereas the turn locations are more difficult to determine from the *C*-terminal half data and assignments can only be made at positions 275, 295, 341 and 465.

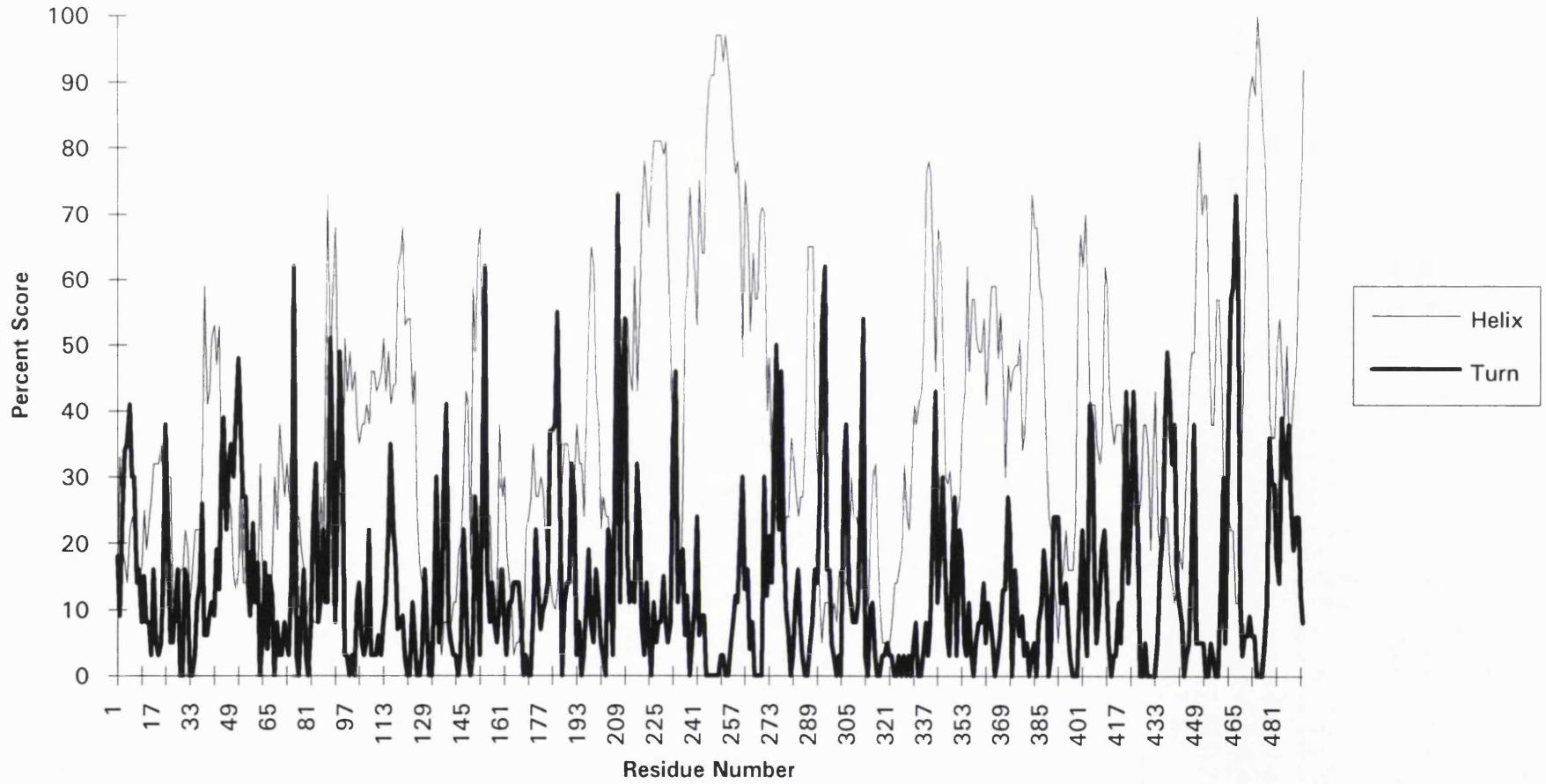


Figure 3.3 Consensus α-helix and β-turn prediction for the sugar transporter family.

3.4.2 Hydropathic analysis of the sugar transporter family

The lipid bilayer has a profound effect upon the permissible conformations, orientations and topologies of secondary structural elements, as well as on the hydrophobicity of the outer surface that interacts with the membrane. A stringent requirement of a membrane-embedded structure is complete main-chain hydrogen bonding so that α-helices, clusters of helices or cylinders of β-sheet are usual, whereas loops, other non-repetitive structures and isolated extended strands seem to be forbidden. In general terms, the folding of a protein is determined by its amino acid sequence, in conjunction with the entropy of removing hydrophobic groups from contact with the solvent. Hydrophobicity is believed to play a major role in the self-assembly of protein molecules because some amino acid residues are abundantly water soluble whereas others are not. Thus, there is a simultaneous attraction of charged and polar amino acid side chains to water and an avoidance of water by apolar side chains which are major factors in dictating the conformation adopted by the polypeptide backbone.

The question, therefore, is how to predict transmembrane segments given that secondary structure prediction in the form of Chou-Fasman analysis is inappropriate. A first step towards the deduction of the possible transmembrane distribution of a protein from its amino acid sequence was made by Kyte and Doolittle (1982) who applied the concept of hydrophobicity in a quantitative manner. As a measure for the hydrophobicity of an amino acid residue, they used the mean value of two quantities: the transfer free energy of a residue between water and vapour phase, and the interior/exterior distribution of the residue in soluble proteins of known structure. Since a membrane spanning helix requires about 20 residues, a search for hydrophobic stretches of about 20 residues along the sequence is made and these stretches are then predicted to be membrane spanning helices. Helices predicted in this way are hydrophobic over the entire surface. If a protein has only one membrane spanning segment then the entire helix must be hydrophobic and such a

membrane anchor can be identified successfully using this procedure.

Specifically, the hydrophobicity profile for a protein is a graph of the average hydrophobicity per residue against position in the sequence. Plotting the curve reveals the loci of minima and maxima in hydrophobicity along the linear polypeptide chain and is relatively simple to construct. It depends upon the choice of hydrophobicity scale and the degree of averaging (the number of consecutive residues considered as a unit). A hydrophobicity scale assigns a hydrophobicity value to each of the 20 amino acids. The profile is then computed by averaging the hydrophobicity within a moving window that is stepped along the sequence.

The importance of hydrophobicity with respect to protein structure is undisputed. However, complications regarding its application have arisen due to an abundance of scales for its determination which is testimony to the structural and functional diversity of proteins. Although the Kyte and Doolittle scale is the most widely used in the literature (Fasman and Gilbert, 1990), it has not always proved to be the most accurate. For example, it has led to wrong predictions for the folding of important families of proteins, such as cytochrome P_{450} and cytochrome *b* (Degli Esposti *et al.*, 1990). Indeed the features that discriminate between different scales include such considerations as whether the aliphatic character or surface location of proline is more important, or whether the hydrogen bonding capacity of tryptophan makes it moderately hydrophilic. The correct weighting of parameters should be different for the purposes of judging membrane insertion, antigenic potential, solubilising ability, or globular protein conformation.

Engelman *et al.* (1986) reviewed the literature on identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins, and considered the arguments in support of the notion that the helical structure will be a dominant motif in integral membrane protein organisation. Consequently, the non-polar properties of the amino acids, as they exist in a helix, were calculated using a semi-theoretical approach that combines separate

experimental values for the polar and non-polar characteristics of groups in the amino acid side chains, in order to develop a new scale (GES). The discriminating feature between this scale and other scales, is the inclusion of hydrophobic and hydrophilic components of the transfer of amino acid side chains from water to a non-aqueous environment of dielectric 2. When this scale was applied to the sequence of the RC of *Rhodopseudomonas viridis*, all putative membrane spanning helices observed in the crystal are predicted. Further, the scale fails to predict the β-sheet structure found in porin (Kleffel *et al.*, 1985).

In contrast to the Kyte and Doolittle scale, the GES scale attempts to address accurately the conformational and environmental aspects of the α-helix. The most striking difference is that the polarities of aspartic acid, glutamic acid, lysine and arginine are not as strong in the Kyte and Doolittle scale as they are in the GES scale. With respect to the hydrophobic amino acids, the two scales are in fairly close agreement. These differences have important consequences in the prediction of transmembrane helices in cases where polar or potentially charged groups are in regions traversing the membrane.

The alignment of the sugar transporter family demonstrates the presence of topologically equivalent secondary structural features of the family. Consequently, it was hoped that a consensus prediction could be achieved using the GES scale that would provide a better description of the hydropathic/transmembrane helix character of the family, than an analysis of single sequences. Figure 3.4 demonstrates the results obtained using a window of 21 residues, and it can be seen clearly that twelve membrane spanning helices are predicted. Further, the GES prediction appears to be superior to the Kyte and Doolittle plot since the hydrophilic turns at the membrane boundaries appear to be located more easily. For example, the Kyte and Doolittle scale does not identify a hydrophilic boundary between helices 3 and 4, or helices 9 and 10, whereas there are clear breaks in the hydrophobic/transmembrane helix prediction at these locations when the GES scale is applied.

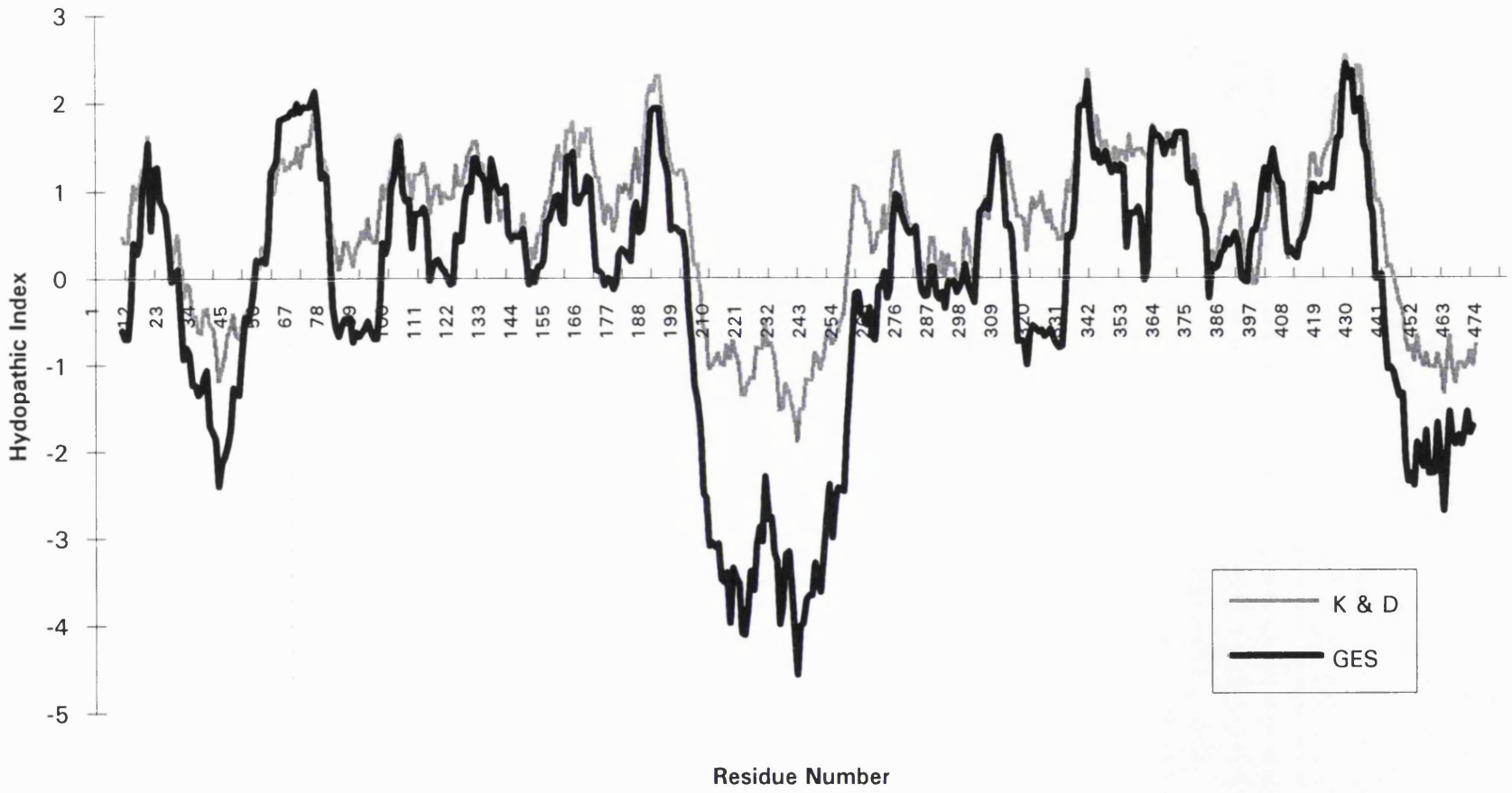


Figure 3.4 Consensus hydropathic analysis of the sugar transporter family.

3.5 Refinement of helix boundaries

A basic feature of protein sequence alignment is that it should demonstrate an alignment of those residues that are performing equivalent structural roles in the proteins. Each member of the sugar transporter family has been predicted to possess the twelve transmembrane helix architecture solely on the basis of hydropathic analysis. Thus, for an alignment that attempts to arrange sequences thought to be comprised of similar structural elements, it was necessary to be fairly confident about the likely locations of the boundaries of the membrane spanning regions since the principle secondary structural elements are believed to be the helices crossing the membrane. An algorithm to aid this determination was provided by a recent method that simultaneously takes into account the prediction of transmembrane secondary structure and the location of topogenic signals (Jones *et al.*, 1994).

This algorithm possesses an advantage over pure hydrophobicity profiles by virtue of its consideration of topogenic signals. That is, a set of statistical tables was compiled from well-characterised membrane protein data that show a definite bias towards certain amino acids existing at different locations with respect to a cell membrane. Consequently, the preference for positively charged residues to be located within the intracellular loops was actively used to guide the topological prediction. In addition, the intriguing abundance of tryptophan and tyrosine residues in outside locations of the RC and bacteriorhodopsin, respectively, as noted by Schiffer *et al.* (1992), was also used in generating the topogenic scoring system. Thus, the aim was to obtain a topological prediction for each sequence that would enable a more enhanced refinement of the membrane boundaries than if hydropathy analysis alone had been performed.

Each sequence of the sugar transporter family was passed through this algorithm and the resultant predicted membrane topologies are described in Table 3.2. This table provides the helix topology prediction for the mammalian,

yeast and bacterial transporters and shows that, in the main, a twelve membrane spanning helix topology is predicted. There are, however, instances of underprediction that do appear difficult to explain. For example, putative helices 7 and 11 of rabbit GLUT1 are not predicted even though there is only one residue difference per helix between these helices and the corresponding ones of human GLUT1. There is also a gross underprediction for the rat GLUT7 sequence where putative helices 6, 7, 11 and 12 are not predicted. Interestingly, the rat SV2 sequence is predicted to possess the full complement of putative transmembrane helices even though its sequence identity to all other members of the family is very low.

Of the yeast transporters, only the HXT1 sequence is predicted to be comprised of twelve membrane-spanning helices, whereas all the others are underpredicted to different extents. Once again, helices 7 and 11 are frequently not predicted, a finding that probably stems from their highly amphipathic nature, that is the abundance of glutamine and asparagine residues, respectively. It seems likely that these residues are being interpreted by the algorithm as topogenic signals rather than as potential inter-helix stabilising or pore-lining residues. This might also account for the misprediction of helices 4 and 5 of the cit+, citA, kgtP and bap3 transporters, where comparison with the other family members reveals a glutamine in the centre of helix 4 and two additional glutamate residues within the *N*-terminal half of helix 5.

In addition to the above examples of underprediction, it can be clearly seen that there are several cases where additional helices have been predicted to exist, in particular for the proteins that provide antibiotic resistance. This was not an unexpected finding since the hydropathic profiles for these proteins have been variously interpreted to be consistent with 12, 13 or 14 transmembrane helices, depending upon the method of analysis (Rouch *et al.*, 1990, Levy, 1992).

Table 3.2 Predicted membrane spanning helix boundaries

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Human GLUT1	12-28	64-87	96-112	121-144	154-176	185-207	272-293	307-328	336-358	367-391	402-422	432-450
Rabbit GLUT1	12-28	64-87	96-112	121-144	154-176	185-207		307-328	338-358	366-390		432-450
Rat GLUT1	12-28	64-87	96-112	121-144	154-176	185-207	272-293	307-328	336-358	376-391	402-422	432-450
Pig GLUT1		23-46	55-71	80-103	113-135	144-166	231-252	266-287	295-317	326-350	361-381	391-409
Mouse GLUT1	12-28	64-87	96-112	121-144	154-176	185-207	272-293	307-328	336-358	367-391	404-422	432-450
Human GLUT2	8-24	95-119	127-144	152-176	194-212	220-239	304-325	339-360	368-390	399-423	434-457	464-482
Rat GLUT2	8-24	93-117	125-141	150-174	192-211	218-237	302-323	337-358	366-388	397-421	433-455	462-480
Mouse GLUT2	8-24	94-118	126-142	151-175	193-212	219-238	303-324	338-359	367-389	398-422	434-456	463-381
Rat GLUT7	8-24	93-117	125-141	150-174	184-206			337-358	366-388	397-421		
Human GLUT3	9-26	62-86	94-110	119-143	160-178	186-205	270-291	305-326	334-355	364-388	400-420	427-448
Mouse GLUT3	10-26	61-85	94-111	119-143	160-177	185-204	269-290	304-325	333-354	363-387	399-419	426-447
Human GLUT4	20-36	80-102	111-129	137-160	170-192	201-223	288-309	323-344	352-374	383-407	418-438	448-466
Rat GLUT4	20-36	80-102	111-129	137-160	170-192	201-223	288-309	323-344	352-374	383-407	418-438	448-466
Mouse GLUT4	22-38	82-104	113-131	139-162	180-198	206-225	290-311	325-346	354-376	385-409	419-439	449-467
Human GLUT5	13-30	69-93	100-117	127-151	160-182	192-213	278-299	316-336	343-365	375-399	411-433	440-460
Rat SV2	27-51	67-87	95-113	124-141	154-176	196-212	307-329	347-370	378-395	402-425	436-459	466-482
Yeast SNF3	23-42	72-93	101-117	124-148	159-179	194-213		320-341	349-371	379-402	417-438	451-471
Yeast GAL2	25-44	79-96	108-127	134-155	165-183	200-219		327-344	353-372	391-414	436-453	460-476
Yeast HXT1	25-44	79-97	107-127	134-155	165-183	200-219	304-320	327-344	352-372	392-415	430-453	460-476
Yeast HXT2	22-44	79-97	107-125	133-155	165-183	202-219			352-376	391-415	429-453	460-477
Yeast RAG1	25-44	79-96	107-127	134-155	165-183	200-219		327-344		394-417	438-455	462-478
Yeast MAL61		72-92	106-124			206-227	292-310	325-346	355-374	387-408	427-443	455-471
Yeast Lac12	26-50	71-91	99-116	128-144	157-177	192-211	283-305	321-342	351-370	383-399		448-469
Yeast ITR1	22-40	70-90	98-116	127-144	156-178	187-206		314-334	342-366	373-397		453-471
Yeast ITR2	22-42	70-90	99-116	127-144		187-206		314-334	342-366	383-407		453-471
Hup1	23-43	87-106	113-129	136-153	179-195	204-223	299-316	323-344	352-374	391-414	426-450	457-476
STP1	19-39	82-103	111-127	134-158	169-185	203-222	286-310	321-342	358-375	388-412	428-448	455-474
glcP	16-33	54-78	87-108	115-133	150-170	187-206	263-287	304-325	333-357	370-394	405-426	433-457
Pro-1	17-38	130-151	158-174	183-207	220-243	262-280	312-331	350-368	375-396	408-432	451-467	484-505
qa-y	21-37	72-89	100-117	127-149	161-179	194-215		323-346	354-373	388-406		460-482
qutD	21-37	95-114		128-145		190-211		321-343	351-370	386-402		456-475
AraE	23-47	63-82	90-108	115-137	149-166	179-198	258-289	297-318	329-346	359-383	395-419	428-446
GalP	16-40	56-77	84-100	107-130	142-159	172-191	251-273	290-311	320-342	349-373	385-409	418-436
XylE	9-26	56-78	88-104	129-152	169-186	201-220	270-287	313-336	346-364	371-395	407-424	443-460
glf	13-35	56-80	87-104	118-142	159-176	197-219	259-278	305-323	335-356	364-386	396-417	434-451

Table 3.2 cont.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
cit+	35-52	60-84	92-110			192-211	243-265	281-301	310-329	336-360	375-396	408-424
citA		49-73	85-103			185-204	236-257	274-294	303-319	331-353	364-388	400-418
kgtP	38-54		96-116		166-184	195-214	244-262	280-302	312-330	337-360	371-387	405-423
bap3		57-73	93-110			192-211	246-262	281-303	310-327	334-358	371-394	405-421
CmlA	10-30	53-72	81-98	105-127	139-162	170-187	217-241	249-266	281-298	307-328	343-366	374-391
TetC	7-31	45-66	75-97	104-121	133-157	164-181	211-235	243-262	279-297	304-320	340-361	369-385
TetA	7-31	45-66	75-97	104-121	133-154	163-181	211-235	248-272	279-297	304-320	340-361	369-385
TetB	7-30	43-62	73-95	102-119	130-152	161-179	212-234	244-266	277-295	302-324	336-357	367-388
norA	7-30	40-62	69-85	93-117	129-151	159-177	203-221	240-259	269-286	293-309	327-351	359-375
Bmr	8-28	38-62	72-88	96-120	131-154	162-180	202-224	243-262	272-289	296-312	337-354	362-379

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
TetL	15-33	53-69	81-98	105-129	140-163	170-186	201-217	224-240	256-279	294-317	324-343	350-373	388-412	432-451
pNS	15-33	53-69	81-98	105-129	140-163	170-186	201-217	224-240	256-279	294-317	324-343	350-373	388-412	432-451
BS9	15-33	53-69	81-98	105-129	140=163	170-186	201-217	224-240	256-279	292-316	323-342	349-365	388-412	432-451
TetK	11-28	53-69	81-98	105-129	140-163	170-186	201-217	224-240	256-279	298-316	323-344	353-369	388-410	
Mmr	28-52	65-85	93-111	125-143	152-175	183-199	210-230	237-258	282-303	316-333	344-362	370-394	415-431	438-457
qacA	18-42	57-78	86-102	109-132	145-168	176-195	210-227	238-255	276-297	314-334	341-360	369-392	412-429	438-461
ACT		43-67	80-100	109-126	137-161	173-192	205-223	232-250	274-297	317-333	340-360	371-395	415-432	
ATR1	26-42	66-83	94-111	129-148	156-180	189-209	224-241	252-271	294-313	325-349	356-374	386-407	417-441	461-479
LacY	10-34	46-66	75-96	103-125	145-162	169-187	222-239	260-283	291-313	321-337	349-370	385-409		
LacY	15-39	51-71	81-100	108-130	150-167	175-191	227-244		265-288	296-318	386-405			
RafB	13-37	49-69	79-99	106-128	148-165	172-190		225-242	263-286	295-312	386-410			
MelB		32-50	75-96	103-127	146-163	177-194		229-253	266-282	293-315	322-346	377-393	402-426	

Although it is possible that this alternative procedure is more accurate than hydropathy plots alone, it is reasonable to suggest that these data illustrate the difficulty involved in forming hypotheses from predictions when not all of the rules of protein folding and assembly are known, particularly when the statistical tables are generated from proteins of unknown structure. It is, therefore, not surprising that there is a consensus of prediction in support of the notion that these proteins possess twelve membrane spanning helices. Also, even where there are instances of underprediction, subsequent helix locations do appear to correlate with those for the other members of the family, which suggest would that the weighting for the topogenic signals is perhaps inadequate to be applied generally. This fact serves to illustrate the need for care when interpreting data from just one particular methodology, especially when that methodology is speculative.

The principal purpose for obtaining consensus predictive data regarding the secondary structure and hydropathic profiles of the sugar transporter family, was to enable a refinement of the existing topological model of the transporter (Figure 1.2), with respect to the locations of the transmembrane segments and the interconnecting turn/loop regions. On the basis of the data presented, subtle revisions of the current transporter model have been made and are shown in Figure 3.5.

The most notable revisions have been made in the intracellular regions of the transporter, principally because of the greater degree of accuracy in the prediction of soluble protein structure. Two, long α-helical stretches, separated by a sharp β-turn between residues 232-236, have been imposed on the central cytoplasmic loop and the *C*-terminus. In addition, the membrane boundaries have been altered to account for the better predictions of hydropathy and the turn/loop regions. However, the consensus predictions have not resulted in a vast modification of vast modification to the original (Figure 1.2). It was hoped, therefore, that an additional prediction scheme would justify, or enhance, these alterations. This was found in a periodicity analysis of the transporter family.

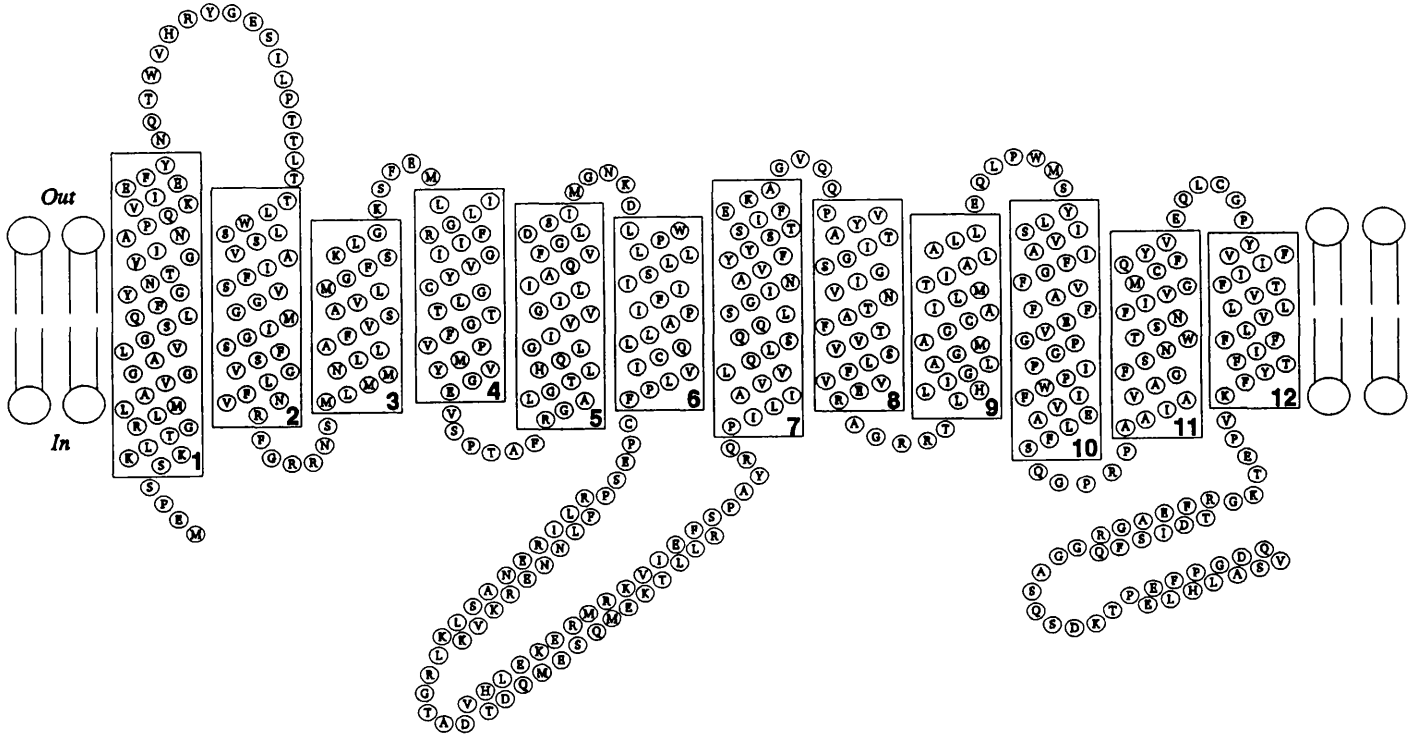


Figure 3.5 Refined topological model of GLUT1 based upon several consensus predictive schemes.

3.6 Periodicity analysis of the sugar transporter family

If a membrane protein consists of several helices that are clustered to form a pore for the translocation of hydrophilic solutes, the helices will not be hydrophobic over their entire surfaces. On the side facing the lipids they will be hydrophobic, but on the side facing the pore, they are likely to be at least partially hydrophilic. Such helices are called amphipathic and, because of the hydrophilic nature of the substrates, they are envisaged to play a major role in the structure of the sugar transporters. The oldest method of illustrating such helices is to construct a helical wheel (Schiffer and Edmundson, 1967), but a more quantitative plot was proposed by Eisenberg *et al.* (1982) who introduced the hydrophobic moment. These plots are well suited to demonstrate the amphipathic nature of an individual helix, but not appropriate to identify amphipathic helices on an amino acid sequence of several hundred residues, such as GLUT1. Consequently, an additional method for evaluating the consensus structure of the sugar transporter family was required that possessed the ability to detect the relative accessibility of residues to the surrounding solvent, plus helical periodicity.

One of the features of an amino acid alignment is that information can be gained regarding the constraints that have been placed upon each residue throughout evolution. Such constraints become apparent in the different character that is exhibited between residues buried within the core of a protein and those found at the surface. In essence, buried residues are more conserved than those that are exposed. Moreover, the cores of water soluble proteins tend to be more hydrophobic than their surface positions which are in contact with aqueous solution, whereas the cores of membrane proteins tend to be more polar than their lipid facing exteriors (Rees *et al.*, 1989). These differences can be used to predict the extent to which each position in a protein is buried by examining the residues present at each position of a sequence alignment.

The periodicity of residues on the face of an α-helical structure that are exposed to a membrane, coupled with the increased sequence variability of exposed residues suggests the possibility of identifying exposed residues by analysing the sequence alignments of homologous proteins. Assuming (a) that the sequence represents a transmembrane helix and (b) that the helix is positioned on the exterior of a helix bundle, the residues in contact with the lipid bilayer may be identified from the pattern of hypervariable positions occurring with a periodicity of about 3.6 residues in a family of sequence alignments. Even a superficial examination of the sugar transporter sequence alignment reveals an apparent periodicity of certain residues, perhaps the most prominent of which is the presence of glycine residues separated by three or four residues in helices 2, 3, 4 and 9.

This type of approach has been used previously with hydrophobicity scales (Cornette *et al.*, 1987) and variability characteristics (Donnelly *et al.*, 1989) found within amino acid alignments. In addition, the relative directions of the conserved and variable faces of a membrane-spanning helix can be used to predict whether an exposed face is in contact with a lipid or aqueous environment. Another method of predicting the faces of helices in contact with the lipid bilayer from sequence alignments, is to predict the accessibility of each residue position in the alignment from the substitution pattern at that position. The structural environment of an amino acid residue provides constraints on the evolutionary diversity of that residue. The amino acid substitution patterns are characteristic of their structural environment so that the mutational properties of an exposed residue are different to those of a buried one. Consequently, it is possible to predict the structural environment of residues from a sequence alignment, for which only a substitution pattern is known.

Environment-dependent substitution tables derived from accessible and inaccessible residues in aligned protein structures (Overington *et al.*, 1992) are used to predict whether the substitution patterns in sequence alignments are more typical of buried or exposed residues. Substitution tables derived from

residues that are accessible to a lipid environment are used to predict and orientate transmembrane helices (Donnelly *et al.*, 1993).

Fourier transform methods provide a quantitative approach for characterising the periodicity of conserved and variable residues in a family of aligned sequences (Komiya *et al.*, 1988). First. the variability profile is constructed from aligned sequences of the helical regions. Next, the residue positions with greatest variability consistent with an α-helical periodicity are determined by fitting a cosine curve to the variability profile. The residue positions for which this Fourier series has the greatest amplitude correspond to the most variable positions. Calculation of these positions for the 11 RC transmembrane helices shows a strong correlation between the most variable positions and the exposed positions (Rees *et al.*, 1989). The variablity profile may also be used to predict the presence of α-helical segments which are usually identified from hydropathy plots or hydrophobic moment analysis.

The periodicity of hydrophobicity (H), conservation (C) and substitution (S) for the putative membrane spanning helices of the sequence alignment were calculated using PERCON, PERHYD and PERSCAN software (Donnelly *et al.*, 1989, 1993) at University College London. Table 3.3 provides some of the data obtained.

It is clear from the data presented in Table 3.3 that the periodicity of residues throughout the putative transmembrane spanning domains indicates α-helical structure for each of the regions analysed. Although not shown in Table 3.3, the programs also calculate a property of the alignment termed the alpha periodicity (AP). AP is analogous to ψ used by Komiya *et al.* (1988) and to the amphipathic index AI used by Cornette *et al.* (1987), although the precise boundaries of the helical regions of the power spectrum differ in the latter. Komiya *et al.* (1988) suggest that a value of AP greater than 2 indicates that the helical periodicity is significant. Larger values of AP correspond to a greater fraction of the P(ϖ) curve in the α-helical region. If peripheral helices, in a helix

Table 3.3 Predicted number of residues per turn for each putative transmembrane helix.
Periodicity in the patterns of residue substitution (**S**), conservation (**C**) and hydrophobicity (**H**) were calculated from the alignment of the sugar transporter family.

	Family I			**Families I, II, III and IV**		
	S	C	H	S	C	H
1	3.33	3.40	3.43	3.00	3.50	3.33
2	3.60	3.36	3.71	3.75	3.46	3.60
3	3.64	3.56	3.46	3.53	3.19	3.50
4	3.50	3.53	3.71	3.19	3.43	3.71
5	4.00	3.33	3.87	3.64	3.33	3.83
6	3.46	3.40	3.43	3.64	3.36	3.40
7	3.03	3.64	3.27	3.21	3.13	3.40
8	3.53	3.43	3.36	3.46	3.13	3.53
9	3.36	3.24	3.30	3.43	3.00	3.71
10	3.64	3.36	3.00*	3.67	3.60	3.00
11	3.71	3.05	3.79	3.71	3.10	3.56
12	3.08	3.08	3.67	3.43	3.24	3.50

* AP < 2

AP is the ratio of the extent of the periodicity in the helical region of the spectrum compared with that over the whole spectrum.

bundle, have greater AP values than core helices, then it is consistent with the analysis that membrane exposed residues are more poorly conserved than buried residues. Consequently, the AP provides a good measure of the surface exposure of the α-helix and is helpful, therefore, in deriving information from sequence data about the three-dimensional structure.

All AP values for each putative helix of the sugar transporter family, whether determined for Family I alone or the entire sugar transporter family, were greater than 2, except for the one noted in Table 3.3. Consequently, it is predicted that each of the regions of sequence analysed, that is, the putative transmembrane domains, demonstrate a significant periodicity that is consistent with α-helical structure. The most probable explanation for the lower AP value for the periodicity of hydrophobic residues in helix 10 is the presence of the highly conserved GPGPIPW motif. It seems possible that this region, perhaps, represents a deviation from the normal helical structure or, a systematic shift in exposed residues due to interactions with adjacent helices.

In order to generate information regarding the faces of helices which may be exposed to the surrounding lipid bilayer, or those faces in direct contact with adjacent helices, the periodicity of patterns of substitution, conservation and hydrophobicity were calculated. Vectors for each residue position were calculated from statistical tables derived from sequence alignments and summed for each putative transmembrane helix analysed. The resulting moments are illustrated on Figures 3.6 and 3.7 for the assessment of Family I alone, and the whole sugar transporter family, respectively.

For a membrane protein possessing a bundle of α-helices, it would be expected that the most hydrophobic residues would be on the outside of the helices. These residues would also represent the least conserved. In contrast, most residues on the inner faces of the helices, that is, in contact with other helices or forming a hydrophilic pore, would not be expected to be hydrophobic, but would show the greatest degree of conservation. Further, it would be expected that the mutational

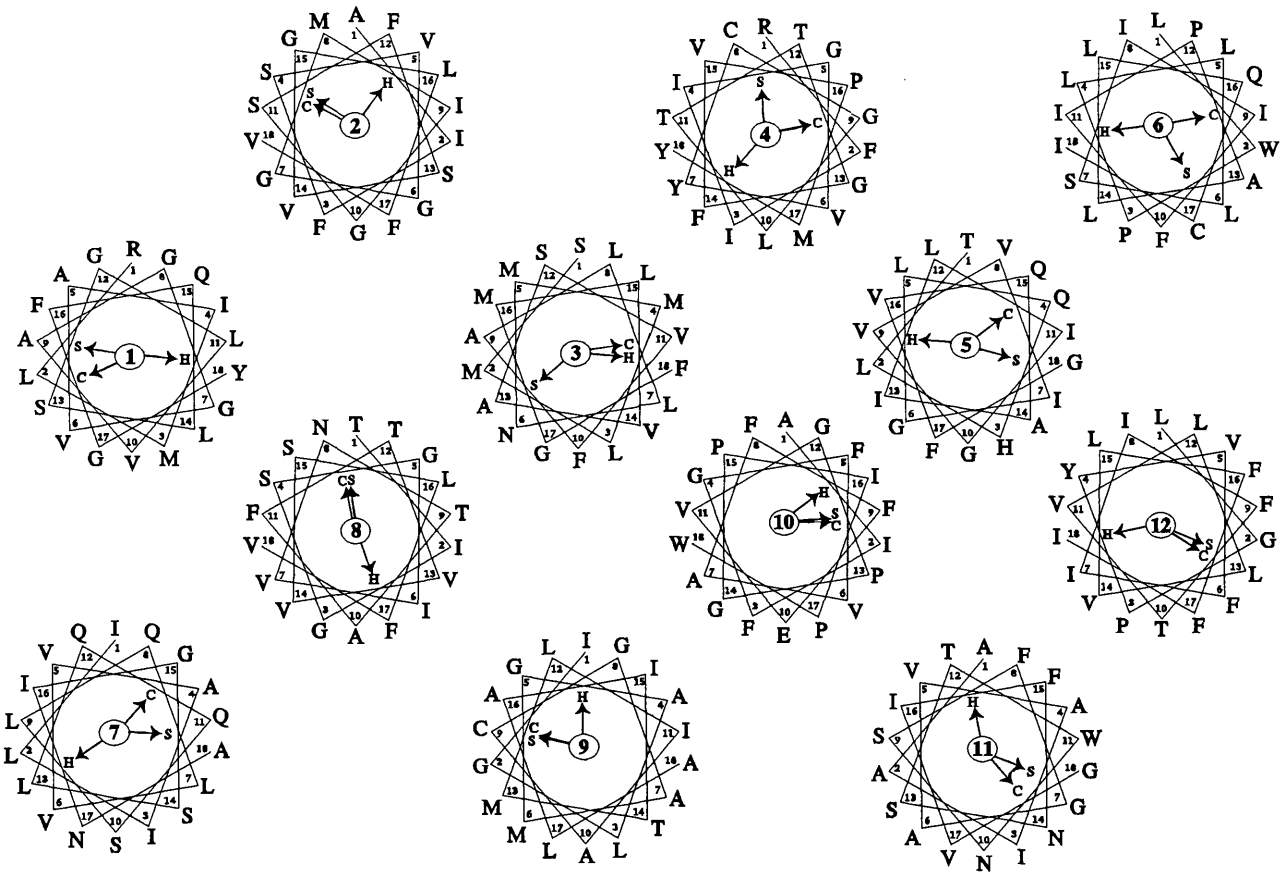


Figure 3.6 Patterns in periodicity of Family I.

Moments of substitution (**S**), conserved (**C**) and hydrophobic (**H**) patterns superimposed upon helical wheel plots for the putative transmembrane helices of Family I.

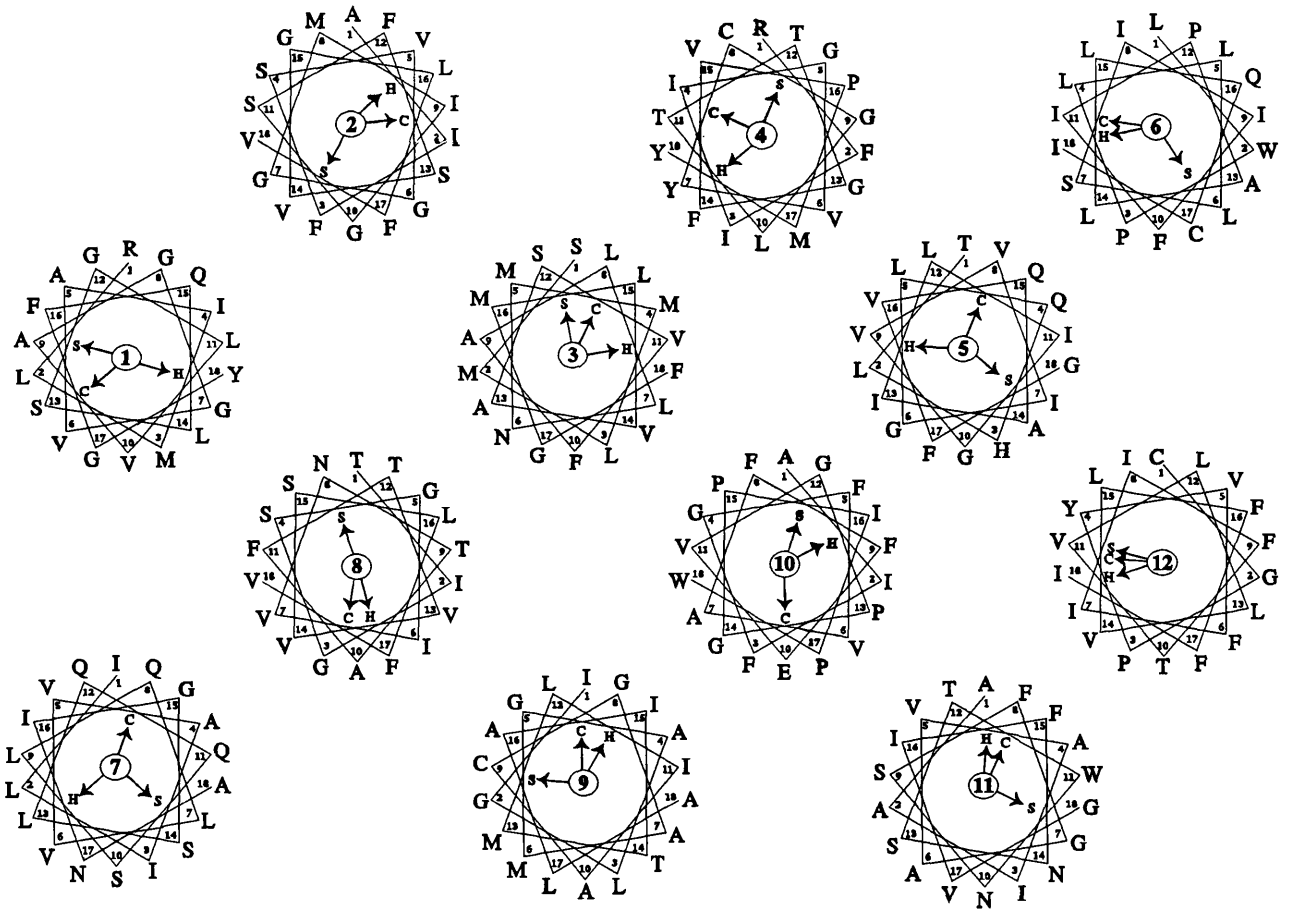


Figure 3.7 Patterns in periodicity of the sugar transporter family.

Moments of substitution (**S**), conserved (**C**) and hydrophobic (**H**) patterns superimposed upon helical wheel plots for the putative transmembrane helices of Families I, II, III and IV.

character of the lipid-exposed faces would be different from the inner faces, which would be less susceptible to mutation. Therefore, for a uniform, amphipathic helix, the hydrophobic moment would be in the opposite direction to the moments of substitution and conservation.

From Figure 3.6, the analysis of moments for Family I alone, it can be seen that in all but helices 3 and 10, the moments of substitution and conservation point to the same face of each helix, and that the hydrophobic moment identifies a separate face. Although the above principles are visible in the analysis of the whole sugar transporter family (Figure 3.7), the data are more difficult to interpret, which probably reflects the inclusion of distantly related sequences. For example, helix 12 of the Family I analysis conforms to expectation, but the analysis of this helix for the whole transporter family is clearly difficult to explain. However, the differences in moments between the analyses could reflect important features of individual helices. For example, helices 2 and 3, and 8 and 9 demonstrate equivalent moments in the patterns of residue conservation and hydrophobicity that are not apparent from the analysis of Family I. At the *C*-terminal ends of these helices, there are highly conserved motifs which are possibly involved in the formation of salt bridges. It is possible that the presence of the salt bridge affects local secondary structure, perhaps by tightening or bending helices, which is reflected in this type of analysis.

Since the highest degrees of homolgy exist between the members of Family I, the moments for this series of sequences only were used to assemble the helices of GLUT1 into a model for this protein. Using the moments of substitution and conservation to identify buried faces of helices, an arrangement of helices for GLUT1 was produced and is shown in Figure 3.8. This model of GLUT1 shows the protein to be composed of two bundles of six helices, arranged with a two-fold axis of symmetry, as proposed by Baldwin (1993).

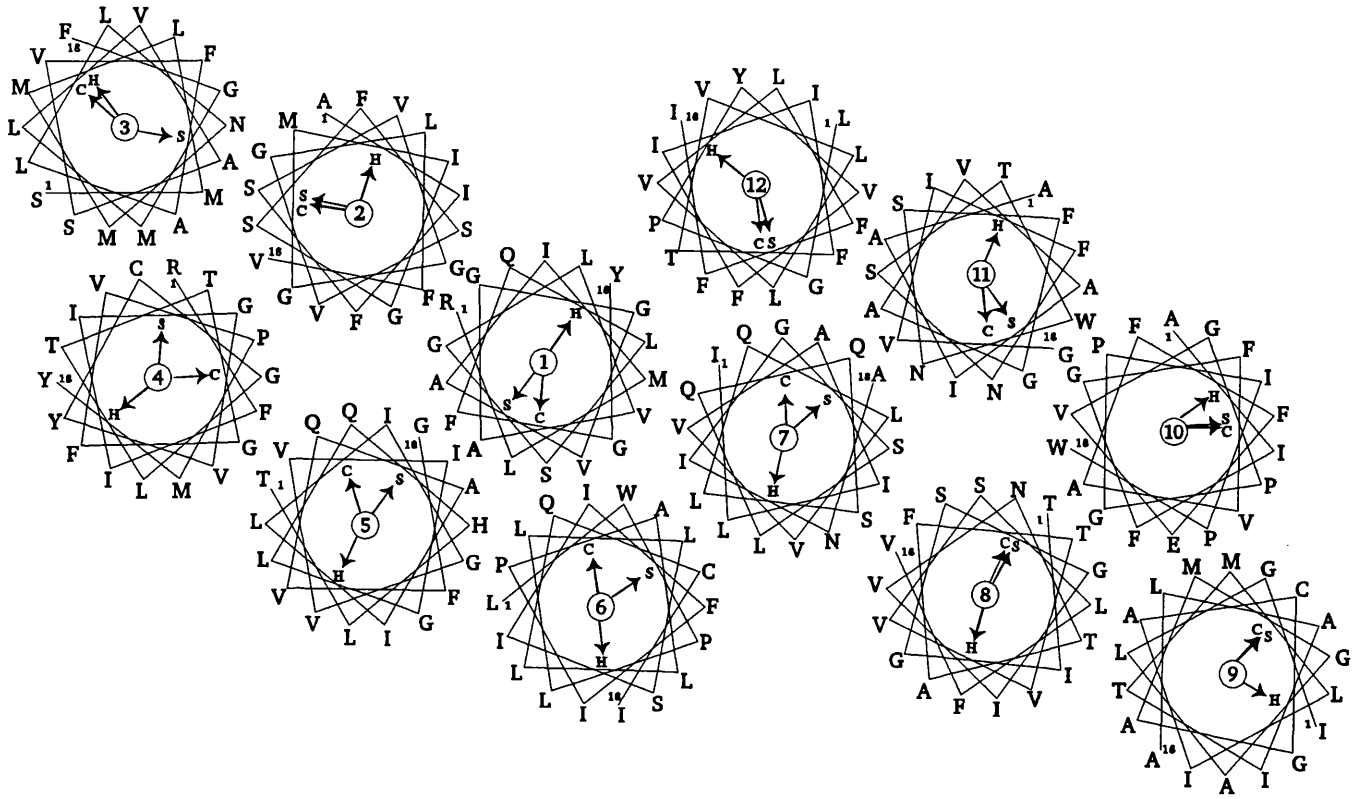


Figure 3.8 Possible arrangement of helices of GLUT1 in the lipid bilayer.

3.7 Three-dimensional modelling of GLUT1

Having refined the two-dimensional topology of GLUT1, and also generated a possible arrangement of the transmembrane helices within the membrane, it was then feasible to incorporate all of these features into a three-dimensional model of GLUT1. However, the danger of placing too much emphasis upon such predictions is extreme, thus it was decided to generate a non-minimised tertiary structure of GLUT1 purely for the purpose of experimental design.

The sequence similarities of the sugar transporter family indicate that they probably have similar three-dimensional structures. It follows that, at the molecular level, the mechanisms of these passive transporters, active symporters and antiporters must share many features. Studies of particular members of the family have also added support to the topological model (Figure 3.5). For example, chemical labelling of the native and mutated tetracycline transporter has confirmed the cytoplasmic location of the *N*-terminus and the loop connecting transmembrane helices 2 and 3 (Eckert and Beck, 1989, Yamaguchi *et al.*, 1990). Protease digestion experiments on this protein have also provided preliminary evidence for the cytoplasmic locations of the loops connecting helices 4 and 5, and 10 and 11 (Eckert and Beck, 1989). Analysis of a series of 36 *lac* permease-alkaline phosphatase fusions by Calamia and Manoil (1990) has also provided strong evidence in favour of the twelve-helix architecture. The information obtained from experiments such as these can be used to construct models for the three-dimensional arrangement of the membrane spanning helices within the transport proteins.

The 3-dimensional model of GLUT1 was produced with a Silicon Graphics Indigo machine, using InsightII software (Biosym). In essence, the predicted secondary structural features indicated in Figure 3.5 were imposed on the human GLUT1 amino acid sequence, and the putative transmembrane domains were assembled into the arrangement illustrated in Figure 3.8. Although the model was not to be subjected to energy minimisation, it was intended to pack the helices as close to each other as possible. In order to facilitate this, the common four-helix bundle was

used as a template for helices 1, 6, 7 and 12. The rest of the model was then constructed around this core.

The model is illustrated in Figure 3.9 and corresponds with the measured dimensions of lacY (Li and Tooth, 1987). However, this model is obviously only one of many possibilities which must be tested by direct experimentation. The helices have been arranged so that those connected by short loops are adjacent, for example 2 and 3, and 9 and 10. Helices 7 and 11 have also been placed together because recent mutagenesis studies on *lac* permease have suggested that Asp_{237} and Lys_{358} of these two helices are close together, possibly forming a salt bridge (King *et al.*, 1991). It is possible that helices 7, 8 and 11, which are highly amphipathic, might be involved in the formation of a substrate-binding cleft in the *C*-terminal domain.

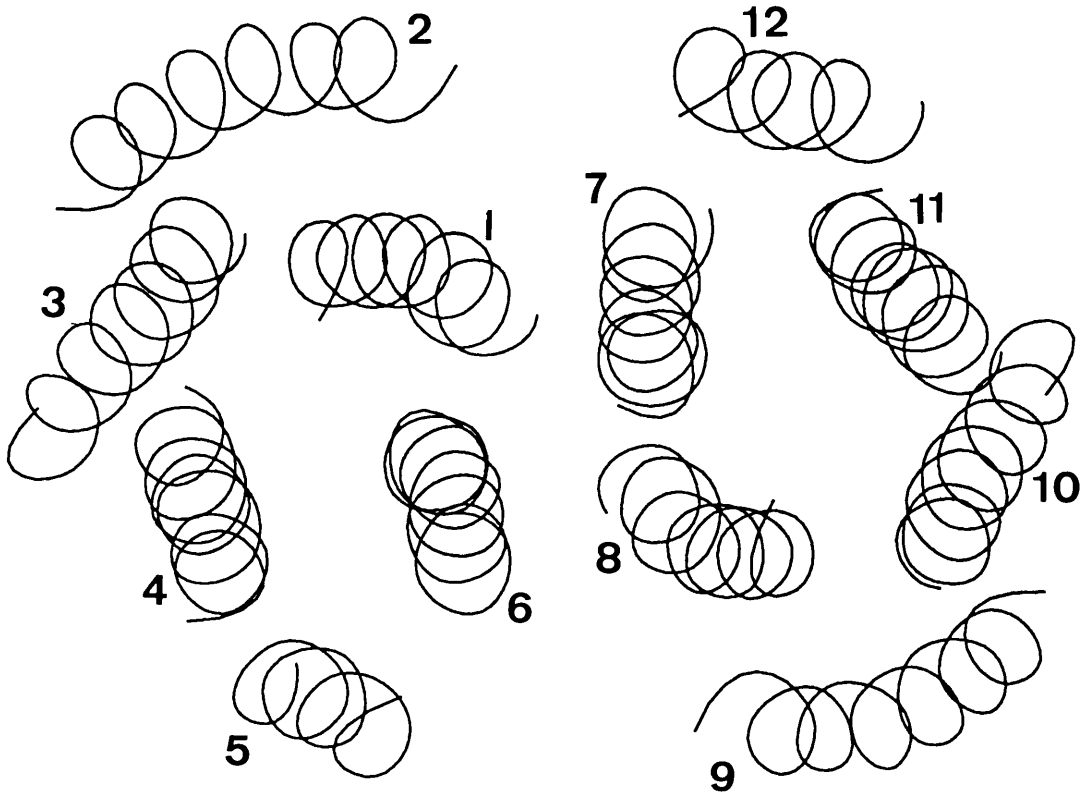


Figure 3.9 Three-dimensional model of GLUT1.

3.8 Discussion

The evidence described in this chapter provides strong support for the notion that seemingly dissimilar transporters have a common origin. The three-dimensional structures of these proteins are likely, therefore, to be similar with relatively subtle structural differences accounting for the recognition of different substrates. This implies a fundamentally similar mechanism of transport at the molecular level, despite the apparent differences in vectorial mechanism, that is, symport versus antiport versus uniport, and import versus export.

A multiple sequence alignment of the sugar transporter family was constructed and assessed by a variety of predictive algorithms in an attempt to derive as much information as possible about the secondary and tertiary structure of GLUT1. Possibly the best argument in favour of the widely held view that these transporters belong to a family of proteins consisting of twelve membrane-spanning helices, is the ability to align the homologous sequences using these regions as a guide. That is, no insertions or deletions are seen in the predicted transmembraene regions but are evident in the external turn/loop regions. However, it was required to proceed as far along the path from sequence to structure as possible and so, consensus secondary structure prediction, hydrophobicity and periodicity analyses were used to refine the two-dimensional model of GLUT1, and then to build a speculative 3-dimensional model of the protein. The purpose of this model is purely to aid the design of experiments and, hence, test the accuracy of the models.

Since all of the rules governing protein folding are not known, there is a long way to go before a model of the tertiary structure of an integral membrane protein can be predicted with any degree of precision. However, the requirement of methods to generate such models is undisputed, as they facilitate the design of experiments testing their features. Experiments to investigate certain features of the model for GLUT1 illustrated here are described in the following chapters.

CHAPTER 4. PROBING THE TOPOLOGY OF GLUT1 USING A MEMBRANE-IMPERMEANT REAGENT

4.1 Introduction

Due to a very limited amount of crystallographic data for integral membrane proteins, models describing their function are heavily dependent upon the predicted two-dimensional topology for such proteins. The studies described in Chapter 3 demonstrate that it is possible to build up a model of a membrane protein from the information contained within the amino acid sequence alone. However, since this type of data is largely hypothetical, a number of experimental approaches have been adopted to test models of membrane protein topology in the absence of diffraction analysis. For example, the location of sites relative to the membrane can be identified by their interaction with other proteins of known cellular location, a strategy which can be extended further by studies using vectorial antibody binding or proteolysis. Alternatively, reagents that react from within the lipid bilayer can be used to identify sequences spanning the membrane. Several such methodologies have been applied to the study of GLUT1 and some aspects of the topology have been confirmed.

Direct evidence in support of the predicted highly α-helical nature of the sugar transporter family has come from CD (Chin *et al.*, 1987) and infrared (Alvarez *et al.*, 1987) spectroscopic studies of purified GLUT1. The cytoplasmic location of the *C*-terminus of GLUT1 has been demonstrated by studies employing vectorial proteolytic digestion and site-directed antibodies as topological probes (Cairns *et al.*, 1987, Davies *et al.*, 1987). Similarly, evidence has been obtained for the cytoplasmic location of a large, hydrophilic loop connecting putative helices 6 and 7 of GLUT1, and for the extracellular location of the loop connecting helices 1 and 2, which bears the site of glycosylation in this glycoprotein (Cairns *et al.*, 1987, Davies *et al.*, 1990).

A feature of *N*-linked glycosylation is that it occurs only on one side of the

membrane, corresponding to the lumenal surface of the endoplasmic reticulum and Golgi apparatus, and this has been utilised recently in a glycosylation scanning mutagenesis procedure to confirm the locations of extramembranous regions of the transporter. In these studies (Hresko *et al.*, 1994), an epitope bearing an *N*-linked glycosylation site was introduced into each of the hydrophilic, putatively extramembranous domains of an aglyco-GLUT1 mutant. The cytoplasmic or exofacial orientation of each hydrophilic domain was then inferred from the glycosylation state of the corresponding insertion mutant, and it was found that the data obtained from expressing these mutants in *Xenopus* oocytes were in complete agreement with the proposed twelve-helix model. Further, 2-deoxy-D-glucose uptake studies revealed that insertion of the epitope into the *N*-terminus, the large central cytoplasmic loop, or the second, third, or fifth exofacial loop had no dramatic effect upon the activity of the transporter, suggesting that these regions probably play no role in sugar transport.

The reactivity of certain exofacial residues has been used previously to gain information about the topology of the transporter. The suggestion that Cys_{429} is the exofacial sulphydryl residue labelled with bis(maleimidomethyl) ether-L-[^{35}S]-cysteine was inferred from the apparent molecular weights and immunoreactivity of fragments of the labelled transporter resulting from chemical cleavage (May *et al.*, 1990). This was confirmed by site-directed mutagenesis experiments involving the application of a membrane-impermeant thiol-group specific reagent pCMBS (*p*-chloromercuribenzenesulphonate) to either the external or internal face of the transporter expressed in *Xenopus* oocytes. It was demonstrated that the residues involved in the inhibition of sugar transport by pCMBS are the exofacial cysteine residue at position 429 and the endofacial cysteine residue at position 207 (Wellner *et al.*, 1994).

A method which relies upon the reactivity of the protein but uses an alternative scheme for detection has also been applied to study the topology of GLUT1. Deziel and Mau (1990) have used the strength of the biotin-avidin interaction to show that, after trypsinisation of the transporter into two domains,

preparations reacted with galactose oxidase/biotin hydrazide were labelled on a glycosylated fragment of apparent M_r 25,000, but not on a carbohydrate-free fragment of apparent M_r 19,000.

However, although these data are in agreement with the model of the sugar transporter that contains twelve transmembrane α-helices, and locates both the *N*- and *C*-termini on the cytoplasmic face of the membrane (Figure 3.5), there is very little further direct evidence for such a predicted topology. The aim of the present study, therefore, was to provide additional evidence in support of this model through the use of a membrane-impermeant derivative of biotin. This amino-group-specific biotinylating reagent, sulphosuccinimidyl-6-(biotinamido)hexanoate (NHS-LC-Biotin, Pierce), is depicted in Figure 4.1, together with an illustration of its reaction with a protein. The presence of the sulfo group renders it water soluble and it reacts with unprotonated amino groups, that is, α- and ε-amino groups at alkaline pH, to yield an amide bond, releasing *N*-hydroxysulfosuccinimide. Using this reagent, it was intended to exploit the reactivity of lysine residues on the exofacial surface of GLUT1 and identify them with probes conjugated to streptavidin.

On the basis of the current model for the topology of GLUT1, there are 5 lysine residues which may be accessible to the biotin reagent and these are highlighted in Figure 4.2. Most of the potentially reactive lysines are within the *N*-terminal half of the protein at positions 38, 114, 117 and 183, whereas only one such exofacial lysine is thought to reside within the *C*-terminal half, at position 300. Clearly, local folding of the polypeptide chain and/or the presence of the carbohydrate moiety may be factors in determining the relative accessibility of each lysine residue. However, it was hoped to obtain experimental evidence in favour of the current topological model for GLUT1 with respect to locations of exofacial turn/loop regions.

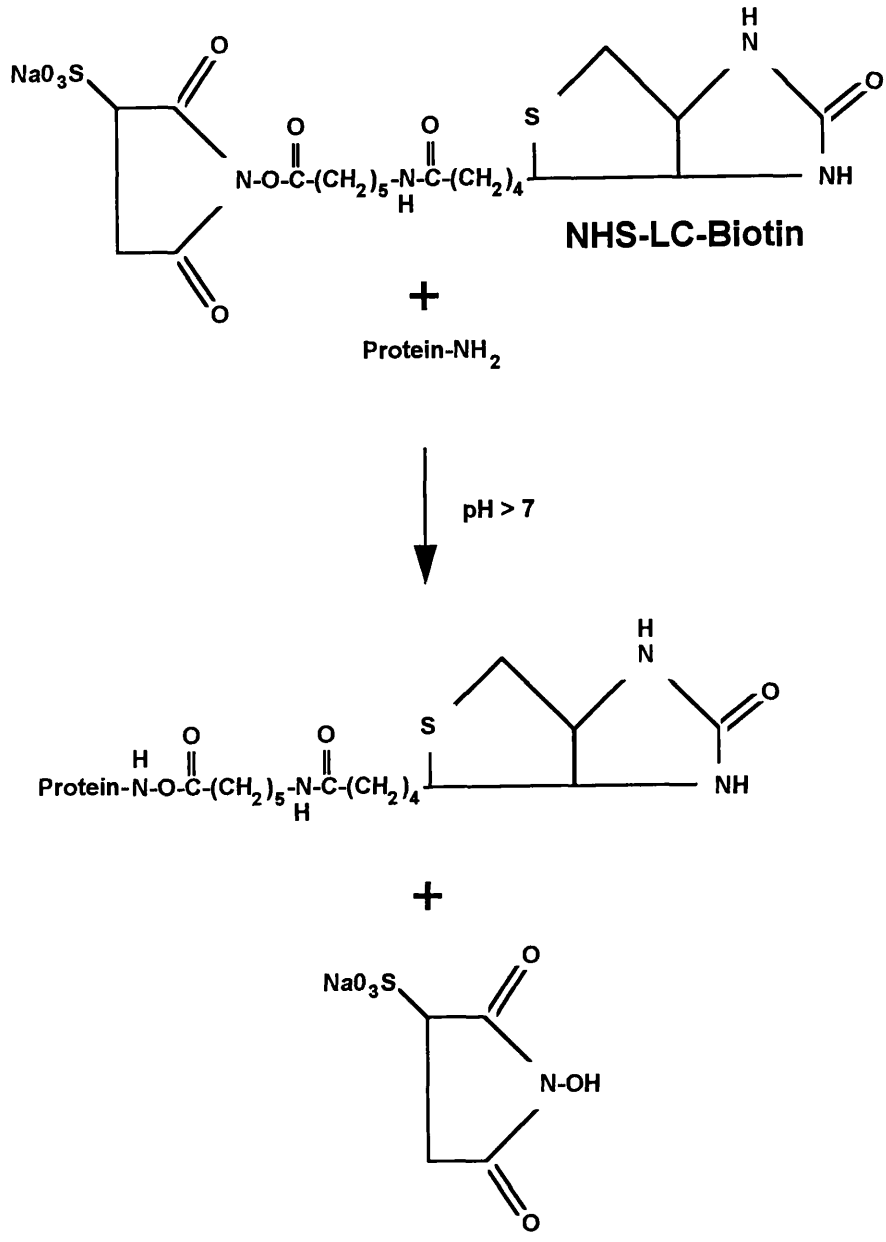


Figure 4.1 Reaction of NHS-LC-Biotin with a protein.
The reagent has a long chain spacer arm that reduces steric hindrances associated with the binding of biotinylated molecules to avidin.

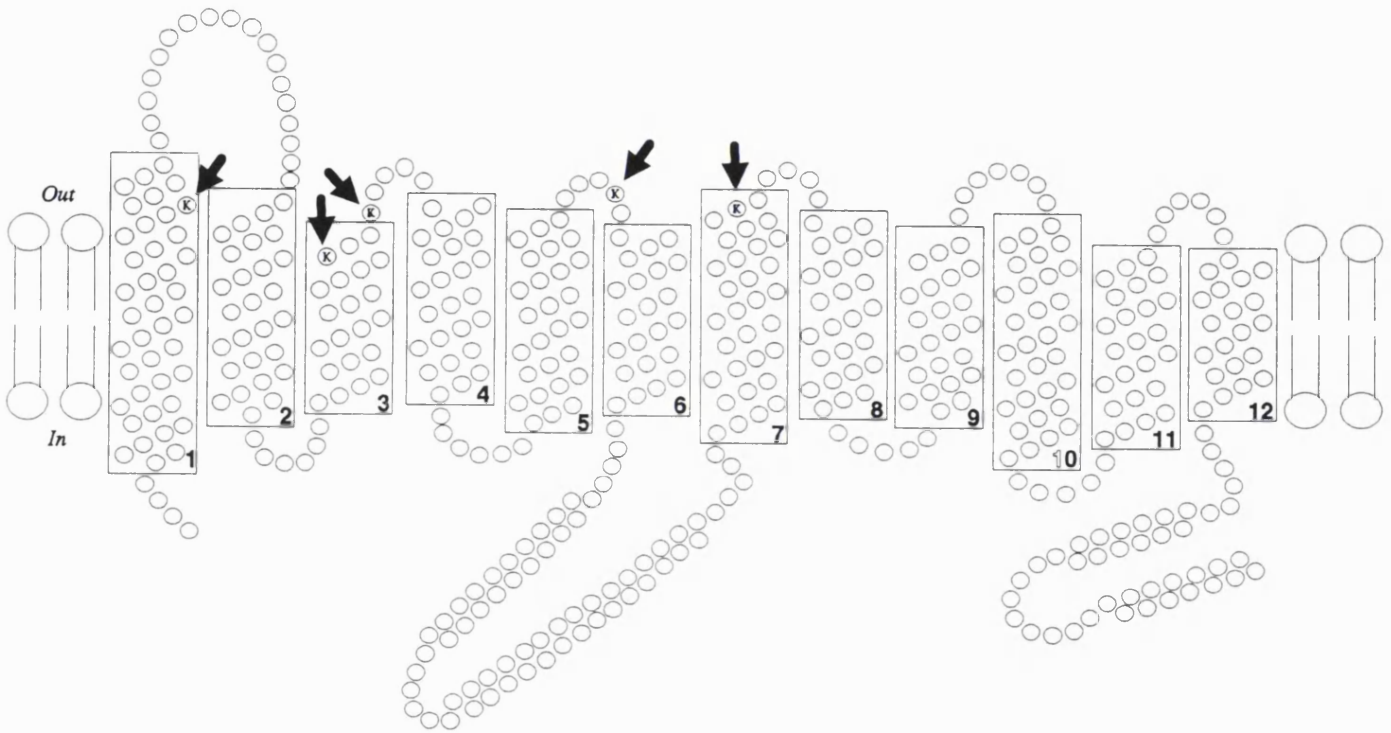


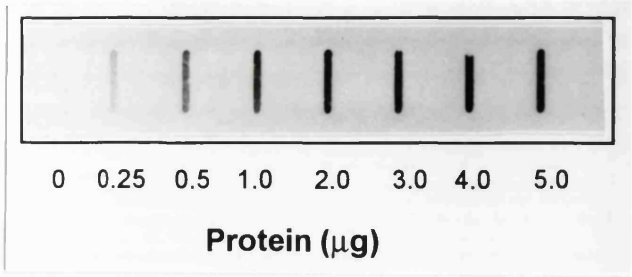
Figure 4.2 Potential sites of exofacial biotinylation.

4.2 Optimisation of biotinylation procedure

In order to ensure that all accessible lysine residues on the exofacial surface of GLUT1 did become biotinylated, it was necessary to establish conditions under which maximum labelling was achieved with NHS-LC-Biotin. For this, a quantitative slot-blotting approach was used. Human erythrocytes were obtained from fresh blood by repeated centrifugation and resuspension in ice-cold PBS (10 mM NaH_2PO4, 145 mM NaCl, pH 7.2). After the leukocytes had been aspirated; the erythrocytes were resuspended at a haematocrit of 25% in 200 mM sodium 4-(2-hydroxyethyl)-1-piperazine-ethanesulphonate (HEPES), pH 8.1. Exofacial biotinylation was then achieved by incubation of the erythrocytes with 1 mM NHS-LC-Biotin at 0°C for time periods up to 2 hours. Excess reagent was quenched after 0, 10, 20, 30, 40, 60 and 120 minutes by the addition of a 10-fold molar excess of glycine and the cells were washed twice with ice-cold PBS. To obtain ghost membranes, the erythrocytes were incubated with 5P8 [5 mM NaH_2PO4, 1 mM EDTA, pH 8.0 containing 0.2 mM phenylmethanesulphonyl fluoride (PMSF)] for 20 minutes on ice. The ghost membranes were sedimented and washed twice with ice-cold 5P8 by centrifugation at 18,000 x *g* for 15 minutes. Finally, the membranes were resuspended in 1 ml 5P8 and assayed for protein content as described in Section 2.2.2.

Ghost membranes (3 µg) were applied to nitrocellulose using a standard slot blotting apparatus. In addition, dilutions of biotinylated ghost membranes (0-5 µg) were also applied to nitrocellulose for the purpose of constructing a calibration curve to demonstrate a linear response between the amount of biotinylated protein and the intensity of the stained band. The membranes were then removed and probed with streptavidin-alkaline phosphatase for 1 hour prior to visualisation of the biotinylated proteins using a colorimetric procedure (Section 2.2.5.2). The intensities of the stained bands were then quantified by reflectance densitometry. As can be seen from Figure 4.3, there is a linear relationship between the amount of biotinylated protein applied to nitrocellulose

A



B

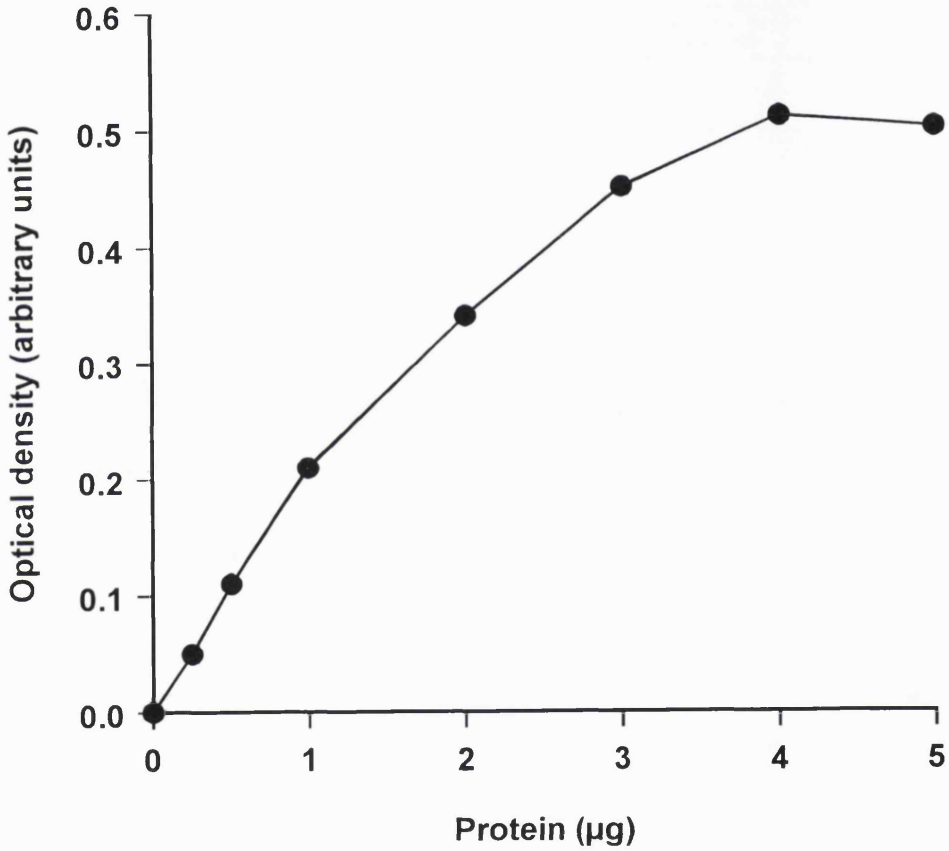


Figure 4.3 Demonstration of linear relationship between amount of biotinylated protein applied to nitrocellulose and the intensity of the stained band.

Biotinylated ghost membranes (0-5 μg) were applied to nitrocellulose and visualised with streptavidin alkaline phosphatase (**A**). The intensities of the stained bands were quantified by reflectance densitometry (**B**).

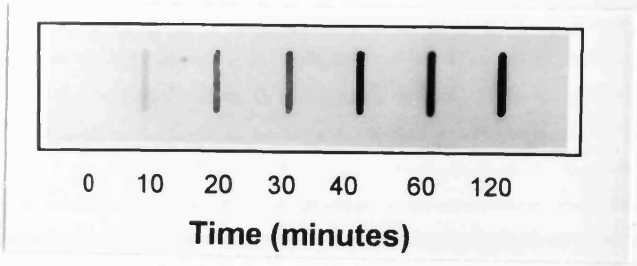
and the intensity of the stained band. The amount of protein (3 μg) applied to the nitrocellulose for the assessment of maximum labelling with NHS-LC-Biotin was demonstrated to lie within the linear range of the detection procedure.

The slot blot analysis for the time course of biotinylation is shown in Figure 4.4A. This blot was quantified by reflectance densitometry (Figure 4.4B) and used to determine the point of maximum incorporation of biotin. It is clear that incorporation of NHS-LC-Biotin into intact erythrocytes occurs in a linear manner up to 60 minutes, after which point no further increase in labelling can be detected. This result was confirmed by Western blot analysis. Ghost membrane samples from each time point were run on a 10% SDS/polyacrylamide gel (Section 2.2.4) and transferred to nitrocellulose (Section 2.2.5.1). The blot was then stained by incubation with streptavidin-alkaline phosphatase according to the procedure described in Section 2.2.5.2. Figure 4.5 is a typical blot and shows clearly that maximum incorporation of NHS-LC-Biotin has occurred by 60 minutes. This incubation period was used in all subsequent experiments to ensure maximum biotinylation.

4.3 Confirmation of the membrane-impermeant character of NHS-LC-Biotin

After having established the conditions for maximum biotinylation of red cells, it was necessary to demonstrate that no lysine residues inside the intact cell membrane were being tagged. That is, it was imperative to confirm the membrane-impermeant character of the NHS-LC-Biotin reagent by comparing the pattern of labelling seen in intact cells with that of unsealed erythrocyte membranes under the same conditions. Ghost membranes were prepared from intact cells that had been biotinylated for 1 hour, according to the procedure described above. In addition, ghost membranes were prepared from erythrocytes not incubated with NHS-LC-Biotin. Biotinylated and non-biotinylated ghost membranes were electrophoresed on 10% SDS/polyacrylamide gels (Section 2.2.4) and transferred to nitrocellulose (Section 2.2.5.1). The blot was then incubated for 1 hour at room temperature

A



B

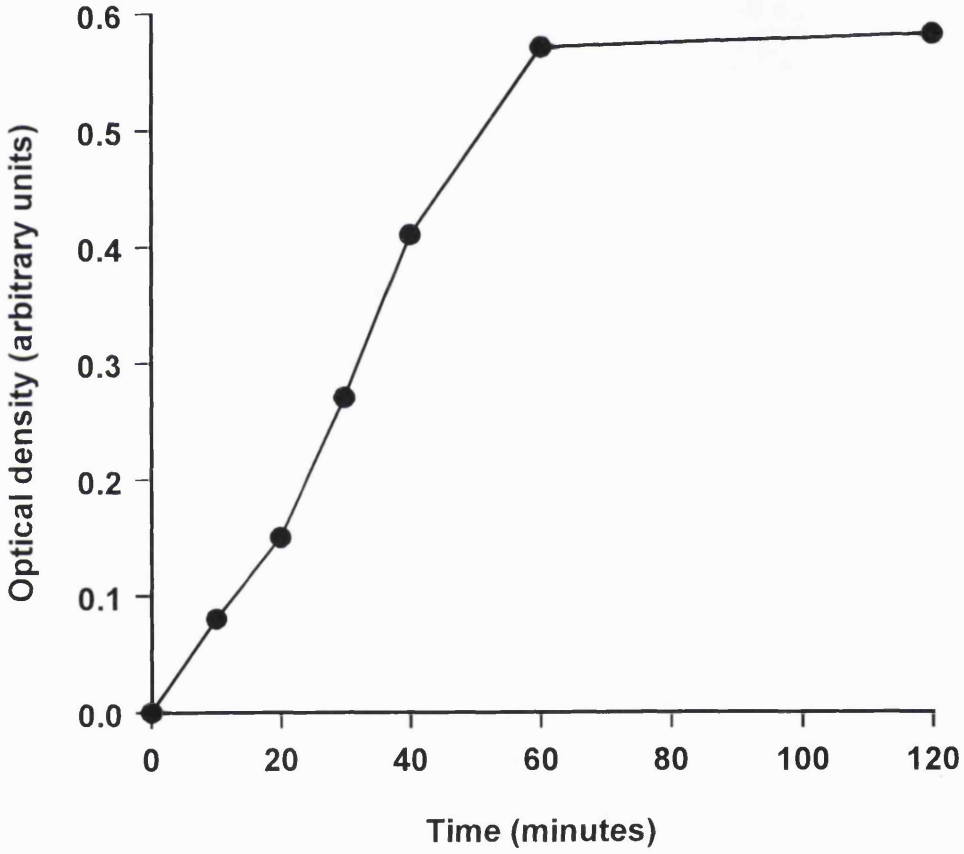


Figure 4.4 Time course for biotinylation of intact erythrocytes. (**A**) Slot blot analysis. (**B**) Calibration graph derived from reflectance densitometry of slot blot.

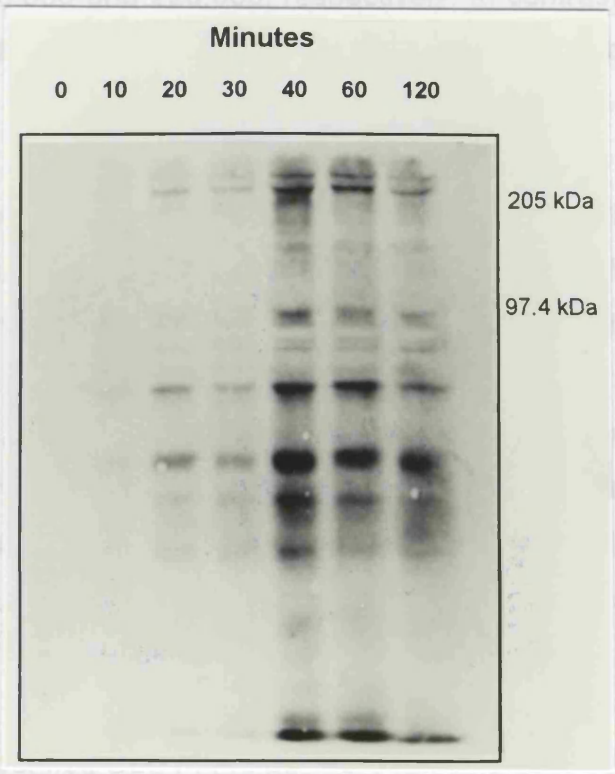


Figure 4.5 Western blot analysis of biotinylated erythrocytes. Ghost membranes were prepared from intact erythrocytes that had been biotinylated for 0, 10, 20, 30, 40, 60 and 120 minutes.

with streptavidin conjugated to alkaline phosphatase, prior to visualisation of the biotinylated proteins as described in Section 2.2.5.2. Figure 4.6 demonstrates clearly the absence of any biotinylation of cytoplasmic components. That is, in the case of intact cells, no labelling was observed of the major peripheral protein spectrin. This protein lies on the cytoplasmic surface of the membrane and migrates on SDS/polyacrylamide gels as a doublet of bands with apparent M_r values of 240,000 and 220,000, respectively. In contrast, this protein was strongly labelled in unsealed membranes, confirming the impermeability of the reagent.

The major protein band labelled by the reagent in intact cells appeared to be the anion transporter, which migrates with an apparent M_r of 100,000 on SDS/polyacrylamide gels. However, faint labelling of a broad band of apparent M_r 50,000, equivalent to that of the glucose transporter, was also apparent. Since GLUT1 represents approximately 6% of the protein of the total erythrocyte membrane protein (Baldwin, 1993), and in order to demonstrate more clearly that the transporter was indeed biotinylated, GLUT1 was purified from the biotinylated membranes by the procedure of Cairns *et al.* (1984). The purified protein was assayed for protein (Section 2.2.2).

Samples of the purified GLUT1 (10 µg) were run on a 10% SDS/polyacrylamide gel (Section 2.2.4), transferred to nitrocellulose (Section 2.2.5.1) and probed with either streptavidin-alkaline phosphatase (Figure 4.7, Lane 1) to detect the biotin label, or affinity-purified antibody raised against residues 477-492 of human GLUT1 to confirm the identity of the purifed protein (Figure 4.7, Lane 2). Figure 4.7 shows that the biotinylated protein migrated with a mobility corresponding to the authentic glucose transporter on a Western blot.

The next stage was to ensure that the purified GLUT1 had retained functional activity, which was examined by its ability to bind cytochalasin B, performed using standard procedures essentially as described by Zoccoli *et al.* (1978) (Section 2.2.3). Lack of an effect of biotinylation on the tertiary structure and

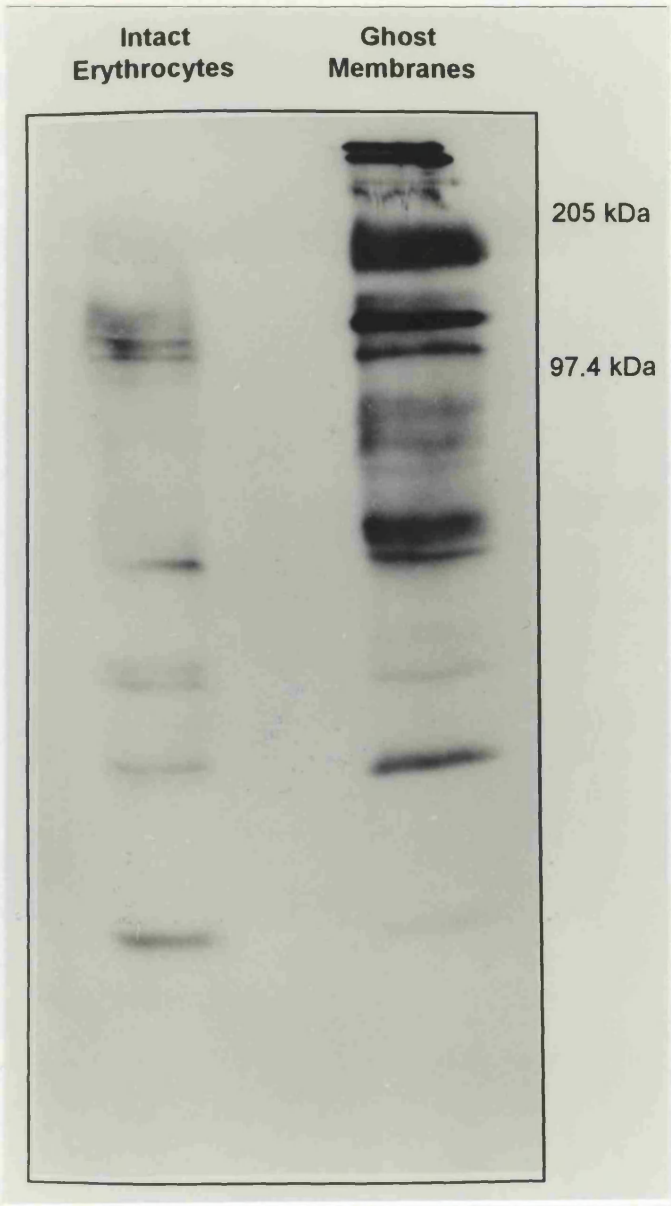


Figure 4.6 Western blot demonstrating biotinylation of exofacial amino groups.
Lane 1 contains ghost membranes prepared from biotinylated intact cells, whereas lane 2 contains biotinylated unsealed erythrocyte membranes.

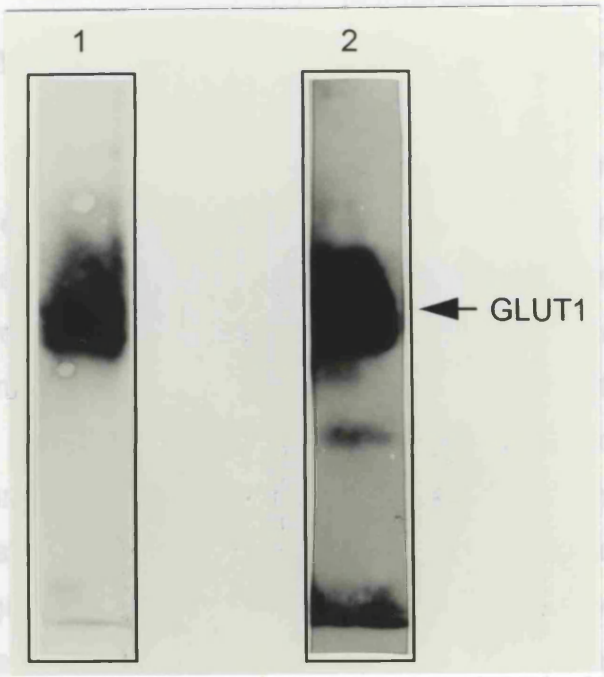


Figure 4.7 Demonstration of biotinylation of purified GLUT1. Western blots of biotinylated GLUT1 probed with streptavidin alkaline phosphatase (Lane 1) or an antibody raised against residues 477-492 of human GLUT1 (Lane 2).

membrane topology of the transporter was indicated by the finding that the purified, biotinylated protein retained its ability to bind the transport inhibitor cytochalasin B (55 B/F per mg/ml). Thus, it had been shown that exofacially-biotinylated GLUT1 could be purified to near homogeneity, and also, that the labelling had not affected the topology of the purified protein.

4.4 Proteolytic digestion of the biotinylated transporter

The results of Figure 4.7 had shown that at least one lysine residue of GLUT1 was accessible to the biotinylating reagent, but in order to determine which of the putative exofacial lysine residues (Figure 4.2) were labelled, limited proteolytic digestion of the biotinylated transporter was performed. Digestion of the purified, membrane-bound transporter in 50 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA, pH 7.4 was achieved by addition of 5% (w/w) diphenylcarbamyl chloride-treated trypsin at 0, 2 and 4 hours. After a total of 6 hours incubation at 25°C, digestion was terminated by addition of bovine lung aprotinin to give a final concentration equal to that of trypsin by weight. Samples were then electrophoresed on 12% SDS/polyacrylamide gels and subjected either to staining with Coomassie blue (Section 2.2.4) or electrophoretic blotting onto nitrocellulose (Section 2.2.5.1). Biotinylated proteins were detected on the blots either colorimetrically following incubation with an alkaline-phosphatase conjugate of streptavidin, or using a peroxidase conjugate of streptavidin and ECL, as described in Section 2.2.5.2.

Digestion with trypsin yielded a prominent labelled fragment that migrated with an apparent of M_r 18,000 (Figure 4.8), that is known from previous studies to comprise residues 270 to 456 of the transporter sequence (Cairns *et al.*, 1987, Davies *et al.*, 1990). Labelling of this fragment is consistent with the prediction that Lys_{300}, in the hydrophilic linker between putative transmembrane helices 7 and 8, is exofacial (Mueckler *et al.*, 1985). Labelled fragments that migrated as sharp band of lower M_r were also seen on the blot, and may have resulted from further cleavage of the fragment. In addition, a broad region of labelling of

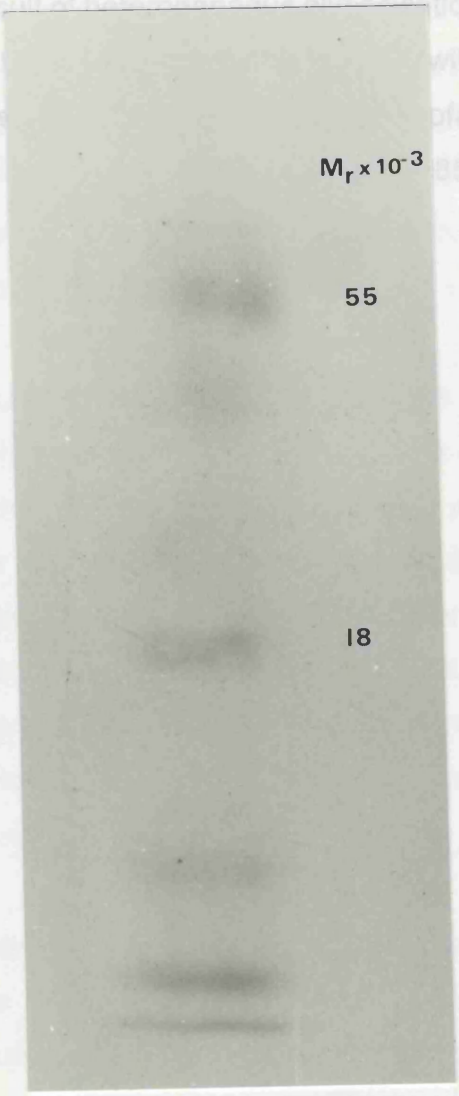


Figure 4.8 Trypsin digest of the biotinylated, membrane-bound transporter.

apparent M_r 30,000-50,000 was observed on the gels. This region contains a fragment comprising residues 1-212 of the transporter, which migrates as a broad band as the result of heterogeneous glycosylation at Asn_{45} (Cairns *et al.*, 1987, Davies *et al.*, 1990). Its labelling is consistent with the prediction that the *N*-terminal half of the transporter contains four exofacial lysine residues, at positions 38, 114, 117 and 183 (Mueckler *et al.*, 1985).

4.5 Discussion

Biotin-conjugated protein-modifying reagents possess several advantages over commonly employed radiolabelled analogues. They provide a simple, rapid and sensitive means of detecting labelled peptides derived from proteins compared to modifications with ^{3}H- or ^{14}C-labelled reagents. Biotinylated reagents are also cost-effective compared to radioisotopes and require no special handling or disposal. This technology was applied to a study of GLUT1 structure.

The labelling of GLUT1 in intact erythrocytes by a membrane-impermeant derivative of biotin, NHS-LC-biotin, indicated that the protein contains at least one exofacial lysine residue. The site(s) of labelling were further investigated by limited proteolysis of the native, membrane-bound transporter followed by blotting to identify biotinylated fragments. Trypsinisation of the native transporter revealed that the biotin label had been incorporated into both halves of the transporter. Only one lysine residue (Lys_{300}) is predicted to be in an exofacial location within the C-terminal half of the transporter, and this finding is therefore consistent with the proposed model of GLUT1.

It is clearly necessary to expand the proteolytic strategy in order to identify which of the lysine residues in the *N*-terminal half of GLUT1 can be biotinylated. For example, in the future it might be possible to identify more precisely the sites of labelling by using immobilised streptavidin to purify smaller biotinylated fragments resulting from more extensive digestion of the transporter, followed by the use of a panel of site-specific anti-peptide antibodies (Davies *et al.*, 1990) to identify the sequence origin of the labelled peptides.

Unfortunately, the lack of additional exofacial lysine residues within the *C*-terminal half of GLUT1 hindered a further structural examination of accessible sites within this region. As a consequence, a strategy of site-directed mutagenesis was embarked upon that would create additional sites for

biotinylation. This is described in the next chapter and the fundamental difference between it and the study of Hresko *et al.* (1994), is that minimal changes to the structure and function of the transporter are made. Consequently, it is hoped to obtain topographical data about the *C*-terminal half of GLUT1 from functional molecules.

CHAPTER 5. GENERATION OF GLUT1 MUTANTS FOR STRUCTURAL STUDIES AND THE IDENTIFICATION OF BINDING-SITES FOR TRANSPORT INHIBITORS

5.1 Introduction

A fundamental feature of the topological model of GLUT1 (Figure 3.5) is the presence of a large, central cytoplasmic loop which separates the structure into two halves; the *N*-terminal domain and the *C*-terminal domain. Within these domains of six helices each, there are a series of interconnecting loops that are generally slightly longer on the exofacial surface than on the cytoplasmic surface. It is the short length of these loops that suggests there are severe constraints on the possible tertiary structures of GLUT1, with perhaps a slightly less compact packing arrangement at the external surface. In addition, conserved motifs exist between helices 2 and 3 in the *N*-terminal half of the protein, and at the equivalent location in the *C*-terminal half between helices 8 and 9. Similarly, putative salt bridges that could help to maintain conformation may exist between helices 4 and 5 and, correspondingly, between helices 10 and 11. It is these features that have led to the proposal that the transporter evolved by duplication of a gene that encoded an ancestral six-membrane spanning helical protein to produce the twelve-membrane spanning helical protein possessing a bilobular structure (Maiden *et al.*, 1987). Although there is very little direct evidence in favour of the model for GLUT1, more experimental data exist regarding the binding of transport inhibitors to the protein.

Labelling studies have shown that all high-affinity ligands, including ATB-BMPA (specific for external site), cytochalasin B and forskolin (which are specific for internal sites) bind to the *C*-terminal half of the protein (Cairns *et al.*, 1987, Clark and Holman, 1990, Wadzinski *et al.*, 1990). In each case, trypsin cleavage of the photolabelled transporters clearly shows that the labelling occurs in a fragment of apparent M_r 18,000 derived from the *C*-terminal half

(residues 270-456). However, more detailed mapping investigations have suggested a structural separation of external and internal substrate binding sites. The endofacial ligand, [^{125}I]-IAPS-forskolin, appears to label the transporter between residues Ile_{369} and Phe_{389} within helix 10 (Wadzinski *et al.*, 1990). The site of labelling by the other endofacial ligand, cytochalasin B, is suggested to reside somewhere between residues Phe_{389} and Trp_{412}, that is, between the end of helix 10 and half way through helix 11 (Holman and Rees, 1987). In contrast, the exofacial inhibitor, ATB-BMPA, is thought to label the transporter within putative transmembrane helix 8, that is, between residues Ala_{301} and Arg_{330} (Davies, 1991).

The evidence for the locations of binding-sites has relied mostly on fragment size and the recognition of fragments by anti-peptide antibodies, rather than the direct sequencing or immuno-isolation of labelled fragments. Until GLUT1 has been crystallised, site-directed mutagenesis represents the most efficient approach to study the functional consequences of structural changes which are designed on the basis of the hypothetical two-dimensional arrangement of the transporter polypeptide within the plasma membrane (Figure 3.5). However, although such studies have revealed important information regarding GLUT1 and glucose transport, the methodology is not without drawbacks. With respect to the analysis of GLUT1, no single mutation will necessarily provide all the answers to a specific problem, since a particular residue change may induce latent effects in other areas of the protein that may or may not be reflected in kinetic analysis or ligand binding studies. For example, GLUT1 has been demonstrated to be photoaffinity-labelled with [^{3}H]-forskolin (Shanahan *et al.*, 1987), and also with [^{125}I]-IAPS-forskolin with a higher affinity (Wadzinski *et al.*, 1990). Chemical and proteolytic digestion and isoelectric focusing experiments suggested that the binding of the latter occurs between residues 369-389 within helix 10. The most likely residue for the labelling was believed to be Trp_{388}, since it is highly conserved in the H^{+}/arabinose transporter (AraE) of *E.coli* (Maiden *et al.*, 1988) which is also labelled with [^{125}I]IAPS-forskolin. The findings of Katagiri *et al.* (1993) did not substantiate this hypothesis as a

substitution of leucine for tryptophan at position 388 did not abolish photolabelling with forskolin.

Mutants of GLUT1 have also been used to probe the photolabelling of the transporter with cytochalasin B. Since the optimal wavelength for photolabelling is 280nm, it has been suggested that it proceeds via photoactivation of an aromatic residue on GLUT1, rather than via activation of the ligand (Shanahan, 1983). Based on this premise, both Trp_{388} and Trp_{412} have been proposed as possible sites of photolabelling. In support of this hypothesis, Katagiri *et al.* (1991) showed that cytochalasin B labelling of a W412L mutant of GLUT1, while not abolished, was decreased by 40% compared to that of the wild-type transporter. Similar findings were shown recently by Inukai *et al.* (1994) for a GLUT1 mutant with Trp_{388} replaced by a leucine residue, while cytochalasin B binding was finally abolished in a mutant possessing leucines at both positions 388 and 412. This raises the possibility that both of these residues may participate in the binding of cytochalasin B.

Thus, although mutagenesis studies do not always provide conclusive data, investigations of the consequences of mutations upon the structure and ligand binding characteristics are helping to derive hypotheses about the mechanism of transport. A recent study by Hashiramoto *et al.* (1992) demonstrated that a glutamine residue within helix 7 (position 282) is important for the binding of ATB-BMPA, but is not critical for either transport activity or the binding of cytochalasin B. It is believed that this transmembrane helix takes part in a substantial conformation change that allows the binding of exofacial ligands. This is consistent with data demonstrating protection against proteolysis of the *N*-terminal end of helix 7 afforded by the binding of ATB-BMPA (Clark and Holman, 1990). Similarly, Oka *et al.* (1990) have reported that truncation of GLUT1 at the *C*-terminus greatly reduced its ability to bind ATB-BMPA. These findings suggest that helices 7 and 12 are close together in the tertiary structure of the transporter (as described in Chapter 3) and contribute jointly to the exofacial ligand binding-site. The likely proximity of helix 12 to helix 7 is

compatible with both regions playing a role in the large conformational change involved in glucose transport.

Clearly, therefore, definitive elucidation of the role of individual residues in GLUT1 will ultimately require the determination of the structure of glucose transporter crystals. Nonetheless, it is clear that studies involving direct sequencing of labelled fragments are required to provide experimental verification of sites of inhibitor binding. This chapter describes strategies that were devised to investigate a) the locations of the binding-site(s) of certain inhibitors of glucose transport, b) the predicted topology and c) the domain structure of the transporter. A key aspect of the mutagenesis strategy involved the removal or introduction of proteolytic cleavage sites in order to assist the isolation and identification of GLUT1 fragments labelled with transport inhibitors. An additional feature of this approach was that the cleavage sites introduced into the putative exofacial loops linking helices 9 and 10, and 11 and 12 were lysine residues that offered potential new targets for amino-group specific membrane-impermeant probes. Thus, it was hoped that the strategy would provide direct experimental evidence for the topography of GLUT1, and supplement that described in Chapter 4. The final aspect of the strategy involved the production of partial length GLUT1 molecules, via introduction of novel start and/or stop codons, in an attempt to assess the plausibility of the bilobular structure of GLUT1 illustrated in Chapter 3.

5.2 Mutagenesis strategy

5.2.1 Mutants to aid identification of inhibitor binding-sites

The indication from the data of Chapter 3 is that an alignment of homologous sequences can provide a wealth of information regarding the relationship of an amino acid sequence to higher orders of structure. An additional benefit of an alignment is the ability to rationalise the design of experiments, particularly those involving site-directed mutagenesis. Many studies have used mutagenic strategies to detect alterations of transport activity in attempts to assess the role(s) of specific amino acids in the transport process. For the purposes of this investigation, however, it was essential that the generation of mutants would not perturb transporter function by a significant amount, since the aim was to retain the ability of GLUT1 to bind site-specific inhibitors of transport. Consequently, the sequence alignment was used to aid the design of mutations that would be tolerated.

The mutagenesis strategy for the identification of inhibitor binding-sites is described in Figure 5.1. Cleavage of the wild-type, membrane-embedded, labelled transporter with trypsin is known to produce a fragment of apparent M_r 18,000 (residues 270-451) that is labelled with ATB-BMPA, cytochalasin B and forskolin (Davies *et al.*, 1992). Within this region, the sites of labelling by ATB-BMPA and cytochalasin B have been approximately located to helix 8 and to helix 9-10, respectively, through further fragmentation of the polypeptide and identification of labelled fragments both by their size and their recognition by site-directed antibodies (Davies, 1991). A more precise identification of the sites of labelling would require the direct sequencing of labelled fragments, but the purification of suitable fragments has been difficult to achieve (data not shown). A particular problem is that proteolytic cleavage sites exist between the site of labelling and epitopes recognised by antibodies capable of recognising the native protein, and so it is not possible to purify the fragments by immunoprecipitation. Furthermore, in order to be able to sequence through the

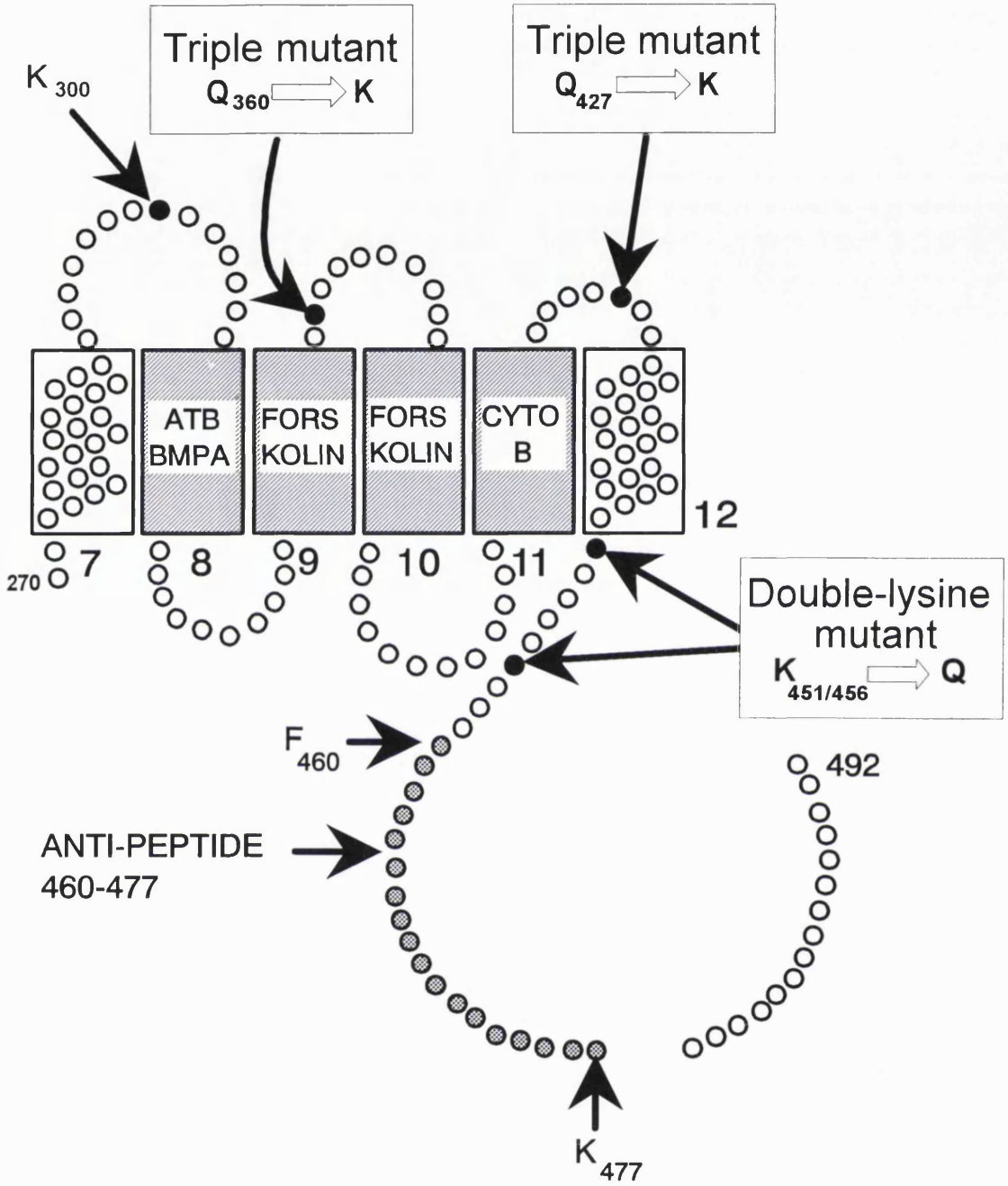


Figure 5.1 Diagram illustrating the mutations designed to facilitate the identification of inhibitor binding-sites.
The substitutions of arginine for lysine at positions 451 and 456 are present in each of the triple mutants.

sites of labelling, it would be desirable if these sites lay close to the *N*-terminus. Unfortunately, suitable cleavage sites are not located here in the wild-type protein. Thus, it was decided to introduce proteolytic cleavage sites into GLUT1 with the result that labelled fragments of suitable length for sequencing could be obtained.

The initial mutation required was one that would aid the purification, preferably by immunoaffinity methods, of labelled fragments. A mutation was therefore designed that would eliminate two sites recognised by endoproteinase Lys-C which closely follow helix 12, that is, at positions 451 and 456. The removal of these cleavage sites by the conversion of the relevant lysine residues to arginine would allow labelled, proteolytic fragments of the transporter to be isolated using antibodies raised against a peptide comprising Phe_{460}-Lys_{477} (Davies *et al.*, 1990). This interchange of Lys and Arg occurs naturally within closely related pairs of transporters, namely rat and mouse GLUT4 at the position corresponding to Lys_{451}, and human and mouse GLUT3 at the position corresponding to Lys_{456}. It was felt to be likely, therefore, that such mutations would be tolerated and not result in a substantial decrease in transport activity. This mutant, if it did prove to resemble the wild-type transporter in terms of activity, would then serve as a template for the subsequent introduction of proteolytic cleavage sites, also illustrated in Figure 5.1.

The additional mutations involved the creation of endoproteinase Lys-C sites in the exofacial loops linking helices 9 and 10 (Q360K) and helices 11 and 12 (Q427K). It was hoped that these changes would be tolerated also as there is a lysine residue at position 360 in GLUT2 and GLUT7, and at position 427 in XylE. As a consequence of these mutations, digestion of the double-lysine mutant (KK) by endoproteinase Lys-C would produce a fragment of apparent M_r 19,279, comprising of residues Ala_{301}-Lys_{477}. This fragment should be labelled by ATB-BMPA, cytochalasin B and forskolin, and could be imunoaffinity purified by antibodies raised against residues 460-477 of human GLUT1. Upon digesting the GLUT1 protein possessing the additional Q360K mutation

(KK.Q360K), a fragment corresponding to Leu_{361}-Lys_{477} of apparent M_r 13,113 should be retrieved by immunoprecipitation and shown to be labelled by cytochalasin B and forskolin only. A fragment of apparent M_r 6,134, consisting of residues Ala_{301}-Lys_{360} (that is, helices 8 and 9), would also be released on proteolysis that would be labelled by ATB-BMPA only. Finally, digestion of the GLUT1 protein possessing the additional Q427K mutation (KK.Q427K) followed by immunoaffinity purification using the antibody raised against residues 460-477, would recover a fragment of the transporter consisting of helix 12 only. The labelling characteristics of this fragment are not known.

5.2.2 Production of the *N*- and *C*-terminal halves of GLUT1

As elaborated upon in Chapter 3, a fundamental structural feature of the sugar transporter is believed to be a dual domain assembly that has arisen from the duplication of an ancestral six membrane-spanning helix protein. The *N*- and *C*-terminal halves of the protein are thought to be comprised of two bundles of six helices and a goal of this study was to examine this premise. It was intended that mutants of GLUT1 consisting of either the *N*- or *C*-terminal halves of the molecule would be produced and then co-expressed in *Xenopus* oocytes, where functionality of the individual halves and assembly of the expressed species could be tested through sugar uptake experiments.

In order to minimise any perturbation of function, the locations for the division of GLUT1 into its constituent halves were chosen using the consensus secondary structure prediction (Figure 3.5). From this data, two distinct predictions of α-helical structure are present within the central cytoplasmic loop, separated by a strong β-turn prediction between residues 232-236. There is also the suggestion of another turn/loop region between residues 260-264 that probably corresponds to the end of the central, cytoplasmic loop before the polypeptide chain re-enters the lipid bilayer as transmembrane helix 7. Separate GLUT1 half-molecule constructs were therefore generated consisting of residues 1-234, 235-492, 1-263, or 264-492, as shown in Figure 5.2.

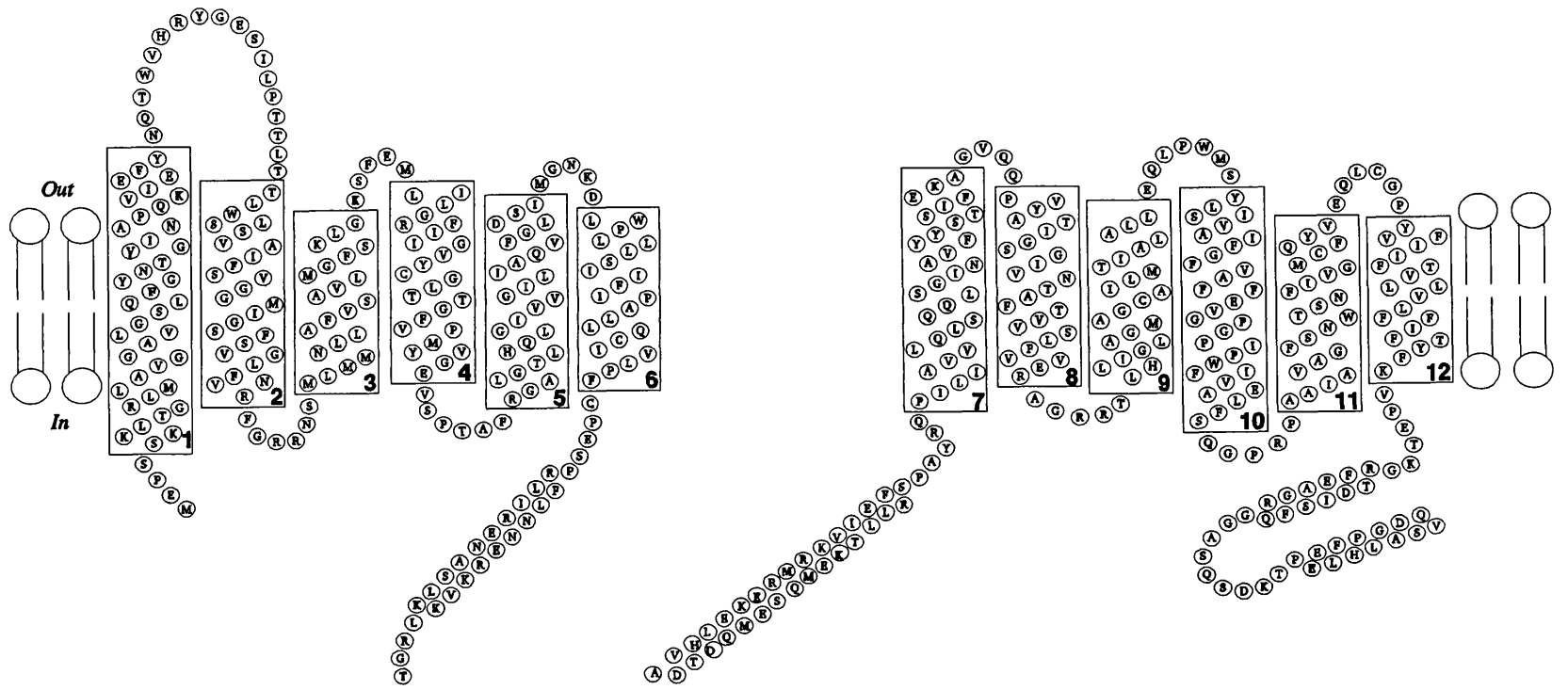


Figure 5.2 Production of the *N*- and *C*-terminal halves of GLUT1. Secondary structure prediction was used to divide GLUT1 into *N*-terminal halves consisting of residues 1-234 and 1-263, and *C*-terminal halves comprising residues 235-492 and 264-492.

5.3 PCR mutagenesis

The methodology adopted to generate the halves of the transporter, and the topological mutants, is a polymerase chain reaction (PCR) procedure termed recombinant PCR (Higuchi, 1990). The strategy is illustrated in Figure 5.3. PCR is an *in vitro* method of nucleic acid synthesis by which a particular segment of DNA can be replicated specifically using two oligonucleotide primers that flank the DNA fragment to be amplified. The process involves repeated cycles of heat denaturation of the DNA, annealing of the primers to their complementary sequences, and extension of the annealed primers with DNA polymerase. The primers hybridise to opposite strands of the target DNA and are oriented so that DNA synthesis by the polymerase proceeds across the region between the primers. Since the extension products themselves are complementary to, and capable of binding primers, successive cycles of amplification essentially double the amount of target DNA synthesised in the previous cycle.

Recombinant PCR is a two-step procedure requiring complementary mutant oligonucleotides that are utilised in two separate primary PCRs to generate products overlapping in sequence. These primary products can then be denatured and allowed to reanneal together. Of the two possible heteroduplex products, those with recessed 3' termini can be extended to produce a fragment that is the sum of the two overlapping products. A subsequent reamplification of this fragment using the right- and left-most primers results in the enrichment of the full-length, secondary product containing the desired mutation.

Traditional oligonucleotide-directed mutagenesis protocols create a mismatch at the desired position between the template DNA strand and a complementary strand synthesised *in vitro*. When DNA containing the mismatch is transformed into bacteria, the bacteria correct the mismatch, either introducing the desired mutation or restoring the original nucleotide. Recombinant PCR, however, uses overlapping mutagenic primers to amplify the template and combine segments of DNA which eliminates the need for restriction sites.

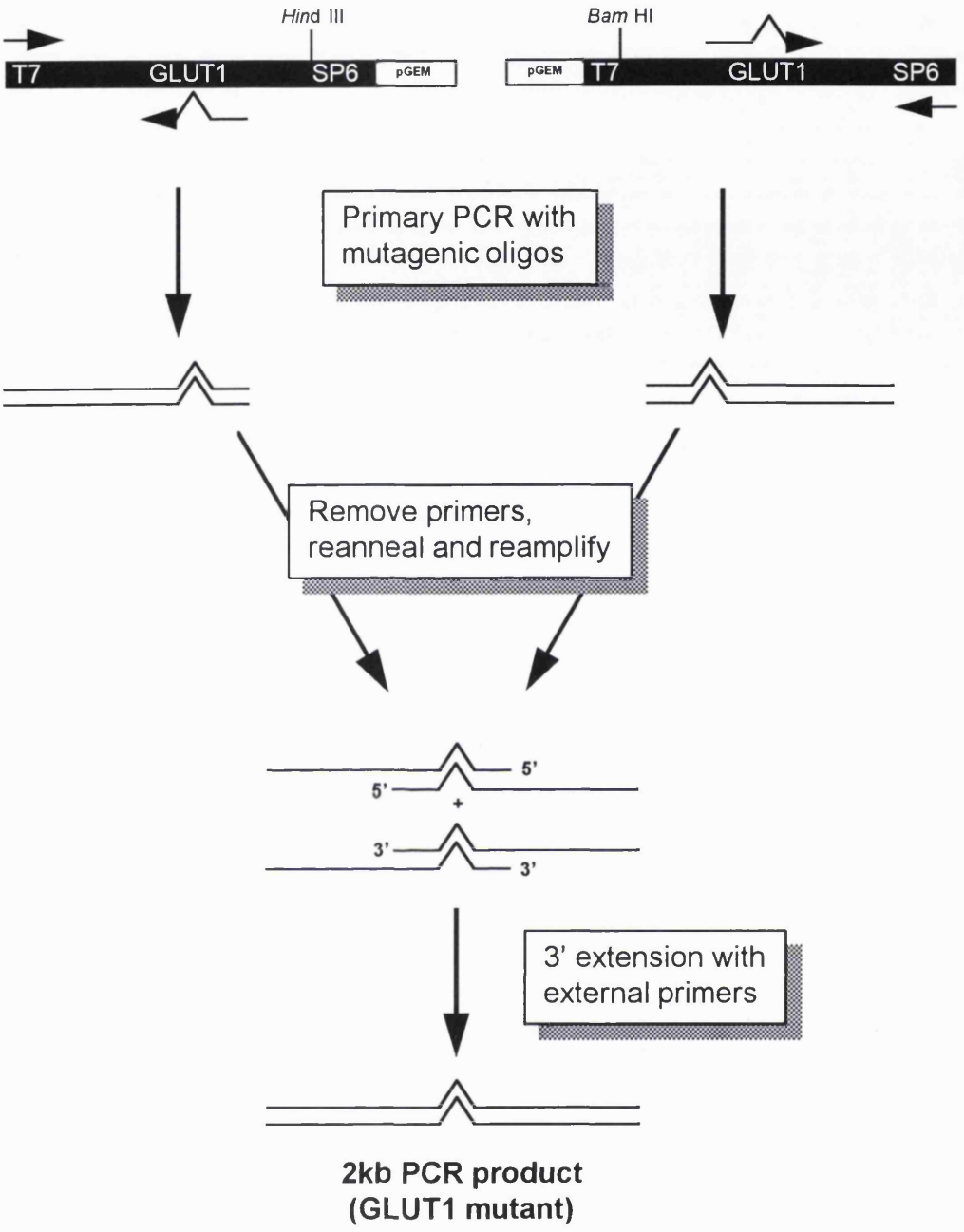


Figure 5.3 Recombinant PCR.

The strategy used to generate all of the mutants described in this study.

Figure 5.4 describes the primers that were designed to generate the topographical mutants. Codon usage already apparent in the GLUT1 cDNA was obeyed, and additional restriction sites were incorporated for screening purposes. The KK451/456RR mutant primer is a 39mer and possesses four mismatches to enable the Lys to Arg changes, simultaneously generating an additional *Xho* I restriction site. In contrast, the Q360K and Q427K primers required additional changes to those generating the amino acid change in order to permit screening. The former involved a two base pair alteration that caused the deletion of a *Nhe* I site whereas a single base change in the Q427K mutant primer resulted in the creation of an additional *Apa* I site.

The design of the primers to produce the halves of GLUT1 (Figure 5.5) included 18 base pair insertions coding for a translation termination codon, *Bam* H1and *Hin*d III restriction sites, plus a translation initiation site. Successful incorporation of these features would enable i) the resulting secondary PCR product to be digested and subcloned with ease, and ii) satisfy the fundamental conditions required for correct translation of the half molecules. These GLUT1 half molecule mutants were constructed by Dr. A.J. Sami and verified by automated sequencing. They were then kindly donated by Dr. A.J. Sami for subsequent functional expression studies in *Xenopus* oocytes (Section 5.7).

5.4 Generation of mutants for inhibitor binding-site identification

The success of PCR techniques is dependent upon the ability of the method to yield a specific amplified product, that is, generate a single band of DNA on a gel. Non-specific products, visible as multiple bands on gels, can arise and are often detrimental particularly with respect to subcloning and expression studies. If non-specific extensions do occur in the early cycles of PCR, the products then serve as templates in the remaining cycles giving rise to a mixture of non-specific and specific products. Thus, in order to maximise the specificity of a PCR, it is necessary to maximise the annealing of the primers to their proper positions on the DNA template. A number of different factors can

A. $K_{451/456} \longrightarrow R$

L F T Y F K_{451} V P Q T K_{456} G R

Wild-type 5' ATCTTCACCTACTTCAAAGTTCCTGAGACTAAAGGCCGG 3'

R_{451} R_{456}

Mutagenic 5' ATCTTCACCTACTTC**CG**AGTTCCTGAGACT**CG**AGGCCGG 3'

Xho I

B. $Q_{360} \longrightarrow K$

I A L A L L E Q_{360} L P W M

Wild-type 5'ATCGCGCTAGCACTGCTGGAGCAGCTACCCTGGATG 3'

Nhe I

K_{360}

Mutagenic 5'ATCGCCCTGGCACTGCTGGAGAAACTACCCTGGATG 3'

C. $Q_{427} \longrightarrow K$

F Q Y V E Q_{427} L C G P Y

Wild-type 5'TTCCAGTATGTGGAGCAACTGTGTGGTCCCTAC 3'

K_{427}

Mutagenic 5' TTCCAGTATGTGGAGAAACTGTGTGGGCCCTAC 3'

Apa I

Figure 5.4 Design of the primers to generate the mutants for subsequent inhibitor binding-site identification.

A.

K_{230} L_{231} R_{232} G_{233} T_{234} ... A_{235} D_{236} V_{237} T_{238} H_{239}

5'AAGCTGCGCGGGACA.....GCTGACGTGACCCAT 3'

stop *Hind* III *Bam* HI start

5'AAGCTGCGCGGGACA**TAGAAGCTTGGATCCATG**GCTGACGTGACCCAT 3'

B.

I_{259} L_{260} E_{261} L_{262} F_{263} ... R_{264} S_{265} P_{267} A_{268} Y_{269}

5' ATCCTGGAGCTGTTC.....CGCTCCCCCGCCTAC 3'

stop *Hind* III *Bam* HI start

5' ATCCTGGAGCTGTTC**TAGAAGCTTGGATCCATG**CGCTCCCCCGCCTAC 3'

Figure 5.5 Design of the primers used to produce the *N*- and *C*-terminal halves of GLUT1.

contribute to this desired result, including optimisation of the magnesium ion concentration, primer concentration, template concentration, optimisation of the annealing temperature and the use of 'hot start' PCR. Optimisation of each PCR was therefore attempted for each of the mutants. Certain features of the procedure, such as the relative amounts of template and primer, did not appear to have a substantial effect upon the efficiency of individual reactions, and thus represent properties of the 'standard PCR amplification protocol' outlined below. However, the component resulting in the greatest effect upon specificity and yield was shown to be the free magnesium ion concentration. Consequently, it was necessary to optimise this for each set of mutagenic primers.

Standard PCR Amplification Protocol:

Linearised plasmid (*Hin*d III or *Bam* H1) template DNA (5 ng)
Mutagenic primer (20 pmol)
External primer, T7 or SP6 (20 pmol)
10 µl reaction buffer (20 mM Tris-HCl, pH 8.0, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol, 0.5% Nonidet-P40, 0.5% Tween-20)
dNTP (50 µM of each)
Free Mg^{2+} (optimum concentration)
(Final volume 100 µl, overlaid with 100 µl mineral oil)

'Hot Start' PCR using the following temperature profile:

	Initial denaturation	94°C, 4 minutes
		85°C, 4 minutes (*Taq* addition, 2units)
25 cycles of,	Denaturation	94°C, 2 minutes
	Primer annealing	60°C, 30 seconds
	Primer extension	72°C, 2 minutes
	Final extension	72°C, 5 minutes

The benefit of 'hot start' PCR is that any non-specific primer-template complexes that form at low or intermediate temperatures prior to denaturation cannot be extended. Also, preliminary experiments indicated that an annealing temperature of 60°C ensured optimal replication fidelity and an efficient PCR.

The template that was used for the initial mutagenesis procedure is illustrated in Figure 5.6 and consists of the pGEM11zf(-) vector (Promega) containing GLUT1 cDNA. The features that make the vector suitable for the work described in this study are ampicillin selection, SP6 and T7 RNA polymerase promoter sequences useful for sequencing, and the ability to produce *in vitro* RNA transcripts.

5.4.1 Construction of KK451/456RR mutant GLUT1 cDNA

The exchange of Lys_{451} and Lys_{456} for arginine residues would yield a GLUT1 cDNA that would serve as a template for subsequent mutagenesis steps, and generate a GLUT1 protein central to the protocol for the identification of inhibitor binding-sites. It was therefore critical that such a mutant be constructed correctly and shown to be active. Recombinant PCR (Section 5.3) was used to construct the KK451/456RR GLUT1 mutant, and amplification of the primary PCR products was performed using the standard protocol, as described above but with the following modifications. Depending upon which of the complementary mutant oligonucleotides (coding or non-coding) were being used to generate a primary PCR product, the template DNA (pGEM.GT, Figure 5.6) was linearised with either *Bam* H1 or *Hin*d III, respectively. The combination of *Bam* H1 linearised template, the coding mutagenic oligonucleotide and the SP6 external primer, together with the components of the standard amplification protocol, would be expected to produce a primary PCR product of 597bp. In contrast, a fragment of 1451bp would be expected with *Hin*d III linearised template, the non-coding mutagenic oligonucleotide plus the T7 external primer in the standard PCR mixture.

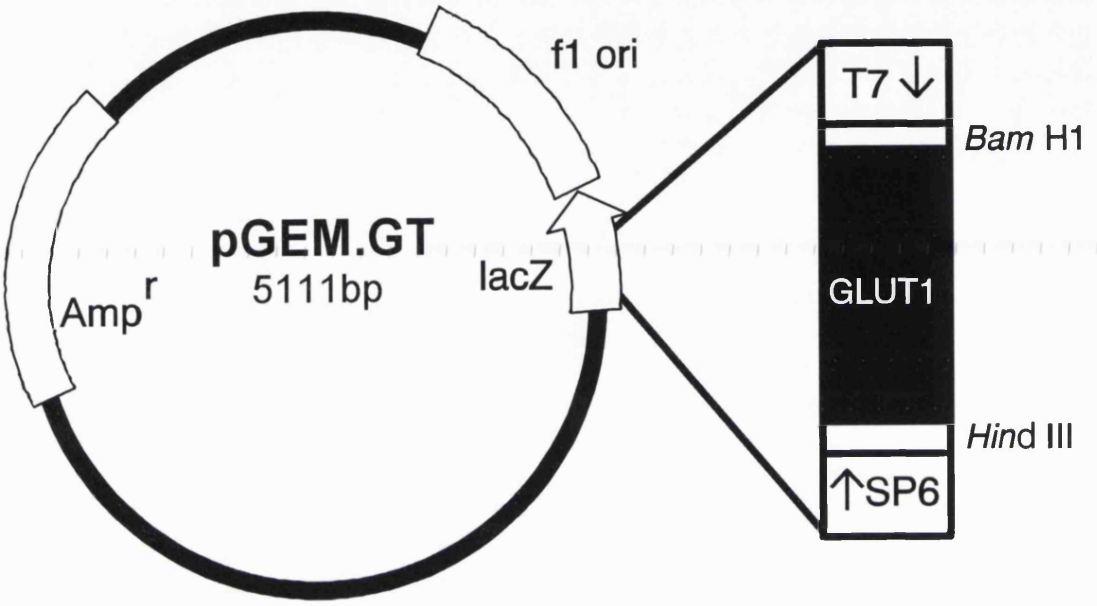


Figure 5.6 Diagram of the pGEM.GT construct.
This construct consists of the *Bam* HI/*Hin*d III fragment containing GLUT1 cDNA contained within the multiple cloning site of pGEM11-zf(-) vector (Promega).

The results of the steps to generate the KK451/456RR GLUT1 mutant by PCR mutagenesis are shown in Figure 5.7. Figure 5.7A shows the successful production of both of the primary PCR products at the predicted sizes of 597bp and 1451bp. At the initial Mg^{2+} concentration employed, it can be clearly seen that amplification of the 1451bp PCR product occurred with greater efficiency. In order to increase the yield of the 597bp product, a titration of Mg^{2+} ion concentration was performed, the results of which are demonstrated in Figure 5.7B. The standard amplification procedure was followed, except with the addition of increasing amounts of $MgCl_2$ to final concentrations of 0.5 mM, 1.0 mM, 1.5 mM, 2.0 mM, 2.5 mM and 3.0 mM (lanes 2 to 7, respectively). Yields of the PCR product were reduced at the extremes of the Mg^{2+} ion concentrations used. However, at a Mg^{2+} ion concentration of 2.5 mM, a single product of 597bp was obtained in sufficient amounts to be purified and used in subsequent PCR procedures. Optimisation of the 1451bp PCR was not necessary.

In order to ensure the removal of the mutagenic oligonucleotides, and thus increase the efficiency of the secondary PCR by preventing re-amplification of the primary PCR products, the 597bp and 1451bp PCR products were gel purified (Section 2.3.7). Equimolar amounts of each were then used in a secondary round of PCR utilising just the external primers (SP6 and T7) to generate the 2009bp (2kb) PCR product containing the entire mutant GLUT1 coding sequence. Optimisation of this PCR with respect to the concentration of free Mg^{2+} ions was necessary, as demonstrated by Figure 5.7C. Lanes 1, 2 and 3 show clearly the effect of increasing the Mg^{2+} from 1.0 mM to 1.5 mM and 2.0 mM respectively. For all subsequent secondary PCRs, therefore, a free Mg^{2+}concentration of 2.0 mM was employed.

If the mutagenesis procedure had been successful, then restriction digestion of the 2kb PCR product with *Xho* I would be expected to produce two bands on an agarose gel corresponding to 1441bp and 568bp. Figure 5.7D shows the result of a partial digestion of the 2kb PCR product with *Xho* I, and confirms

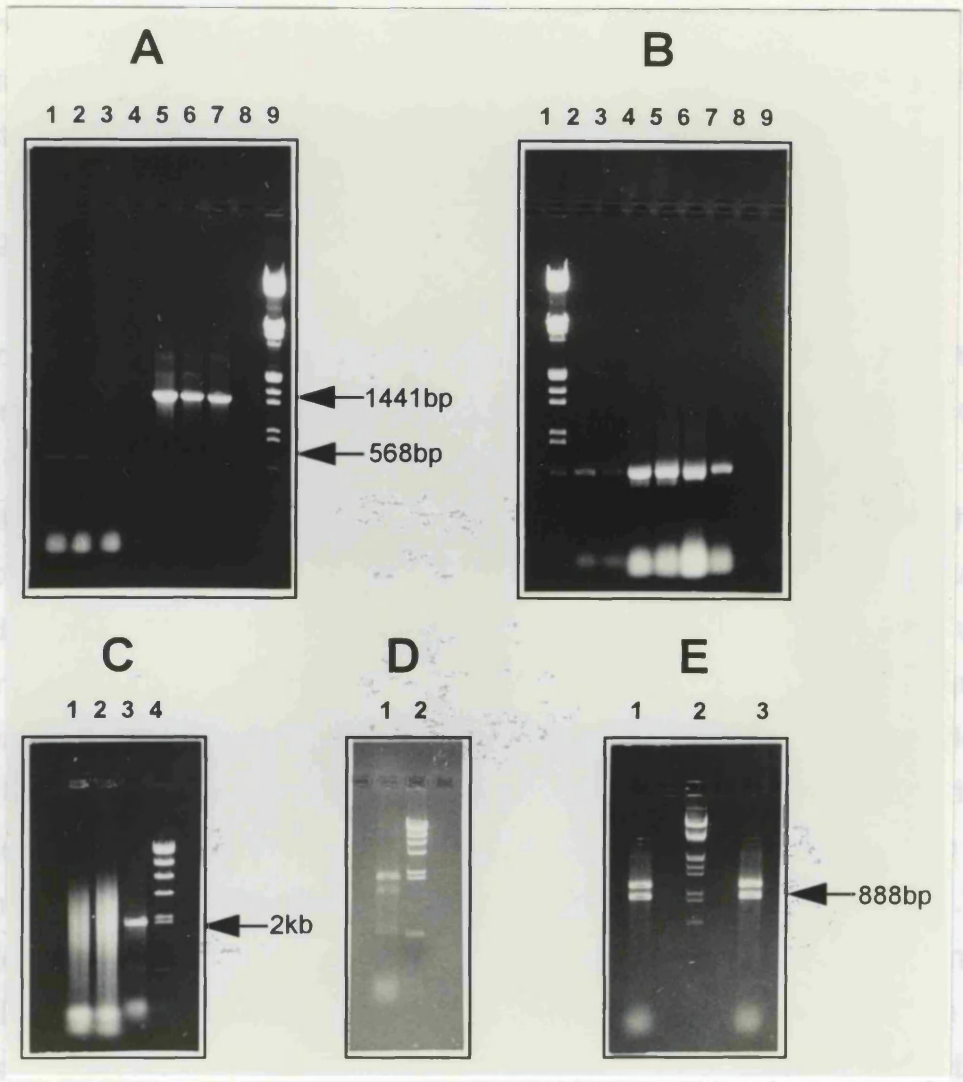


Figure 5.7 PCR of the double-lysine mutant of GLUT1 cDNA.

A. Primary PCR products (triplicate) of 597bp and 1451bp.

B. Optimisation of Mg^{2+} concentration for 597bp PCR product.

C. Optimisation of Mg^{2+} concentration for 2kb PCR product.

D. Partial *Xho* I digestion of 2kb secondary PCR product.

E. *Stu* I / *Hin*d III digestion of 2kb secondary PCR product.

that the mutation is in place as bands corresponding to 2kb (uncut PCR product), 1441bp and 568bp are visible. The 2kb PCR product was then used to prepare a plasmid construct containing the GLUT1 mutant, as described in Section 5.4.2, which would be verified with respect to sequence integrity and functional activity.

5.4.2 Preparation and subcloning of GLUT1 mutant cassette

In order to reduce the possibility of unwanted mutations appearing in the final plasmid construct, an additional step in the subcloning strategy was included that would reduce the amount of screening and sequencing necessary. Figure 5.8 illustrates this extra manipulation in the preparation of pGEM.KK. The 2kb secondary PCR product (mutant GLUT1 cDNA) was digested with *Stu* I and *Hin*d III prior to electrophoresis on a 0.8% agarose gel. There is a unique restriction site for *Stu* I within the GLUT1 cDNA at position 1033. Consequently, digestion of the 2kb secondary PCR product with *Stu* I and *Hin*d III would yield two DNA fragments (Figure 5.7E), one of which (888bp) would comprise the nucleotide sequence coding for the GLUT1 protein from within helix 9 to the *C*-terminus. This mutant 'cassette' was, therefore, gel purified (Section 2.3.7) and ligated (Section 2.3.8) into a similarly digested pGEM.GT construct. Since the GLUT1 termination codon lies at position 1493, only a 500bp portion of the mutant GLUT1 cDNA would need to be sequenced in order to verify that no errors in the coding region had occurred as a result of *Taq* polymerase infidelity.

5.4.3 Construction of GLUT1 mutants KK.Q360K and KK.Q427K

The generation of the other mutants to be used in the identification of substrate-binding sites (KK.Q360K and KK.Q427K) was performed in an analogous manner to that shown to be effective for the generation of the KK451/456RR mutant. That is, primary PCR products, using appropriate mutant oligonucleotides, were amplified using optimum free Mg^{2+} ion concentrations

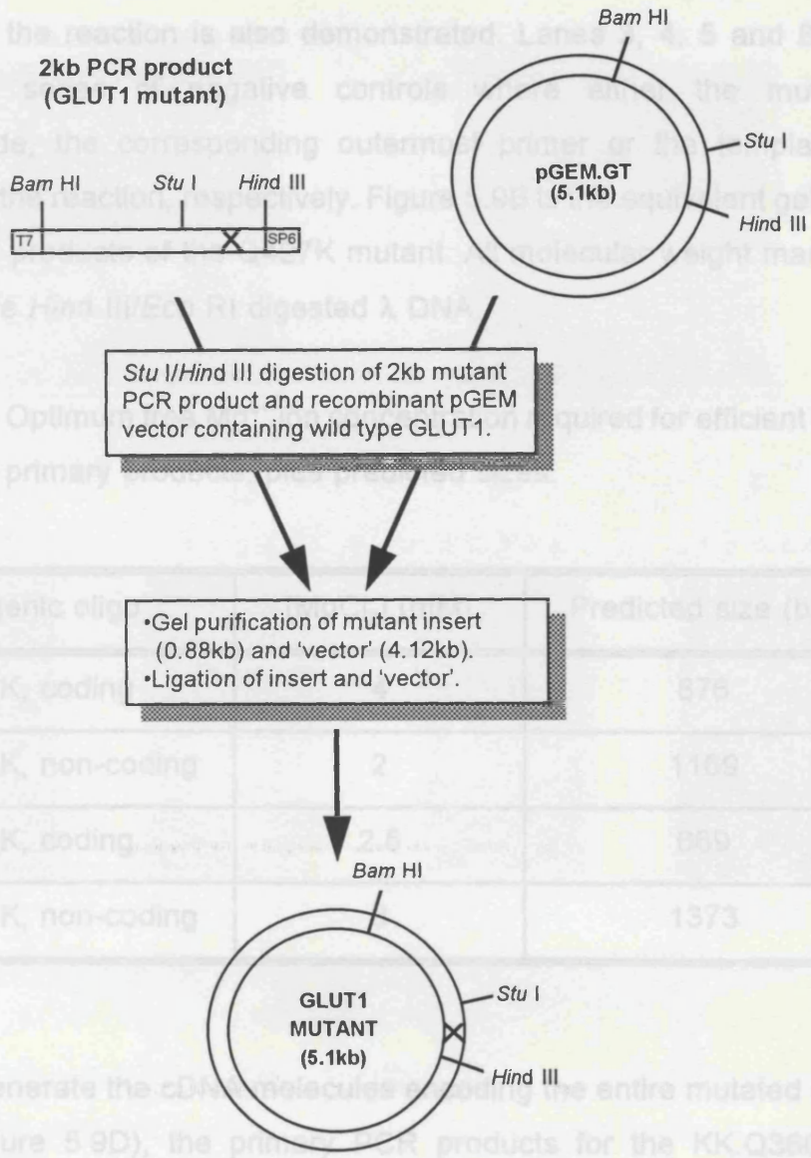


Figure 5.8 The subcloning strategy used to prepare the GLUT1 mutant constructs.

A *Stu* I / *Hind* III fragment of the PCR-derived GLUT1 mutant was purified and ligated as a 'cassette' back into the 'wild-type' GLUT1 cDNA.

(Table 5.1). The results of the PCR generating the primary products of the Q360K mutant are shown in Figure 5.9A, lanes 2 and 7, where either the combination of coding mutagenic oligonucleotide plus SP6 primer, or non-coding mutagenic oligonucleotide plus T7 primer were used, respectively. The specificity of the reaction is also demonstrated. Lanes 3, 4, 5 and 8, 9, 10 represent a series of negative controls where either the mutagenic oligonucleotide, the corresponding outermost primer or the template was omitted from the reaction, respectively. Figure 5.9B is the equivalent gel for the primary PCR products of the Q427K mutant. All molecular weight markers in Figure 5.9 are *Hin*d III/*Eco* RI digested λ DNA.

Table 5.1 Optimum free Mg^{2+} ion concentration required for efficient PCR of primary products, plus predicted sizes.

Mutagenic oligo	[$MgCl_2$] (mM)	Predicted size (bp)
KK.Q360K, coding	4	876
KK.Q360K, non-coding	2	1169
KK.Q427K, coding	2.5	669
KK.Q427K, non-coding	3	1373

In order to generate the cDNA molecules encoding the entire mutated GLUT1 proteins (Figure 5.9D), the primary PCR products for the KK.Q360K and KK.Q427K mutants were gel purified (Figure 5.9C) and combined in a second round of amplification using the external primers (SP6 and T7). Figures 5.9E and 5.9F show the restriction analysis of the KK.Q360K and KK.Q427K 2kb mutant secondary PCR products, where each product has been incubated at 37°C for 2 hours in the presence of no enzyme (lane 2), *Apa* I (lane 3), *Nhe* I (lane 4) and *Xho* I (lane 5). From these figures it is clear that the desired mutations have been created as *Nhe* I no longer cleaves the KK.Q360K 2kb

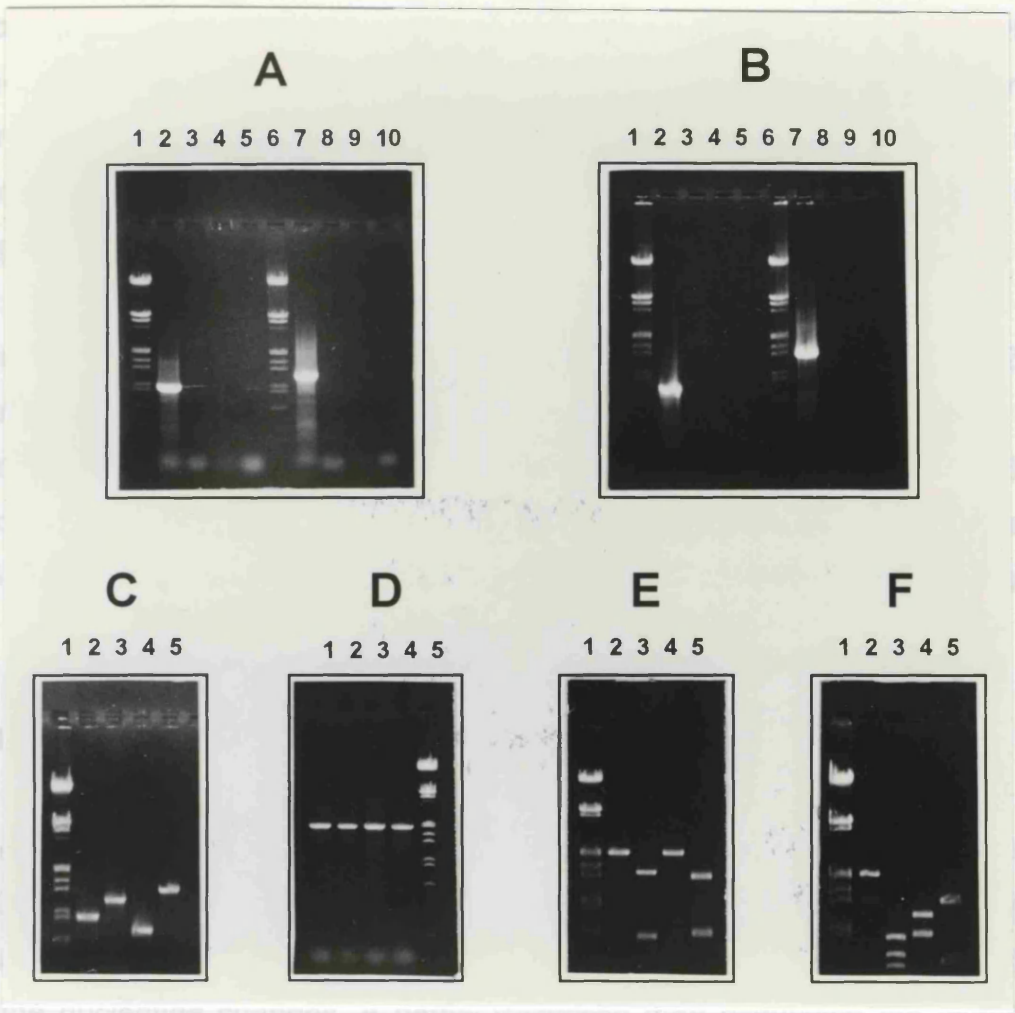


Figure 5.9 Production of Q360K and Q427K triple mutants of GLUT1.

A.Q360K primary PCR products plus negative controls.

B.Q427K primary PCR products plus negative controls.

C.Purified Q360K and Q427K primary products.

D.Secondary PCR (2kb) products of Q360K and Q427K mutants.

E.Restriction analysis of purified Q360K 2kb PCR product.

F.Restriction analysis of purified Q427K 2kb PCR product.

PCR product, whereas an identical incubation with the KK.Q427K PCR product results in two bands of 1139bp and 870bp. Digestion of the KK.Q360K PCR product with *Apa* I produces two bands corresponding to 1467bp and 542bp, whereas an identical digestion of the KK.Q427K mutant produces three bands corresponding to 827bp, 640bp and 542bp, demonstrating the creation of the additional Apa I restriction site. The digestion with *Xho* I demonstrates that the original KK451/456RR mutation is still present.

The 2kb GLUT1 PCR products containing the KK.Q360K and KK.Q427K mutations were restricted with *Stu* I and *Hin*d III and the 888bp fragments gel purified (data not shown). These cassettes were ligated back (Section 2.3.8) into the wild-type pGEM.GT to generate plasmid constructs pGEM.K3 and pGEM.K4, respectively, which were then sequenced (Section 5.5) and assayed for functional expression in *Xenopus* oocytes (Section 5.7).

5.4.4 Restriction analysis of mutant plasmid constructs

The plasmid constructs pGEM.KK, pGEM.K3 and pGEM.K4, possessing the KK451/456RR, KK451/456RR plus Q360K and KK451/456RR plus Q427K mutations, were analysed by restriction digestion to confirm the incorporation of the nucleotide changes. A partial restriction map combining the changes made in all of these mutants is shown in Figure 5.10A. Restriction enzyme digestion of pGEM.KK (Figure 5.10B) with *Bam* HI (lane 1) linearised the construct, producing a single band corresponding to the recombinant plasmid at 5111bp. The presence of an insert corresponding to the entire GLUT1 coding sequence was detected by the release of a 1921bp fragment from the recombinant plasmid via double-digestion with *Bam* HI and *Hin*d III (Figure 5.10B, lane 2). A less mobile band corresponding to the vector, pGEM11, was also visible at 3190bp. In order to confirm that the recombinant PCR procedure had generated the desired mutations, a restriction digestion of the mutant recombinant plasmid was performed with *Xho* I. Restriction digestion of pGEM.KK by *Xho* I produces a band pattern that differs from that yielded with

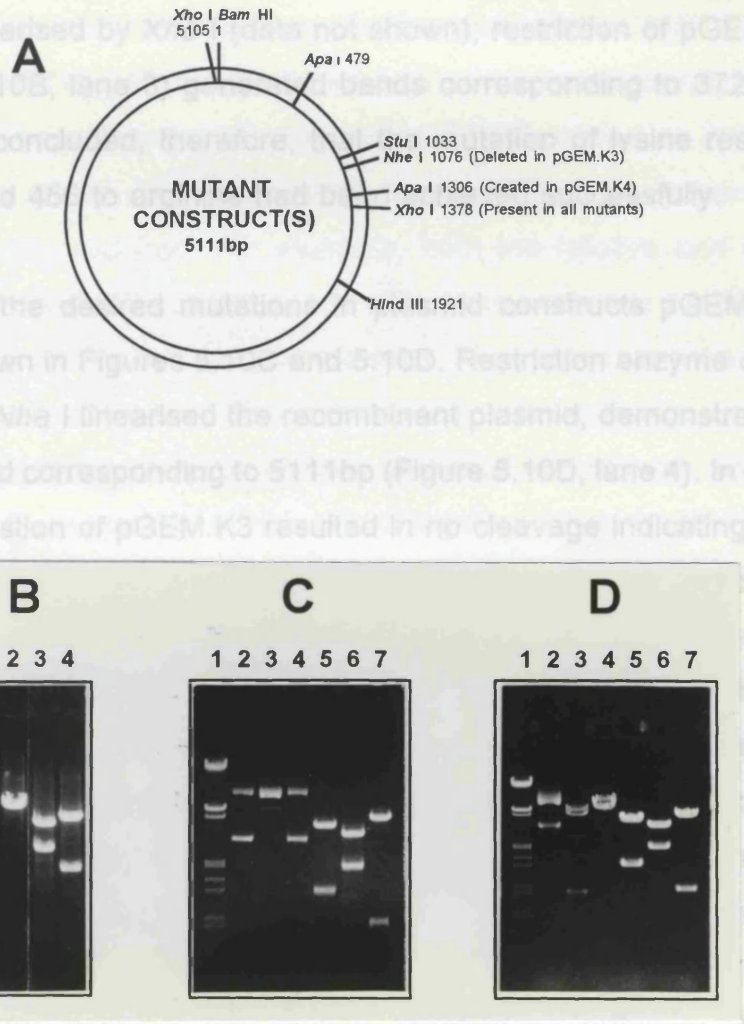


Figure 5.10 Partial restriction map and restriction analysis of GLUT1 mutant constructs.

(**A**). GLUT1 mutant restriction map. (**B**). Lane 1, λ *Hin*d III markers. Lane 2, *Bam* HI restricted pGEM.KK. Lane 3, *Bam* HI/*Hin*d III restricted pGEM.KK. Lane 4, *Xho* I restricted pGEM.KK. (**C**). Lane 1, λ *Hin*d III/*Eco* RI markers. Uncut plasmid DNA, Lane 2. pGEM.K3 plasmid DNA restricted with *Apa* I, Lane 3, *Nhe* I, Lane 4, *Xho* I, Lane 5, *Bam* HI / *Hin*d III, Lane 6, *Stu* I / *Hin*d III, Lane 7. (**D**). Digestions of pGEM.K4 as for pGEM.K3.

an identical digestion of the wild-type construct, pGEM.GT. That is, whereas pGEM.GT is linearised by *Xho* I (data not shown), restriction of pGEM.KK by *Xho* I (Figure 5.10B, lane 3) generated bands corresponding to 3727bp and 1384bp. It was concluded, therefore, that the mutation of lysine residues at positions 451 and 456 to arginine had been achieved successfully.

Confirmation of the desired mutations in plasmid constructs pGEM.K3 and pGEM.K4 is shown in Figures 5.10C and 5.10D. Restriction enzyme digestion of pGEM.K4 by *Nhe* I linearised the recombinant plasmid, demonstrated by a single sharp band corresponding to 5111bp (Figure 5.10D, lane 4). In contrast, an identical digestion of pGEM.K3 resulted in no cleavage indicating that the *Nhe* I site had been removed (Figure 5.10C, lane 4); the band pattern resembled that of the uncut plasmid DNA (Figure 5.10C, lane 2). Restriction enzyme digestion of pGEM.K3 with *Apa* I linearised the recombinant plasmid (Figure 5.10C, lane 3) whereas an identical digestion of pGEM.K4 revealed the additional *Apa* I site created by recombinant PCR, as bands corresponding to 4284bp and 827bp are visible (Figure 5.10D, lane 3) together with a band at 5111bp indicative of partial cleavage. Digestion of the mutant constructs with *Bam* HI / *Hin*d III and *Stu* I / *Hin*d III (Figures 5.10C and 5.10D, lanes 6 and 7) demonstrated that both of the mutant constructs contained an insert corresponding in size with the GLUT1 cDNA, and that accurate ligation of the mutant cassettes back into the remainder of the GLUT1 cDNA had occurred. Having established that the desired mutations had been produced, the next step was to sequence the constructs to ensure that no additional mutations had arisen due to *Taq* polymerase infidelity (Section 5.5).

5.5 Partial sequencing of GLUT1 mutants

Sequencing of PCR products is essential to identify any unwanted mutations generated during the process. There is a low frequency (≤0.25%) of such mutations per nucleotide (Saiki *et al.*, 1988) that can be attributed to the misincorporation of deoxynucleotides by *Taq* polymerase during template-

dependent chain elongation. In addition, *Taq* polymerase possesses a terminal transferase-like activity (Clark, 1988) that preferentially adds dAMP to double-stranded DNA, and *Taq* polymerase tolerates 3' primer/template mismatches. However, there are modifications to the standard PCR procedure that can be used to minimise errors, and these were employed in order to reduce the amount of screening required. For example, both the relative and absolute concentrations of the dNTP substrates are critically important in the production of base substitution errors by DNA polymerases. High concentrations of dNTP substrates generally increases the polymerase error rate by driving the enzymatic reaction in the direction of DNA synthesis, thereby decreasing the amount of error discrimination in the extension step. In contrast, low dNTP concentrations will tend to increase fidelity by influencing the rate at which the polymerase extends misprimed termini. Further, maximum fidelity is achieved by performing reactions with equal concentrations of all four dNTP substrates. In addition, *Taq* polymerase fidelity can be improved by decreasing the magnesium chloride concentration relative to the total concentration of dNTPs present in the reaction (Eckert and Kunkel, 1990). Also, the minimisation of reaction times and number of cycles results in an enhancement of fidelity.

The sequence integrity of the constructs coding for the halves of GLUT1 was verified by automated sequencing at the University of Sussex (data not shown), but the sequencing of the point mutants of GLUT1 was performed manually by the chain termination method (Sanger *et al.*, 1977) as described in Section 2.3.10. The coding regions of the *Stu* I/*Hin*d III cassettes containing the mutations of GLUT1 were sequenced using a combination of complementary oligonucleotides as primers for sequencing, shown in Figure 5.11A. That is, complementary oligonucleotides were used to sequence outwards from the middle of the cassette to the 5' and 3' ends. In addition, the reverse-complement sequence information was generated by using the KK451/456RR and Q360K mutagenic oligonucleotides. In this way, it was possible to obtain the forward and reverse sequence of each of the cassette coding regions. Five potential clones of each of the mutants were sequenced manually and it was

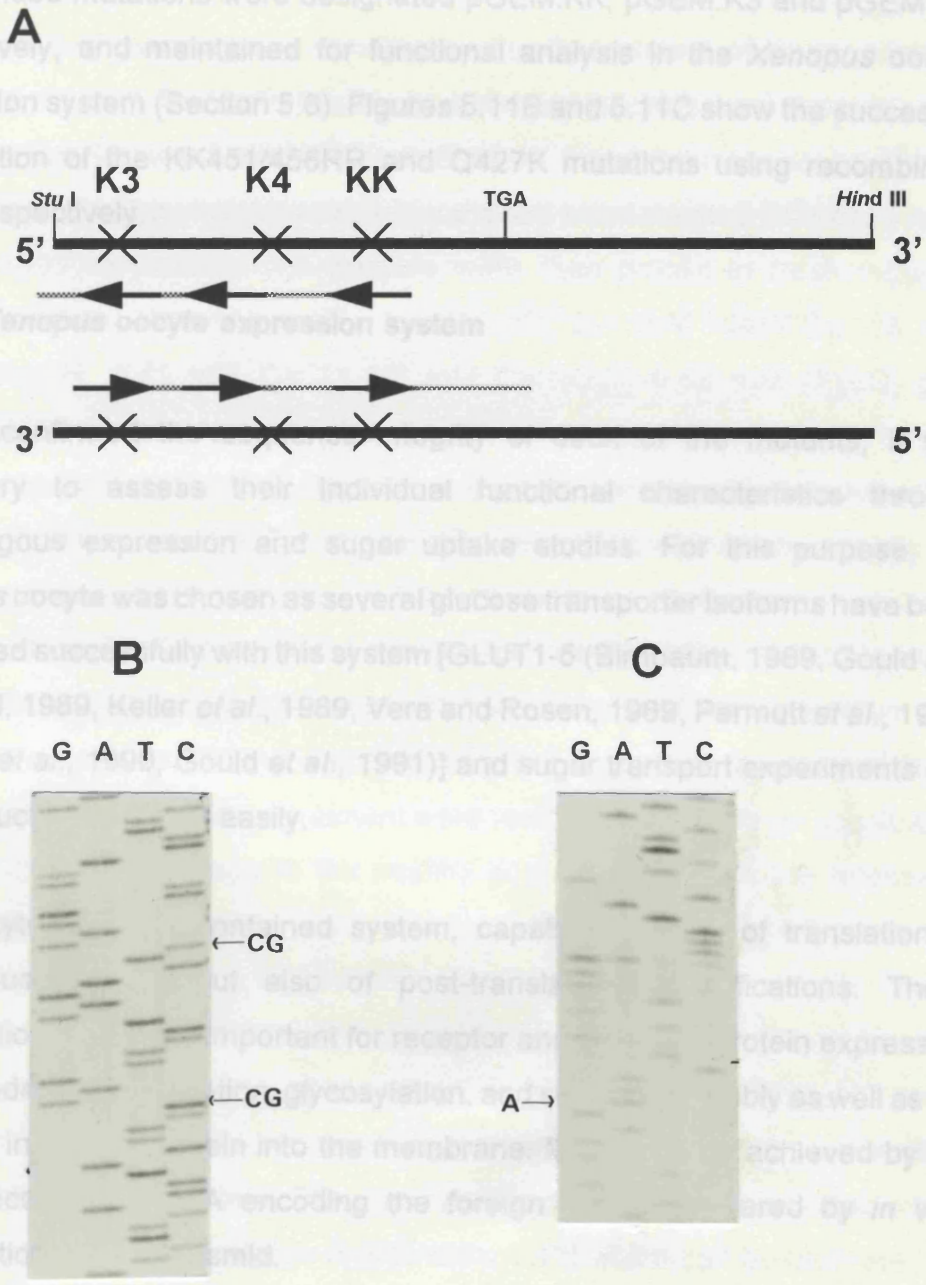


Figure 5.11 Sequencing of GLUT1 mutant constructs.
(**A**) Strategy for manual sequencing of pGEM.KK and pGEM.K4. Incorporation of KK451/456RR (**B**) and Q427K (**C**) mutations. The sequence integrity of pGEM.K3 was also confirmed by manual sequencing (data not shown).

determined that 20% of the clones of each mutant did not possess any unwanted *Taq* polymerase-induced mutations. Those clones possessing only the intended mutations were designated pGEM.KK, pGEM.K3 and pGEM.K4, respectively, and maintained for functional analysis in the *Xenopus* oocyte expression system (Section 5.6). Figures 5.11B and 5.11C show the successful introduction of the KK451/456RR and Q427K mutations using recombinant PCR, respectively.

5.6 *Xenopus* oocyte expression system

Having confirmed the sequence integrity of each of the mutants, it was necessary to assess their individual functional characteristics through heterologous expression and sugar uptake studies. For this purpose, the *Xenopus* oocyte was chosen as several glucose transporter isoforms have been expressed successfully with this system [GLUT1-5 (Birnbaum, 1989, Gould and Lienhard, 1989, Keller *et al.*, 1989, Vera and Rosen, 1989, Permutt *et al.*, 1989, Kayano *et al.*, 1990, Gould *et al.*, 1991)] and sugar transport experiments can be conducted relatively easily.

The oocyte is a self-contained system, capable not only of translation of exogenous mRNAs but also of post-translational modifications. These modifications are often important for receptor and transport protein expression and include phosphorylation, glycosylation, and subunit assembly as well as the ability to insert the protein into the membrane. Expression is achieved by the microinjection of mRNA encoding the foreign protein prepared by *in vitro* transcription from a plasmid.

5.6.1 Isolation of oocytes for expression studies

It is important to isolate oocytes at the correct stage of development: oocytes should be in stage V or VI, which is characterised as having a distinct boundary between the animal and vegetal poles (dark brown and light green/yellow,

respectively). The pigmentation of the animal pole must be uniform, that is, a mottled appearance should not be evident, or microinjection will prove futile.

Xenopus laevis ovaries were donated by the Department of Pharmacology, University of Leeds for use in this study. Individual oocytes were isolated using 'watchmakers' forceps under a Nikon SMZ-1B binocular microscope (10x23 eyepieces) with illumination provided by a Scott swan-necked fibre-optic lamp unit (KL1500 electronic). The oocytes were then placed in fresh modified Barth's medium (88 mM NaCl, 1 mM KCl, 2.4 mM $NaHCO_3$, 15 mM HEPES/NaOH, 0.41 mM $CaCl_2$, 0.3 mM $Ca(NO_3)_2$, 0.82 mM $MgSO_4$ plus penicillin and streptomycin sulphate, each at 10 µg/ml) containing 1mg/ml freshly prepared collagenase (Type II) and left at room temperature for 40-50 minutes. This treatment, which removes the tough follicular cell layer, was not totally necessary but it did allow the injections to be performed much more easily, and also prevented blockage of the needle. After the collagenase treatment, the oocytes were washed with five changes of Barth's medium and then left overnight at 18°C. It was extremely important that oocytes which did not survive the collagenase treatment were removed as quickly as possible, in order to minimise damage to the healthy population through the release of proteases.

5.6.2 Injection of mRNA

The injection of mRNA into oocytes requires a microinjector with a needle of about 10-20 µm in diameter. The apparatus used to pull such needles was a PUL-1 system (World Precision Instruments, WPI) which can be calibrated for consistency. The WPI nanoliter microinjector uses 7" Drummond capillaries (0.0285 inches internal diameter, 0.048 inches external diameter). Once pulled, the needles were filled with mineral oil (Sigma) and then attached to the microneedle of the microinjector; it was critical at this stage to avoid the introduction of air bubbles which would have rendered quantitative injections impossible. The needles were then loaded with mRNA (0.5 mg/ml, prepared as

described in Section 5.6.3) and used to inject individual oocytes.

5.6.3 Preparation of 5'-capped mRNA by *in vitro* transcription of plasmid DNA

One advantage of using the pGEM series of vectors from Promega is the presence of opposing T7 and SP6 promoters which allow for *in vitro* transcripts to be prepared from both strands of the cloned insert. All of the GLUT1 mutants described in this study were oriented with the T7 promoter at the 5' end of the insert (Figure 5.6). Synthetic mRNA suitable for translation in *Xenopus* oocytes was produced via a modification of the plasmid DNA transcription method described by Gould and Lienhard (1989). To avoid degradation of the RNA, gloves were worn throughout, and all solutions and materials were treated with diethyl pyrocarbonate (DEPC) to ensure RNase-free conditions.

The main features of the *in vitro* transcription procedure are described in Figure 5.12. It was particularly important to ensure that linearisation of the plasmid by *Hin*d III was complete, since a small amount of undigested plasmid can give rise to very long transcripts. Consequently, linearisation was confirmed routinely by running a small amount of the digestion mixture on a 0.8% agarose gel. Upon complete digestion of the plasmid, the digestion was phenol extracted as described in Section 2.3.1 and the DNA was precipitated by the addition of 0.3 volumes of 5 M potassium acetate and 2.5 volumes of ice-cold absolute ethanol. This mixture was incubated at -20°C overnight then centrifuged for 30 minutes at top speed in a microfuge. The supernatant was removed and the pellet resuspended in about 500 µl ice-cold 70% ethanol by vortexing, followed by a second 30 minute spin at 4°C. The supernatant was then discarded and the pellet was allowed to dry before being resuspended in 10 µl RNase-free water.

For *in vitro* transcription of RNA, a total of 5 µg DNA was made up to a final volume of 100 µl with 40 mM Tris-HCl buffer, pH 7.5, containing 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 0.5 mM each of ATP, CTP, UTP, 0.1 mM GTP,

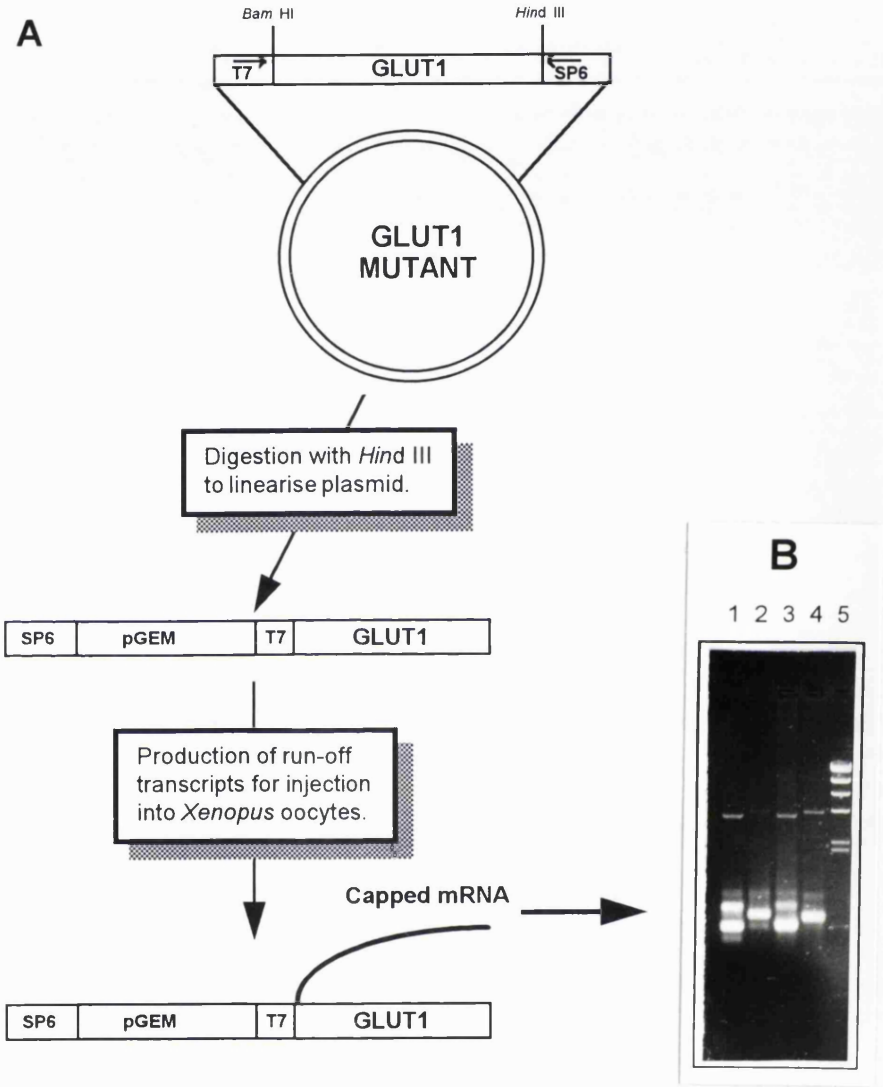


Figure 5.12 *In vitro* transcription.

A. Schematic illustration of *in vitro* transcription procedure.

B. mRNA preparation for halves of GLUT1, corresponding to residues 1-234 (Lane 1), 235-492 (Lane 2), 1-263 (Lane 3) and 264-492 (Lane 4). Lane 5 contains *Hin*d III-restricted λ DNA as reference.

and 0.5 mM diguanosine triphosphate (5' cap analogue). Thirty units of T7 RNA polymerase were then added and the mixture was incubated at 39°C. After 1 hour, an additional 10 units of T7 polymerase were added and followed by a further incubation period of 30 minutes.

The mixture was then extracted once with phenol and twice with chloroform before the RNA was recovered with two rounds of ethanol precipitation with potassium acetate. Quantification and visualisation of the RNA transcripts was achieved by UV spectroscopy and denaturing agarose gel electrophoresis, the latter being performed in the following manner. An aliquot of mRNA (2-5 µl) was added to 15 µl of sample buffer [10:3.5:2 (v/v) ratio of deionised formamide, 37% formaldehyde, and 3-(*N*-morpholino)propanesulfonic acid (MOPS) buffer (0.2 M MOPS, pH 7.0, 50 mM sodium acetate, 5 mM EDTA, pH 8.0)]. 2-5 µl of loading buffer (50% glycerol, 1 mM EDTA, 0.4% bromophenol blue, 1 µg/µl ethidium bromide) was then added and the sample was heated for 10 minutes at 65°C prior to loading on an agarose gel. The gel was run under normal conditions for the analysis of DNA samples. Figure 5.12B is a typical example of such a gel which, in this case, illustrates mRNA preparations derived from template cDNA molecules corresponding to the halves of GLUT1. The mRNA samples were then adjusted to 0.5mg/ml prior to injection.

For expression experiments, oocytes (typically 15 for each species of mRNA) were injected with 50 nl of mRNA (0.5mg/ml) or RNase-free water to give an indication of basal transport. Immediately after injection, oocytes were incubated at 18°C for 72 hours with changes of Barth's medium every 24 hours. Unhealthy oocytes (as judged by a mottled appearance of the animal pole) were discarded as soon as possible, although usually only a very small proportion (<5%) of the microinjected oocytes did not survive until assay.

5.6.4 Analysis of functional expression

In order to assess whether the mutants of GLUT1 created in this study were capable of transport, functional expression was assayed by 2-deoxy-D-[^{3}H(G)]-glucose, ([^{3}H]-2DG, 8.1 Ci/mmol, DuPont-New England Nuclear) uptake experiments, essentially as described by Gould and Lienhard (1989). After the 72 hour expression period, groups of five oocytes were incubated in 25 µM 2-deoxy-D-glucose, 0.5 µCi [^{3}H]-2DG in 500 µl Barth's medium for 1 hour at room temperature with gentle agitation. The uptake was terminated by three rapid washes with ice-cold PBS containing 0.1 mM phloretin, a potent transport inhibitor (Krupka, 1971). Oocytes were then dispensed to individual scintillation vials containing 500 µl of 1% SDS and incubated at room temperature for 1 hour with gentle agitation, before the addition of scintillation fluid and measurement of radioactivity in a Packard TR1900 liquid scintillation analyser.

5.6.5 Western blot analysis of *Xenopus* oocytes injected with wild-type and mutant GLUT1 mRNA

Oocytes (usually 10) believed to be expressing human GLUT1, or mutants thereof, were subjected to Western blot analysis to verify expression. The procedure used was adapted from that described previously by Geering *et al.* (1989). Oocytes were placed in a small petri dish containing 1 ml of ice-cold homogenisation buffer (83 mM NaCl, 1 mM $MgCl_2$, 10 mM HEPES, pH 7.9, 0.5 mM PMSF, 1 µg/ml pepstatin A, 0.2 mM di-isopropyl fluorophosphate and 10 µM L-*trans*-epoxysuccinyl-leucylamido-3-methylbutane). Using 'watchmaker's forceps, oocytes were squeezed gently and the plasma membrane 'ghosts' were separated from the yolk/cellular contents, which were expelled into the buffer. The resulting oocyte ghosts were placed in 1 ml of ice-cold homogenisation buffer with the final NaCl concentration raised to 0.5 M and gently stirred for 10 minutes on ice. After stirring, this buffer was centrifuged at 1,000 x *g* for 10 minutes at 4°C. The supernatant containing plasma membrane sheets was centrifuged further at 10,000 x *g* for 20 minutes at 4°C. The

resulting pellet was resuspended in 50 µl, assayed for protein concentration (Section 2.2.2), electrophoresed (Section 2.2.4) and transferred to nitrocellulose membrane (Section 2.2.5).

5.6.6 Time course of [^{3}H]-2DG uptake in *Xenopus* oocytes injected with mRNA encoding rat GLUT1

A time course for [^{3}H]-2DG uptake in *Xenopus* oocytes was performed in order to assess certain features of the procedure. This was achieved by injecting oocytes with 25 ng of rat GLUT1 mRNA (prepared as described in section 5.6.3). The oocytes were incubated in Barth's medium for 72 hours prior to sugar transport uptake studies (Section 5.6.4), whereupon sugar uptake was terminated after 0, 20, 40 and 60 minutes. Individual oocytes were counted for radioactivity in the usual manner. The results of this experiment are shown in Figure 5.13, and it is evident that the uptake of sugar occurs in a linear manner for up to 60 minutes. Subsequent incubation of the oocytes in the presence of radiolabelled 2-deoxy-D-glucose was thus performed for 60 minutes. Further, it was shown that sugar uptake could be inhibited efficiently by the addition of ice-cold PBS containing 0.1 mM phloretin. Further, no transport activity was detected in oocytes that had been injected with DEPC-treated water alone. The results, therefore, illustrate the utility of the *Xenopus* oocyte expression system in the analysis of sugar transport.

After having demonstrated the negligble endogenous sugar transport capability of the *Xenopus* oocyte, and also established optimal conditions for many aspects of the expression system, it was then possible to proceed with the injection of mRNA encoding the GLUT1 mutants in order to test for functionality (Section 5.7). For each series of injections, negative and positive controls were also included in the form of DEPC-treated water and wild-type rat GLUT1, respectively. Oocytes injected with wild-type rat GLUT1 were assayed for transport activity in the presence and absence of 0.1 mM phloretin. A further 10 oocytes were injected for Western blot analysis.

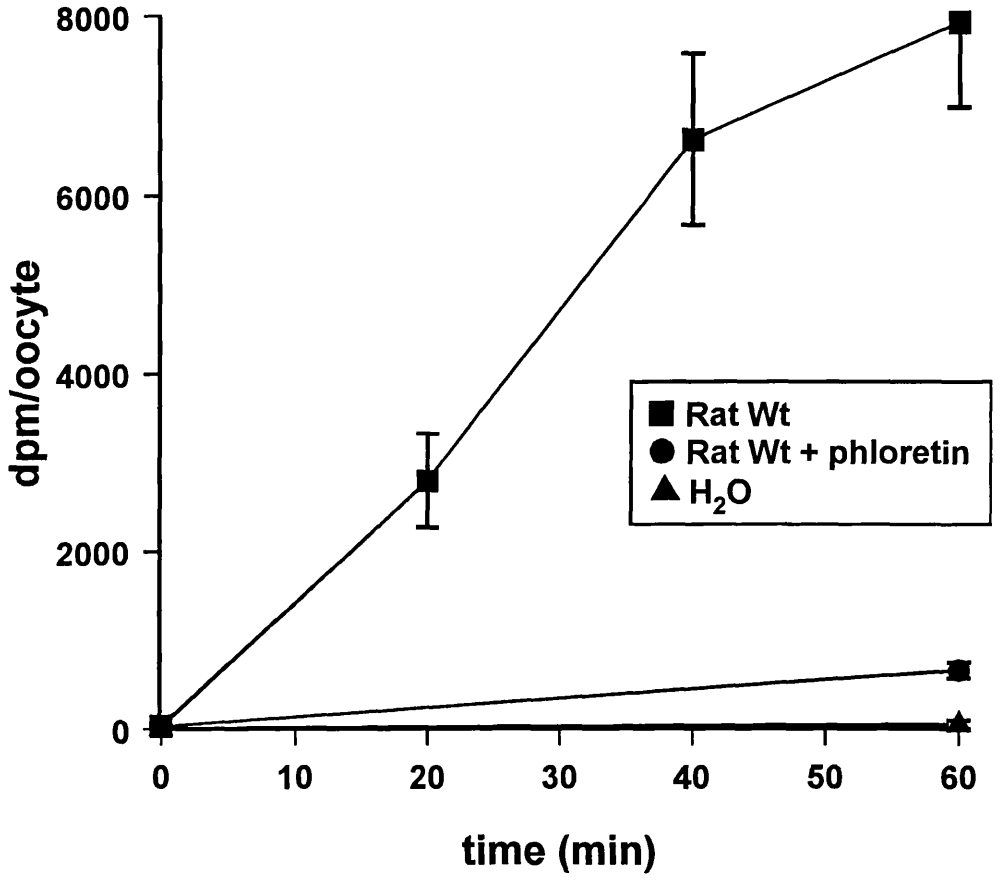


Figure 5.13 Time course of sugar uptake in oocytes expressing rat GLUT1.

5.7 Functional expression of GLUT1 mutants

5.7.1 Expression of point mutants designed for inhibitor binding-site studies

Xenopus oocytes were injected with 25 ng (50 nl) of mRNA encoding either the double-lysine mutant (KK), the KK.Q360K triple mutant, or the KK.Q427K triple mutant. The oocytes were incubated in Barth's medium for 72 hours prior to sugar transport uptake studies (Section 5.6.4) or Western blot analysis (Section 5.7.3). The results of a typical experiment are shown in Figure 5.14, which demonstrates that injection of each of the GLUT1 mutant mRNA species yielded a detectable sugar transport activity. Oocytes injected with either the KK mutant GLUT1 mRNA, or the KK.Q360K mutant GLUT1 mRNA demonstrated a level of sugar transport very similar to that of wild-type human GLUT1. These data suggest, therefore, that these mutations had not affected the function of the expressed protein to any significant extent. Oocytes that had been injected with mRNA encoding the KK.Q427K mutant GLUT1 transporter were also shown to be functionally active, but demonstrated an activity 35% of that of wild-type GLUT1. In all cases, uptake was sensitive to inhibition by phloretin (data not shown).

The results of the Western blotting experiments (Figure 5.16) demonstrate the successful expression of GLUT1 mutants in *Xenopus* oocytes, although the blots do also show that degradation of the expressed transporter has occurred. Use of an antibody raised against the *C*-terminus of GLUT1 revealed the presence of bands corresponding to intact GLUT1 at an apparent M_r 55,000 for the KK and KK.Q360K mutants, together with immunoreactive degradation products at an apparent M_r 27,000. These data demonstrate that the double-lysine mutant and the Q360K triple mutant have been expressed in the plasma membrane of the *Xenopus* oocyte. However, mainly degradation products of the Q427K triple mutant were detected by this procedure, with only a very small proportion of immunoreactive protein appearing as intact protein. Thus, it is possible, that the much lower transport activity observed for this mutant can be

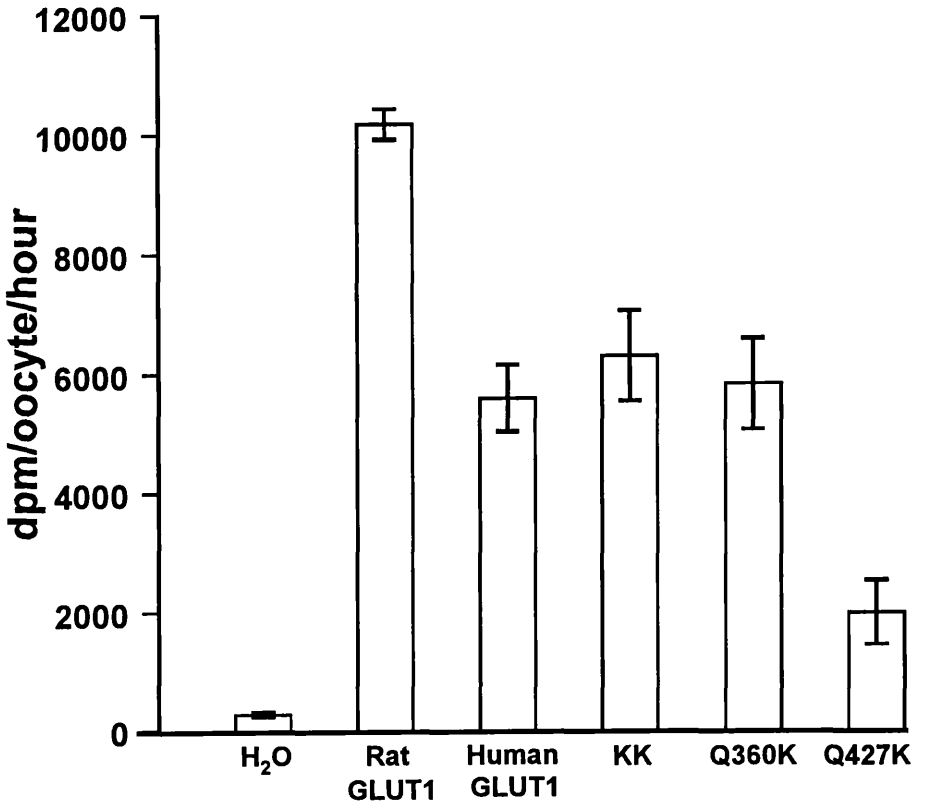


Figure 5.14 Functional expression of GLUT1 mutants in *Xenopus* oocytes. [^{3}H]-2DG uptake in *Xenopus* oocytes injected with mRNAs encoding GLUT1 mutants to be used for inhibitor binding-site studies. Datapoints are ± SEM (n=15). 2000 dpm corresponds to 12.5 pmoles.

attributed to reduced expression in the plasma membrane compared with the double-lysine mutant of GLUT1. However, it is unclear whether the presence of this exofacial lysine residue has disrupted the targeting process to the plasma membrane, or decreased the general stability of the protein in the oocyte.

5.7.2 Co-injection of mRNA encoding the *N*- and *C*-terminal halves of GLUT1

Xenopus oocytes were injected either separately with mRNAs (25 ng) encoding each of the GLUT1 half molecules or with a mixture of mRNAs encoding two of the half molecules (25 ng each). Since oocytes cannot readily tolerate an injected volume greater than 50 nl, it was necessary to concentrate the mRNA species to be co-injected by a factor of two, in order that equivalent amounts would be present compared with the separately injected mRNA species. The mRNA was concentrated by ethanol precipitation with potassium acetate, followed by resuspension at half the original volume. The injected oocytes were incubated in Barth's medium for 72 hours prior to sugar transport uptake studies (Section 5.6.4) or Western blot analysis (Section 5.6.5). Typical data from such uptake experiments are shown in Figure 5.15. None of the GLUT1 half molecules when injected separately into *Xenopus* oocytes produced a detectable increase in transport activity. Further, no increase in transport activity was observed upon the co-injection of mRNA encoding residues (1-263) plus (264-492), or the co-injection of mRNA encoding residues (1-234) plus (264-492) of GLUT1. However, the rate of glucose uptake into oocytes injected with the mixture of mRNAs encoding the halves of GLUT1 corresponding to residues (1-234) plus (235-492) was significantly greater than the control rate in water-injected oocytes and, in fact, was more than 50% of that seen in oocytes injected with wild-type human GLUT1 mRNA. These results, therefore, suggest that although the *N*- and *C*-terminal halves of GLUT1 are thought to have arisen from a gene duplication event, both halves of the GLUT1 molecule are required for its transport activity. Furthermore, these halves represent separate domains that are sufficiently stable to associate in the membrane after

synthesis to form a functional whole.

Figure 5.16 also shows the expression of the *C*-terminal halves of GLUT1 corresponding to residues 235-492 and 264-492 as detected by Western blotting. A sharp band of apparent M_r 25,000 is present in Lane 1 and corresponds to the predicted size of the *C*-terminal half of GLUT1 comprising residues 264-492. Also, an additional immunoreactive band running close to the degradation products of lanes 5, 6 and 7 is visible. This band is presumably a breakdown product from the aggregation of the GLUT1 halves 1-263 and 264-492, although no band corresponding to the M_r of intact GLUT1 is visible. Such a band, however, is evident in lane 3 which is the plasma membrane preparation from oocytes injected with GLUT1 halves corresponding to residues 1-234 and 235-492. An immunoreactive band of apparent M_r 28,500, corresponding to the predicted size of residues 235-492 of GLUT1 is also visible. Further, no cross-reactivity of this antibody to the endogenous oocyte transporter was observed (lane 8). Although it is likely that the *N*-terminal halves of GLUT1 are expressed in *Xenopus* oocytes to the same level as the C-terminal halves, conclusive evidence for this needs to be obtained through Western blotting studies using antibodies directed against *N*-terminal regions of the protein.

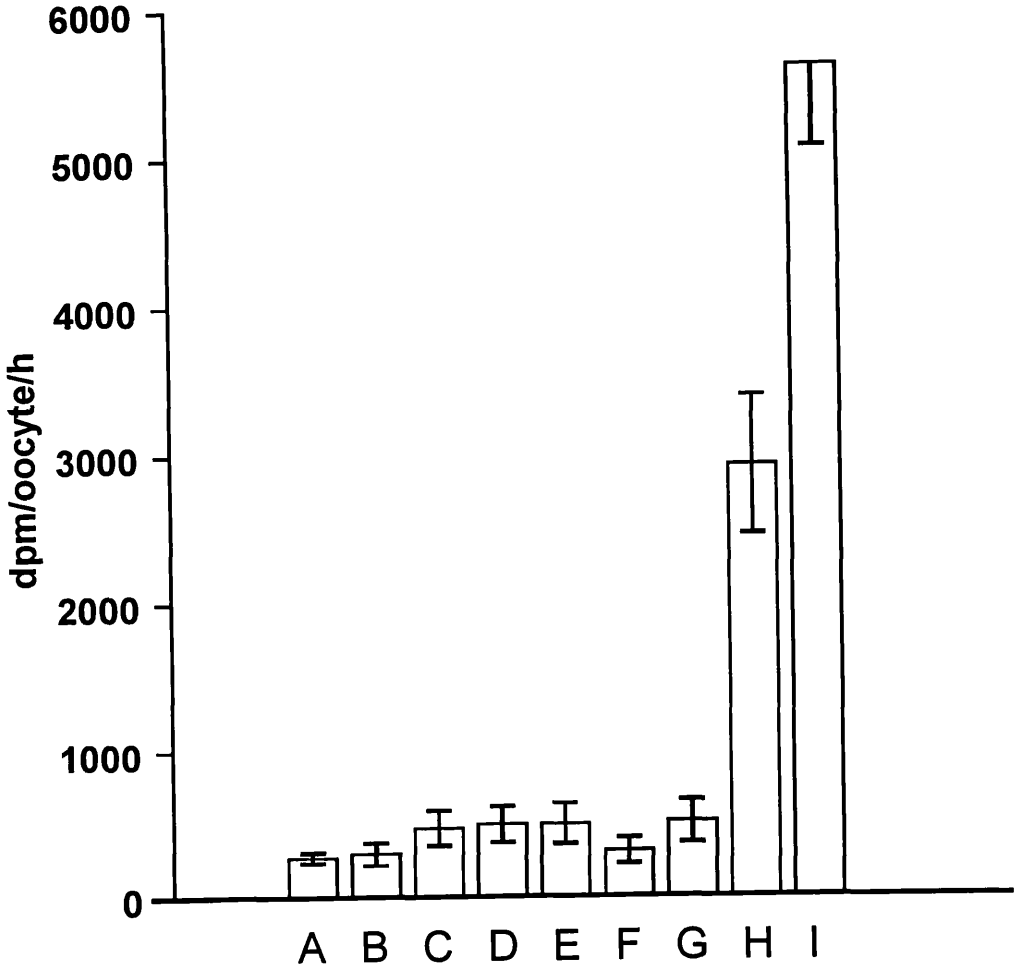


Figure 5.15 Functional expression of GLUT1 half molecules in *Xenopus* oocytes.

[^{3}H]-2DG uptake in *Xenopus* oocytes injected with mRNAs encoding the intact GLUT1 molecule and with mRNAs encoding the two half molecules, either separately or together. Results are expressed as mean ± SEM (n=15). 2000 dpm corresponds to 12.5 pmoles.

(**A**) Basal transport activity (H_20). (**B**) *N*-terminal half comprising residues 1-234. (**C**) *C*-terminal half comprising residues 235-492. (**D**) *N*-terminal half comprising residues 1-263. (**E**) *C*-terminal half comprising residues 264-492. (**F**) Residues (1-234) + (264-492). (**G**) Residues (1-263) + (264-492). (**H**) Residues (1-234) + (235-492). (**I**) Wild-type human GLUT1.

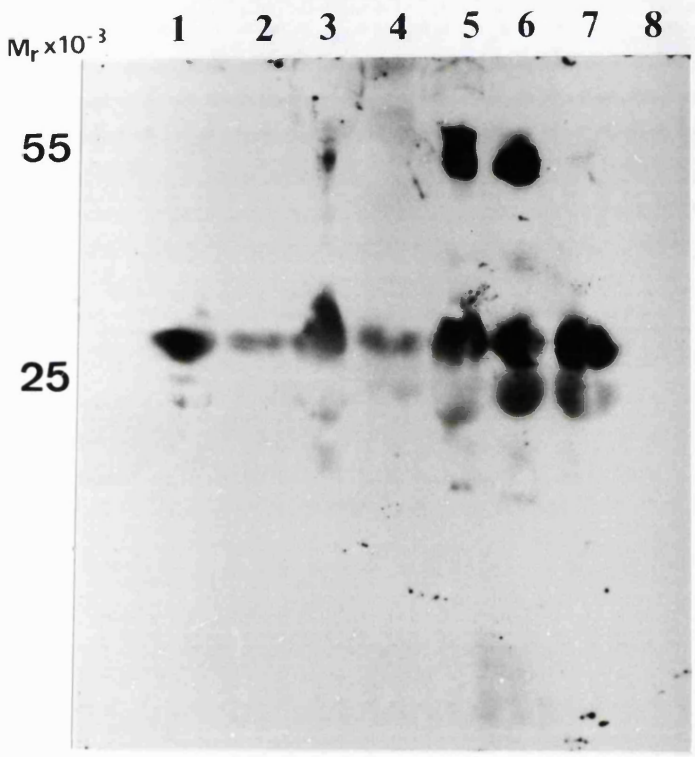


Figure 5.16 Western blot analysis of membranes prepared from *Xenopus* oocytes injected with GLUT1 mutant mRNA.

Membrane samples from oocytes injected with mRNA encoding GLUT1 residues 1-263 plus 264-492 (Lane **1**), 235-492 (Lane **2**), 1-234 plus 235-492 (Lane **3**), as lane 3, but from a separate experiment (Lane **4**), KK (Lane **5**), KK.Q360K (Lane **6**), KK.Q427K (Lane **7**) and DEPC-treated water (Lane **8**).

5.8 Discussion

In the absence of detailed structural information for any member of the sugar transporter family, site-directed mutagenesis has proven to be a powerful tool in providing structural and functional data about GLUT1. However, to be of use, all potential mutations need to be chosen carefully. For this study, a comprehensive multiple alignment of the sugar transporter family, and its subsequent computational analysis, were used to design oligonucleotide primers that would minimise damage to the resultant GLUT1 structural and functional characteristics. That is, for the mutants to be of subsequent use, it was imperative that functional activity could be demonstrated.

Although many different methodologies exist for the generation of mutants, this chapter describes an effective and efficient approach whereby PCR-derived mutated cDNA cassettes are religated into 'wild-type' cDNAs to yield mutants of potential structural and functional significance. The first mutant to be constructed required an interchange of lysine residues at positions 451 and 456 for arginine. The *C*-terminus of such a GLUT1 molecule should then be resistant to proteolytic cleavage by endoproteinase Lys-C, and thus provide an epitope to allow immunoaffinity purification of the protein using an immobilised polyclonal antibody raised against residues 460-477. Functional activity of this double-lysine mutant was demonstrated to be comparable with that of native GLUT1 when expressed in *Xenopus* oocytes, from which it was deduced that the structure of the protein had not been affected to a significant extent. The most important consequence of the functional expression of this mutant was that it was now possible to attempt to gain information regarding the exofacial substrate binding site through the identification of the site of labelling by ATB-BMPA. In the meantime, however, this mutant could be used as a template for the incorporation of additional mutations.

Such mutations took the form of substitutions of glutamine residues at positions 360 and 427 for lysine residues, and were designed with the aim of facilitating

the dissection of the *C*-terminal half of GLUT1 to enable the binding sites of specific inhibitors of transport to be determined. It was critical for these triple mutants to exhibit functional activity also, and this was indeed shown to be the case by the observation of sugar uptake in *Xenopus* oocytes. The implication of a failure of these mutants to induce sugar uptake would be a structural or mechanistic alteration, possibly resulting in the inability to bind site-specific ligands.

The Q360K triple mutant exhibited a transport activity approaching that of native GLUT1, whereas oocytes injected with Q427K triple mutant mRNA appeared to transport sugar less efficiently. From the Western blot analysis, it appears that this mutant is expressed at a significantly lower level than the double-lysine and Q360K triple mutants, as equivalent amounts of protein were loaded on to the gel. Since total oocyte membrane preparations were not assayed, it is possible that a large amount of this particular mutant was not targeted efficiently to the plasma membrane and thus remained trapped within the oocyte. Such a phenomenon has been noted with the expression of mutations at Trp_{388} and Trp_{412} in *Xenopus* oocytes (Garcia *et al.*, 1992).

The technique of recombinant PCR was also used to generate mutants of GLUT1 that were designed to test the dual-domain assembly hypothesis of GLUT1. It was found that phloretin-sensitive sugar transport activity could be regenerated by the co-expression of halves corresponding to residues 1-234 and 235-492 of the native GLUT1 protein. Transport activity was not observed, however, by expression of the individual halves or upon co-expression of halves corresponding to 1-263 plus 264-492 or 1-234 plus 264-492. Western blot analysis of plasma membrane preparations from injected oocytes indicated that the lack of transport could not be attributed to a lack of expression. Consequently, the reasoning behind the failure of the pair of halves corresponding to residues 1-263 plus 264-492 to induce sugar uptake is unclear although, taken with the lack of transport by co-expression of halves corresponding to 1-234 plus 264-492, it does suggest that the central

cytoplasmic loop plays a role in the adoption of a functional conformation. Whether this region plays a role in stabilising the inter-domain connections or is directly involved in the transport process remains to be established.

A similar study has been performed very recently by Cope *et al.* (1994) with the use of the baculovirus/*Sf*9 system to examine the co-expression of separate halves of rabbit GLUT1. They showed that when either half is expressed alone, no binding of ATB-BMPA or cytochalasin B is observed but when *Sf*9 cells are doubly-infected with virus constructs producing both halves of the transporter, ligand labelling is detected. It does appear, therefore, that the two halves of GLUT1 can assemble and produce a stable complex possessing the ability to bind site-specific ligands.

The assessment of 'functionality' in the study of Cope *et al.* (1994) was achieved through the binding of transport inhibitors as, unfortunately, the use of the baculovirus/*Sf*9 expression system precludes experiments on sugar transport. The reason for this is that by the time detectable levels of expressed transporter are appearing at the cell surface (about 2 days post-infection), the plasma membranes of the insect cells are becoming damaged. The use of *Xenopus* oocytes as the expression system of choice in this study allowed functionality to be examined directly at the level of transport.

In conclusion, the data presented in this chapter demonstrates the successful application of a methodology for the construction of mutants in conjunction with an excellent assay for functional activity. GLUT1 cDNA has been engineered in several ways to generate mutants to facilitate the gathering of structural and functional information about this sugar transporter. Further, point mutations of GLUT1 were generated with a dual purpose in mind. That is, lysine residues at positions proposed to exist on the exofacial surface of GLUT1 were created at positions 360 and 427, within the loops linking helices 9 and 10, and 11 and 12, respectively. Therefore, not only were proteolytic cleavage sites introduced into GLUT1 without a gross effect upon the structure of the protein, but the

accessibility of these sites to membrane-impermeant probes can now be tested directly. In addition, the data presented in this chapter, together the independent evidence of Cope *et al.* (1994), provide good evidence in support of the dual-domain assembly hypothesis for GLUT1. This, alone, should help to simplify the process of model building and the subsequent design of experiments to test the features of such models.

CHAPTER 6. APPLICATION OF THE BACULOVIRUS EXPRESSION VECTOR SYSTEM TO THE IDENTIFICATION OF TRANSPORT INHIBITOR BINDING SITES

6.1 Introduction

The baculovirus expression vector system is a helper virus-independent system which has been used to express genes from many different sources, such as eukaryotes, fungi, plants, bacteria and viruses, reviewed in Luckow and Summers (1989). Expression of proteins in this system has several advantages over prokaryotic and transfection-based eukaryotic expression systems, a major one being the ability of the system to produce biologically active proteins in substantial amounts. However, to understand the utility of the baculovirus expression system, an appreciation of the normal virus infection cycle is required.

The Baculoviridae are a large family of ds DNA viruses that infect many different species of insects as their natural hosts. Baculovirus strains are highly species-specific and are not known to infect any non-invertebrate host. The baculovirus genome is replicated and transcribed in the nuclei of infected host cells where the large baculovirus DNA (between 80kbp and 200kbp) is packaged into rod-shaped nucleocapsids. Since the size of these nucleocapsids is flexible, recombinant baculovirus particles can accomodate large amounts of foreign DNA. *Autographa californica* nuclear polyhedrosis virus (AcNPV) is the most extensively studied strain of baculovirus and its entire genome has been sequenced (Ayres *et al.*, 1994). Infectious virus particles enter susceptible insect cells by facilitated endocytosis or fusion and the viral DNA is uncoated in the nucleus (Figure 6.1). DNA replication starts about 6 hours post-infection (p.i.). From then on, the baculovirus infection cycle can be divided into two different phases, regardless of whether it is under tissue culture conditions or in infected larvae. During the early phase, the infected insect cell is releasing extracellular virus particles (ECV) budding off the cell membrane of infected

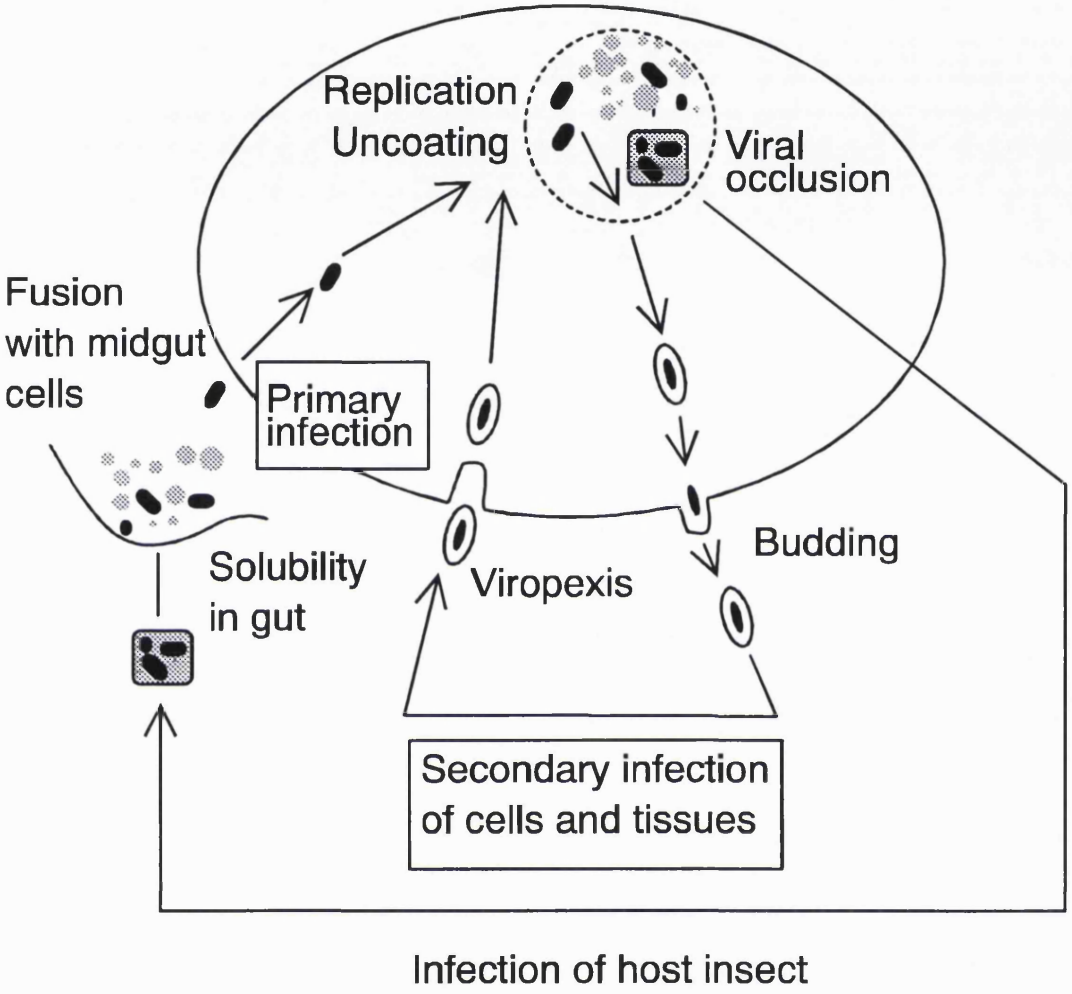


Figure 6.1 Schematic representation of the baculovirus life cycle. The occluded virus is responsible for horizontal transmission among susceptible insects and the extracellular virus is responsible for secondary and cell to cell infection in cultured cells or in the insect host.

cells. Later in the infection cycle occluded virus particles are assembled in the nucleus.

These particles are embedded in a homogeneous matrix made predominantly of a single protein, polyhedrin, and are released when the infected cells are being lysed during the last phase of the infection cycle. Whereas the first ECV particles are detectable 10 hours p.i., the first viral occlusion bodies of wild-type AcNPV develop 2 days p.i., but continue to accumulate and reach a maximum 5-6 days p.i.. Occlusion bodies are visible under light microscopy where they appear as dark polygonal-shaped bodies filling up almost the whole nucleus of the infected cell. Although not all known baculoviruses form occlusion bodies, the AcNPV virus is a representative of the group of occlusion body-positive baculoviruses. The polyhedrin protein which is the major component of the occlusion bodies has an apparent M_r of 29,000 (Summers and Smith, 1978). During the late phases of infection, the polyhedrin protein accumulates to very high levels (up to 1 mg/ml of polyhedrin protein per 1-2x10^6 infected cells) and may account for 30-50% of the total insect cell protein. *In vivo*, viral occlusion bodies are an important part of the natural baculovirus life cycle, essential for horizontal transmission of the virus in its native environment. When infected larvae die, millions of occlusion bodies are released by the decomposing material. Viral occlusion bodies protect the embedded virus particles from inactivation by environmental factors, such as heat and desiccation, that would otherwise rapidly inactivate unprotected ECV particles. When insect larvae feed on contaminated plants, they ingest the polyhedrin occlusion bodies. These occlusions dissolve in the alkaline environment of the insect gut, releasing infectious virus particles which invade the cells of the midgut tissue and replicate the virus DNA. After this initial phase of infection, secondary infection spreads to other insect tissues, via the ECV form, through the insect haemolymph.

Since the polyhedrin gene product itself is not essential for baculovirus replication in cell culture (Smith *et al.*, 1983), replacement of its coding

sequence by that of a foreign gene (in a cDNA form) allows expression of the foreign gene under the control of the retained polyhedrin promoter. Artificial deletion or insertional inactivation of the polyhedrin gene of AcNPV wild-type virus results in the production of occlusion body-negative viruses, a phenomenon which was exploited early on for the identification of recombinant viruses. Newer modified AcNPV viruses either allow colour selection to identify recombinants or even permit positive survival selection for recombinants (BaculoGold™ system, PharMingen) which has rendered the old occlusion body-based visual screening method obsolete.

The main attributes of the baculovirus system have been well described recently (Summers and Smith, 1987) and will not be detailed here, although several features are noteworthy. For example, the generation of recombinants is a relatively rapid procedure and high-level expression of biologically active products is possible. The system also supports a variety of common eukaryotic post-translational modifications including: protein folding, disulphide formation, myristoylation, phosphorylation, amidation, *N*-terminal processing, *O*- and *N*-linked glycosylation, intracellular targeting and secretion (Luckow and Summers, 1988). In addition, the capacity to express unspliced genes and even the simultaneous expression of multiple genes have made it the system of choice for many applications.

In a previous study from this laboratory, expression of wild-type GLUT1 in insect cells using the baculovirus system had been shown to yield considerable quantities of the transport protein (Yi *et al.*, 1992). The protein was clearly detectable by Western blotting 24 hours after infection and synthesis of the polypeptide became significant 2 days p.i.. The transporter was stable and accumulated until the cells lysed, 4-5 days p.i., a time course resembling that for the synthesis of polyhedrin protein in wild-type AcNPV-infected insect cells (Summers and Smith, 1987). The study showed that 4 days p.i. up to 1.47 nmol of transporter was present per mg of membrane protein, equivalent to almost 8% (w/w) of the total membrane protein.

Comparable levels of expression of other membrane glycoproteins, such as the Shaker K^+ channel, has been reported in the baculovirus system (Klaiber *et al.*, 1990), although the mammalian Na^+/glucose cotransporter was found to be expressed at a somewhat lower level, corresponding to 5% of the membrane protein (Smith *et al.*, 1992). Some non-membrane proteins have been expressed to levels equating to polyhedrin itself, i.e. 35-50% of total cell protein (Emery, 1991b). Thus, it is difficult to generalise how efficiently foreign genes will be expressed in the baculovirus expression system.

In the study by Yi *et al.* (1992), transport by the human glucose transporter expressed in insect cells was not demonstrated, although it was shown to possess native-like biological activity with respect to the binding of cytochalasin B. However, the concentration of cytochalasin B-binding sites obtained by ligand binding assay was less than the concentration of immunologically cross-reactive protein. A phenomenon similar to this has been reported for other membrane glycoproteins (Germann *et al.*, 1990, Klaiber *et al.*, 1990). Although the reason for this discrepancy is not clear, it may stem from the high level of expression of the transport protein. However, even though all of the expressed GLUT1 appears not to be active, Cope *et al.* (1994) have demonstrated that sufficient protein is produced to enable the characterisation of GLUT1 at the level of site-specific inhibitor binding. Indeed, GLUT1 expressed in insect cells was shown to be labelled by cytochalasin B and ATB-BMPA.

The rationale behind the construction of the GLUT1 mutants to be expressed in insect cells, and their subsequent functional expression in *Xenopus* oocytes, was reported in Chapter 4. The work described in this chapter details the attempt to generate recombinant baculovirus particles containing these GLUT1 mutants, and achieve expression levels sufficient for the identification of inhibitor binding sites. The baculovirus expression vector system was chosen for this purpose on the basis of its proven ability to express substantial amounts of functional GLUT1 in insect cells (Yi *et al.*, 1992, Cope *et al.*, 1994).

6.2 Generation of recombinant pAcYM1 transfer vectors containing GLUT1 cDNA mutants

In order to take advantage of the baculovirus expression system, foreign genes need to be positioned immediately downstream of a strong viral gene promoter. Since recombinant viruses containing foreign genes cannot be obtained by a direct cloning route, the foreign gene has to be first subcloned into a baculovirus transfer vector. The transfer vector utilised in the present study, pAcYM1, uses the polyhedrin promoter to drive foreign gene expression (Matsuura *et al.*, 1987). This vector contains all of the polyhedrin 5' leader sequence up to, and including, the 'A' of the normal polyhedrin ATG translation initiation codon, but lacks the rest of the polyhedrin coding sequence, plus 13 nucleotides downstream from the translation termination codon of the gene. The insertion site for pAcYM1 is a *Bam* HI restriction site (Figure 6.2).

The strategy used to achieve the construction of recombinant transfer vectors is described in Figure 6.3. Standard recombinant DNA techniques were used throughout, as described by Sambrook *et al.* (1989) with some modifications. The transfer vector pAcYM1 was kindly donated by Dr. V. Emery (Division of Communicable Diseases, Royal Free Hospital, University of London). The vector DNA was prepared by the alkaline lysis procedure (Section 2.3.2), linearised by restriction digestion with the enzyme *Bam* HI (Section 2.3.4) and dephosphorylated (Section 2.3.8). The strain JM109 was used for propagation of the pAcYM1, according to the methods described in Section 2.3.9.

An examination of the GLUT1 cDNA restriction map revealed that it was possible to subclone a 1536bp *Bst* YI fragment containing the entire GLUT1 coding sequence directly into the unique *Bam* HI cloning site of pAcYM1. All of the GLUT1 mutants, described in Chapter 4, were subcloned into the pGEM11-zf(-) vector which, unfortunately, possesses multiple sites recognised by *Bst* YI. Thus, in order to obtain the *Bst* YI fragment it was first necessary to digest the pGEM.GT recombinant constructs with *Bam* HI and *Hin*d III, and gel

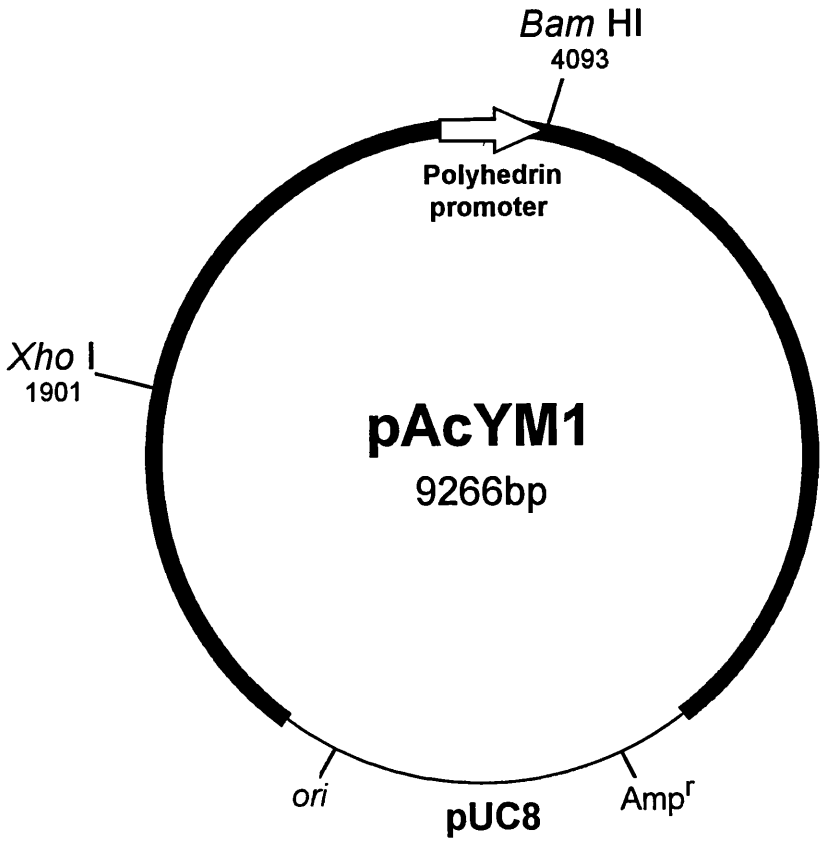


Figure 6.2 Partial restriction enzyme map of transfer vector pAcYM1.

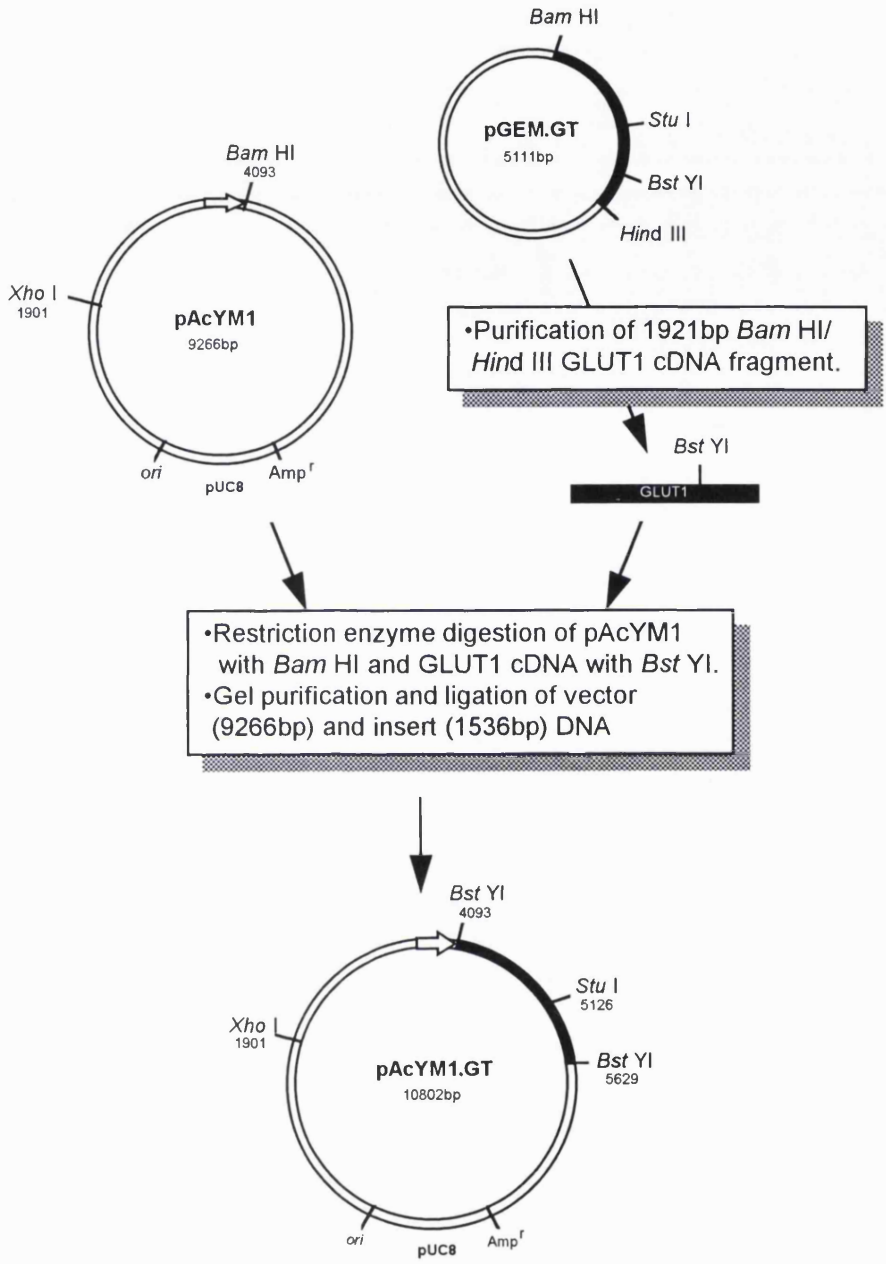


Figure 6.3 Strategy used for the subcloning of GLUT1 mutants into the transfer vector pAcYM1.

purify the resultant 1921bp GLUT1 cDNA fragments (data not shown) prior to digestion with *Bst* YI at 60°C for 2 hours. The 1536bp mutant GLUT1 cDNA molecules were then gel purified and ligated with *Bam* HI linearised pAcYM1.

In order to assess the orientation of the GLUT1 cDNA fragments within the recombinant transfer vectors, a double restriction digestion with *Xho* I (which cuts at position 1901 within the vector sequence) and *Stu* I (which cuts the wild-type GLUT1 cDNA at position 1033) was performed. Recombinants possessing a correctly oriented, wild-type insert would be expected to produce a pattern on a 0.8% agarose gel with bands corresponding to 7577bp and 3225bp, whereas bands corresponding to 8107bp and 2695bp would be expected for an incorrectly oriented wild-type insert. All of the mutants, however, possess an additional *Xho* I restriction site at position 1378 and, as a consequence, would produce a slightly different band pattern on an agarose gel. Correctly oriented GLUT1 mutants, upon digestion with *Xho* I and *Stu* I, would be expected to produce a band pattern corresponding to 7232bp, 3225bp and 345bp. An incorrectly oriented recombinant transfer vector digested in an identical manner would reveal DNA fragments corresponding to 7762bp, 2695bp and 345bp. It was, therefore, the appearance of bands corresponding to 3225bp or 2695bp that determined the correct or incorrect orientation of the GLUT1 inserts. Figure 6.4 shows a series of restriction digests of recombinant transfer vectors that confirm the presence and correct orientation of the wild-type GLUT1, the double-lysine mutant, the KK.Q360K and KK.Q427K triple mutant cDNA inserts. These constructs are hereafter designated pAcYM1.GT, pAcYM1.KK, pAcYM1.K3 and pAcYM1.K4, respectively.

Once the recombinant transfer vector containing the gene of choice under the control of one of the strong baculovirus promoters has been generated, recombinant baculovirus particles are then prepared via the co-transfection of purified transfer vector and AcNPV baculovirus DNA into insect cells.

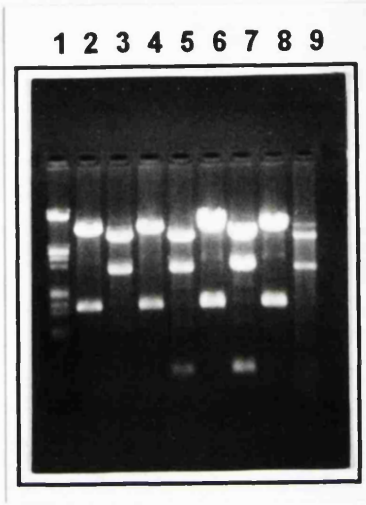


Figure 6.4 Restriction digests of recombinant transfer vectors containing wild-type and mutant GLUT1 cDNA fragments.
pAcYM1.GT (Lanes 2 & 3) pAcYM1.KK (Lanes 4 & 5), pAcYM1.K3 (Lanes 6 & 7) and pAcYM1.K4 (Lanes 8 & 9) digested with *Bam* HI and *Xho* I / *Stu* I, respectively. Lane 1 contains *Hin*d III / *Eco* RI restricted λ DNA.

6.3 Generation of recombinant baculovirus

The technology employed for the production of recombinant baculovirus in this study was the BaculoGold™ system (PharMingen) which provides a tool for the selection of recombinants and achieves recombination efficiencies of virtually 100%. The principle of this technique, referred to as allelic replacement, is illustrated in Figure 6.5, and utilises a modified type of baculovirus DNA (BaculoGold™) containing a lethal deletion and, therefore, does not code for viable virus itself. Co-transfection of the BaculoGold™ DNA with a complementing plasmid construct rescues the lethal deletion of this virus DNA and reconstitutes viable virus particles inside the transfected insect cells. For this recombination event to take place, the baculovirus transfer plasmid must be polyhedrin locus-based, that is, the flanking sequences of its promoter region must be derived from the polyhedrin locus of the AcNPV wild type virus.

6.3.1 Preparation of transfer vector DNA

High quality DNA is required for the purpose of transfection. It was decided, therefore, to prepare the transfer vector DNA using a 'spin-column' method, rather than by alkaline lysis (Section 2.3.1) which can result in less pure DNA preparations. The recombinant transfer plasmids of the GLUT1 mutants designed for inhibitor binding-site studies (pAcYM1.GT, pAcYM1.KK, pAcYM1.K3 and pAcYM1.K4) were propagated by transforming 100 ng of the plasmid DNA into competent *E. coli* (JM109) cells (Section 2.3.9). A small volume (5 ml) of LB medium was inoculated with a single colony and grown at 37°C overnight in the presence of ampicillin (50 µg/ml). Plasmid DNA was then prepared by using a Magic Miniprep DNA preparation kit (Promega). Cells from 2 ml of the overnight-grown culture were pelleted by centrifugation using a microfuge and resuspended in 20 µl of Resuspension Solution (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 100 µg/ml RNase A). The cells were lysed by adding 200 µl of Lysis Solution (0.2 M NaOH, 1% SDS), mixed by inverting the tube several times, and neutralised by the addition of 200 µl of Neutralisation

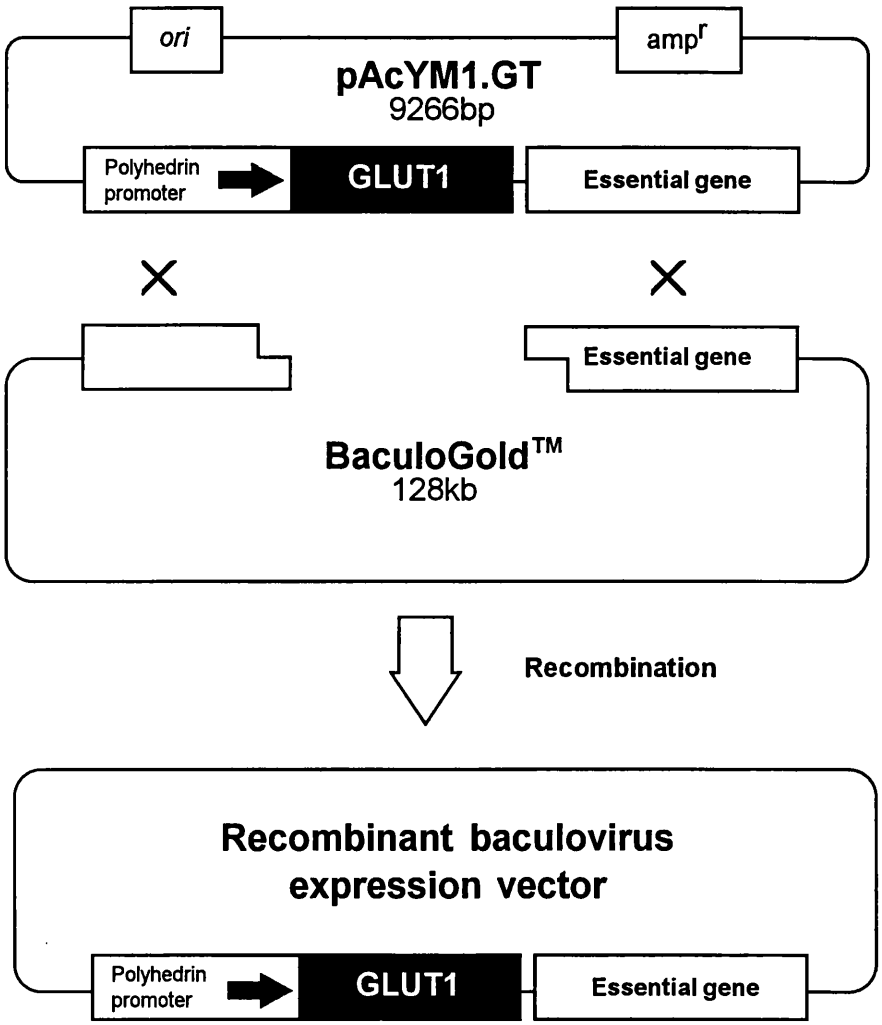


Figure 6.5 Allelic replacement reaction between recombinant transfer vector and viral DNA.

Solution (2.55 M potassium acetate, pH 4.8). The mixture was centrifuged for 5 minutes at 12,000 x *g*. The supernatant was mixed well with 1 ml of the Magic Minipreps DNA purification Resin, and then passed through a Magic Minicolumn. The column was washed with 2 ml of column wash solution (200 mM NaCl, 20 mM Tris-HCl, pH 7.5, 5 mM EDTA) and centrifuged for 30 seconds at 12,000 x *g* in a microfuge to dry the resin. The DNA was eluted by applying 40 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) onto the column and centrifuging for 20 seconds at 12,000 x *g*. The plasmid DNA was then analysed by restriction digestion and electrophoresis on a 0.8% agarose gel.

6.3.2 Co-transfection of insect cells

To introduce the mammalian glucose transporter gene into the AcNPV genome, the permissive insect cell line *Sf*9 was co-transfected with the recombinant transfer vectors and the viral DNA BaculoGold™ using lipofectAMINE™ (Gibco BRL). This reagent is a polycationic liposome transfection reagent that incorporates a spermine head group and is reported to increase the frequency and activity of eukaryotic transfection efficiency in most cell types, achieving up to 30-fold higher activity than monocationic reagents (Hawley-Nelson *et al.*, 1993). Insertion of the GLUT1 cDNA into the baculovirus genome was then achieved via a cell-mediated homologous recombination between the flanking sequences of the vector and the wild-type viral DNA.

*Sf*9 cells (2 x 10^6) were seeded into 25 cm^2 tissue culture flasks and allowed to attach for 15 minutes at room temperature. Meanwhile, 500 ng (5 µl) of the linearised BaculoGold™ viral DNA was combined with 2 µg of each of the purified GLUT1 mutant transfer vectors. The mixture was added to 50 µl of lipofectAMINE™ solution (diluted 1:10 to achieve a final concentration of 100 µg/ml) and left at room temperature for 15 minutes. During this time, the culture medium was removed from the cells which were then washed twice with serum-free TC100 medium before the addition of 1.5 ml of serum-free TC100 medium.

The DNA-lipofectAMINE™ complexes were then added dropwise to the medium covering the cells and the flasks were incubated at 28°C overnight. After this time, 1.5 ml of complete TC100 growth medium was added to each flask which were then incubated at 28°C for 48 hours. The co-transfection supernatants were collected by centrifugation at 1,000 x *g* for 5 minutes in order to remove cellular debris. Since cytopathic effects due to virus infection were not visible at this stage, the co-transfection supernatants were used to re-infect *Sf*9 cells in order to both amplify any virus present and, assess visually for signs of infection.

The virus amplification was performed as described in Section 2.4.2, and it transpired that signs of infection were only apparent in flasks to which co-transfection supernatants from cells transfected with pAcYM1.GT and pAcYM1.KK DNA had been added. Such cytopathic effects consisted of enlargment of the cell nuclei resulting in an irregular cell morphology. Plaque assays for each transfection were performed (Section 2.4.1) which confirmed that recombinant baculovirus particles bearing wild-type GLUT1 and the double-lysine mutant of GLUT1 had been generated. Although the BaculoGold™ system achieves recombination efficiencies of virtually 100%, two plaques from each successful co-transfection were purified in order to ensure that the recombinant viruses were derived from a single clone. This was performed by picking plaques using a sterile micropipette tip and adding them to 0.5 ml TC100 complete growth medium in a sterile microcentrifuge tube. The tubes were then rotated overnight at 4°C in order to elute the virus from the agarose. The subsequent low titre recombinant virus stocks were then amplified several times (Section 2.4.1) until a high virus titre ($\approx 3 \times 10^8$ pfu/ml) was obtained, as determined by plaque assay (Section 2.4.2).

Upon re-infection of *Sf*9 cells with high virus titre stocks, within 2 days cytopathic effects were evident in cells infected with putative recombinant virus particles possessing wild-type GLUT1 and the double-lysine mutant. Furthermore, no occlusion bodies could be detected indicating that the infecting

viruses were indeed recombinant species. Unfortunately, the presence of recombinant baculovirus particles possessing either of the triple mutants could not be detected. One of the high virus titre stocks from the transfection of the wild-type GLUT1 transfer vector was retained and designated AcNPV.GT, whereas one of the high virus titre stocks resulting from transfection of pAcYM1.KK was retained and designated AcNPV.KK.

6.4 Production of GLUT1 via infection of *Sf*9 cells with wild-type and recombinant baculovirus

6.4.1 Immunoblot analysis of wild-type and double-lysine mutant GLUT1 over-expression in insect cells

Having generated recombinant baculovirus particles possessing wild-type GLUT1 and the double-lysine mutant of GLUT1, it was then necessary to determine if the transporter could be over-expressed in insect cells. For this purpose, Western blot analysis of insect cell membranes was performed. The *Sf*9 cells were counted using a haemocytometer (BDH) and seeded into flasks or dishes at the appropriate density (e.g. 6-8 X 10^6 cells/75 cm^2). The cells were then allowed to attach by leaving the dishes for 1 hour in a laminar flow cabinet. Following attachment, the medium was removed and the appropriate amount of virus was added to the cells to achieve a MOI of 5. Four flasks of cells were prepared and infected with no virus, wild-type AcNPV, recombinant AcNPV.GT or recombinant AcNPV.KK. After incubating for 1 hour at 28°C or room temperature, the inoculum was removed. Fresh complete medium (e.g. 10 ml for a 75 cm^2 flask) was then added to the cells, followed by incubation at 28°C for 2 days, reported to represent maximum expression levels (Yi *et al.*, 1992, Cope *et al.*, 1994). After 2 days, signs of infection were apparent in all infected flasks, the marked difference between the cells infected with wild-type AcNPV and those infected with recombinant virus particles being the appearance of occlusion bodies in the former. Following incubation, the culture medium was collected and centrifuged to remove residual cells at 1,000 x *g* for

10 minutes. The supernatant, i.e. the extracellular virus, was then stored at 4°C. The cells were then solubilised in 1 ml of detergent buffer (2% octaethylene glycol monododecyl ether ($C_{12}E_8$) in 5 mM sodium phosphate buffer, pH 7.4, 1 mM EDTA) containing proteinase inhibitors aprotinin, leupeptin and pepstatin A at 1 µg/ml. Following the removal of unsolubilised material by centrifugation, the resulting supernatant was assayed for protein content (Section 2.2.2), electrophoresed on a 12% SDS/polyacrylamide gel (Section 2.2.4) and transferred to nitrocellulose (Section 2.2.5.1). Immunoreactivity towards an antibody raised against the *C*-terminus was determined using enhanced chemiluminescence (Section 2.2.5.2).

Figure 6.6 demonstrates the successful over-expression of wild-type GLUT1 and the double-lysine mutant, as bands corresponding to the expected apparent M_r 43,000 to 50,000 are evident. Although the precise levels of over-expression were not quantitated in this study, the relative abundances of the wild-type and mutated GLUT1 do appear to be comparable as equivalent amounts of solubilised protein were loaded into each lane (10 µg). In contrast, immunostaining of the lane corresponding to insect cells infected with wild-type AcNPV is not apparent in this blot.

6.5 Purification of polyclonal antibodies to the peptide corresponding to residues 460-477 of human GLUT1

Having generated recombinant baculovirus particles (AcNPV.GT and AcNPV.KK) possessing wild-type and the double lysine mutant of GLUT1, respectively, it was then possible to proceed with the purification of the antibody to be used in obtaining fragments of GLUT1 labelled with site-specific inhibitors of transport. This was performed using the protocol described in Section 2.2.6. The purified antibody was then analysed by ELISA (Section 2.2.7) to determine if it was capable of recognising the original peptide to which it was raised and, most importantly, the native glucose transporter.

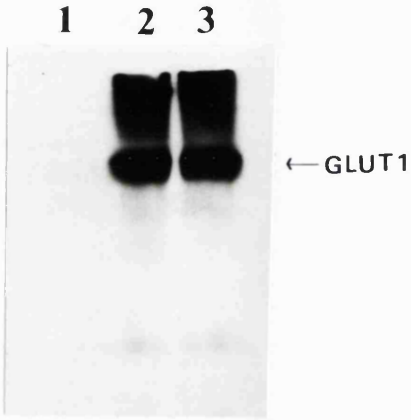


Figure 6.6 Western blot demonstrating expression of wild-type GLUT1 and the double-lysine mutant of GLUT1 in *Sf*9 cells.
Insect cells not infected with baculovirus (Lane 1), infected with AcNPV.GT (Lane 2) and AcNPV.KK (Lane 3) were solubilised and run on a 12% SDS/polyacylamide gel, transferred to nitrocellulose which was then probed with an antibody directed against residues 477-492 of GLUT1.

As is evident from the results of the ELISA (Figure 6.7), the affinity purified antibody is capable of recognising both the original synthetic peptide and the purified glucose transporter, with maximum effectiveness being retained at a dilution of 1:3,000. Further, half maximal recognition of the synthetic peptide has not been achieved even at a dilution of 1:12,800, indicating that the antibody should prove to be adequate for immunoprecipitation purposes.

The ability of the antibody to recognise the transporter on a Western blot was also analysed. Insect cells infected with no virus, AcNPV.GT or AcNPV.KK were solubilised as described in Section 6.4.1, run on a 12% SDS/polyacrylamide gel, transferred to nitrocellulose and probed with the affinity-purified antibody in the manner detailed in Section 2.2.5.2. Figure 6.8 demonstrates quite clearly that the antibody is recognising both the wild-type and the double-lysine mutant of GLUT1.

6.6 Photolabelling of wild-type and the double-lysine mutant of GLUT1 expressed in insect cells with ATB-BMPA

In order to demonstrate that the expressed wild-type and double-lysine mutant GLUT1 proteins expressed in insect cells possessed the ability to bind the ATB-BMPA, it was necessary to photolabel insect cells expressing these proteins with this exofacial ligand. Insect cells, infected with wild-type AcNPV, AcNPV.GT or AcPV.KK at a MOI of 5, were incubated for 2 days p.i. and then washed twice in PBS before incubation with 100 µCi of ATB-BMPA (Figure 6.9) at room temperature. The cells were then irradiated for 10 minutes with a 100W UV lamp (Model R-52, Ultraviolet Products Inc., San Gabriel, C.A., U.S.A.) at a distance of 10 cm. Following irradiation, the cells were solubilised as described in Section 6.4.1, assayed for protein content (Section 2.2.2), and electrophoresed on a 12% SDS/polyacrylamide gel (Section 2.2.4). Relevant tracks were cut out to a width of 1 cm using a long blade and then cut into 4 mm slices with a gel slicer. Four slices were also taken from a non-radioactive part of the gel to determine backgrounds. The slices were placed in scintillation

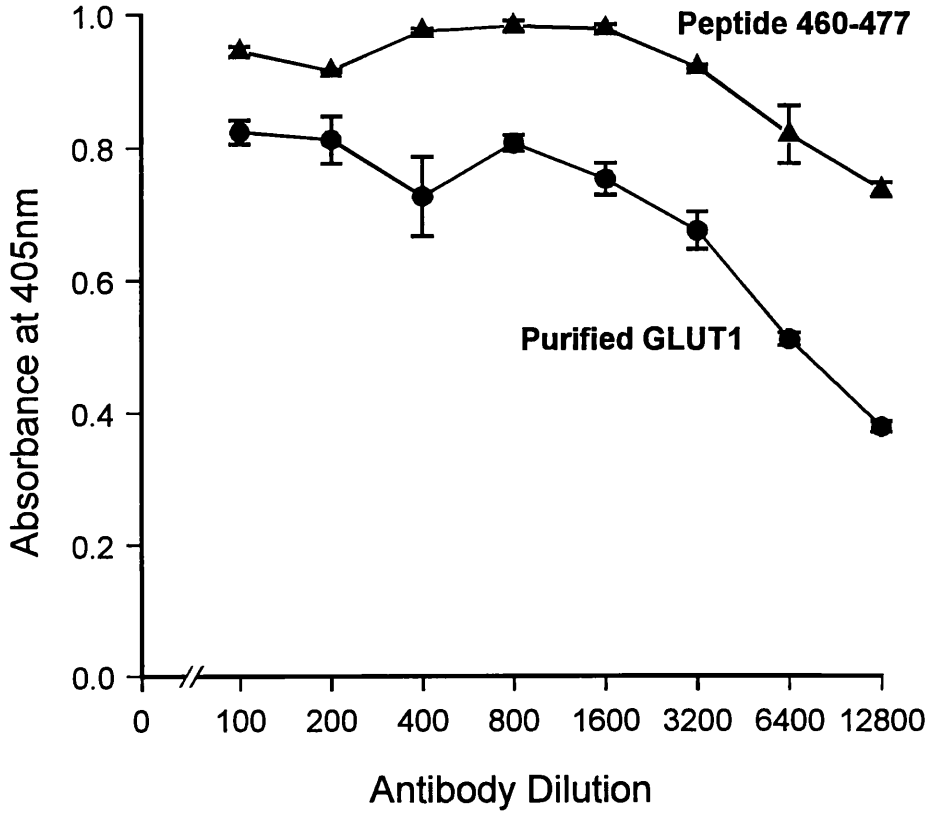


Figure 6.7 Analysis of purified antibody by ELISA.
Each point is the mean of triplicate samples ± SEM.

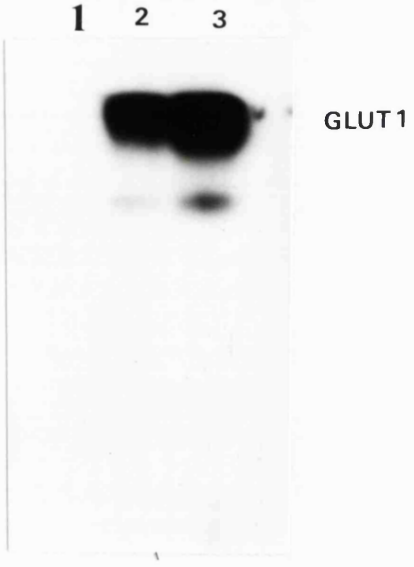


Figure 6.8 Western blot analysis with purified antibody.
Insect cells were infected with wild-type AcNPV (Lane 1), wild-type GLUT1 (Lane 2) or the double-lysine mutant of GLUT1 (Lane 3).

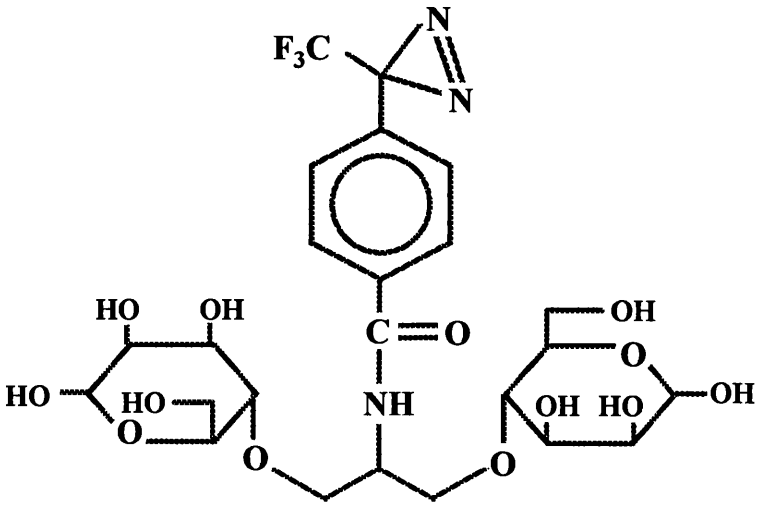


Figure 6.9 Structure of ATB-BMPA.

vials and solubilised by incubation with 1 ml 50% (v/v) Solvable (Du Pont GMBH, Hamburg, Germany) for 3 hours at 50°C. Finally, 4 ml of scintillation fluid was added to each vial and the radioactivity was counted using a Packard 1900TR liquid scintillation analyser.

Figure 6.10 demonstrates clearly the successful expression of the double-lysine mutant of GLUT1 in a form able to bind the exofacial transport inhibitor, ATB-BMPA. No photolabelling of cells infected with wild-type AcNPV could be detected. In addition, the competitive inhibition of photolabelling by the presence of increasing concentrations of D-glucose shows that the mutant form of the transporter is also able to bind the transported substrate. Thus, substrate-binding properties of the expressed transporter do not appear to have been modified by the exchange of lysine for arginine residues at positions 451 and 456. Although not shown, the extent of labelling and the electrophoretic profile of insect cells infected with AcNPV.GT, and then photolabelled with ATB-BMPA, is very similar to that of Figure 6.10, suggesting that the double-lysine mutation has not affected the ability of the expressed protein to adopt the outward-facing conformation.

6.7 Immunopurification of photolabelled GLUT1 expressed in insect cells

Insect cells were infected with AcNPV.KK at a MOI of 5 and incubated at 28°C for 2 days. The cells were then harvested by centrifugation at 1,000 x *g* for 10 minutes. Following two washes in PBS, the cells were photolabelled with ATB-BMPA as described in Section 6.6. A total insect membrane preparation was obtained by lysing the cells using a modification of a nitrogen cavitation procedure (Cezanne *et al.*, 1992). The cells were resuspended in lysis buffer (40 mM Tris-HCl, 90 mM KCl, 2 mM $MgCl_2$, 2 mM ATP, 1.5 mM EGTA, 1 mM PMSF, pH 5.4) at a density of 3 x 10^6 cells/ml, and equilibrated at 4°C with nitrogen in a Parr bomb (Model 4635) at 800 pounds/sq. inch for 10 minutes. The cell suspension was then released dropwise from the bomb. The resulting cell lysate was fractionated by differential centrifugation, initially at 500 x *g* for

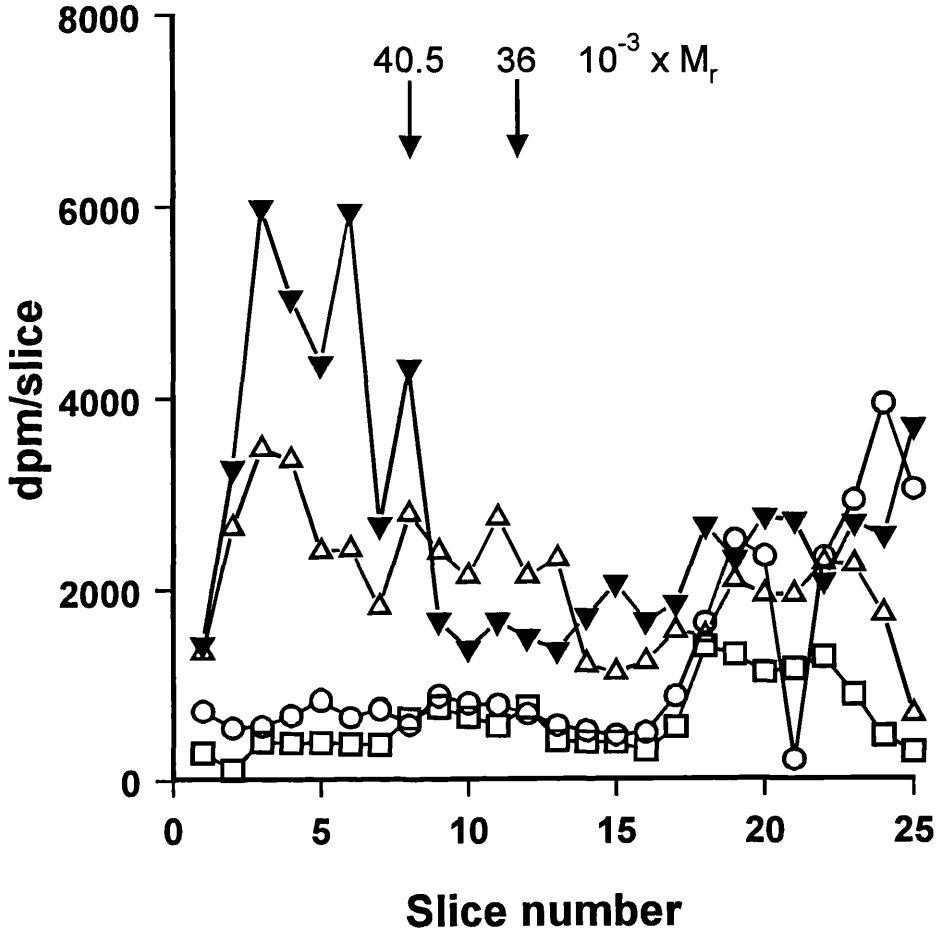


Figure 6.10 D-glucose inhibition of ATB-BMPA photolabelling.
Electrophoretic profile of ATB-BMPA-photolabelled insect cells infected with AcNPV.KK (▾), AcNPV.KK + 5 mM D-glucose (△), AcNPV.KK + 150 mM D-glucose (□), AcNPV (○).

5 minutes to remove intact cells and nuclei, followed by centrifugation at 100,000 x *g* for 45 minutes at 4°C in order to recover a total membrane preparation. Protein concentration was determined as described in Section 2.2.2.

Membrane samples (100 μg) were digested with endoproteinase Lys-C (Promega, 1:50, w/w) in 0.1% SDS at 25°C for 20 hours. The digest, along with an undigested control, was solubilised and immunoprecipitated with an immobilised antibody raised against residues 460-477 of human GLUT1 as described in Sections 2.2.8 and 2.2.9. The samples were then electrophoresed on a 12% SDS/polyacrylamide gel, the relevant lanes were sliced and the radioactivity of each slice counted as described in Section 6.6.

The electrophoretic profiles of the photolabelled, double-lysine mutant of GLUT1 before and after endoproteinase Lys-C digestion are shown in Figure 6.11. The disappearance of the peak of radioactivity corresponding to immunopurified, undigested GLUT1 at an apparent M_r 50,000 in the digested sample, indicates that complete proteolysis has been achieved. Furthermore, the emergence of a peak at an apparent M_r 19,000 provides definitive evidence that the site of photolabelling with ATB-BMPA occurs between residues 301 and 477, within the *C*-terminal half of GLUT1. It is now necessary to confirm the identity of this radioactive band, and the precise location of photolabelling, by *N*-terminal sequencing. Having established that the mutation of lysines 451 and 456 to arginine results in a fully functional form of GLUT1, which retains the ability to be photolabelled with ATB-BMPA, it is feasible to proceed and generate recombinant baculovirus particles possessing the triple mutants of GLUT1, and thus attempt to identify the sites of labelling of cytochalasin B and forskolin in a similar manner.

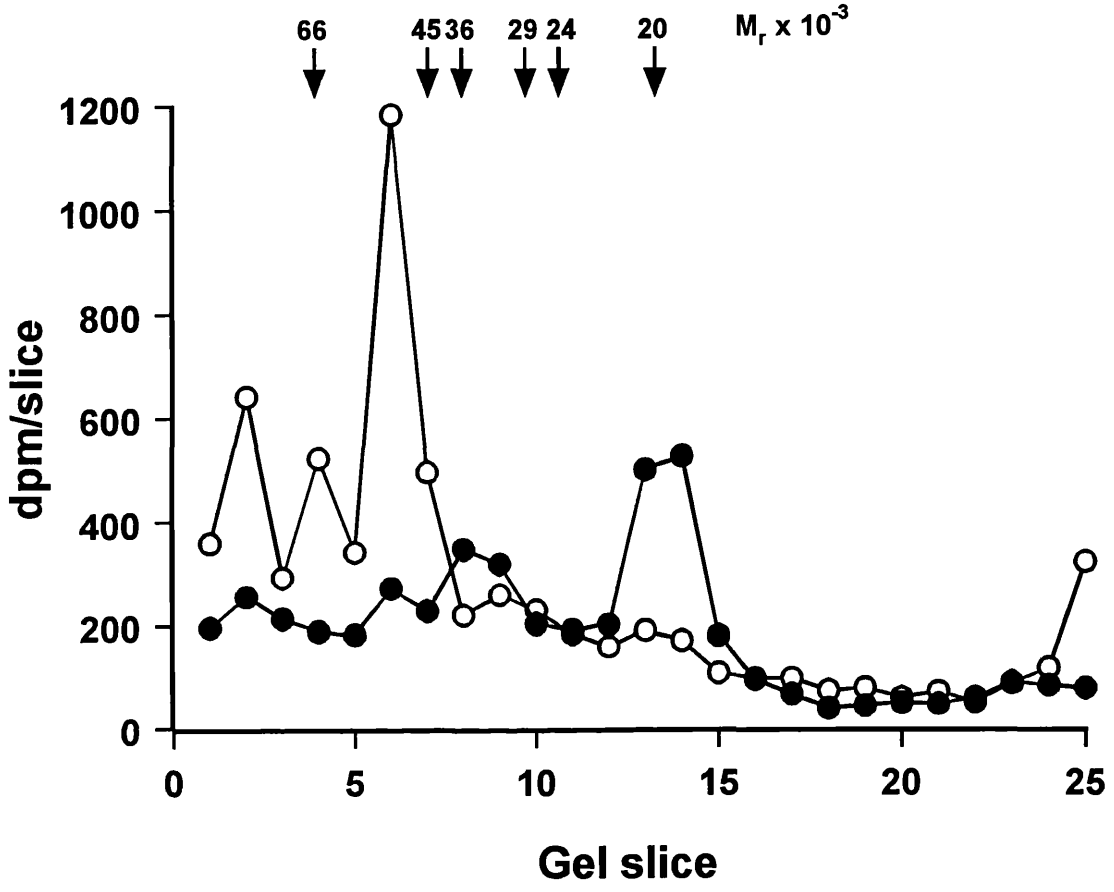


Figure 6.11 Electrophoretic profile of ATB-BMPA-photolabelled insect cell membranes.

Control membranes (○) and membranes digested with endoproteinase Lys-C in 0.1% SDS (●).

6.8 Discussion

The large size of the AcNPV genome makes it impossible to construct a recombinant baculovirus by direct subcloning methods. Therefore, in order to introduce the coding sequence of GLUT1 cDNA into the genome of wild-type AcNPV, recombinant transfer vectors, pAcYM1.GT and pAcYM1.KK, were first constructed (Figure 6.3). The transfer vectors contained the entire coding region of the wild-type or double-lysine mutant GLUT1 cDNA under the control of the polyhedrin promoter. Furthermore, by subcloning *Bst* YI GLUT1 cDNA fragments, the lengths of the 5' and 3' untranslated regions were kept to a minimum, 15 and 47 nucleotides, respectively, thereby ensuring that maximum expression would be achieved (Matsuura *et al.*, 1987).

Recombinant baculoviruses were generated by introducing DNA from the recombinant transfer vector pAcYM1.GT or pAcYM1.KK into *Sf*9 cells, together with the linearised wild-type AcNPV viral DNA, via polycationic liposome mediated transfection. The AcNPV sequences present in the transfer vector underwent homologous recombination with the wild-type AcNPV genomic DNA and gave rise to recombinant AcNPV.GT and AcNPV.KK baculoviruses. Recombinant baculoviruses were subsequently plaque-purifed, amplified and then used to generate large amounts of recombinant GLUT1 protein.

The study of Yi *et al.* (1992) demonstrated that GLUT1 could be expressed in insect cells to a very high level, although 75% of the immunoreactive protein was shown to be inactive, as determined by cytochalasin B binding assays. However, since the identification of inhibitor binding sites is dependent upon functional protein, the presence of inactive GLUT1 confusing such studies was overcome through the expression of a mutant of GLUT1 which could be immunopurified. The double-lysine mutant of GLUT1 was previously shown to be functionally active in *Xenopus* oocytes (Chapter 5), hence, it was expected that up to 2% of the total membrane protein from insect cells infected with the double-lysine mutant of GLUT1 would comprise functionally active GLUT1.

Indeed, photolabelling of this mutant with ATB-BMPA indicated that significant amounts of correctly-folded protein had been expressed.

A fundamental requirement of this strategy was the ability of an antibody, raised against residues 460-477 of human GLUT1, to recognise the expressed mutant GLUT1 protein. Antibodies to this epitope were affinity-purified and shown to recognise both the original peptide and the native glucose transporter by ELISA. In addition, the immunoreactivity of expressed GLUT1 (wild-type and double-lysine mutant) was demonstrated by Western blotting. These affinity-purified antibodies were cross-linked to protein A-Sepharose CL-4B and used to immunopurify an ATB-BMPA-labelled fragment of the transporter released by digestion with endoproteinase Lys-C. Polyacrylamide gel electrophoresis revealed that this radiolabelled fragment possessed an apparent M_r of 19,000, corresponding to residues Ala_{301}-Arg_{477}.

In summary, the mutagenesis strategy for the identification of sites of labelling of transport inhibitors described in Chapter 5 has been verified. Definitive evidence for the labelling of GLUT1 between residues 301 and 477 by ATB-BMPA has been provided, and precise identification of the site of labelling by *N*-terminal sequencing is now possible. Furthermore, it should be possible for the triple mutants of GLUT1 to be expressed in insect cells and, hence, determine the sites of labelling of cytochalasin B and forskolin using this methodology.

CHAPTER 7. CHARACTERISATION OF GLUT1 EXPRESSED IN CHINESE HAMSTER OVARY (CHO) CELLS

7.1 Introduction

Although high level expression of GLUT1 has been achieved in insect cells, they are unsuitable for an examination of topography using amino group-specific exofacial probes for several reasons. Firstly, since approximately 75% of the GLUT1 protein expressed in insect cells is non-functional (Yi *et al.*, 1992), it is likely that this protein possesses an altered topography. However, this does not cause a problem for the studies described in Chapter 6, as only the transporters capable of adopting the outward-facing conformation were photolabelled with ATB-BMPA and immunopurified. A second reason for the unsuitability is that much of the protein is intracellular (Yi *et al.*, 1992) and, therefore, inaccessible to probes. Finally, upon infection with baculovirus, the insect cells become leaky which would thus undermine the use of membrane-impermeant reagents as probes of topology. Consequently, an expression system involving CHO cells was investigated in an attempt to overcome these difficulties.

CHO cells are used widely for recombinant gene expression partly because they can be readily transfected and grow well both in suspension culture and attached to plastic. Perhaps the most important reason for their use is the availability of well-established efficient expression systems based on the use of vector amplification. It is these factors that have led to the involvement of expression systems involving CHO cells in the analysis of mutants of GLUT1.

CHO cells have been used previously to examine the properties of mutants of GLUT1 (Katagiri *et al.*, 1991, Asano *et al.*, 1992). However, the levels of expression achieved have only been in the order of several-fold greater than the endogenous activity, that is, much less than in the erythrocyte. It was felt that higher expression levels would be needed to perform topological

experiments upon the mutants of GLUT1 of the type desribed for the native transporter in Chapter 4. It was decided, therefore, to assess a eukaryotic expression system that had been developed at Celltech, Ltd. (Slough, UK). The vector contains an amplifiable selection marker, glutamine synthetase, that allows for a high copy number of expression vectors to be incorporated into the genome of the mammalian cell (Bebbington, 1991).

An amplifiable system that is commonly used for the expression of eukaryotic proteins requires the presence of methotrexate to select for dihydrofolate reductase genes in cell lines lacking endogenous enzyme. This methodology, however, is very time-consuming since it requires multiple rounds of amplification and selection in order to reach maximal expression. The principal advantage of the expression system investigated in this study is the facility to produce glutamine synthetase. Expression of this enzyme allows selection and amplification to be achieved under stringent conditions in the presence of methionine sulphoximine (MSX), which inhibits vector encoded and endogenous glutamine synthetase. As a consequence, very high levels of expression are obtainable upon initial selection of clones and amplification of expression is usually achieved after one round.

This system has been applied successfully to the expression of secreted proteins (Davis *et al.*, 1990) and a membrane protein, TSHr, the human thyrotropin receptor (Harfst and Johnstone, 1992). The successful high level expression of TSHr was achieved in CHO cells at levels at least 10-fold greater than has been achieved in any other system, including the baculovirus expression vector system. Consequently, the potential of this sytem for the high level expression of GLUT1, with a view to performing topographical analyses of the GLUT1 mutants described in Chapter 5, was of great interest. This chapter, therefore, describes an analysis of a CHO cell line that was stably transfected with wild-type GLUT1 cDNA using the glutamine synthetase/MSX vector system, as a prelude to the transfection of GLUT1 mutants.

7.2 Generation of CHO cell line over-expressing GLUT1

This section of the work was performed in collaboration with Dr. A. Johnstone (St. George's Hospital Medical School, London (SGHMS)) who was provided with a 2.5kb fragment of DNA containing the entire coding region of GLUT1. The source of this fragment was the plasmid pSGT (Mueckler and Lodish, 1986), a gift from Dr. M. Mueckler, Washington University, USA. The 2.5kb fragment was excised from pSGT by restriction digestion with *Bam* HI, and the 3' recessed ends were blunted prior to ligation into the *Sma* I restriction site of the pEE14 glutamine synthetase vector (Figure 7.1). Transfection of the recombinant vector into CHO-K1 cells had to be performed at SGHMS under the terms of an agreement with Celltech.

The pEE14 vector containing the GLUT1 cDNA was introduced into CHO-K1 cells by calcium phosphate-DNA co-precipitation. The cells were transferred to the selection medium (GMEM-S) consisting of glutamine-free Glasgow Modified Eagle's Medium (GMEM, Gibco-BRL), 10% dialysed foetal calf serum, 1% antibiotics (penicillin 5000 units/ml plus streptomycin 5000 µg/ml, Gibco-BRL) and 25 µM MSX. The GMEM was supplemented with non-essential amino acids, 33 mM $NaHCO_3$, 7 µM L-glutamic acid, 7 µM L-asparagine, 1 mM sodium pyruvate, 28 µM each of adenosine, guanosine, uridine, cytidine and 10 µM thymidine. Although CHO cells possess endogenous glutamine synthetase activity, low levels of MSX (25 µM) will kill non-transfected cells, but allow those cells transfected with the vector to survive. A typical figure for the efficiency of transfection is 1-3 cells per 10^5 plated. MSX-resistant cell lines producing significant amounts of GLUT1 (determined by quantitative RNA blotting experiments) were isolated and incubated with MSX at concentrations from 100 µM to 1 mM for 2 weeks. Distinct colonies from the highest MSX concentration applied were then isolated, and the cells with high expression levels were cloned by limiting dilution. The clone expressing the highest levels of GLUT1 (hereafter designated OE-CHO), plus non-transfected CHO-K1 cells (CHO), were then kindly donated by Dr. A. Johnstone for subsequent analysis.

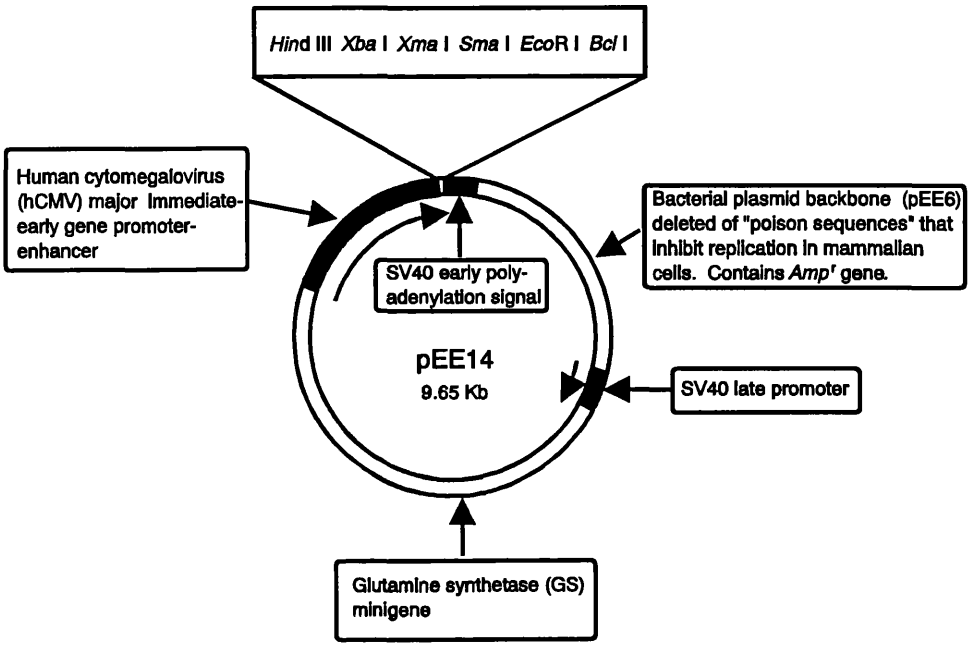


Figure 7.1 Diagram of vector for stable transfection of CHO cells with GLUT1 cDNA.

7.3 CHO cell culture

Wild-type CHO cells were grown in the absence of MSX, whereas MSX was included in the GMEM-S for the OE-CHO cells. CHO cells exhibited a doubling time of approximately 24 hours in GMEM-S. The CHO cells required 5% CO_2 and were passaged every 3/4 days in the following manner. The GMEM-S medium was removed and the confluent monolayers were washed twice with PBS, before incubation with 0.25% trypsin-EDTA for 1 minute at room temperature. The cells were then collected by centrifugation at 1000 x *g* for 5 minutes and resuspended in 5 ml of fresh medium. Non-transfected CHO cells were split 1:10, whereas the OE-CHO cells were split 1:5 and then transferred into new flasks containing a suitable quantity of medium. Although the wild-type CHO cells appeared to possess a doubling time half that of the OE-CHO cells, no morphological distinction between the clones could be discerned by phase-contrast light microscopy (Figures 7.5A and 7.6A).

7.3.1 Long-term storage of CHO cells.

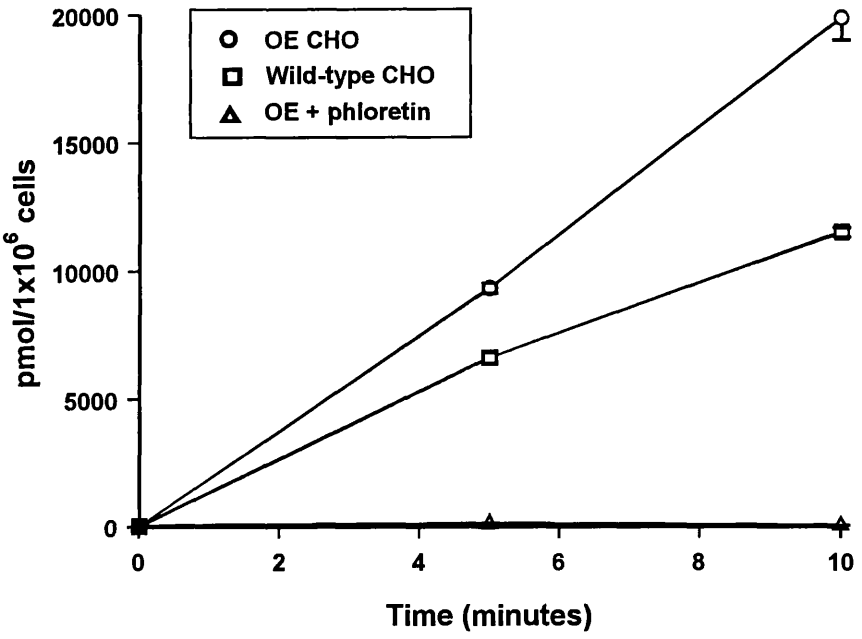
Healthy log-phase cultures were collected after trypsinisation (Section 7.3) by centrifugation at 1000 x *g* for 5 min and resuspended in fresh medium at a density of 4 to 5 x 10^6 cells/ml. The cell suspension was diluted with an equal volume of fresh freezing medium [20% (v/v) DMSO in GMEM-S, 10% FCS] to yield a final DMSO concentration of 10% and maintained on ice. The diluted cell suspension was then dispensed into 1 ml aliquots. The cells were frozen slowly by placing freezing vials in an insulated container at -20°C for 1 hour and then at -70°C overnight. Finally, the cells were stored in liquid nitrogen. Recovery of CHO cells from frozen stocks was achieved by rapid thawing in a 37°C waterbath. The outside of the vial was decontaminated quickly with 70% ethanol and the cells were placed into a 75 cm^2 flask containing 5 ml of fresh complete medium. After overnight incubation at 37°C, the old medium was discarded and replaced with fresh medium.

7.4 2-Deoxy-D-glucose uptake in CHO cells

In order to assess the degree of over-expression of GLUT1 at the cell surface, 2-deoxy-D-glucose uptake studies were performed on confluent and pre-confluent (approximately 80%) monolayers of CHO cells in 35mm dishes. Dishes of control (non-transfected) and transfected CHO cells were transferred to a 37°C waterbath, and the cells were washed three times with 3 ml Krebs Ringer HEPES (KRH; 128 mM NaCl, 4.7 mM KCl, 1.25 mM $CaCl_2$, 1.25 mM $MgSO_4$, 10 mM NaH_2PO_4, 10 mM HEPES, pH 7.4), and maintained at 37°C in 950µl of KRH for 10 minutes. The uptake was begun by the addition of 50 µl of 4 mM [^{3}H]-2-deoxy-D-glucose in KRH to each dish, giving a final 2-deoxy-D-glucose concentration of 0.2 mM. Uptake was terminated after 0, 5 and 10 minutes by rapidly aspirating the medium and washing three times with 2 ml of ice-cold PBS. The cells were then permeabilised by the addition of 500 µl of 1% Triton X-100 (Pierce) to each dish for 5 minutes. Samples (250 µl) were then removed into scintillation vials to which was added 10 ml scintillation fluid. All samples, together with negative (250 µl Triton X-100) and positive controls (5 µl aliquots of 4 mM [^{3}H]-2-deoxy-D-glucose plus 245 µl 1% Triton X-100), were then counted in a Packard 1900TR liquid scintillation analyser. Aliquots of the detergent-dissolved cells were also taken for protein estimation (Section 2.2.2), which indicated that 1 x 10^6 CHO cells yielded approximately 0.5 mg of protein.

The results from the sugar uptake analyses are shown in Figures 7.2A and 7.2B for pre-confluent and confluent dishes of CHO cells, respectively. Almost a two-fold increase in transport in the OE-CHO cell clone was observed over the endogenous transport activity when sugar uptake was assayed at either state of confluence. Further, transport capability was totally eliminated by the presence of 0.1 mM phloretin. However, perhaps the most interesting feature of this data is the huge difference in transport observed between CHO cells at pre-confluence and cells at confluence. Although the increase in transport activity observed in OE-CHO cells over the endogenous level appears to be the

A



B

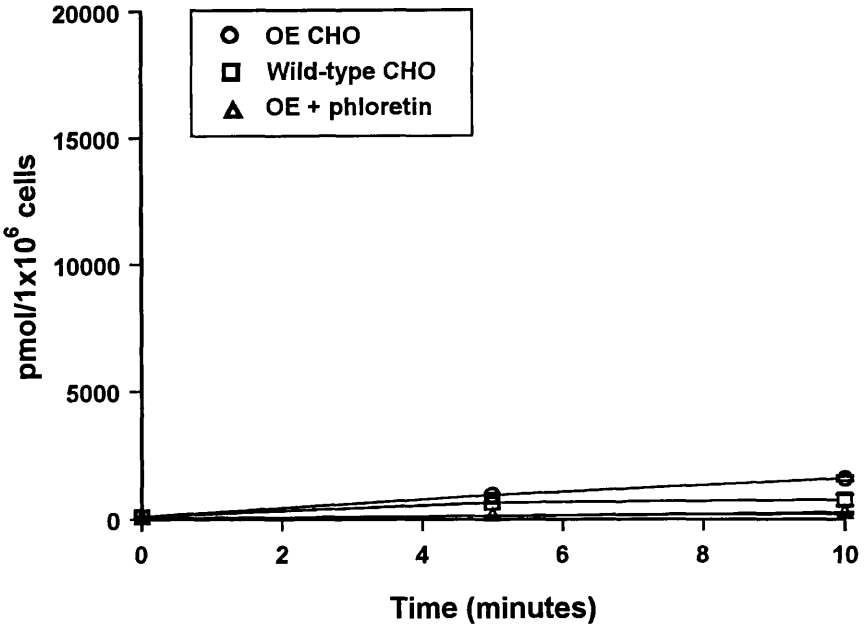


Figure 7.2 Sugar uptake by wild-type and OE-CHO cells.
Uptake studies were performed at pre-confluence (**A**) and confluence (**B**). Datapoints ± SEM (n=3).

same at both levels of confluence, the rate of 2-deoxy-D-glucose uptake measured in cells at pre-confluence was approximately 10 times greater than that determined in cells at confluence. One possibility that could account for this difference is a reduction in the total glucose transporter population upon the cells attaining confluence. Although, a quantification of GLUT1 at different stages from cell seeding to passaging was not performed in this study, the most probable reason for the decrease in sugar uptake at confluence is the internalisation of the transporter, a phenomenon noted by Hresko *et al.* (1994). Clearly, experiments designed to investigate the recycling of the transporter in CHO and OE-CHO cells would be required to confirm this.

From information provided by Dr. A. Johnstone regarding the quantification of GLUT1 mRNA levels in the OE-CHO clone (determined from quantitative RNA blotting studies, data not shown), the degree of over-expression suggested that transport activity might be increased by up to 15-fold. The two-fold increase in transport activity observed in this study was, therefore, particularly disappointing, as other researchers have demonstrated increases in sugar uptake from 3- to 8-fold over endogenous activity upon transfection of GLUT1 cDNA in CHO cells (Katagiri *et al.*, 1991, Asano *et al.*, 1992, Katagiri *et al.*, 1992). Such a low activity may have been due to one or a combination of several factors. For example, the GLUT1 mRNA may not have been translated efficiently leading to low levels of expression, or the protein may have been expressed at a high level although in a denatured or inactive form. Alternatively, inefficient targeting of GLUT1 to the plasma membrane may have been the cause of the poor increase in sugar uptake observed over the endogenous level of transport. As a consequence, several techniques were applied in order to determine which, if any, of these possibilities were responsible for the limited sugar transport capability.

7.5 Quantification of GLUT1 expressed in CHO cells

In order to quantify the amount of GLUT1 expressed in CHO cells, quantitative immuno-blotting was carried out, essentially as described by Madon *et al.* (1990). Total membrane samples from wild-type CHO and OE-CHO cells at pre-confluence were prepared in the following manner. Cells were lysed using a modification of a nitrogen cavitation procedure (Cezanne *et al.*, 1992) that involved resuspending the cells in lysis buffer (40 mM Tris-HCl, 90 mM KCl, 2 mM $MgCl_2$, 2 mM ATP, 1.5 mM EGTA, 1 mM PMSF, pH 5.4) at a density of 3 x 10^6 cells/ml and equilibrating at 4°C with nitrogen in a Parr bomb (Model 4635) at 800 psi for 10 minutes. The cell suspension was then released dropwise from the bomb. The resulting cell lysate was fractionated by differential centrifugation, initially at 500 x *g* for 5 minutes to remove intact cells and nuclei, followed by centrifugation at 100,000 x *g* for 45 minutes at 4°C in order to recover a total membrane preparation.

Membrane samples (4 µg) from wild-type CHO and OE-CHO cells were electrophoresed on a 12% SDS/polyacrylamide gel, transferred electrophoretically to nitrocellulose and immunolabelled with affinity-purified antibodies against residues 477-492 of human GLUT1. From the studies described in Chapter 3, the *C*-terminal amino acid sequences of human, rat, rabbit, pig and mouse GLUT1 are known to be identical. Therefore, although the sequence of the endogenous transporter present in CHO cells is unknown, it was expected that the primary antibody (raised against residues 477-492 of human GLUT1) would detect the endogenous and expressed transporter isoforms with equal efficiency.

The bound primary antibody was detected by incubation with an anti-rabbit IgG conjugated to peroxidase as described in Section 2.2.5.2, and a typical blot is shown in Figure 7.3A. From this figure, it is apparent that the primary antibody is capable of recognising a protein present in CHO cell membranes with electrophoretic properties very similar to human GLUT1, and which, therefore,

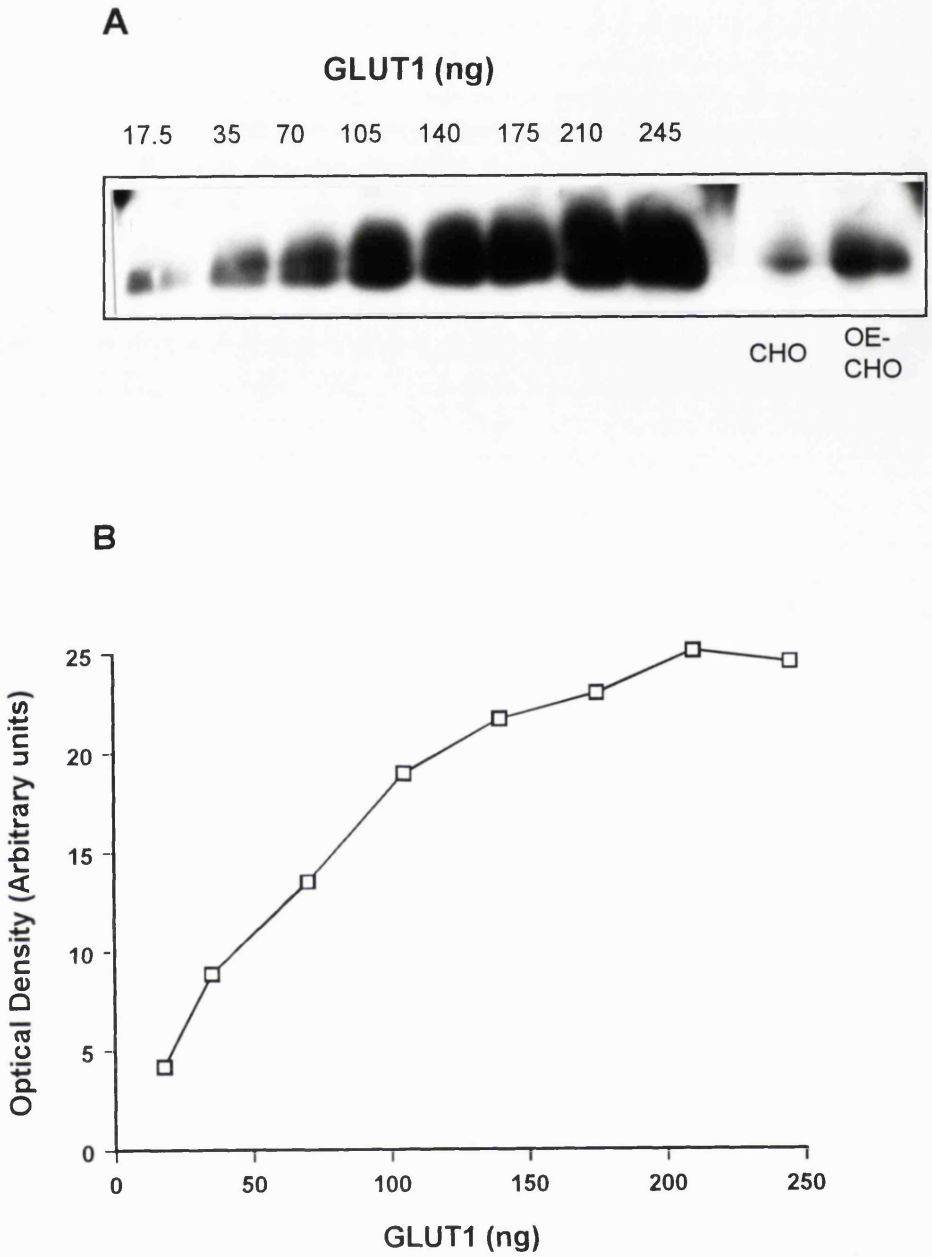


Figure 7.3 Quantification of GLUT1 in CHO and OE-CHO cells. Western blot analysis using known amounts of purified GLUT1 were probed with an affinity-purified antibody to residues 477-492 (**A**) in order to generate a calibration graph (**B**).

is probably the endogenous CHO glucose transporter. Further, there appears also to be a substantial increase in the amount of transporter present in the OE-CHO cells compared to the wild-type CHO cells. To estimate the amount of expressed protein, samples of purified human erythrocyte GLUT1 (17.5-245 ng) were included on the same gel to act as standards. Following the enhanced chemiluminescence procedure, the films were subjected to scanning densitometry. As the signal resulting from the erythrocyte transporter band was proportional to the amount of transporter applied to the gel, a calibration curve could be constructed (Figure 7.3B) and used to determine the amount of GLUT1 present in the CHO samples. From the densitometric analysis, the amount of GLUT1 present within the OE-CHO cells was calculated to represent 678 pmol/mg of membrane protein (based upon the predicted M_r for GLUT1 of 54,117). The amount of endogenous transporter from CHO cells, however, was calculated to be only 46 pmol/mg of membrane protein. Consequently, the OE-CHO cells were shown to be expressing human GLUT1 to a level comprising 3.13% of the total membrane protein, that is, 12.5-fold above the level of the endogenous transporter.

It is also noteworthy that the glycosylation state of GLUT1 expressed in CHO cells (Figure 7.3A) resembles the erythrocyte counterpart more closely than does GLUT1 expressed in insect cells (Figure 6.6). Whether the differences in glycosylation are due to differences between the glycosylation machinery of vertebrates and invertebrates and/or the different levels of over-expresion achieved by the two systems is unclear.

7.6 Photoaffinity labelling of membranes from CHO cells with [^{3}H]-cytochalasin B

Due to the discrepancy between the large increase in GLUT1 content but much smaller increase in transport capacity of the transfected CHO cells, another means of assessing the function of the expressed transporter was sought. The binding of cytochalasin B provides a good means of quantifying the function of

the transport protein, and it was decided to photolabel the transporter with this inhibitor in order to assess whether the low transport activity was due to expression of inactive protein.

The photoaffinity labelling of GLUT1 expressed in CHO cells was performed using [^{3}H]-cytochalasin B, essentially as described previously (Kasanicki *et al.*, 1987). CHO membrane samples at 1 mg protein/ml in 50 mM sodium phosphate, pH 7.4, 100 mM NaCl, 1 mM EDTA and 500 mM D- or L-Glucose, were incubated with 0.51 µM [^{3}H]-cytochalasin B on ice for 30 minutes to allow the attainment of binding equilibrium. Cytochalasin E (10 µM) was also included to inhibit cytochalasin B binding to cytoskeletal elements not associated with glucose transport. The samples were transferred to 1 ml quartz cuvettes, flushed with N_2, stoppered and then irradiated on ice for 10 minutes with a 100W UV lamp (Model R-52, Ultraviolet Products Inc., San Gabriel, C.A., U.S.A.) at a distance of 10 cm. In order to remove non-covalently bound [^{3}H]-cytochalasin B, the irradiated samples were transferred to ultracentrifuge tubes (Beckman) and washed twice with 50 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA pH 7.4, containing 20 µM unlabelled cytochalasin B, by centrifugation at 126,000 x *g* for 10 minutes at 4°, and the supernatants were discarded each time. After washing, samples (100 µg) were electrophoresed on a 1.5 mm thick, 12% SDS/polyacrylamide gel. The gel was fixed, stained with coomassie blue and then destained (Section 2.2.4). Relevant tracks were cut out to a width of 1 cm using a long blade and then cut into 4 mm slices with a gel slicer. Four blank slices were also taken from a non-radioactive part of the gel to determine backgrounds. The slices were then placed in scintillation vials and solubilised by incubation with 1 ml 50% (v/v) Solvable for 3 hours at 50°C. Finally, 4 ml of scintillation fluid were added to each vial and the radioactivity counted using a Packard 1900TR liquid scintillation analyser.

The results of the photolabelling procedure are shown in Figure 7.4. From this figure, it is evident that, there is a substantial increase in the binding of the inhibitor to the membranes prepared from the OE-CHO cell clone, compared

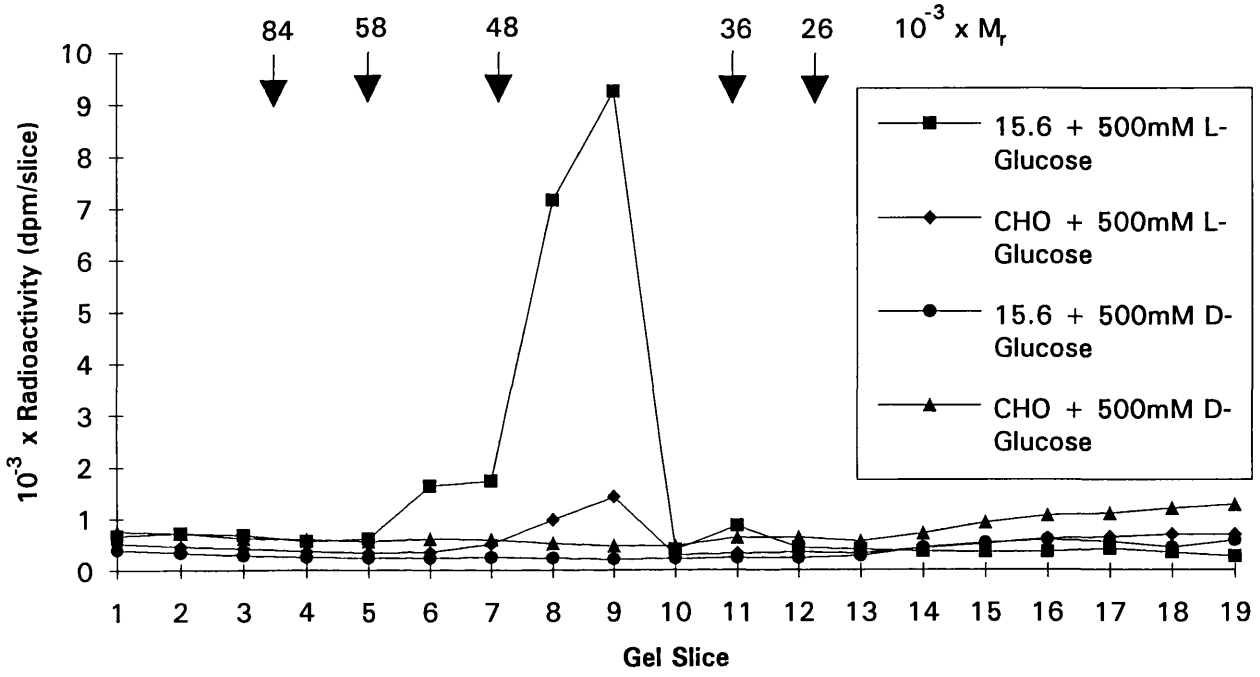


Figure 7.4 Electrophoretic profile of photoaffinity-labelled GLUT1 expressed in CHO cells.

Membranes from wild-type CHO cells and CHO cells overexpressing GLUT1 (OE-CHO) were labelled with [^{3}H]-cytochalasin B, as described in Section 7.6. Samples (100 µg) were electrophoresed on a 12% SDS/polyacrylamide gel and the radioactivity of 4 mm slices was determined by liquid scintillation counting. Arrows indicate the positions of M_r markers.

to the membranes prepared from wild-type CHO cells. In fact, the OE-CHO cell clone appears to possess a 15.2-fold greater amount of GLUT1 over the wild-type CHO cells. In addition, the binding of cytochalasin B to wild-type CHO cell membranes and membranes derived from the OE-CHO cell clone was completely inhibited by 500 mM D-glucose, but not 500 mM L-glucose. This finding indicates that the expressed protein bound not only the inhibitor, but also the transported substrate.

Although there is a slight discrepancy between fold-increase in the amount of immunologically cross-reactive protein present within the OE-CHO cells (Figure 7.3A) and the fold-increase in the amount of transporter that can be photolabelled, it appears likely that all of the over-expressed protein is biologically active. The data from the Western blot analysis (Figure 7.3A), taken together with the photoaffinity labelling data, shown in Figure 7.4, appeared to be inconsistent with the low transport activity demonstrated in Section 7.4. A plausible reason that could have accounted for these observations was that the over-expressed transporter was not being targeted correctly to the plasma membrane. It was decided, therefore, to undertake an immunocytochemical analysis of the clones.

7.7 Subcellular location of GLUT1 in wild-type and over-expressing CHO cell clones

In most mammalian cells that express GLUT1, this isoform is located predominantly at the cell surface, although many cells contain an additional, intracellular population of transporters. To examine the subcellular location of GLUT1 within CHO cells, indirect immunofluorescence studies were carried out. CHO cells were grown on 1 cm diameter, HCl-washed (1 N HCl), circular glass coverslips in multi-well plates, until pre-confluent. The coverslips were sterilised by dipping in 100% ethanol and then flaming. The cells were fixed with 4% paraformaldehyde in PBS for 20 minutes at room temperature and washed rapidly three times with PBS. Excess fixative was quenched by incubating the

cells in PBS containing 100 mM glycine for 15 minutes. The cells were then washed three times for 10 minutes with PBS, and stored in PBS containing 0.02% sodium azide at 4°C. Permeabilisation of the cells was achieved by placing the coverslips in 500 µl 0.1% Triton X-100 in PBS, pH 7.2 in a 24-well plate at 37°C. Following a 15 minute incubation, with gentle shaking, the cells were washed three times for 15 minutes with PBS at 37°C. Non-specific binding of the primary antibody was prevented by a blocking step where the cells were incubated for 1 hour at 37°C with shaking in 500 µl 10% foetal calf serum in PBS containing 0.02% azide. The cells were then washed twice with 0.5 ml PBS before incubation with primary antibody (affinity purified anti-GLUT1 C-terminus) at 10 µg/ml for 1 hour at 37°C. After washing the cells seven times by picking the cover slips up with tweezers and dipping successively into seven beakers of PBS, the secondary antibody, a 1:50 dilution of Pierce goat anti-rabbit IgG conjugated to fluorescein, was applied to the cells for 1 hour at 37°C. The cells were then washed briefly seven times with PBS. Finally, a drop of Citifluor (anti-fade mountant) was placed on the coverslips which were then mounted on slides with DPX mountant prior to fluorescence light microscopy. The results from this series of experiments are shown in Figures 7.5 and 7.6 and illustrate the subcellular location of GLUT1 within the wild-type and OE-CHO cell clones. The specificity of labelling was demonstrated by omitting the primary antibody of duplicate samples of CHO and OE-CHO cells (data not shown). In addition, the fluorescent images were photographed with equivalent exposures to compensate for the automatic metering system of the camera, thus preventing the 'appearance' of a greater amount of endogenous transporter than actually present.

These results are consistent with the findings of the quantitative Western blotting experiments described in Section 7.5, and the photolabelling data shown in Section 7.6. That is, the OE-CHO cell clone expressed the glucose transporter (Figure 7.6B) to a level considerably greater than the endogenous CHO cell transporter (Figure 7.5B). Figure 7.6B illustrates also the presence of substantial amounts of cell-surface fluorescence, indicating that a proportion of

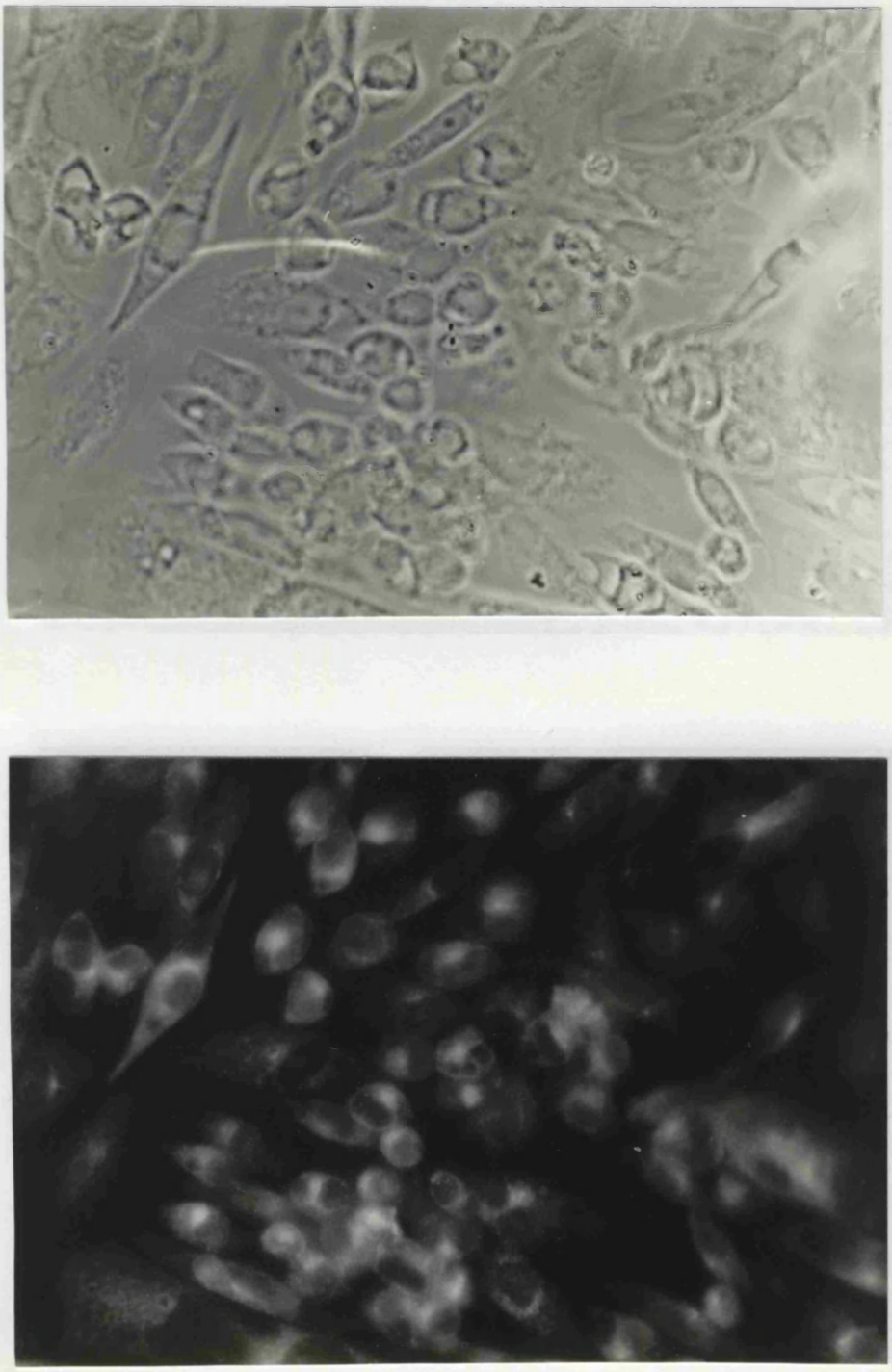


Figure 7.5 Subcellular location of GLUT1 within wild-type CHO cells. CHO cells were subjected to indirect immunofluorescence studies (Section 7.7) and viewed by phase-contrast (**A**) and fluorescence light microscopy (**B**). Magnification X 400.

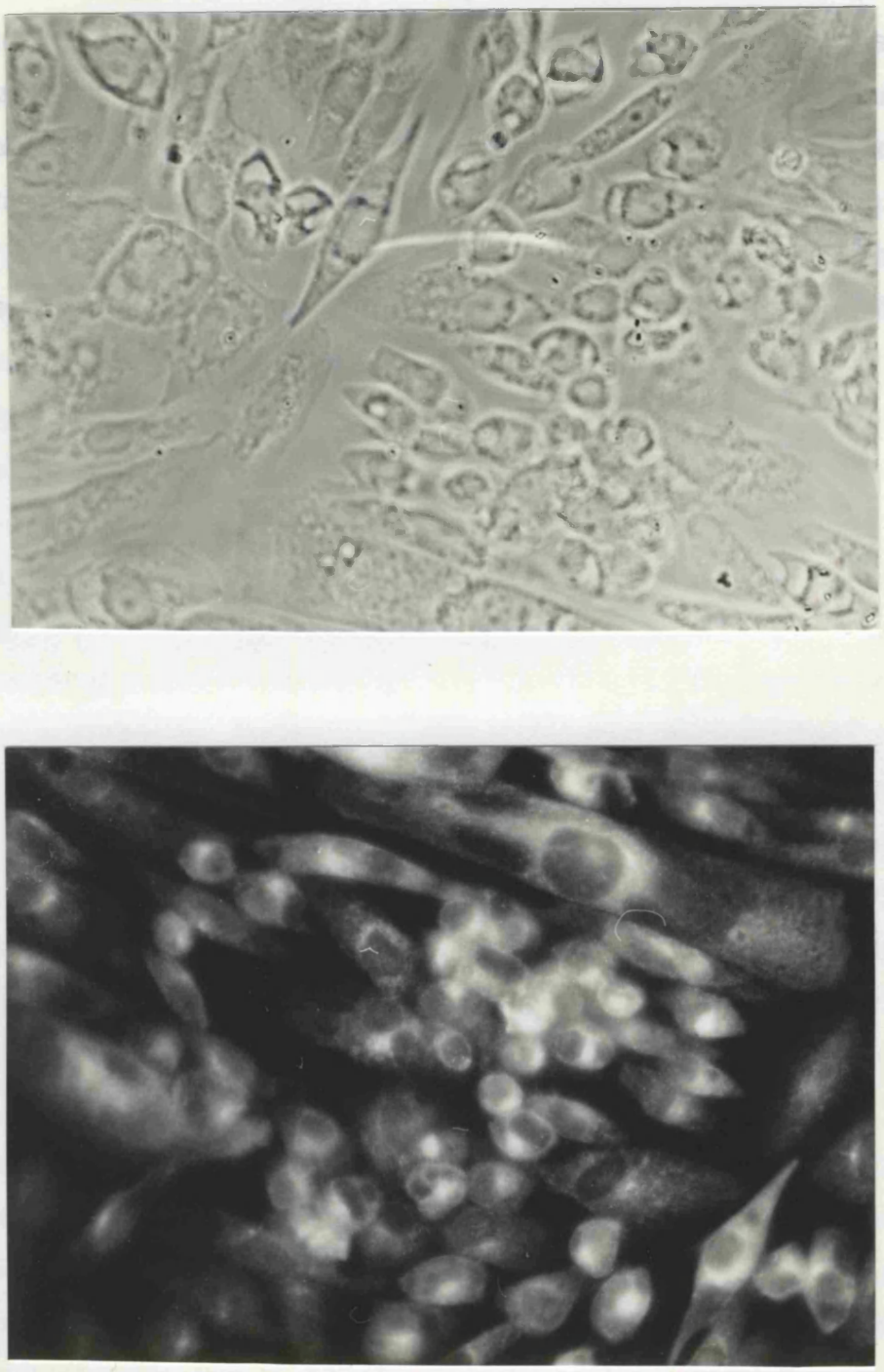


Figure 7.6 Subcellular location of GLUT1 within OE-CHO cells.
OE-CHO cells were subjected to indirect immunofluorescence studies (Section 7.7) and viewed by phase-contrast (**A**) and fluorescence light microscopy (**B**). Magnification X 400.

the expressed GLUT1 is indeed located at the plasma membrane. However, it does appear that the majority of the fluorescence is present throughout the cytoplasm, and appears to be particularly concentrated in the peri-nuclear region of the cells. This is indicative of most of the over-expressed transporter remaining trapped within the cell and not being targeted to the cell surface. These data are, therefore, in accord with the findings of Section 7.4 which describe the low increase in transport capacity observed in the OE-CHO cell clone.

7.8 Discussion

For a foreign gene to be expressed in a functional form, it needs to be transcribed and translated. In many cases, post-translational modifications and compartmentalisation of the nascent polypeptide are also required. A failure to perform correctly any one of these processes can result in a lack of gene expression. The purpose of the work described in this chapter was to assess the functional expression of human GLUT1 in CHO cells produced as a result of transfection of GLUT1 cDNA contained within a high-expression vector (pEE14) developed by Celltech Ltd., Slough.

A CHO cell clone transfected with GLUT1 cDNA was prepared by Dr. A. Johnstone (SGHMS). From quantitative RNA blotting studies, this clone was believed to be capable of over-expressing GLUT1 at levels 15-fold greater than the endogenous transporter. Sugar uptake studies, however, revealed that the OE-CHO cells possessed a transport activity only 2-fold greater than wild-type CHO cells. This relative difference in transport capacity was shown to exist at pre-confluence and confluence, although the transport process was more efficient at pre-confluence. Recycling of transporters was not investigated in this study, but it seems probable that the internalisation of transporters could account for the decreased transport activity observed at confluence.

Photolabelling of membranes derived from OE-CHO cells revealed an abundance of GLUT1 protein compared to CHO cells that was not evident from transport studies. The OE-CHO cells appeared to be over-expressing functionally-active GLUT1 at the expected level of 15.2-fold greater than the endogenous transporter, that is, half the amount of GLUT1 present in the erythrocyte membrane. However, fluorescence studies demonstrated that the over-expressed transporter was not located primarily at the plasma membrane, but was present within subcellular compartments and, hence, not detected by the transport assay employed in this study. Unfortunately, such findings precluded the use of this system in a topographical analysis of the GLUT1

mutants described in Chapter 5. It seems likely that, due to overproduction of the glycoprotein, the efficiency of the translocation process bringing newly synthesised protein to the plasma membrane might have been limiting, thereby causing some protein to be retained in intracellular membranes.

The possibility that alterations in the asparagine-linked glycosylation may change the cell surface localisation, or acquisition of a functional conformation of the glucose transporter, has been suggested previously (Haspel *et al.*, 1988a). Indeed, subsequent site-directed mutagenesis and expression of GLUT1 in CHO cells has added further weight to this claim (Asano *et al.*, 1993). These authors deleted the site of *N*-glycosylation at Asn_{45} and, when compared with CHO cells expressing wild-type GLUT1 which was targeted accurately to the cell surface, expression of the glycosylation-defective GLUT1 protein was not only limited to intracellular compartments, but was shown to possess a much shorter half-life. These results, therefore, are strongly indicative of a role for *N*-glycosylation with respect to intracellular targeting and protein stability. Although the exact glycosylation state of GLUT1 over-expressed in CHO cells was not examined in the present study, in view of the Western blot analysis (Figure 7.3A), it seems unreasonable to attribute the retention of the protein in the cell interior to incorrect glycosylation. Therefore, the reason for the over-expressed GLUT1 not appearing at the cell surface may be solely a function of its expression. That is, the protein may be being produced in such enormous amounts that it is physically impossible for the expressed GLUT1 to be accommodated within the CHO plasma membrane.

In conclusion, the work in the present chapter describes the analysis of a CHO cell clone, stably transfected with GLUT1 cDNA, with the aim of investigating the glutamine synthetase expression vector system as a means of producing substantial quantities of GLUT1 for topographical studies. Although this heterologous expression system has been shown to synthesise large amounts of functionally-active GLUT1, the inability of the CHO cells to incorporate the majority of this protein into the plasma membrane precluded its use for

topographical analysis using exofacial probes. However, the system does exhibit tremendous potential, particularly in view of the higher proportion of functionally-active protein produced compared to the *Sf*9/baculovirus expression system. The apparent 15-fold over-expression of GLUT1 above the endogenous transporter is considerably higher than that achieved by other workers (Katagiri *et al.*, 1991, 1992). Therefore, despite the disappointing levels of cell surface expression, it might be feasible to purify this protein for reconstitution purposes. Such purification could be facilitated by the over-expression of GLUT1 bearing either a poly-His or streptavidin tag. Indeed, it would be interesting to determine the limit of over-expression that is capable with this system. Thus, although not suitable for topological investigations, this CHO cell expression system may, in the future, provide an excellent means of obtaining large amounts of functional human erythrocyte glucose transporter for biochemical and mechanistic studies, as an alternative to the baculovirus/insect cell expression system.

CHAPTER 8. GENERAL DISCUSSION

A complete understanding of the mechanism of glucose transport is dependent upon a detailed knowledge of the structure of GLUT1 at atomic resolution. Unfortunately, since the crystallisation of hydrophobic membrane proteins for X-ray and electron diffraction studies is extremely difficult, this project was aimed at developing a combination of theoretical and experimental techniques to obtain structural and functional information about GLUT1.

Initially, attention was focused upon a detailed analysis of the sugar transporter family at the level of the amino acid sequence. A complete sequence alignment was generated manually for the sugar transporter family that would provide the basis for subsequent computational analyses and rational experimental design. Possibly the most important feature to be realised upon completion of the alignment, was that the putative transmembrane helical segments of each protein sequence could be aligned without gaps, whereas the intervening regions of sequence had numerous insertions and deletions. This, alone, is good evidence for the predicted structural nature of the sugar transporter family; an arrangement of twelve membrane-spanning α-helices. However, it was intended to extract as much information as possible about higher orders of structure from this compilation of primary structures, and several predictive algorithms were utilised to generate data about consensus structural features of the sugar transporter family. As a result, the two-dimensional model of GLUT1 was refined to incorporate these ideas.

It is evident, even from a superficial examination of the alignment, that a periodicity of residues within the putative transmembrane helical regions exists, a typical example being glycine residues in helices 2 and 4. The periodicity of amino acid residues with respect to hydrophobicity, substitution and conservation at each residue position within the alignment was assessed using a suite of programs developed by Dr. D. Donnelly (University of Leeds) to provide a more quantitative appraisal of this phenomenon. On the basis of

these data, possible helical arrangements of the GLUT1 protein were constructed that led to the development of a three-dimensional arrangement of the protein.

Naturally, such models are highly speculative due to the severe lack of experimental data and the intrinsic limitations of predictive schemes. However, as shown in this study, the utility of a model lies in the ability to design experiments to test its features. One characteristic of the model proposed in Chapter 3, is that GLUT1 is comprised of two domains of six helices. Sequence similarities between regions in the *N*- and *C*-terminal halves of the sugar transporter family suggest that they evolved via an internal gene duplication event of an ancestral six-helix protein. Thus, to examine the 'dual-domain assembly' hypothesis of GLUT1, PCR was used to amplify cDNAs corresponding to each of the halves which were then expressed in *Xenopus* oocytes. Stop codons were inserted at positions corresponding to residues 235 or 264 in order to produce the *N*-terminal halves, whereas start codons were inserted at positions corresponding to residues 234 or 263 for the production of the *C*-terminal halves. The rationale for the choice of these sites was determined by the consensus secondary structural analysis of the sugar transporter family, which suggested the presence of short turn or random coil structures at these locations. It was hoped, therefore, that minimal damage to local secondary structural elements would result from these mutations

Xenopus oocytes were injected either separately with mRNAs encoding each of the half molecules or with a combination of the two halves. None of the half molecules on its own induced sugar uptake by the oocytes. However, the rate of 2-deoxy-D-glucose uptake into oocytes injected with a mixture of mRNAs encoding fragments 1-234 and 235-492 was significantly greater than that seen in water-injected oocytes and, in fact, was more than 50% of that seen in oocytes injected with mRNA encoding native GLUT1. Consequently, these results suggest that both halves of the GLUT1 molecule are necessary for its transport activity and, further, that the halves represent separate domains which

are sufficiently stable to associate in the membrane after synthesis to form a functional whole. Thus, it is probable that the presence of the *N*-terminal domain provides a stabilising influence upon the structure of the *C*-terminal domain in the restoration of functional activity. Independent evidence for this conclusion has been provided recently by Cope *et al.* (1994) who demonstrated reconstitution of ligand-binding activity by co-expression of separate halves in insect cells.

It is likely that the degree of contact between the two halves of GLUT1 is quite substantial for ligand-binding and transport activity to be restored. Although the tertiary structure of GLUT1, as modelled by Hodgson *et al.* (1992), illustrates the dual-domain assembly, an intimate association of the two domains is not apparent. A feature of the three-dimensional model of GLUT1 proposed in this study, however, is that helices 1, 6, 7 and 12 do reside in close proximity to each other and have been modelled on a bundle of four helices. Clearly, though, this model is extremely speculative and the precise nature of the contact between the domains needs to be established. Whether this contact is solely inter-helical or in conjunction with the central cytoplasmic loop is unknown, but the capacity of *Xenopus* oocytes to express mutants of membrane proteins could be extended to examine this arrangement by a 'domain-tagging' approach. That is, mutagenesis of residues thought to be on the external surfaces of adjacent helices could be combined with cross-linking experiments in order to gain information about neighbouring helices. It is feasible that such experiments could be used to assess both the sites of contact between the domains as well as the inter-helical relationships within the domains. In addition, the strategy of mutagenesis and co-expression adopted in this study could be extended to determine the requirements of each domain for particular helices. That is, individual helices could be deleted in order to assess, for example, the domain packing arrangements, which helices are essential for stabilising the structure of the *C*-terminal domain for functional activity and, perhaps, those helices involved in the translocation of glucose.

From peptide-mapping studies on photoaffinity-labelled GLUT1 (described in Chapter 5), it has been shown that the *C*-terminal half contains the site-specific binding sites of a number of inhibitors of transport such as ATB-BMPA, cytochalasin B and forskolin. In most cases, the locations of binding have been tentatively suggested by deductions of apparent sizes of labelled transporter fragments separated by SDS-PAGE. Definitive identification of the labelled fragments by their isolation and sequencing was hindered by the difficulty of purifying highly hydrophobic fragments. It was, therefore, necessary to engineer GLUT1 in such a manner that immunoaffinity purification of labelled fragments for sequence analysis could be achieved. This process involved the mutagenesis of lysines 451 and 456 to arginine residues in order to render these sites resistant to proteolysis by endoproteinase Lys-C.

Expression of the double-lysine mutant in *Xenopus* oocytes revealed that it possessed functional activity equivalent to native GLUT1 enabling further characterisation of the construct to be performed using the *Sf*9/baculovirus expression system. This mutant was expressed to a high level in insect cells and was shown to retain the ability to be photolabelled by the membrane-impermeant transport inhibitor, ATB-BMPA. Digestion of this mutant GLUT1 with endoproteinase Lys-C produced a labelled fragment of apparent M_r 19,000 that was successfully immunopurified using antibodies to residues 460-477 immobilised on protein A-Sepharose. The use of the mutagenesis strategy described in this study for the identification of sites of inhibitor binding was thus verified, and it will now be possible to determine the precise site of photolabelling by ATB-BMPA by the *N*-terminal sequencing of this fragment.

A feature of the mutagenesis strategy employed in this study is its progressive nature. That is, having utilised sequence alignments to generate a functionally-active GLUT1 mutant that could facilitate purification of the transporter, additional mutations could then be introduced into the double-lysine construct. Lysine residues were introduced into the exofacial loops connecting helices 9 and 10, and 11 and 12 for the purposes of identifying transport inhibitor binding

sites and gaining topographical information through exofacial labelling schemes. No detectable difference between the former triple mutant and wild-type GLUT1 could be detected upon expression in *Xenopus* oocytes, whereas the latter mutant did appear to be expressed at a lower level in the plasma membrane, resulting in a concomitant decrease in measurable transport activity. Although the Q360K and Q427K triple mutants have yet to be fully exploited through expression in the *Sf*9/baculovirus system, the methodology of progressive mutagenesis applied in this study shows tremendous potential, and should prove useful to the study of any membrane protein.

The mutagenesis strategies described in this study were dependent upon the sequence analysis data described in Chapter 3. Whereas the construction of the GLUT1 halves utilised the secondary structure prediction data, the design of the point mutations for transport inhibitor studies exploited the patterns of conservation within the sequence alignment. From the subsequent assays for functional activity, the initial confidence in the accuracy of the alignment has been justified since the aim to generate point mutants of GLUT1 possessing wild-type activity was achieved. Further, although predictive data requires careful interpretation, a certain degree of faith can be placed in the structure prediction of the central cytoplasmic loop. However, it would be feasible to over-express the central cytoplasmic loop for crystallisation studies in order to determine the 3-dimensional structure of this region.

A requirement of subsequent exofacial labelling schemes using the triple mutants is an expression system involving intact cells. That is, for such an approach to be successful, it is imperative that only those sites on the exofacial surface of the membrane can be labelled. Since the nature of AcNPV infection process causes insect cell membranes to become leaky towards the point of maximal recombinant protein expression, the *Sf*9/baculovirus system is inadequate. An additional feature preventing the use of the *Sf*9/baculovirus system is the apparent different mode of glycosylation, as it is possible that the carbohydrate moiety of the transporter would have an effect upon the

accessibility of an exofacial probe to certain residues. In an attempt to identify an expression system capable of overcoming these difficulties, and also able to synthesise sufficient recombinant protein for topographical analysis, an expression vector system involving CHO cells was investigated.

A CHO cell clone, transfected with human GLUT1 cDNA, was shown to be over-expressing the transporter at a level 15-fold greater than the endogenous transporter. Furthermore, the majority of this protein was functionally-active, as assessed by photolabelling with cytochalasin B. However, immunocytochemical analysis of the clone revealed that most of the transporter remained within intracellular compartments of the cell, accounting for the observed low transport activity. As a consequence, the inadequate targeting of the over-expressed transporter to the plasma membrane precluded the use of this system in the topographical analysis of the point mutations. This finding was quite disappointing since the combination of the pEE14 vector and CHO cells initially held much appeal. Although the overall amount of over-expression obtained by the CHO clone analysed in this study is less than that achieved by insect cells, it is apparent that the proportion of over-expressed GLUT1 which is functionally-active is much greater in CHO cells. Furthermore, the ultimate level of over-expression in CHO cells needs to be established, since it may be that other transfectants would produce GLUT1 to even higher levels. However, it would also be interesting to determine why the majority of the over-expressed protein remains in subcellular compartments. Although the system was shown to be inadequate for the purposes of this study, it is clear that the over-expression of GLUT1 in CHO cells, with an apparently more authentic glycosylation compared to GLUT1 over-expressed in insect cells, will provide an extremely useful alternative for biochemical and mechanistic studies of GLUT1.

In essence, this project has demonstrated the power in the application of several methodologies to a particular problem, such as the use of theoretical studies to direct experimentation. Furthermore, until the development of a system that is capable of expressing membrane proteins to a high level, which

are also correctly targeted and functionally active, the choice of expression system is dependent upon the chosen application. That is, since each system possesses its own limitations, factors such as the limits of over-expression, the proportion of active protein and its cellular distribution need to be considered.

REFERENCES

Adang, M.J. and Miller, L.K. (1982). Molecular cloning of DNA complementary to mRNA of the baculovirus *Autographa californica* nuclear polyhedrosis virus: location and gene products of RNA transcripts found late in infection. J. Virol. *44*, 782-793.

Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., and Rees, D.C. (1988). Structure of the reaction center from *Rhodobacter sphaeroides* R-26: protein-cofactor (quinones and Fe^{2+}) interactions. Proc. Natl. Acad. Sci. USA. *85*, 8487-8491.

Alvarez, J., Lee, D.C., Baldwin, S.A., and Chapman, D. (1987). Fourier transform infrared spectroscopic study of the structure and conformational changes of the human erythrocyte glucose transporter. J. Biol. Chem. *262*, 3502-3509.

Asano, T., Shibasaki, Y., Kasuga, M., Kanazawa, Y., Takaku, F., Akanuma, Y., and Oka, Y. (1988). Cloning of a rabbit brain glucose transporter cDNA and alteration of glucose transporter mRNA during tissue development. Biochem. Biophys. Res. Commun. *154*, 1204-1211.

Asano, T., Shibasaki, Y., Ohno, S., Taira, H., Lin, J.L., Kasuga, M., Kanazawa, Y., Akanuma, Y., Takaku, F., and Oka, Y. (1989). Rabbit brain glucose transporter responds to insulin when expressed in insulin-sensitive Chinese hamster ovary cells. J. Biol. Chem. *264*, 3416-3420.

Asano, T., Katagiri, H., Takata, K., Tsukuda, K., Lin, J.L., Ishihara, H., Inukai, K., Hirano, H., Yazaki, Y., and Oka, Y. (1992). Characterisation of GLUT3 protein expressed in Chinese hamster ovary cells. Biochem. J. *288*, 189-193.

Asano, T., Takata, K., Katagiri, H., Ishihara, H., Inukai, K., Anai, M., Hirano, H., Yazaki, Y., and Oka, Y. (1993). The role of *N*-glycosylation in the targeting and stability of GLUT1 glucose transporter. FEBS Lett. *324*, 258-261.

Aslanidis, C., Schmid, K., and Schmitt, R. (1989). Nucleotide sequences and operon structure of plasmid-borne genes mediating uptake and utilisation of raffinose in *Escherichia coli*. J. Bacteriol. *171*, 6753-6763.

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., and Possee, R.D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virol. *202*, 586-605.

Baldwin, J.M., Lienhard, G.E., and Baldwin, S.A. (1980). The monosaccharide transport system of the human erythrocyte. Orientation upon reconstitution. Biochim. Biophys. Acta *599*, 699-714.

Baldwin, S.A., Baldwin, J.M., and Lienhard, G.E. (1982). Monosaccharide transporter of the human erythrocyte. Characterisation of an improved preparation. Biochemistry *21*, 3836-3842.

Baldwin, S.A. (1992). Mechanisms of active and passive transport in a family of homologous sugar transporters found in both prokaryotes and eukaryotes. In Molecular Aspects of Transport Proteins. J.J.H.M. de Pont and E.M. Wright, eds. (Elsevier Science Publishers, B.V.), pp. 169-217.

Baldwin, S.A. (1993). Mammalian passive glucose transporters: members of an ubiquitous family of active and passive transport proteins. Biochim. Biophys. Acta *1154*, 17-49.

Barnell, W.O., Yi, K.C., and Conway, T. (1990). Sequence and genetic organisation of a *Zymomonas mobilis* gene cluster that encodes several enzymes of glucose metabolism. J. Bacteriol. *172*, 7227-7240.

Barnett, J.E.G., Holman, G.D., and Munday, K.A. (1973). Structural requirements for binding to the sugar-transport system of the human erythrocyte. Biochem. J. *131*, 211-221.

Bebbington, C.R. (1991). Expression of antibody genes in nonlymphoid mammalian cells. Methods *2*, 136-145.

Birnbaum, M.J., Haspel, H.C., and Rosen, O.M. (1986). Cloning and characterisation of a cDNA encoding the rat brain glucose-transporter protein. Proc. Natl. Acad. Sci. USA *83*, 5784-5788.

Birnbaum, M.J. (1989). Identification of a novel gene encoding an insulin-responsive glucose transporter protein. Cell *57*, 305-315.

Bissonnette, L., Champetier, S., Buisson, J.-P., and Roy, P.H. (1991). Characterisation of the nonenzymatic chloramphenicol resistance (cmlA) gene of the In4 integron of Tn1696: similarity of the product to transmembrane transport proteins. J. Bacteriol. *173*, 4493-4502.

Burant, C.F., Takeda, J., Brot-Laroche, E., Bell, G.I., and Davidson, N.O. (1992). Fructose transporter in human spermatozoa and small intestine is GLUT5. J. Biol. Chem. *267*, 14523-14526.

Burant, C.F. and Bell, G.I. (1992a). Mammalian facilitative glucose transporters: evidence for similar substrate recognition sites in functionally monomeric proteins. J. Biol. Chem. *31*, 10414-10420.

Burant, C.F. and Bell, G.I. (1992b). Mammalian facilitative glucose transporters: evidence for similar substrate recognition sites in functionally monomeric proteins. Biochemistry *31*, 10414-10420.

Burchell, A. and Waddell, I.D. (1991). The molecular basis of the hepatic microsomal glucose-6-phosphatase system. Biochim. Biophys. Acta *1092*, 129-137.

Büchel, D.E., Gronenborn, B., and Müller-Hill, B. (1980). Sequence of the lactose permease gene. Nature *283*, 541-545.

Calamia, J. and Manoil, C. (1990). lac permease of *Escherichia coli*: Topology and sequence elements promoting membrane insertion. Proc. Natl. Acad. Sci. USA *87*, 4937-4941.

Cairns, M.T., Elliot, D.A., Scudder, P.R., and Baldwin, S.A. (1984). Proteolytic and chemical dissection of the human erythrocyte glucose transporter. Biochem. J. *221*, 179-188.

Cairns, M.T., Alvarez, J., Panico, M., Gibbs, A.F., Morris, H.R., Chapman, D., and Baldwin, S.A. (1987). Investigation of the structure and function of the human erythrocyte glucose transporter by proteolytic dissection. Biochim. Biophys. Acta *905*, 295-310.

Carruthers, A. (1990). Facilitated diffusion of glucose. Physiol. Rev. *70*, 1135-1176.

Carter-Su, C., Pessin, J.E., Mora, R., Gitomer, W., and Czech, M.P. (1982). Photoaffinity labeling of the human erythrocyte D-glucose transporter. J. Biol. Chem. *257*, 5419-5425.

Celenza, J.L., Marshall Carlson, L., and Carlson, M. (1988). The yeast SNF3 gene encodes a glucose transporter homologous to the mammalian protein. Proc. Natl. Acad. Sci. USA *85*, 2130-2134.

Cezanne, L., Navarro, L., and Tocanne, J.F. (1992). Isolation of the plasma membrane and organelles from Chinese hamster ovary cells. Biochim. Biophys. Acta *1112*, 205-214.

Chang, Y.-D. and Dickson, R.C. (1988). Primary structure of the lactose permease gene from the yeast *Kluyveromyces lactis*. Presence of an unusual transcript structure. J. Biol. Chem. *263*, 16696-16703.

Charron, M.J., Brosius, F.C., Alper, S.L., and Lodish, H.F. (1989). A glucose transport protein expressed predominantly in insulin-responsive tissues. Proc. Natl. Acad. Sci. USA *86*, 2535-2539.

Chin, J.J., Jung, E.K.Y., and Jung, C.Y. (1986). Structural basis of human erythrocyte glucose transporter function in reconstituted vesicles. α-helix orientation. J. Biol. Chem. *261*, 7101-7104.

Chin, J.J., Jung, E.K.Y., Chen, V., and Jung, C.Y. (1987). Structural basis of

human erythrocyte glucose transporter function in proteoliposome vesicles: circular dichroism measurements. Proc. Natl. Acad. Sci. USA *84*, 4113-4116.

Chou, P.Y. and Fasman, G.D. (1974). Prediction of protein conformation. Biochemistry *13*, 222-245.

Clark, J.M. (1988). Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases.. Nuc. Acid. Res. *16*, 9677-9686.

Clark, A.E. and Holman, G.D. (1990). Exofacial photolabelling of the human erythrocyte glucose transporter with an azitrifluoroethylbenzoyl-substituted bismannose. Biochem. J. *269*, 615-622.

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

Cope, D.L., Holman, G.D., Baldwin, S.A., and Wolstenholme, A.J. (1994). Domain assembly of the GLUT1 glucose transporter. Biochem. J. *300*, 291-294.

Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A., and DeLisi, C. (1987). Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. J. Mol. Biol. *195*, 659-685.

Cuppoletti, J., Jung, C.Y., and Green, F.A. (1981). Glucose transport carrier of human erythrocytes. Radiation target size measurement based on flux inactivation. J. Biol. Chem. *256*, 1305-1306.

Cushman, S.W. and Wardzala, L.J. (1980). Potential mechanism of insulin action on glucose transport in the isolated rat adipose cell. Apparent translocation of intracellular transport systems to the plasma membrane. J. Biol. Chem. *255*, 4758-4762.

Davidson, N.O., Hausman, A.M.L., Ifkovits, C.A., Buse, J.B., Gould, G.W., Burant, C.F., and Bell, G.I. (1992). Human intestinal glucose transporter expression and localisation of GLUT5. Am. J. Physiol. *262*, C795-C800.

Davies, A., Meeran, K., Cairns, M.T., and Baldwin, S.A. (1987). Peptide-specific antibodies as probes of the orientation of the glucose transporter in the human erythrocyte membrane. J. Biol. Chem. *262*, 9347-9352.

Davies, A., Ciardelli, T.L., Lienhard, G.E., Boyle, J.M., Whetton, A.D., and Baldwin, S.A. (1990). Site-specific antibodies as probes of the topology and function of the human erythrocyte glucose transporter. Biochem. J. *266*, 799-808.

Davies, A.F. (1991). A biochemical and biophysical investigation into the structure and function of the human erythrocyte glucose transport protein (Ph.D. Thesis, University of London).

Davies, A.F., Cairns, M.T., Davies, A., Preston, R.A.J., Clark, A., Holman, G., and Baldwin, S.A. (1992). Studies of membrane topology and substrate binding-site location in the human erythrocyte glucose transport protein. In Molecular Mechanisms of Transport. E. Quagliariello and F. Palmieri, eds. (Elsevier Science Publishers B.V.), pp. 125-131.

Davis, E.O. and Henderson, P.J.F. (1987). The cloning and DNA sequence of the gene xylE for xylose-proton symport in *Escherichia coli* K12. J. Biol. Chem. *262*, 13928-13932.

Davis, S.J., Ward, H.A., Puklavec, M.J., Willis, A.C., Williams, A.F., and Barclay, A.N. (1990). High level expression in Chinese hamster ovary cells of soluble forms of CD4 T lymphocyte glycoprotein including glycosylation variants. J. Biol. Chem. *265*, 10410-10418.

Degli Esposti, M., Crimi, M., and Venturoli, G. (1990). A critical evaluation of the hydropathy profile of membrane proteins. Eur. J. Biochem. *190*, 207-219.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985). X-ray structure analysis at 3Å resolution of a membrane protein complex: Folding of the protein subunits in the photosynthetic reaction center from *Rhodopseudomonas viridis*. Nature *318*, 618-624.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nuc. Acid. Res. *12*, 387-395.

Devés, R. and Krupka, R.M. (1978). Cytochalasin B and the kinetics of inhibition of biological transport. A case of asymmetric binding to the glucose carrier. Biochim. Biophys. Acta *510*, 339-348.

Deziel, M.R. and Mau, M.M. (1990). Biotin-conjugated reagents as site-specific probes of membrane protein structure - Application to the study of the human erythrocyte hexose transporter. Anal. Biochem. *190*, 297-303.

Deziel, M.R. and Rothstein, A. (1984). Proteolytic cleavages of cytochalasin B binding components of band 4.5 proteins of the human red blood cell membrane. Biochim. Biophys. Acta *776*, 10-20.

Donnelly, D., Johnson, M.S., Blundell, T.L., and Saunders, J. (1989). An analysis of the periodicity of conserved residues in sequence alignments of G-protein coupled receptors. Implications for three-dimensional structure. FEBS Lett. *251*, 109-116.

Donnelly, D., Overington, J.P., Ruffle, S.V., Nugent, J.H., and Blundell, T.L. (1993). Modeling alpha-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. Prot. Sci. *2*, 55-70.

Downer, N.W., Bruchman, T.J., and Hazzard, J.H. (1986). Infrared spectroscopic study of photoreceptor membrane and purple membrane. Protein secondary structure and hydrogen deuterium exchange. J. Biol. Chem. *261*, 3640-3647.

Eckert, B. and Beck, C.F. (1989). Topology of the transposon Tn10-encoded tetracycline resistance protein within the inner membrane of *Escherichia coli.* J. Biol. Chem. *264*, 11663-11670.

Eckert, K.A. and Kunkel, T.A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nuc. Acid. Res. *18*, 3729

Eisenberg, D., Weiss, R.M., and Terwilliger, T.C. (1982). The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature *299*, 371-374.

Elbrink, J. and Bihler, I. (1975). Membrane transport: its relation to cellular metabolic rates. Science *188*, 1177-1184.

Emery, V.C. (1991a). Methods In Molecular Biology. M. Collins, ed. (Clifton, N.J.: The Human Press Inc.), pp. 309-318.

Emery, V.C. (1991b). Baculovirus expression vectors for the production of viral proteins. Rev. Med. Virol. *1*, 11-17.

Engelman, D.M., Steitz, T.A., and Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Ann. Rev. Biophys.Biophys. Chem. *15*, 321-353.

Everitt, B.S. (1980). Cluster Analysis. Anonymous, ed. (London: Heineman Educational Books),

Fasman, G.D. and Gilbert, W.A. (1990). The prediction of transmembrane protein sequences and their conformation: an evaluation. Trends Biochem. Sci. *15*, 89-92.

Fernandez-Moreno, M.A., Caballero, J.L., Hopwood, D.A., and Malpartida, F. (1991). The act cluster containing regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of Streptomyces. Cell *66*, 769-780.

Fischbarg, J., Kuang, K.Y., Vera, J.C., Arant, S., Silverstein, S.C., Loike, J., and Rosen, O.M. (1990). Glucose transporters serve as water channels. Proc. Natl. Acad. Sci. USA *87*, 3244-3247.

Foster, D.L., Boublik, M., and Kaback, H.R. (1983). Structure of the lac carrier protein of Escherichia coli. J. Biol. Chem. *258*, 31-34.

Fukumoto, H., Seino, S., Imura, H., Seino, Y., Eddy, R.L., Fukushima, Y., Byers, M.G., Shows, T.B., and Bell, G.I. (1988). Sequence, tissue distribution, and chromosomal localisation of mRNA encoding a human glucose transporter-like protein. Proc. Natl. Acad. Sci. USA *85*, 5434-5438.

Fukumoto, H., Kayano, T., Buse, J.B., Edwards, Y., Pilch, P.F., Bell, G.I., and Seino, S. (1989). Cloning and characterisation of the major insulin-responsive glucose transporter expressed in human skeletal muscle and other insulin-responsive tissues. J. Biol. Chem. *264*, 7776-7779.

Garcia, J.C., Strube, M., Leingang, K., Keller, K., and Mueckler, M.M. (1992). Amino acid substitutions at tryptophan 388 and tryptophan 412 of the HepG2 (GLUT1) glucose transporter inhibit transport activity and targeting to the plasma membrane in *Xenopus* oocytes. J. Biol. Chem. *267*, 7770-7776.

Garnier, J. and Robson, B. (1989). Prediction of Protein Structure and the Principles of Protein Conformation (New York: Plenum Press).

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

Geering, K., Theulaz, I., Verrey, F., Hauptle, M.T., and Rossier, B.C. (1989). A role for the beta-subunit in the expression of functional Na^+-K^+-ATPase in *Xenopus* oocytes. Am. J. Phys. *257*, C851-C858.

Geever, R.F., Huiet, L., Baum, J.A., Tyler, B.M., Patel, V.B., Rutledge, B.J., Case, M.E., and Giles, N.H. (1989). DNA sequence, organisation and regulation of the qa gene cluster of *Neurospora crassa*. J. Mol. Biol. *207*, 15-34.

Germann, U.A., Willingham, M.C., Pastan, I., and Gottesman, M.M. (1990). Expression of the human multidrug transport in insect cells by a recombinant baculovirus. Biochemistry *29*, 2295-2303.

Goffrini, P., Wesolowski Louvel, M., Ferrero, I., and Fukuhara, H. (1990). RAG1 gene of the yeast *Kluyveromyces lactis* codes for a sugar transporter. Nuc. Acid. Res. *18*, 5294

Gould, G.W., Derechin, V., James, D.E., Tordjman, K., Ahern, S., Gibbs, E.M., Lienhard, G.E., and Mueckler, M. (1989). Insulin-stimulated translocation of the HepG2/erythrocyte-type glucose transporter expressed in 3T3-L1 adipocytes. J. Biol. Chem. *264*, 2180-2184.

Gould, G.W., Thomas, H.M., Jess, T.J., and Bell, G.I. (1991). Expression of human glucose transporters in *Xenopus* oocytes: Kinetic characterisation and substrate specificities of the erythrocyte, liver, and brain isoforms. Biochemistry *30*, 5139-5145.

Gould, G.W. and Bell, G.I. (1990). Facilitative glucose transporters: an expanding family. Trends Biochem. Sci. *15*, 18-23.

Gould, G.W. and Lienhard, G.E. (1989). Expression of a functional glucose transporter in *Xenopus* oocytes. Biochemistry *28*, 9447-9452.

Griffith, J.K., Baker, M.E., Rouch, D.A., Page, M.G.P., Skurray, R.A., Paulsen, I., Chater, K.F., Baldwin, S.A., and Henderson, P.J.F. (1992). Evolution of transmembrane transport. Curr. Opin. Cell Biol. *4*, 684-695.

Gunge, N. (1983). Yeast DNA plasmids. [Review]. Ann. Rev. Microbiol. *37*, 253-276.

Harfst, E. and Johnstone, A.P. (1992). Characterisation of the glutamine synthetase amplifiable eukaryotic expression system applied to an integral membrane protein--the human thyrotropin receptor. Anal. Biochem. *207*, 80-84.

Harris, D.S., Slot, J.W., Geuze, H.J., and James, D.E. (1992). Polarised distribution of glucose transporter isoforms in Caco-2 cells. Proc. Natl. Acad. Sci. U. S. A. *89*, 7556-7560.

Harrison, S.A., Buxton, J.M., Clancy, B.M., and Czech, M.P. (1990). Insulin regulation of hexose transport in mouse 3T3-L1 cells expressing the human HepG2 glucose transporter. J. Biol. Chem. *265*, 20106-20116.

Hashiramoto, M., Kadowaki, T., Clark, A.E., Muraoka, A., Momomura, K., Sakura, H., Tobe, K., Akanuma, Y., Yazaki, Y., Holman, G.D., and Kasuga, M. (1992). Site-directed mutagenesis of GLUT1 in helix 7 residue 282 results in perturbation of exofacial ligand binding. J. Biol. Chem. *267*, 17502-17507.

Haspel, H.C., Revillame, J., and Rosen, O.M. (1988a). Structure, biosynthesis, and function of the hexose transporter in Chinese hamster ovary cells deficient in *N*-acetylglucosaminyltransferase 1 activity. J. Cell Physiol. *136*, 361-366.

Haspel, H.C., Rosenfeld, M.G., and Rosen, O.M. (1988b). Characterisation of antisera to a synthetic carboxyl-terminal peptide of the glucose transporter protein. J. Biol. Chem. *263*, 398-403.

Hawkins, A.R., Lamb, H.K., Smith, M., Keyte, J.W., and Roberts, C.F. (1988). Molecular organisation of the quinic acid utilisation gene cluster in *Aspergillus nidulans.* Mol. Gen. Genet. *214*, 224-231.

Hawley-Nelson, P., Ciccarone, V., Gebeyehu, G., Jessee, J., and Felgner, P.L. (1993). LipofectAMINE reagent: A new, higher efficiency polycationic liposome transfection reagent. FOCUS *15 (3)*, 73-79.

Hebert, D.N. and Carruthers, A. (1991). Cholate-solubilised erythrocyte glucose transporters exist as a mixture of homodimers and homotetramers. Biochemistry *30*, 4654-4658.

Hebert, D.N. and Carruthers, A. (1992). Glucose transporter oligomeric structure determines transporter function. Reversible redox-dependent interconversions of tetrameric and dimeric GLUT1. J. Biol. Chem. *267*, 23829-23838.

Hediger, M.A., Coady, M.J., Ikeda, T.S., and Wright, E.M. (1987). Expression cloning and cDNA sequencing of the Na+/glucose co-transporter. Nature *330*, 379-381.

Henderson, P.J.F., Baldwin, S.A., Cairns, M.T., Charalambous, B.M., Dent, H.C., Gunn, F.J., Liang, W.-.J., Lucas, V.A., Martin, G.E., McDonald, T.P., McKeown, B.J., Muiry, J.A.R., Petro, K.R., Roberts, P.E., Shatwell, K.P., Smith, G., and Tate, C.G. (1992). Sugar-cation symport systems in bacteria. In Molecular Biology of Receptors and Transporters. Bacterial and Glucose Transporters, Volume 1. M. Friedlander and M. Mueckler, eds. (Orlando, Florida: Academic Press, Inc.), pp. In press

Henderson, P.J.F. and Maiden, M.C.J. (1990). Homologous sugar transport proteins in Escherichia coli and their relatives in both prokaryotes and eukaryotes. Phil. Trans. Roy. Soc. London. B. *326*, 391-410.

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

Higuchi, R. (1990). Recombinant PCR. In PCR Protocols. A Guide to Methods and Applications.. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White, eds. (San Diego, California 92101: Academmic Press, Inc.), pp. 177-183.

Hillen, W. and Schollmeier, K. (1983). Nucleotide sequence of the Tn10 encoded tetracycline resistance gene. Nucl. Acid. Res. *11*, 525-539.

Hodgson, P.A., Osguthorpe, D.J., and Holman, G.D. (1992). Molecular modelling of the human erythrocyte glucose transporter. 11th Annual Molecular Graphics Society Meeting.

Holman, G.D., Parkar, B.A., and Midgley, P.J.W. (1986). Exofacial photoaffinity labelling of the human erythrocyte sugar transporter. Biochim. Biophys. Acta *855*, 115-126.

Holman, G.D. and Rees, W.D. (1987). Photolabelling of the hexose transporter at external and internal sites: fragmentation patterns and evidence for a conformational change. Biochim. Biophys. Acta *897*, 395-405.

Hoshino, T., Ikeda, T., Tomizuka, N., and Furakowa, K. (1985). Nucleotide sequence of the tetracycline resistance gene of pTHT15, a thermophilic Bacillus plasmid: comparison with Staphylococcal Tcr controls. Gene *37*, 131-138.

Hresko, R.C., Kruse, M., Strube, M., and Mueckler, M. (1994). Topology of the Glut1 glucose transporter deduced from glycosylation scanning mutagenesis. J. Biol. Chem. *269*, 20482-20488.

Inukai, K., Asano, T., Katagiri, H., Anai, M., Funaki, M., Ishihara, H., Tsukuda, K., Kikuchi, M., Yazaki, Y., and Oka, Y. (1994). Replacement of both tryptophan residues at 388 and 412 completely abolished cytochalasin B photolabelling of the GLUT1 glucose transporter. Biochem. J. *302*, 355-361.

James, D.E., Brown, R., Navarro, J., and Pilch, P.F. (1988). Insulin-regulatable tissues express a unique insulin-sensitive glucose transport protein. Nature *333*, 183-185.

James, D.E., Strube, M., and Mueckler, M. (1989). Molecular cloning and characterisation of an insulin-regulatable glucose transporter. Nature *338*, 83-87.

Jarvis, S.M., Ellory, J.C., and Young, J.D. (1986). Radiation inactivation of the human erythrocyte nucleoside and glucose transporters. Biochim. Biophys. Acta *855*, 312-315.

Jennings, M.L. (1989). Topography of membrane proteins. [Review]. Ann. Rev. Biochem. *58*, 999-1027.

Johnson, J.H., Newgard, C.B., Milburn, J.L., Lodish, H.F., and Thorens, B. (1990). The high Km glucose transporter of islets of Langerhans is functionally similar to the low affinity transporter of liver and has an identical primary sequence. J. Biol. Chem. *265*, 6548-6551.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1994). A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry *33*, 3038-3049.

Joost, H.G., Habberfield, A.D., Simpson, I.A., Laurenza, A., and Seamon, K.B. (1988). Activation of adenylate cyclase and inhibition of glucose transport in rat adipocytes by forskolin analogues: structural determinants for distinct sites of action. Mol. Pharmacol. *33*, 449-453.

Jung, C.Y., Hsu, T.L., Hah, J.S., Cha, C., and Haas, M.N. (1980). Glucose transport carrier of human erythrocytes. Radiation-target size of glucose-sensitive cytochalasin B binding protein. J. Biol. Chem. *255*, 361-364.

Jung, E.K.Y., Chin, J.J., and Jung, C.Y. (1986). Structural basis of human erythrocyte glucose transporter function in reconstituted system. Hydrogen exchange. J. Biol. Chem. *261*, 9155-9160.

Kaback, H.R., Frillingos, S., Jung, H., Jung, K., Prive, G.G., Ujwal, M.J., Weitzman, C., Wu, J., and Zen, K. (1994). The lactose permease meets Frankenstein. J. Exp. Biol. *196*, 183-195.

Kaback, H.R. (1992a). In and out and up and down with lac permease. Int. Rev. Cytol. *137*, 97-125.

Kaback, H.R. (1992b). The lactose permease of *Escherichia coli*: a paradigm for membrane transport proteins. Biochim. Biophys. Acta *1101*, 210-213.

Kanazawa, S., Driscoll, M., and Struhl, K. (1988). ATR1, a *Saccharomyces cerevisiae* gene encoding a transmembrane protein required for aminotriazole resistance. Mol. Cell. Biol. *8*, 664-673.

Kaestner, K.H., Christy, R.J., McLenithan, J.C., Braiterman, L.T., Cornelius, P., Pekala, P.H., and Lane, M.D. (1989). Sequence, tissue distribution, and differential expression of mRNA for a putative insulin-responsive glucose transporter in mouse 3T3-L1 adipocytes [published erratum appears in Proc Natl Acad Sci U S A 1989 Jul;86(13):4937]. Proc. Natl. Acad. Sci. USA *86*, 3150-3154.

Kasahara, M. and Hinkle, P.C. (1977). Reconstitution and purification of the D-glucose transporter from human erythrocytes. J. Biol. Chem. *252*, 7384-7390.

Kasanicki, M.A., Cairns, M.T., Davies, A., Gardiner, R.M., and Baldwin, S.A. (1987). Identification and characterisation of the glucose transporter of the bovine blood-brain barrier. Biochem. J. *247*, 101-108.

Katagiri, H., Asano, T., Shibasaki, Y., Lin, J.L., Tsukuda, K., Ishihara, H., Akanuma, Y., Takaku, F., and Oka, Y. (1991). Substitution of leucine for tryptophan 412 does not abolish cytochalasin B labeling but markedly decreases the intrinsic activity of GLUT1 glucose transporter. J. Biol. Chem. *266*, 7769-7773.

Katagiri, H., Asano, T., Ishihara, H., Tsukuda, K., Lin, J.L., Inukai, K., Kikuchi, M., Yazaki, Y., and Oka, Y. (1992). Replacement of intracellular C-terminal domain of GLUT1 glucose transporter with that of GLUT2 increases V_{max} and K_m of transport activity. J. Biol. Chem. *267*, 22550-22555.

Katagiri, H., Asano, T., Ishihara, H., Lin, J.L., Inukai, K., Shanahan, M.F., Kikuchi, M., Yazaki, Y., and Oka, Y. (1993). Role of tryptophan-388 of GLUT1 glucose transporter in glucose-transport activity and photoaffinity-labelling with forskolin. Biochem. J. *291*, 861-867.

Kayano, T., Fukumoto, H., Eddy, R.L., Fan, Y.S., Byers, M.G., Shows, T.B., and Bell, G.I. (1988). Evidence for a family of human glucose transporter-like proteins. Sequence and gene localisation of a protein expressed in fetal skeletal muscle and other tissues. J. Biol. Chem. *263*, 15245-15248.

Kayano, T., Burant, C.F., Fukumoto, H., Gould, G.W., Fan, Y.S., Eddy, R.L., Byers, M.G., Shows, T.B., Seino, S., and Bell, G.I. (1990). Human facilitative glucose transporters. Isolation, functional characterisation, and gene localisation of cDNAs encoding an isoform (GLUT5) expressed in small intestine, kidney, muscle, and adipose tissue and an unusual glucose transporter pseudogene-like sequence (GLUT6). J. Biol. Chem. *265*, 13276-13282.

Keller, K., Strube, M., and Mueckler, M. (1989). Functional expression of the human HepG2 and rat adipocyte glucose transporters in *Xenopus* oocytes. Comparison of kinetic parameters. J. Biol. Chem. *264*, 18884-18889.

Khan, S.A. and Novick, R.P. (1983). Complete nucleotide sequence of pT181, a tetracycline-resistance plasmid from *Staphylococcus aureus*. Plasmid *10*, 251-259.

King, S.C., Hansen, C.L., and Wilson, T.H. (1991). The interaction between aspartic acid 237 and lysine 358 in the lactose carrier of Escherichia coli. Biochim. Biophys. Acta *1062*, 177-186.

Klaiber, K., Williams, N., Roberts, T.M., Papazian, D.M., Jan, L.Y., and Miller, C. (1990). Functional expression of Shaker K+ channels in a baculovirus-infected insect cell line. Neuron. *5*, 221-226.

Kleffel, B., Garavito, R.M., Baumeister, W., and Rosenbusch, J.P. (1985). Secondary structure of a channel-forming protein: porin from E. coli outer membranes. EMBO Journal *4*, 1589-1592.

Komiya, H., Yeates, T.O., Rees, D.C., Allen, J.P., and Feher, G. (1988). Structure of the reaction center from *Rhodobacter sphaeroides* R-26 and 2.4.1: symmetry relations and sequence comparisons between different species. Proc. Natl. Acad. Sci. USA *85*, 9012-9016.

Kruckeberg, A.L. and Bisson, L.F. (1990). The HXT2 gene of *Saccharomyces cerevisiae* is required for high-affinity glucose transport. Mol. Cell. Biol. *10*, 5903-5913.

Krupka, R.M. (1971). Inhibition of sugar transport in erythrocytes by fluorodinitrobenzene. Biochemistry *10*, 1148-1153.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105-132.

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

Langford, C.K., Little, B.M., Kavanaugh, M.P., and Landfear, S.M. (1994). Functional expression of two glucose transporter isoforms from the parasitic protozoan *Leishmania enriettii*. J. Biol. Chem. *269*, 17939-17943.

Leblanc, G., Pourcher, T., and Zani, M.L. (1993). The melibiose permease of *Escherichia coli*: importance of the NH2-terminal domains for cation recognition by the Na+/sugar cotransporter. Soc. Gen. Phys. Ser. *48*, 213-227.

LeFevre, P.G. (1961). Sugar transport in the red blood cell: structure-activity relationships in substrates and antagonists. Pharmacol. Rev. *13*, 39-70.

LeFevre, P.G. and Marshall, J.K. (1958). Conformational specificity in a biological sugar transport system. Am. J. Physiol. *194*, 333-337.

Levy, S.B. (1992). Active efflux mechanisms for antimicrobial resistance. Anti. Agents Chemo. *36*, 695-703.

Lewis, D.A. and Bisson, L.F. (1991). The HXT1 gene product of *Saccharomyces cerevisiae* is a new member of the family of hexose transporters. Mol. Cell. Biol. *11*, 3804-3813.

Li, J. and Tooth, P. (1987). Size and shape of the *Escherichia coli* lactose permease measured in filamentous arrays. Biochemistry *26*, 4816-4823.

Lienhard, G.E., Slot, J.W., James, D.E., and Mueckler, M.M. (1992). How cells absorb glucose. Sci. Am. *266*, 86-91.

Liu, M.L., Olson, A.L., Moye-Rowley, W.S., Buse, J.B., Bell, G.I., and Pessin, J.E. (1992). Expression and regulation of the human GLUT4/muscle-fat facilitative glucose transporter gene in transgenic mice. J. Biol. Chem. *267*, 11673-11676.

Lowe, A.G. and Walmsley, A.R. (1986). The kinetics of glucose transport in human red blood cells. Biochim. Biophys. Acta *857*, 146-154.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. J. Biol. Chem. *193*, 265-275.

Luckow, V.A. and Summers, M.D. (1988). Signals important for high-level expression of foreign genes in *Autographa californica* nuclear polyhedrosis virus expression vectors. Virology *167*, 56-71.

Luckow, V.A. and Summers, M.D. (1989). High level expression of nonfused foreign genes with *Autographa californica* nuclear polyhedrosis virus expression vectors. Virology *170*, 31-39.

Lund-Andersen, H. (1979). Transport of glucose from blood to brain. Physiol. Rev. *59*, 305-352.

Lundahl, P., Mascher, E., Andersson, L., Englund, A.K., Greijer, E., Kameyama, K., and Takagi, T. (1991). Active and monomeric human red cell glucose transporter after high performance molecular-sieve chromatography in the presence of octyl glucoside and phosphatidylserine or phosphatidylcholine. Biochim. Biophys. Acta *1067*, 177-186.

Madon, R.J., Martin, S., Davies, A., Fawcett, H.A., Flint, D.J., and Baldwin, S.A. (1990). Identification and characterisation of glucose transport proteins in plasma membrane- and Golgi vesicle-enriched fractions prepared from lactating rat mammary gland. Biochem. J. *272*, 99-105.

Maher, F., Vannucci, S., Takeda, J., and Simpson, I.A. (1992). Expression of mouse-GLUT3 and human-GLUT3 glucose transporter proteins in brain. Biochem. Biophys. Res. Commun. *182*, 703-711.

Maher, F. and Harrison, L.C. (1991). Stimulation of glucose transporter (GLUT1) mRNA and protein expression by inhibitors of glycosylation. Biochim. Biophys. Acta *1089*, 27-32.

Maiden, M.C.J., Davis, E.O., Baldwin, S.A., Moore, D.C.M., and Henderson, P.J.F. (1987). Mammalian and bacterial sugar transport proteins are homologous. Nature *325*, 641-643.

Maiden, M.C.J., Jones-Mortimer, M.C., and Henderson, P.J.F. (1988). The cloning, DNA sequence, and overexpression of the gene araE coding for arabinose-proton symport in *Escherichia coli* K12. J. Biol. Chem. *263*, 8003-8010.

Mantych, G.J., James, D.E., and Devaskar, S.U. (1993). Jejunal/kidney glucose transporter isoform (Glut-5) is expressed in the human blood-brain barrier. Endocrinology *132*, 35-40.

Matsuura, Y., Possee, R.D., Overton, H.V., and Bishop, D.H.L. (1987). Baculovirus expression vectors: the requirements for high level expression of proteins, including glycoproteins. J. Gen. Virol. *68*, 1233-1250.

May, J.M., Buchs, A., and Carter-Su, C. (1990). Localisation of a reactive exofacial sulfhydryl on the glucose carrier of human erythrocytes. Biochemistry *29*, 10393-10398.

McMorrow, I., Chin, D.T., Fiebig, K., Pierce, J.L., Wilson, D.M., Reeve, E.C.R., and Wilson, T.H. (1988). The lactose carrier of *Klebsiella pneumoniae* M5a1; the physiology of transport and the nucleotide sequence of the lacY gene. Biochim. Biophys. Acta *945*, 315-323.

Mori, H., Hashiramoto, M., Clark, A.E., Yang, J., Muraoka, A., Tamori, Y., Kasuga, M., and Holman, G.D. (1994). Substitution of tyrosine-293 of GLUT1 locks the transporter into an outward facing conformation. J. Biol. Chem. *269*, 11578-11583.

Mueckler, M., Caruso, C., Baldwin, S.A., Panico, M., Blench, I., Morris, H.R., Allard, W.J., Lienhard, G.E., and Lodish, H.F. (1985). Sequence and structure of a human glucose transporter. Science *229*, 941-945.

Mueckler, M. and Lodish, H.F. (1986). The human glucose transporter can insert posttranslationally into microsomes. Cell *44*, 629-637.

Nagamatsu, S., Kornhauser, J.M., Burant, C.F., Seino, S., Mayo, K.E., and Bell, G.I. (1992). Glucose transporter expression in brain. cDNA sequence of mouse GLUT3, the brain facilitative glucose transporter isoform, and identification of sites of expression by in situ hybridisation. J. Biol. Chem. *267*, 467-472.

Neal, R.J. and Chater, K.F. (1987). Nucleotide sequence analysis reveals similarities between proteins determining methylenomycin A resistance in Streptomyces and tetracycline resistance in eubacteria. Gene *58*, 229-241.

Nehlin, J.O., Carlberg, M., and Ronne, H. (1989). Yeast galactose permease is related to yeast and mammalian glucose transporters. Gene *85*, 313-319.

Neyfakh, A. (1992). The multidrug efflux transporter of Bacillus subtilis is a structural and functional homolog of the *Staphylococcus aureus* NorA protein. Anti. Agents Chemo. *36*, 484-485.

Nikawa, J., Tsukagoshi, Y., and Yamashita, S. (1991). Isolation and characterisation of two distinct myo-inositol transporter genes of *Saccharomyces cerevisiae*. J. Biol. Chem. *266*, 11184-11191.

Oka, Y., Asano, T., Shibasaki, Y., Kasuga, M., Kanazawa, Y., and Takaku, F. (1988). Studies with antipeptide antibody suggest the presence of at least two types of glucose transporter in rat brain and adipocyte. J. Biol. Chem. *263*, 13432-13439.

Oka, Y., Asano, T., Shibasaki, Y., Lin, J.L., Tsukuda, K., Katagiri, H., Akanuma, Y., and Takaku, F. (1990). *C*-terminal truncated glucose transporter is locked into an inward-facing form without transport activity. Nature *345*, 550-553.

Osborne, H.B. and Nabedryk-Viala, E. (1978). The conformation of membrane-bound and detergent-solubilised bovine rhodopsin. A comparative hydrogen-isotope exchange study. European Journal of Biochemistry *89*, 81-88.

Overington, J., Donnelly, D., Johnson, M.S., Sali, A., and Blundell, T.L. (1992). Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. Prot. Sci. *1*, 216-226.

Palmiter, R.D., Chen, H.Y., and Brinster, R.L. (1982). Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell *29*, 701-710.

Perkins, S.J., Haris, P.I., Sim, R.B., and Chapman, D. (1988). A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. Biochemistry *27*, 4004-4012.

Permutt, M.A., Koranyi, L., Keller, K., Lacy, P.E., Scharp, D.W., and Mueckler, M. (1989). Cloning and functional expression of a human pancreatic islet glucose-transporter cDNA. Proc. Natl. Acad. Sci. USA *86*, 8688-8692.

Pessino, A., Hebert, D.N., Woon, C.W., Harrison, S.A., Clancy, B.M., Buxton, J.M., Carruthers, A., and Czech, M.P. (1991). Evidence that functional erythrocyte-type glucose transporters are oligomers. J. Biol. Chem. *266*, 20213-20217.

Poolman, B. and Konings, W.N. (1993). Secondary solute transport in bacteria. Biochim. Biophys. Acta *1183*, 5-39.

Raibaud, A., Zalacain, M., Holt, T.G., Tizard, R., and Thompson, C.J. (1991). Nucleotide sequence analysis reveals linked N-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialaphos biosynthetic gene cluster of *Streptomyces hygroscopicus*. J. Bacteriol. *173*, 4454-4463.

Rampal, A.L., Jung, E.K.Y., Chin, J.J., Deziel, M.R., Pinkofsky, H.B., and Jung, C.Y. (1986). Further characterisation and chemical purity assessment of the human erythrocyte glucose transporter preparation. Biochim. Biophys. Acta *859*, 135-142.

Rees, D.C., Komiya, H., Yeates, T.O., Allen, J.P., and Feher, G. (1989). The bacterial photosynthetic reaction center as a model for membrane proteins. Ann. Rev. Biochem. *58*, 607-633.

Rouch, D.A., Cram, D.S., DiBerardino, D., Littlejohn, T.G., and Skurray, R.A. (1990). Efflux-mediated antiseptic resistance gene qacA from *Staphylococcus aureus*: common ancestry with tetracycline- and sugar-transport proteins. Mol. Microbiol. *4*, 2051-2062.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R., Horn, R.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.. Science *239*, 487-491.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Sakaguchi, R., Hitoshi, A., and Shishido, K. (1988). Nucleotide sequence homology of the tetracycline resistance determinant naturally maintained in *Bacillus subtilis* Marburg 168 chromosome and the tetracycline resistance gene of B. subtilis plasmid pNS1981. Biochim. Biophys. Acta *950*, 441-444.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning (A laboratory manual) (NY: Cold Spring Harbour Laboratory, Cold Spring Harbour).

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

Santalucia, T., Camps, M., Castello, A., Munoz, P., Nuel, A., Testar, X., Palacin, M., and Zorzano, A. (1992). Developmental regulation of GLUT-1 (erythroid/Hep G2) and GLUT-4 (muscle/fat) glucose transporter expression in rat heart, skeletal muscle, and brown adipose tissue. Endocrinology *130*, 837-846.

Sarkar, H.K., Thorens, B., Lodish, H.F., and Kaback, H.R. (1988). Expression of the human erythrocyte glucose transporter in *Escherichia coli.* Proc. Natl. Acad. Sci. USA *85*, 5463-5467.

Sasatsu, M., Misra, T.K., Chu, L., Laddaga, R., and Silver, S. (1985). Cloning and DNA sequence of a plasmid-determined citrate utilisation system in *Escherichia coli.* J. Bacteriol. *164*, 983-993.

Sauer, N., Friedlander, K., and Graml-Wicke, U. (1990). Primary structure, genomic organisation and heterologous expression of a glucose transporter from *Arabidopsis thaliana*. EMBO J. *9*, 3045-3050.

Sauer, N. and Tanner, W. (1989). The hexose carrier from Chlorella. cDNA cloning of a eucaryotic H+-cotransporter. FEBS Lett. *259*, 43-46.

Schiffer, M., Chang, C.H., and Stevens, F.J. (1992). The functions of tryptophan residues in membrane proteins. Prot. Eng. *5*, 213-214.

Schiffer, M. and Edmundson, A.B. (1967). Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. *7*, 121-135.

Schmetterer, G.R. (1990). Sequence conservation among the glucose transporter from the Cyanobacterium synechocystis sp PCC 6803 and mammalian glucose transporters. Plant. Mol. Biol. *14*, 697-706.

Schürmann, A., Keller, K., Monden, I., Brown, F.M., Wandel, S., Shanahan, M.F., and Joost, H.G. (1993). Glucose transport activity and photolabelling with 3-[i-125]iodo-4-azidophenethylamido-7-o-succinyldeacetyl (IAPS)-forskolin of 2 mutants at tryptophan-388 and tryptophan- 412 of the glucose transporter GLUT1 - dissociation of the binding domains of forskolin and glucose. Biochem. J. *290*, 497-501.

Seol, W. and Shatkin, A.J. (1991). *Escherichia coli* kgtP encodes an alpha-ketoglutarate transporter. Proc. Natl. Acad. Sci. USA *88*, 3802-3806.

Sergeant, S. and Kim, H.D. (1985). Inhibition of 3-O-methylglucose transport in human erythrocytes by forskolin. J. Biol. Chem. *260*, 14677-14682.

Shanahan, M.F. (1983). Characterisation of cytochalasin B photoincorporation into human erythrocyte D-glucose transporter and F-actin. Biochemistry *22*, 2750-2756.

Shanahan, M.F., Morris, D.P., and Edwards, B.M. (1987). [3H] Forskolin. Direct photoaffinity labeling of the erythrocyte D-glucose transporter. J. Biol. Chem. *262*, 5978-5984.

Sipos, L. and von Heijne, G. (1993). Predicting the topology of eukaryotic membrane proteins. Eur. J. Biochem. *213*, 1333-1340.

Smith, G.E., Fraseer, M.J., and Summers, M.D. (1983). Molecular engineering of the *Autographa californnica* nuclear polyhedrosis virus genome: deletion mutations within the polyhedrin gene.. J. Virol. *46*, 584-593.

Smith, C.D., Hirayama, B.A., and Wright, E.M. (1992). Baculovirus-mediated expression of the Na+/glucose cotransporter in *Sf*9 cells. Biochim. Biophys. Acta *1104*, 151-159.

Sofue, M., Yoshimura, Y., Nishida, M., and Kawada, J. (1992). Possible multifunction of glucose transporter. Transport of nicotinamide by reconstituted liposomes. Biochem. J. *288*, 669-674.

Sogin, D.C. and Hinkle, P.C. (1978). Characterisation of the glucose transporter from human erythrocytes. J. Supramol. Struct. *8*, 447-453.

Stark, G.R. and Wahl, G.M. (1984). Gene Amplification. Ann. Rev. Biochem. *53*, 447 -491.

Summers, M.D. and Smith, G.E. (1978). Baculovirus structural polypeptides. Virology *84*, 390-402.

Summers, M.D. and Smith, G.E. (1987). A manual of methods for Baculovirus vectors and insect cell culture procedures. Texas Agricultural Experiment Station Bulletin No. 1555

Takata, K., Kasahara, T., Kasahara, M., Ezaki, O., and Hirano, H. (1990). Erythrocyte/HepG2-type glucose transporter is concentrated in cells of blood-tissue barriers. Biochem. Biophys. Res. Commun. *173*, 67-73.

Tamori, Y., Hashiramoto, M., Clark, A.E., Mori, H., Muraoka, A., Kadowaki, T., Holman, G.D., and Kasuga, M. (1994). Substitution at Pro^{385} of GLUT1 perturbs the glucose transport function by reducing conformational flexibility. J. Biol. Chem. *269*, 2982-2986.

Taylor, W.R. (1990). Hierarchical Method to Align Large Numbers of Biological Sequences. In Methods in Enzymology. R.F. Doolittle, ed. (San Diego, California: Academic Press, Inc.), pp. 456-474.

Thorens, B., Sarkar, H.K., Kaback, H.R., and Lodish, H.F. (1988). Cloning and functional expression in bacteria of a novel glucose transporter present in liver, intestine, kidney, and beta-pancreatic islet cells. Cell *55*, 281-290.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA *76*, 4350-4354.

van der Rest, M.E., Schwarz, E., Oesterhelt, D., and Konings, W.N. (1990). DNA sequence of a citrate carrier of *Klebsiella pneumoniae*. Eur. J. Biochem. *189*, 401-407.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., and McCawley, P. (1977). The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). In Vitro *13*, 213-217.

Vera, J.C., Rivas, C.I., Fischbarg, J., and Golde, D.W. (1993). Mammalian facilitative hexose transporters mediate the transport of dehydroascorbic acid. Nature *364*, 79-82.

Vera, J.C. and Rosen, O.M. (1989). Functional expression of mammalian glucose transporters in *Xenopus* laevis oocytes: evidence for cell-dependent insulin sensitivity. Mol. Cell Biol. *9*, 4187-4195.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. *225*, 487-494.

Waddell, I.D., Scott, H., Grant, A., and Burchell, A. (1991). Identification and characterisation of a hepatic microsomal glucose transport protein. T3 of the glucose-6-phosphatase system?. Biochem. J. *275*, 363-367.

Waddell, I.D., Zomerschoe, A.G., Voice, M.W., and Burchell, A. (1992). Cloning and expression of a hepatic microsomal glucose transport protein. Comparison with liver plasma-membrane glucose-transport protein GLUT2. Biochem. J. *286*, 173-177.

Wadzinski, B.E., Shanahan, M.F., and Ruoho, A.E. (1987). Derivatisation of the human erythrocyte glucose transporter using a novel forskolin photoaffinity label. J. Biol. Chem. *262*, 17683-17689.

Wadzinski, B.E., Shanahan, M.F., Seamon, K.B., and Ruoho, A.E. (1990). Localisation of the forskolin photolabelling site within the monosaccharide transporter of human erythrocytes. Biochem. J. *272*, 151-158.

Wallace, B.A., Cascio, M., and Mielke, D.L. (1986). Evaluation of methods for the prediction of membrane protein secondary structures. Proc. Natl. Acad. Sci USA *83*, 9423-9427.

Walmsley, A.R. (1988). The dynamics of the glucose transporter. Trends Biochem. Sci. *13*, 226-231.

Waters, S.H., Rogowsky, J., Grinstead, J., Altenbuchner, J., and Schmitt, R. (1983). The tetracycline resistance determinants of RP1 and Tn1721: nucleotide sequence analysis. Nucl. Acid. Res. *11*, 6089-6140.

Weber, P.C. and Salemme, F.R. (1980). Structural and functional diversity in 4-alpha-helical proteins. Nature *287*, 82-84.

Wellner, M., Monden, I., and Keller, K. (1994). The role of cysteine residues in glucose-transporter-GLUT1-mediated transport and transport inhibition. Biochem. J. *299*, 813-817.

Weiss, M.S. and Schulz, G.E. (1992). Structure of porin refined at 1.8 A resolution. J. Mol. Biol. *227*, 493-509.

Yamaguchi, A., Ono, N., Akasaka, T., Noumi, T., and Sawai, T. (1990). Metal-tetracycline/H+ antiporter of Escherichia coli encoded by a transposon, Tn10. The role of the conserved dipeptide, Ser65-Asp66, in tetracycline transport. J. Biol. Chem. *265*, 15525-15530.

Yano, H., Seino, Y., Inagaki, N., Hinokio, Y., Yamamoto, T., Yasuda, K., Masuda, K., Someya, Y., and Imura, H. (1991). Tissue distribution and species difference of the brain type glucose transporter (GLUT3). Biochem. Biophys. Res. Commun. *174*, 470-477.

Yao, B., Sollitti, P., and Marmur, J. (1989). Primary structure of the maltose-permease-encoding gene of *Saccharomyces carlsbergensis*. Gene *79*, 189-197.

Yazyu, H., Shiota-Niiya, S., Shimamoto, T., Kanazawa, H., Futai, M., and Tsuchiya, T. (1984). Nucleotide sequence of the *melB* gene and characteristics of deduced amino acid sequence of the melibiose carrier in *Escherichia coli.* J. Biol. Chem. *259*, 4320-4326.

Yi, C.-K., Charalambous, B.M., Emery, V.C., and Baldwin, S.A. (1992). Characterisation of functional human erythrocyte-type glucose transporter (GLUT1) expressed in insect cells using a recombinant baculovirus. Biochem. J. *283*, 643-646.

Yoshida, H., Bugaki, M., Nakamura, K., Ubukata, K., and Konno, M. (1990). Nucleotide sequence and characterisation of the *Staphylococcus aureus* norA gene which confers resistance to quinolones. J. Bacteriol. *172*, 6942-6949.

Zoccoli, M.A., Baldwin, S.A., and Lienhard, G.E. (1978). The monosaccharide transport system of the human erythrocyte. Solubilisation and characterisation on the basis of cytochalasin B binding. J. Biol. Chem. *253*, 6923-6930.

APPENDIX SUGAR TRANSPORTER FAMILY ALIGNMENT

```
                                                                  Helix 1
CONSENSUS                                P     T                  A  GSF FGYDTGVI N     I
   FAMILY I                                     .                  .   L    .       .
Human    GLUT1                       1 MEPSSKKLTG-RL|MLAVGGAVLGSLQFGYNTGVI|NAPQKVIEEFYNQTW 48
Rabbit   GLUT1                         MEPSSKKVTG-RL|MLAVGGAVLGSLQFGYNTGVI|NAPQKVIEEFYNQTW
Rat      GLUT1                         MEPSSKKVTG-RL|MLAVGGAVLGSLQFGYNTGVI|NAPQKVIEEFYNQTW
Pig      GLUT1                         ************ |********************|********EFYNQTW
Mouse    GLUT1                         MEPSSKKVTG-RL|MLAVGGAVLGSLQFGYNTGVI|NAPQKVIEEFYNQTW
Human    GLUT2                           MTEDKVTG-TL|VFTVITAVLGSFQFGYDIGVI|NAPQQVIISHYRHVL
Rat      GLUT2                           MSEDKITG-TL|AFTVFTAVLGSFQFGYDIGVI|NAPQEVIISHYRHVL
Mouse    GLUT2                           MSEDKITG-TL|AFTVFTAVLSSFQFGYDIGVI|NAPQEVIISHYRHVL
Rat      GLUT7                           MSDDSLTA-TL|SLFVFTAVLGSFQFGYDIGVI|NAPQEVIISHYRHVL
Human    GLUT3                           MGTQKVTP-AL|IFAITVATIGSFQFGYNTGVI|NAPEKIIKEFINKTL
Mouse    GLUT3                           MGTTHVTP-SL|VFAVTVATIGSFQFGYNTGVI|NAPETILKDFLNYTL
Human    GLUT4        MPSGFQQIGSE--DGEPPQQRVTG-TL|VLAVFSAVLGSLQFGYNIGVI|NAPQKVIEQSYNETW
Rat      GLUT4        MPSGFQQIGSE--DGEPPQQRVTG-TL|VLAVFSAVLGSLQFGYNIGVI|NAPQKVIEQSYNATW
Mouse    GLUT4        MPSGFQQIGSDVKDGEPPRQRVTG-TL|VLAVFSAVLGSLQFGYNIGVI|NAPQKVIEQSYNATW
Human    GLUT5                   MEQQDQSMKEGRLTLVLA|LATLIAAFGSSFQYGYNVAAV|NSPALLMQQFYNETY
Rat      SV2    140..QRRKDREELAQQYETILRECGHG-RFQ|WTLYFVLGLALMADGVEVFVV|GFVLPSAEKDMCLSD
Yeast    SNF3    72...TDDISTIDDNSILFSEPP-QKQSMMMS|ICVGVFVAVGGFLFGYDTGLI|NSITSMNYVKSHVAP
Yeast    GAL2    44..KAGESGPEGSQSVPIEIPKKPMSEYVTV|SLLCLCVAFGGFMFGWDTSTI|SGFVVQTDFLRRFGM
Yeast    HXT1    36.NESFHDNLSESQVQPAVAPPNTGKGVYVT|VSICGMVAFGGFIFGWDTGTI|SGFVAQTDFLRRFGM
Yeast    HXT2    29..TDESPIQTKSEYTNAELPAKPIAAYWTV|ICLCLMIAFGGFVFGWDTGTI|SGFVNQTDFKRRFGQ
Yeast    RAG1        SHEEDLNDLEKTAEETLQQKPAKEYIFV|SLCCVMVAFGGFVFGWDTGTI|SGFVNQTDFLRRFGQ
Yeast    MAL61   74....MQDAKEADESERGMPLMTALHTYPKA|AAWSLLVSTTLIQEGYDTAIL|GAFYALPVFQKKYGS
Yeast    LAC12   47....INGVPIEDAREEVLLPGYLSKQYYKL|YGLCFITYLCATMQGYDIGALM|GSIYTEDAYLKYYHL
Yeast    ITR1    61...IQIKPVNDEDDTSVMITFNQSLSP-FI|ITLTFVASISGFMFGYDTGYI|SSALISIGTDLDHKV
Yeast    ITR2    87...IVIKPVNDEDDTSVIITFNQSISP-FI|ITLTFVASISGFMFGYDTGYI|SSALISINRDLDNKV
Chlorella HUP1      MAGGGVVVVSGRGLSTGDYRGGLTV-YV|VMVAFMAACGGLLIGYDNGVT|GGVVSLEAFEK-FFP
Arabidopsis STP1    MPAGGFVVGDGQ----KAYPGKLTP-FV|LFTCVVAAMGGLIFGYDIGIS|GGVTSMPSFLKRFFP
Synechocystis glcP              MNPSSSPSQSTANVKFV|LLISGVAALGGFLFGYDTAVI|NGAVAALQKHFQTDS
Leishmania Pro-1 31DDQEDAPPFMTANNARVMLVQAIGGSLN|GYSIGFVGVYSTLFGYSTNCA|SFLQENSCTTVPNAD
Neurospora qa-y             MTLLALKEDRPTPKAVYNWRV|YTCAAIATFASCMIGYDSAFI|GTTLALPSFTKEFDF
Aspergillus qutD            MSILALVEDRPTPREVYNWRV|YLLAAVASFTSCMIGYDSAFI|GTTLSLQSFQNEFNW
E. coli  AraE             MVTINTESALTPRSLRDTRRMN|MFVSVAAAVAGLLFGLDIGVI|AGALPFITDHFVLTS
E. coli  GalP                     MPDAKKQGRSNKAMT|FFVCFLAALAGLLFGLDIGVI|AGALPFIADEFQITS
E. coli  XylE                          MNTQYNSSYI|FSITLVATLGGLLFGYDTAVI|SGTVESLNTVFVAPQ
Z. mobilis glf                         MSSESSQGLV|TRLALIAAIGGLLFGYDSAVI|AAIGTPVDIHFIAPR
   FAMILY IV
K. pneumoniae cit'  8...ASSTAPVRNMATAGGARIGAILRV|TSGNFLEQFDFFLFGFYATYI|AHTFFPASSEFA---
E. coli  citA                   MTQQPSRAGTFGAILRV|TSGNFLEQFDFFLFGFYATYI|AKTFFPAESEFA---
E. coli kgtP        MAESTVTADSKLTSSDTRRRIW-AIVGA|SSGNLVEWFDFYVYSFCSLYF|AHIFFPSGNTTT---
S. hygroscopicus bap3   MTVRESDRTTVPRSRSLRQLVCG|GIGHTVESHDWYVYTFLAVYF|SDDIFPESSGDPLV-
   FAMILY II
Tn1546 smiA                        MSSKNFSWRYSLAA|TVLLLSPFDLLASLGTDMYLP|AVPFMPNALGTTAST
E. coli pBR322 (Tet C)                    MKSNNAL|IVILGTVTLDAVGIGLVMPVL|PGLLRDIVHSDS---
E. coli pRP1 (Tet A)                      MKPNIPL|IVILSTVALDAVGIGLIMPVL|PGLLRDLVHSND---
E. coli  Tn10 (Tet B)                       MNSST|KIALVITLLDAMGIGLIMPVL|PTLLREFIASED---
S. aureus norA                              MNKQI|FVLYFMIFLIFLGIGLVIPVL|PVYLKDLGLT-----
Bacillus subtilis Bmr                     MEKKNITL|TILLTNLFIAFLGIGLVIPVT|PTIMNELHLS-----
   FAMILY III
Bacillus pTHT15 (Tet L)                 MNTSYSQSNLRH|NQILIWLCILSFFSVLNEMVL|NVSLPDIANDFNKPP
Bacillus pNS1981                        MNTSYSQSNLRH|NQILIWLCILSFFSVLNEMVL|NVSLPDIANDFNKPP
Bacillus BS908                          MNTSYSQSTLRH|NQVLIWLCVLSFFSVLNEMVL|NVSLPDIANEFNKLP
S. aureus pT181 (Tet K)                 MFSLYKKFKGLF|YSVLFWLCILSFFSVLNEMVL|NVSLPDIANHFNTTP
Streptomyces mmr          MTTVRTGGAQTAEVPAGGRRDVPS|GVKITALATGFVMATLDVTVV|NVAGATIQESLDTTL
S. aureus qacA                  MISFFTKTTIMMTSKKR|WTALVVLAVSLFVVTMDMTIL|IMALPELVRELEPSG
S. coelicolor ActII                   ...TAAGPPPYARR|WAALGVILGAEIMDLLDGTVM|NVAAPAVRADLGGSL

Yeast    ATR1   44...QSEDEMVDSNQKWQNPNYFKYAWQEYL|FIFTCMISQLLNQAGTTQTLS|IMNILSDSFGSEGNS
   DISACCHARIDE

E. coli  LacY                           MYYLKNTNF|WMFGLFFFFYFFIMGAYFPFF|PIWLHDINHISK---
K. pneumoniae                        MKLSELAPRERHNF|IYFMLFFFFYFFIMSAYFPFF|PVWLAEVNHLTK---
E. coli  RafB                          MNSASTHKNTDF|WIFGLFFFLYFFIMATCFPFL|PVWLSDVVGLSK---
E. coli  MelB                                  MTTK|LSYGFGAFGKDFAIGIVYMYL|MYYYTDVVGLSV---
```

```
CONSENSUS
                                  *         *
Human     GLUT1  49 VHRYGESILPTT-------------------------------------------------LTTLWS  66
Rabbit    GLUT1     IHRYGERILPTT-------------------------------------------------LTTLWS
Rat       GLUT1     NHRYGESIPSTT-------------------------------------------------LTTLWS
Pig       GLUT1     LHRYGESISPAT-------------------------------------------------LTTLWS
Mouse     GLUT1     NHRIGEPIPSTT-------------------------------------------------LTTLWS
Human     GLUT2     GVPLDDRKAINNYVINSTDELLTISYSMNPKPTPWAEEETVAA--------------AQLITMLWS
Rat       GLUT2     GVPLDDRRATINYDINGTDTPLIVTPAHTT-PDAW-EEETEGS--------------AHIVTMLWS
Mouse     GLUT2     GVPLDDRKAAINYDVNGTDTPLTVTPAYTT-PAPWDEEETEGS--------------AHIVTMLWS
Rat       GLUT7     GVPLDDRRATINYDINGTDTPLIVTPAHTT-PDAW-EEETEGS--------------AHIVTMLWS
Human     GLUT3     TDKGNAPPSEVL-------------------------------------------------LTSLWS
Mouse     GLUT3     EERLEDLPSEGL-------------------------------------------------LTALWS
Human     GLUT4     LGRQGPEGPSSIPPGT---------------------------------------------LTTLWA
Rat       GLUT4     LGRQGPGGPDSIPQGT---------------------------------------------LTTLWA
Mouse     GLUT4     LGRQGPGGPDSIPQGT---------------------------------------------LTTLWA
Human     GLUT5     YGRTQEPMEDFP-------------------------------------------------LTLLWS
Rat       SV2       SNK----------------------------------------------------------G
Yeast     SNF3      NHDSFTAQQM---------------------------------------------------S
Yeast     GAL2      KHKDGTHYLSNVRT-----------------------------------------------G
Yeast     HXT1      KHHDGSHYLSKVRT-----------------------------------------------G
Yeast     HXT2      MKSDGTYYLSDVRT-----------------------------------------------G
Yeast     RAG1      EKADGSHYLSNVRT-----------------------------------------------G
Yeast     MAL61     LNSNTGDYEI---------------------------------------------SVSWQIG
Yeast     LAC12     DINSSSGT-----------------------------------------------------G
Yeast     ITR1      LTYGEKE------------------------------------------------------
Yeast     ITR2      LTYGEKE------------------------------------------------------
Chlorella HUP1      DVWAKKQEVHEDSPYCTYDNAKLQ-------------------------------------
Arabidopsis STP1    SVYRKQQEDASTNQYCQYDSPTLT-------------------------------------
Synechocystis glcP  LLT--------------------------------------------------------G
Leishmania Pro-1    CKWFVSPTGSSYCGWPEVTCRKEYAYSSPAEMPGALARCEADSRCRWSYSDEECQNPSGYSSSESG
Neurospora qa-y     ASYTPGALA----------------------------------------------------LLQS
Aspergillus qutD    ESLNTD-------------------------------------------------------LISA
E. coli  AraE       RLQE---------------------------------------------------------
E. coli  GalP       HTQE---------------------------------------------------------
E. coli  XylE       NLSESAANG----------------------------------------------------LLG
Z. mobilis glf      HLSATAAAG----------------------------------------------------LSG

K. pneumoniae citt--------------------------------------------------SLMMTF
E. coli  citA       -------------------------------------------------------ALMLTF
E. coli kgtP        -------------------------------------------------------QLLQTA
S. hygroscopicus    -------------------------------------------------------PLLNTF
bap3
Tn1696 cmlA         ----------------------------------------------------------IQ
E. coli pBR322      -------------------------------------------------------IASHYG
E. coli  RP1        -------------------------------------------------------VTAHYG
E. coli  Tn10       -------------------------------------------------------IANHFG
S. aureus norA      ------------------------------------------------------GSDLG
B. subtilis Bmr     ------------------------------------------------------GTAVG

Bacillus pTHT15     -----------------------------------------------------------ASTN
Bacillus pN31981    -----------------------------------------------------------ASTN
Bacillus BS908      -----------------------------------------------------------ASAN
S. aureus pT181     -----------------------------------------------------------GITN
Streptomyces mmr    ----------------------------------------------------------TQLT
S. aureus qacA      ----------------------------------------------------------TQQL
S. coelicol' ActII  ----------------------------------------------------------SVIQ
Yeast     ATR1      ----------------------------------------------------------KS

E. coli  LacY       -----------------------------------------------------------SDTG
K. pneumoniae       -----------------------------------------------------------TETG
E. coli  RafB       -----------------------------------------------------------TDTG
E. coli  MelB       ---------------------------------------------------------GLVGTLFL
```

```
                                  Helix 2                                  Helix 3
CONSENSUS                         L V  F  G   GS   G        GRR   ML  N L   G   MG
                                     .                     .          .         .
Human    GLUT1  67  LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLMMNLLAFVSAVLMGFSKLG KSF--------- 119
Rabbit   GLUT1      LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLMMNLLAFVSAVLMGFSKLG KSF---------
Rat      GLUT1      LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLMMNLLAFVSAVLMGFSKLG KSF---------
Pig      GLUT1      LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLMMNLLAFISAVLMGFSKLG KSF---------
Mouse    GLUT1      LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLMMNLLAFVAAVLMGFSKLG KSF---------
Human    GLUT2      LSVSSFAVGGMTASFFGGWLG DTLGRIKA MLVANILSLVGALLMGFSKLG PSH---------
Rat      GLUT2      LSVSSFAVGGMVASFFGGWLG DKLGRIKA MLAANSLSLTGALLMGCSKFG PAH---------
Mouse    GLUT2      LSVSSFAVGGMVASFFGGWLG DKLGRIKA MLAANSLSLTGALLMGCSKFG PAH---------
Rat      GLUT7      LSVSSFAVGGMVASFFGGWLG DKLGRIKA MLAANSLSLTGALLMGCSKFG PSH---------
Human    GLUT3      LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLIVNLLAVTGGCFMGLCKVA KSV---------
Mouse    GLUT3      LSVAIFSVGGMIGSFSVGLFV NRFGRRNS MLLVNLLAIIAGCLMGFAKIA ESV---------
Human    GLUT4      LSVAIFSVGGMISSFLIGIIS QWLGRRRA MLVNNVLAVLGGSLMGLANAA ASY---------
Rat      GLUT4      LSVAIFSVGGMISSFLIGIIS QWLGRRRA MLANNVLAVLGGALMGLANAA ASY---------
Mouse    GLUT4      LSVAIFSVGGMISSFLIGIIS QWLGRRRA MLANNVLAVLGGALMGLANAV ASY---------
Human    GLUT5      VTVSMFFGGFIGSLLVGPLV NKFGRRGA ILFNNIFSIVPAILMGCSRVA TSF---------
Rat      SV2        MLGLIVYLGMMVGAFLWGGLA DRLGRRQC LLISLSVNSVFAFFSSFVQGY GTF---------
Yeast    SNF3       ILVSFLSLGTFFGALTAPFIS DSYGRRPI IIFSTIFIFSIGNSLQVGAGG ITL---------
Yeast    GAL2       LIVAIFNIGCAFGGIILSKGG DMYGRRKG LSIVVSVYIVGIIIQIASINK WY----------
Yeast    HXT1       LIVSIFNIGCAIGGIVLAKLG DMYGRRIG LIVVVVIYTIGIIIQIASINK WY----------
Yeast    HXT2       LIVSIFNIGCAFGGLTLGRLG DMYGRRIG LMCVVLVYIVGIVIQIASSDK WY----------
Yeast    RAG1       LIVSIFNIGCAVGGIVLSNIG DRWGRRIG LITVIIIYVIGIIIQIASVDK WY----------
Yeast    MAL61      LCL-CYMAGFIVGLQVTGPSV DYMGNRYT LIMALFFLAAFIFILYFCKSL G----------
Yeast    LAC12      LVFSIFNVGQICGAFFVPLM- DWKGRKPA ILIGCLGVVIGAIISSLTTTK S----------
Yeast    ITR1       IVTAATSLGALITSIFAGTAA DIFGRRRC LMGSNLMFVIGAILQVSAHTF W----------
Yeast    ITR2       LITAATSLGALITSVGAGTAA DVFGRRRC LMFSNLMFLIGAILQITAHKF W----------
Chlorella HUP1      LFVSSLFLAGLVSCLFASWIT RNWGRRVT MGIGGAFFVAGGLVNAFAQDM A----------
Arabidopsis STP1    MFTSSLYLAALISSLVASTVT RKFGRRLS MLFGGILFCAGALINGFAKHV W----------
Synechocystis glcP  LSVSLALLGSALGAFGAGPIA DRHGRIKT MILAAVLFTLSSIGSGLPFTI W----------
Leishmania Pro-1    IFAGSMIAGCLIGSVFAGPLA SKIGARLS FLLVGLVGVVASVMYHASCAA DEFW--------
Neurospora qa-y     NIVSVYQAGAFFGCLFAYATS YFLGRRKS LIAFSVVFIIGAAIMLAADGQ GRGI--------
Aspergillus qutD    NIVSLYQRGAFFGALFAYPIG HFWGRRWG LMFGALIFFLGAGMMLGANGD -RGL--------
E. coli  AraE       WVVSSMMLGAAIGALFNGWLS FRLGRRYS LMAGAILFVLGSIGSAFATSV ----------
E. coli  GalP       WVVSSMMFGAAVGAVGSGWLS FKLGRRKS LMIGAILFVAGSLFSAAAPNV ----------
E. coli  XylE       FCVASALIGCIIGGALGGYCS NRFGRRDS LKIAAVLFFISGVGSAWPELG FTSINPDNTVPV
Z. mobilis glf      MVVVAVLVGCVTGSLLSGWIG IRFGRRGG LLMSSICFVAAGFGAALTEKL FGTGGSA-----

K. pne|AMMLAGFLMRPIGAIVLGAYI|DKVGRRKG|LIVTLSIMATGTFLIVLIPSY|QTIGLWAPL---
E. coli  citA       AVFGSGFLMRPIGAVVLGAYI DRLGRRKG LMITLAIMGCGTLLIALVPGY QTIGLLAPV---
E. coli  kgtP       GVFAAGFLMRPIGGWLFGRIA DKFGRRKKS MLLSVGMMCFGSLVIACLPGY ETIGTWAPA---
S. hygroscopicus    AVFALAFAARPVGATVMGWYA DRYGRRSA LIVTILLMGLGSLMIGLTPSY ATAGPVAPV---
bap3
Tn1696 cmlA         LTLTTYLVMIGAGQLLFGPLS DRLGRRFV LLGGGLAYVVASMGLALTSSA ----------
E. coli pBR322      VLLALYALMQFLCAPVLGALS DRFGRRPV LLASLLGATIDYAIMATTPVL W----------
E. coli pRP1        ILLALYALVQFACAPVLGALS DRFGRRPI LLVSLAGATVDYAIMATAPFL W----------
E. coli  Tn10       VLLALYALMQVIFAPWLGKMS DRFGRRPV LLLSLIGASLDYLLLAFSSAL W----------
S. aureus norA      LLVAAFALSQMIISPFGGTLA DKLGRKLI ICIGLILFSVSEFMFAVGHNF ----------
B. subtilis Bmr     YMVACFAITQLIVSPIAGRWV DRFGRKIM IVIGLLFFSVSEFLFGIGKTV ----------

Bacillus pTHT15     WVNTAFMLTFSIGTAVYGKLS DQLGIKRL LLFGIIINCFGSVIGFVGHSF FS----------
Bacillus pNS1981    WVNTAFMLTFSIGTAVYGKLS DQLGIKRL LLFGIIINCFGSVIGFVGHSF FS----------
Bacillus BS908      WVNTAFMLTFSIGTALYGKLS DQLGIKNL LLFGIMVNGLGSIIGFVGHSF FP----------
S. aureus pT181     WVNTAYMLTFSIGTAVYGKLS DYINIKKL LIIGISLSCLGSLIAFIGHNH FF----------
Streptomyces mmr    WIVDGYVLTFASLLMLAGGLA NRIGARTV YLWGMGVFFLASLACALAPTA ET----------
S. aureus qacA      WIVDIYSLVLAGFIIPLSAFA DKWGRRKA LLTGFALFGLVSLAIFFAESA EF----------
S. coelicol' ActII  WITVGYTLAFAVLLVVGGRLG DIYGRKRM FVVGAVGFTAASVLCSVAAGP EM----------
Yeast    ATR1       WLMASFPLVSGSFILISGRLG DIYGLKKM LLVGYVLVIIWSLICGITKYS G----------

E. coli  LacY       IIFAAISLFSLLFQPLFGLLS DKLGLRRY LLWIITGMLVMFAPFFIFIFG PLL---------
K. pneumoniae       IVFSCISLFAIIFQPVFGLIS DKLGLRRH LLWTITILLILFAPFFIFVFS PLL---------
E. coli  RafB       IVFSCLSLFAISFQPLLGVIS DRLGLKKN LIWSISLLLVFFAPFFLYVFA PLL---------
E. coli  MelB       VARIWDAINDPIMGWIVNATR SRWGKFKP WILIGTLANSVILFLLFSAHL FEG------TTQ
```

Helix 4

```
CONSENSUS                    LI GR    G   G  S  VPMY  E  AP  L     RGA
                                 .    .                .   .

Human     GLUT1 120 ------EMLIGR|FIIGVYCGLTTGFVPMYVGEV|SPTAF-----RGA---- 155
Rabbit    GLUT1     ------EMLIGR|FIIGVYCGLTTGFVPMYVGEV|SPTAL-----RGA----
Rat       GLUT1     ------EMLIGR|FIIGVYCGLTTGFVPMYVGEV|SPTAL-----RGA----
Pig       GLUT1     ------EMLIGR|FIIGVYCGLTTGFVPMYVGEV|SPTAL-----RGA----
Mouse     GLUT1     ------EMLIGR|FIIGVYCGLTTGFVPMYVGEV|SPTAL-----RGA----
Human     GLUT2     ------ILIAGR|SISGLYCGLISGLVPMYIGEI|APTAL-----RGA----
Rat       GLUT2     ------ALIAGR|SVSGLYCGLISGLVPMYIGEI|APTTL-----RGA----
Mouse     GLUT2     ------ALIAGR|SVSGLYCGLISGLVPMYIGEI|APNTL-----RGA----
Rat       GLUT7     ------ALIAGR|SVSGLYCGLISGLVPMYIGEI|SPHTL-----RGA----
Human     GLUT3     ------EMLIGR|LVIGLFCGLCTGFVPMYIGEI|SPTAL-----RGA----
Mouse     GLUT3     ------EMLIGR|LLIGIFCGLCTGFVPMYIGEV|SPTAL-----RGA----
Human     GLUT4     ------EMLIGR|FLIGAYSGLTSGLVPMYVGEI|APTHL-----RGA----
Rat       GLUT4     ------EILIGR|FLIGAYSGLTSGLVPMYVGEI|APTHL-----RGA----
Mouse     GLUT4     ------EILIGR|FLIGAYSGLTSGLVPMYVGEI|APTHL-----RGA----
Human     GLUT5     ------ELIISR|LLVGICAGVSSNVVPMYLGEL|APKNL-----RGA----
Rat       SV2       ---------LFCR|LLSGVGIGGSIPIVFSYFSEF|LAQEK-----RGEH---
Yeast     SNF3      --------LIVGR|VISGIGIGAISAVVPLYQAEA|THKSL-----RGA----
Yeast     GAL2      -------QYFIGR|IISGLGVGGIAVLCPMLISEI|APKHL-----RGT----
Yeast     HXT1      -------QYFIGR|IISGLGVGGITVLSPMLISEV|APSEM-----RGT----
Yeast     HXT2      -------QYFIGR|IISGMGVGGIAVLSPTLISET|APKHI-----RGT----
Yeast     RAG1      -------QYFIGR|IISGLGVGGITVLSPMLISET|APKHL-----RGT----
Yeast     MAL61     -------MIAVGQ|ALCGMPWGCFQCLTVSYASEI|CPLAL-----RYY----
Yeast     LAC12     -------ALIGGR|WFVAFFATIANAAAPTYCAEV|APAHL-----RGK----
Yeast     ITR1      -------QMAVGR|LIMGFGVGIGSLIAPLFISEI|APKMI-----RGR----
Yeast     ITR2      -------QMAAGR|LIMGFGVGIGSLISPLFISEI|APKMI-----RGR----
Chlorella HUP1      -------MLIVGR|VLLGFGVGLGSQVVPQYLSEV|APFSH-----RGM----
Arabidopsis STP1    -------MLIVGR|ILLGFGIGFANQAVPLYLSEM|APYKY-----RGA----
Synechocystis glcP  -------DFIFWR|VLGGIGVGAASVIAPAYIAEV|SPAHL-----RGR----
Leishmania Pro-1    -------VLIVGR|FVIGLFLGVICVACPVYTDQN|AHPKW-----KRT----
Neurospora qa-y     ------DPIIAGR|VLAGIGVGGASNMVPIYISEL|APPAV-----RGR----
Aspergillus qutD    ------GLIYGGR|VLAGIGVGAGSNLCPIYISEM|APSAI-----RGR----
E. coli  AraE       ------EMLIAAR|VVLGIAVGIASYTAPLYLSEM|ASENV-----RGK----
E. coli  GalP       ------EVLILSR|VLLGLAVGVASYTAPLYLSEI|APEKI-----RGS----
E. coli  XylE       YLAGYVPEFVIYR|IIGGIGVGLASMLSPMQIAEL|APAHI-----RGK----
Z. mobilis glf      -----LQIFCFIR|FLAGLGIGVVSTLIPTYIAEI|RPPDK-----RGQ----

K. pneumoniae       LVLIGR|LLQGFSAGAELGGVSVYLAEI|ATPGR-----RGFYTSW
E. coli  citA       -------LVLVGR|LLQGFSAGVELGGVSVYLSEI|ATPGN-----KGFYTSW
E. coli kgtP        -------LLILAR|LFQGLSVGGEYGTSATYMGEV|AVEGR-----KGFYASF
S. hygroscopicus    -------VLIAAR|LVQGFSLGGEYGAATTFLVES|AAPGR-----RALFSSF
bap3
Tn1696 cmlA         ------EVFLGLR|ILQACGASACLVSTFATVRDI|YAGRE-----ESNV---
E. coli pBR322      -------ILYAGR|IVAGIT-GATGAVAGAYIADI|TDGED-----RAR----
E. coli  RP1        -------VLYIGR|IVAGIT-GATGAVAGAYIADI|TDGDE-----RAR----
E. coli  Tn10       -------MLYLGR|LLSGIT-GATGAVAASVIADT|TSASQ-----RVK----
S. aureus norA      ------SVLMLSR|VIGGMSAGMVMPGVTGLIADI|SPSHQ-----KAK----
B. subtilis Bmr     ------EMLFITR|MLGGISAPFIMPGVTAFIADI|TTIKT-----RPK----

Bacillus pTHT15     -------LLIMAR|FIQGAGAAAFPALVMVVVARY|IPKEN-----RGK----
Bacillus pNS1981    -------LLIMAR|FIQGAGAAAFPALVMVVVARY|IPKEN-----RGK----
Bacillus BS908      -------LLILAR|FIQGIGAAAFPALVMVVVARY|IPKEN-----RGK----
S. aureus pT181     -------ILIFGR|LVQGVGSAAFPSLIMVVVARN|ITRKK-----QGK----
Streptomyces mmr    --------LIAAR|LVQGAGAALFMPSSLSLLVFS|FPEKRQ----RTR----
S. aureus qacA      --------VLAIR|FLLGIAGALIMPTTLSMIRVI|FENPKE----RAT----
S. coelicol' ActII  --------LTAAR|FLQGGLGALMIPQGLGLIKQM|FPPKET-----AA----
Yeast     ATR1      ----SDTFFIISR|AFQGLGIAFVLPNVLGIIGNI|YVGGTF----RKNI---

E. coli  LacY       -QYNILVGSIVGG|IYLGFCFNAGAPAVEAFI-EK|VSRRSNFEFGRAR----
M. pneumoniae       -QMNIMAGALVGG|VYLGIVFSSGSGAVEAYI-ER|VSRANRFEYGKVR----
E. coli  RafB       -HLNIWAGALTGG|VFIGFVFSAGAGAIEAYI-ER|VSRSSGFEYGKAR----
E. coli  MelB       IVFVCVTYILWGM|TYTIMDIPFWSLVPTITL-DK|REREQLVPYPRFF----
```

```
                                   Helix 5                                                  Helix 6
CONSENSUS                          LG   QL I  GIL A   G                                    W  LLGL   PA LQ   L F
                                      .     .                          .                       .     .
Human    GLUT1 156                 LGTLHQLGIVVGILIAQVFGL|-----------DSIMGNKDL|WPLLLSIIFIPALQCIVLPF|- 206
Rabbit   GLUT1                     LGTLHQLGIVVGILIAQVFGL|-----------DSIMGNEDL|WPLLLSVIFVPALQCIVLPL|-
Rat      GLUT1                     LGTLHQLGIVVGILIAQVFGL|-----------DSIMGNADL|WPLLLSVIFIPALQCILLPF|-
Pig      GLUT1                     LGTLHQLGIVVGILIAQVFGL|-----------DSIMGNEEL|WPLLLSVIFIPALQCVLLPF|-
Mouse    GLUT1                     LGTLHQLGIVVGILIAQVFGL|-----------DSIMGNADL|WPLLLSVVFVPALQCILLPF|-
Human    GLUT2                     LGTFHQLAIVTGILISQIIGL|-----------EFILGNYDL|WHILLGLSGVRAILQSLLLF|-
Rat      GLUT2                     LGTLHQLALVTGILISQIAGL|-----------SFILGNQDY|WHILLGLSAVPALLQCLLLF|-
Mouse    GLUT2                     LGTLHQLALVTGILISQIAGL|-----------SFILGNQDH|WHILLGLSAVPALLQCLLLF|-
Rat      GLUT7                     AGTLLQLGITVGITISQIIGL|-----------DNSSGNVNT|WPHLLSLSRIPAALQPAILPF|-
Human    GLUT3                     FGTLNQLGIVVGILVAQIFGL|-----------EFILGSEEL|WPLLLGFTILPAILQSAALPF|-
Mouse    GLUT3                     FGTLNQLGIV-GILVAQIFGL|-----------DFILGSEEL|WPGLLGLTIIPAILQSAALPF|-
Human    GLUT4                     LGTLNQLAIVIGILIAQVLGL|-----------ESLLGTASL|WPLLLGLTVLPAILQLVLLPF|-
Rat      GLUT4                     LGTLNQLAIVIGILVAQVLGL|-----------ESMLGTATL|WPLLLAITVLPALQLLLLPF|-
Mouse    GLUT4                     LGTLNRLAIVIGILVAQVLGL|-----------ESMLGTATL|WPLLLAITVLPALQLILLPF|-
Human    GLUT5                     LGVVPQLFITVGILVAQIFGL|-----------RNLLANVDG|WPILLGLTGVPAIQLLLLPF|-
Rat      SV2                       LSWLCMFWMIGGVYAAAMAWA|-IIPHYGWSFQMGSAYQFHS|WRVFVLVFAFPSVFAIGALTT|-
Yeast    SNF3                      IISTYQWAITWGLLVSSAVSQ|G---------THARNDA-SS|YRIPIGLQYVWSSFLAIGMFF|-
Yeast    GAL2                      LVSCYQLYTTAGIFLGYCTNY|G---------TKSYSNS-VQ|WRVPLGLCFAWSLFMIGALTL|-
Yeast    HXT1                      LVSCYQVMITLGIFLGYCTNF|G---------TKNYSNS-VQ|WRVPLGLCFAWALFMIGGMMF|-
Yeast    HXT2                      CVSFYQLMITLGIFLGYCTNY|G---------TKDYSNS-VQ|WRVPLGLNFAFAIFMIAGMLM|-
Yeast    RAG1                      LVSCYQLMITLGIFLGYCTNY|G---------TKNYSNS-VQ|WRVPLGLCFAWAIFMVLGMMF|-
Yeast    MAL61                     LTTYSNLCWTFGQLFAAGIMK|NS--------QNKYANSELG|YKLPFALQWIWPLPLAVGIFL|-
Yeast    LAC12                     VAGLYNTLWCVGSIVAAFSTY|GT--------NKNFPNSSKA|FKIPLGLQMMFPGLVCIFGWL|-
Yeast    IRT1                      LTVINSLWLTGGLLVAYGCGA|GL----------NYVNN--G|WRILVGLSLIPTAVQFTCLCF|-
Yeast    IRT2                      LTVINSLWLTGGLLIAYGCGA|GL----------NHVKN--G|WRILVGLSLIPTVLQFSFFCF|-
Chlorella HUP1                     LNIGYQLFVTIGILIAGLVNY|AV----------RDWEN--G|WRLSLGLAAAPGAILFLGSLV|-
Arabidopsis STP1                   LNIGFQLITIGILVAFVLNY|FFAKIKGGW----------G|WRLSLGGAVVPALIITIGSLV|-
Synechocystis glcP                 LGSLQQLFIVSGIFIALLSNW|FIALMAGGSAQNPWLFGAAA|WRWMFWTELIPALLYGVCAFL|-
Leishmania Pro-1                   IGVMFQVFTTLGIFVAALMGL|ALGQSI-RFDHDGDQKVMAR|MQGLCVFSTLFSLLTVVLGIV|-
Neurospora qa-y                    LVGIYFLGWQIGGLVGFWINY|GVNTT--------MAPTRSQ|WLIPFAVQLIPAGLLFLGSFW|-
Aspergillus qutD                   LVGVYFLGWQIGGVVGFWINY|GVDET--------LAPSHKQ|WIIPFAVQLIPAGLLIIGALL|-
E. coli  AraE                      MISMYQLMVTLGIVLAFLSDT|AFSYSG----------N|WRAMLGVLALPAVLLIILVVF|-
E. coli  GalP                      MISMYQLMITIGILGAYLSDT|AFSYTG----------A|WRWMLGVIIIPAILLLIGVFF|-
E. coli  XylE                      LVSFNQFAIFGGLLVYCVNY|FIARSGDAS-----WLNTDG|WRYMFASECIPALFLMLLYT|-
Z. mobilis glf                     MVSGQQMAIVTGALTGYIFTW|LLAHFGSID-----WVNASG|WCWSPASEGLIGIAFLLLLLT|-

K. pneumoniae                      QSGNQVAIMVAAAMGFALNA|VLEPSAISDW---------G|WRIPFLFGVLIVPFIFILRRK|-
E. coli  citA                      QSASQQVAIVVAALIGYGLNV|TLGHDEISEW---------G|WRIPFFIGCMIIPLIFVLRRS|-
E. coli  kgtP                      QYVTLIGGQLLALLVVVVLQH|TMEDAALREW---------G|WRIPFALGAVLAVVALWLRRQ|-
S. hygroscopicus                   QYVASSVGHILAGLSTLAASQ|ISGDGMDR-W---------G|WRLPFIWGAVICLAGLALRST|-
bap3
Tn1696 cmlA                        IYGILGGMLAMVPAVGPLIGA|LVDMWL--------------G|WRAIFAFLGLGMIAASAAAWR|F
E. coli pBR322                     HFGLMSACFGVGMVAGPVAGG|LLGAI---------------S|LHAPFLAAAVLNGLNLLLGCF|L-
E. coli  RP1                       HFGFMSACFGFGMVAGPVLGG|LMGGF---------------S|PHAPFFAAAALNGLNFLTGCF|L-
E. coli  Tn10                      WFGWLGASFGLGLIAGPIIGG|FAGEI---------------S|PHSPFFIAALLNIVTFLVVMF|W-
S. aureus NorA                     NFGYMSAIINGGFIIGPGIGG|FMAEV---------------S|HRMPFYFAGALGIIAFIMSIV|L
B. subtilis Bmr                    ALGYMSAFIGTGFIIGPGIGG|FLAEV---------------H|SRLPFFFAAAFALLAAILSIL|T

Bacillus pTHT15                    AFGLIGSIVAMGEGVGPAIGG|MIAHYI--------------H|WSYLLLIPMITIITVPFLMKL|L
Bacillus pNS1981                   AFGLIGSIVAMGEGVGPAIGG|MIAHYI--------------H|WSYLLLIPMITIITVPFLMKL|L
Bacillus BS908                     AFGLIGSLVAMGEGVGPAIGG|MVAHYI--------------H|WSYLLLIPTATIITVPFLIKL|L
S. aureus pT181                    AFGFIGSIVALGEGLGPSIGG|IIAHYI--------------H|WSYLLLLPMITIVTIPFLIKV|M
Streptomyces mmr                   MLGLWSAIVATGSGLGPTVGG|LMVSAF--------------G|WESIFLLNLPIGAIGMAMTYR|Y
S. aureus qacA                     ALAVWSIASSIGAVFGPIIGG|ALLEQF--------------S|WHSAFLINVPFAIIAVVAGLF|L
S. coelicoi' ActII                 AFGAFGPAIGLGAVLGPIVAG|FLVDADLFGT----------G|WRSVFLINLPIGVAVIVGAVL|L
Yeast    ATR1                      VISFVGAMAPIGATLGCLFAG|LIGTEDFK------------Q|WPWAFYAYSIAAFINFVLSIY|A

E. coli  LacY                      MFGCVGWALCASIVGIMFTIN|-------------------NQ|FVFWLGSGCALILAVLLFFAK|-
K. pneumoniae                      VSGCVGWALCASITGILFSID|-------------------PN|ITFWIASGFALILGVLLWVSK|-
E. coli  RafB                      MFGCLGWALCATMAGILFNVD|-------------------PS|LVFWMGSGGALLLLLLLYLAR|-
E. coli  MelB                      ASLAGFVTAGVTLPFVNYVGG|GDRGFGFQ------------M|FTLVLIAFFIVSTIITLRNVH|-
```

```
CONSENSUS            PESPR L      E  AK  L  L G              L E   E              L

Human     GLUT1 207  CPESPRFLLINRNEENRAKSVLKKLRGTADVTH------DLQEMKEESRQMMREKKVTILELF--  263
Rabbit    GLUT1      CPESPRFLLINRNEENRAKSVLKKLRGNADVTR------DLQEMKEESRQMMREKKVTILELF--
Rat       GLUT1      CPESPRFLLINRNEENRAKSVLKKLRGTADVTR------DLQEMKEESRQMMREKKVTILELF--
Pig       GLUT1      CPESPRFLLINRNEENRAKSVLKKLRGTADVTR------DLQEMKEESRQMMREKKVTILELF--
Mouse     GLUT1      CPESPRFLLINRNEENRAKSVLKKLRGTADVTR------DLQEMKEESRQMMREKKVTILELF--
Human     GLUT2      CPESPRYLYIKLDEEVRAKQSLKRLRGYDDVTK------DINEMRKEREEASSEQKVSIIQLF--
Rat       GLUT2      CPESPRYLYLNLEEEVRAKKSLKRLRGTEDITK------DINEMRKEKEEASTEQKVSVIQLF--
Mouse     GLUT2      CPESPRYLYIKLEEEVRAKKSLKRLRGTEDVTK------DINEMKKEKEEASTEQKVSVIQLF--
Rat       GLUT7      PPESPPWLTIDIDDEGNAKRILQSLQGYDEVSH------ELQEIKDESQKEEAETFLTLIELL--
Human     GLUT3      CPESPRFLLINRKEEENAKQILQRLWGTQDVSQ------DIQEMKTESARMSQEKQVTVLELF--
Mouse     GLUT3      CPESPRFLLINKKEEDQATEILAALVGHLGRGQ------EIQEMKDESVRMSQEKQVTVLELF--
Human     GLUT4      CPESPRYLYIIQNLEGPARKSLKRLTGWADVSG------VLAELKLEKRKLERERPLSLLQLL--
Rat       GLUT4      CPESPRYLYIIRNLEGPARKSLKRLTGWADVSD------ALAELKLEKRKLERERPLSLLQLL--
Mouse     GLUT4      CPESPRYLYIIRNLEGPARKSLKPLTGWADVSD------ALAELKLEKRKLERERPMSLLQLL--
Human     GLUT5      PPESPRYLLIQRTTEAAAKRALQTLRGWDSVDR------EVAEIRQEDEAEKAAGFISYLKLF--
Rat       SV2        QPESPRFFLENGKHDEAWMVLKQVHDTNMRAKGHPERVFSVTHIKTIHQEDELIEIQSDTGTWYQ
Yeast     SNF3       IPESPRYYVLKDKLDEAAK-SLSFLRGVPVHDSGLLE--ELVEIKATYDYEASFGSSNFIDCFIS
Yeast     GAL2       VPESPRYLCEVNKVE-DAKRSIAKSNKVSPEDPAVQA--ELDLIMAGIEAEKLAGNASWGELFS-
Yeast     HXT1       VPESPRYLVEAGRID-EARASLAKVNKCPPDHPYIQY--ELETIEASVEEMRAAGTASWGELFT-
Yeast     HXT2       VPESPRFLVEKGRYE-DAKRSLAKSNKVTIEDPSIVA--EMDTIMANVETERLAGNASWGELFS-
Yeast     RAG1       VPESARFLVETDQIE-EARKSLAKTNKVSIDDPVVKY--ELLKIQSSIELEKAAGNASWGELIT-
Yeast     MAL61      APESPWWLVKKGRID-QARRSLERILSGKGPEKELLVSMELDKIKTTIEKEQKMSDEGTYWDCV-
Yeast     LAC12      IPESPRWLVGVGREE-EAREFIIKYHLNGDRTHPLLD-MEMAEIIESFHGTDLSNPLEMLDVRSL
Yeast     IRT1       LPDTPRYYVMKGDLA-RATEVLKRSYTDTSEEIIERKVEELVTLNQSIPGKNVPEKVWNTIKEL-
Yeast     IRT2       LPDTPRYYVMKGDLK-RAKMVLKRSYVNTEDEIIDQKVEELSSLNQSIPGKNPITKFWNMVKEL-
Chlorella HUP1       LPESPNFLVEKGKTE-KGREVLQKLRGTSEVDAEFADIVAAVEIARPITMRQSWASLF-------
Arabidopsis STP1     LPDTPNSMIERGQHE-EAKTKLRRIRGVDDVSQEFDDLVAASKESQSIEHPWRNL----------
Synechocystis glcP   IPESPRYLVAQGQGE-RAAAILWKVEGGDVPSR----IEEIQAT-VSLDHKPRFSDLL-------
Leishmania Pro-1     TPES-RAKFD-GGEEGRAELNPSEYGYVEMI-----------------------------------
Neurospora qa-y      IPESPRWLYANGKRE-EAMKVLCWIRNLEPTDRYIVQEVSFIDADLERYTRQVGNGFWKPFLSL-
Aspergillus qutD     IPESPRWLFLRGNRE-KGIETLAWIRNLPADHIYMVEEINMIEQSLEQQRVKIGLGFWKPFKAA-
E. coli  AraE        LPNSPRWLAEKGRHI-EAEEVLRMLRDTSEKARE-----ELNEIRESLKLKQGG---WALF----
E. coli  GalP        LPISPRWFAAKRRFV-DAERVLLRLRDTSAEAKR-----ELDEIRESLQVKQSG---WALF----
E. coli  XylE        VPESPRWLMSRGKQE-QAEGILRKIMGNTLATQ------AVQEIKHSLDHGRKTGG---------
Z. mobilis glf       APDTPHWLVMKGRHS-EAFKILARLEPQADPNL------TIQKIKAGFDKAMDKSS---------

K. pneumoniae        LMETLAFTARRHHLAMRQ-----------------------------------------------
E. coli  citA        LQETEAFLQRKHRPDTRE-----------------------------------------------
E. coli kgtP         LDETSQQETRALKEAGSL-----------------------------------------------
S. hygroscopicus     AEETLPIGTEGGRKMTRTG----------------------------------------------
Sap3
Tn1696 cmlA          WPETRVQRVAGL-----------------------------------------------------
E. coli pBR322       MQESHKG-ERRPMPL-DAFNPVSSFRWARGMTIVAAL----------------------------
E. coli  RP1         LPESHKG-ERRPLRR-EALNPLGFVRWARGMTVVAAL----------------------------
E. coli Tn10         FPETKNTRDNTDTEV--GVETQSNSVYITLFKTMPIL----------------------------
S. aureus NorA       IHDPKKSTTSGFQKL----EPQLLTKINWKVFIT-------------------------------
B. subtilis Bmr      LPEPERNPENQEIKG----QKTGFKRIFAPMYFI-------------------------------

Bacillus pTHT15      KKEVRIKGHF----DIKGI-ILMSVGIVFFMLFTTSYSISFLIVSVLSFLIFVKHIRKVTDPFVD
Bacillus pNS1981     KKEVRIKGHF----DIKGI-ILMSVGIVFFMLFTTSYSISFLIVSVLSFLIFVKHIRKVTDPFVD
Bacillus BS908       KKEERIRGHI----DMAGI-ILMSAGIVFFMLFTTSYRFSFLIISILAFFIFVQHIRKAQDPFVD
S. aureus pT181      VPGKSTKNTL----DIVGI-VLMSISIICFMLFTTNYNWTFLILFTIFFVIFIKHISRVSNPFIN
Streptomyces mmr IAATESRATRLA-----VPGHLLWIVALAAVSFALIEGPQLGWTAGPVLTAYAVAVTAAALLAL---
S. aureus qacA       LPESKLSKEKGHSWDIPSTILSIAGMIGLVWSIKEFSKEGLADIIPWVVIVLAITMIVIFVK--
S. coelicol' ActII   LPEGKAPVRPKF--DVVGMALVTSGLTLLIFPLVQGRERGWPAWAFVLMLAGAAVLVGFVAHELR
Yeast     ATR1       IPSTIPTNIHHFSMDWIGSVLGVIGLILLNFVWNQAPISGWNQAYIIVILIISVIFLVVFIIYEI

E. coli  LacY        -TDAPSSATVANAVGANHSAFSLKLALELFRQPKLWFLSLY------------------------
K. pneumoniae        -PESSNSAEVIDALGANRQAFSMRTAAELFRMPRFWGFIIY------------------------
E. coli  RafB        -PSTSQTAMVMNALGANSSLISTRMVFSLFRMRQMWMFVLY------------------------
E. coli  MelB        EVFSSDNQPSAEGSHLTLKAIVALIYKNDQL----------------------------------
```

Helix 7

```
CONSENSUS                                          R          LQ  QQLSGIN  FY YST IF  AG
                                                                                      306
Human    GLUT1 264 ----------------------RSPAYRQP|ILIAVVLQLSQQLSGINAVFY|YSTSIFEKAGVQQP----
Rabbit   GLUT1     ----------------------RSPAYRQP|ILSAVVLQLSQQLSGINAVFY|YSTSIFEKAGVQQP----
Rat      GLUT1     ----------------------RSPAYRQP|ILIAVVLQLSQQLSGINAVFY|YSTSIFEKAGVQQP----
Pig      GLUT1     ----------------------RSAAYRQP|ILIAVVLQLSQQLSGINAVFY|YSTSIFEKAGVQQP----
Mouse    GLUT1     ----------------------RSPAYRQP|ILIAVVLQLSQQLSGINAVFY|YSTSIFEKAGVQQP----
Human    GLUT2     ----------------------TNSSYRQP|ILVALMLHVAQQFSGINGIFY|YSTSIFQTAGISKP----
Rat      GLUT2     ----------------------TDPNYRQP|IVVALMLHLAQQFSGINGIFY|YSTSIFQTAGISQP----
Mouse    GLUT2     ----------------------TDANYRQP|ILVALMLHMAQQFSGINGIFY|YSTSIFQTAGISQP----
Rat      GLUT7     ----------------------RSRDTRWP|SLLIAFLMQSQQTSGVNGIFY|YHQHIYKQAGAQDP----
Human    GLUT3     ----------------------RVSSYRQP|IIISIVLQLSQQLSGINAVFY|YSTSIFKDAGVQEP----
Mouse    GLUT3     ----------------------RSPNYVQP|LLISIVLQLSQQLSGINAVFY|YSTSIFKDAGVQEP----
Human    GLUT4     ----------------------GSRTHRQP|LIIAVVLQLSQQLSGINAVFY|YSTSIFETAGVGQP----
Rat      GLUT4     ----------------------GSRTHRQP|LIIAVVLQLSQQLSGINAVFY|YSTSIFELAGVEQP----
Mouse    GLUT4     ----------------------GSRTHRQP|LIIAVVLQLSQQLSGINAVFY|YSTSIFECAGVGQP----
Human    GLUT5     ----------------------RMRSLRWQ|LLSIIVLMGGQQLSGVNAIYY|YADQIYLSAGVPEEHV--
Rat      SV2       RWGVRALSLGGQVWGNFLSCFSPEYRRITL|MMMGVWFTMSFSYYGLTVNFF|DMIRHLQAVDYAARTKVF
(SV2 helix 7 - 8                PGERVEHVTFNFTLENQIHRGGQYFNDKFIGLRLKSVSFEDSLFEECYFEDVTSS
 linker)                        NTFFRNCTFINTVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPLDVTGTGEGAYMV
Yeast    SNF3      ---------------------SKSRPKQTLR|MFTGIALQAFQQFSGINFIFY|YGVNFYNKTGVSNS----
Yeast    GAL2      ----------------------TKTKVFQR|LLMGVFVQMFQQLTGNNYFFY|YGTVIFKSVGLDDS----
Yeast    HXT1      ----------------------GKPAMFQR|TMMGIMIQSLQQLTGNNYFFY|YGTIYFQAVGLSDS----
Yeast    HXT2      ----------------------NKGAILPR|VIMGIMIQSLQQLTGNNYFFY|YGTIIFNAVGMKDS----
Yeast    RAG1      ----------------------GKPSMFRR|TLMGIMIQSLQQLTGNNYFFY|YGTIIFQSVGMDDS----
Yeast    MAL61     ----------------------KDGINRR|RTRIACLCWLGQFSCGASLIG|YSTYFYEFAGVSTDTA--
Yeast    LAC12     ---------------------FRTRSDRYR|AMLVILMAWFGQFSGNNVCSY|YLPTMLRNVGMKSVSLN-
Yeast    IRT1      ---------------------HTVPSNLRA|LIIGCGLQAIQQFTGWNSLMY|FSGTIFETVGFKNS----
Yeast    IRT2      ---------------------HTVPSNFRA|LIIGCGLQAIQQFTGWNSLMY|FSGTIFETVGFKNS----
Chlorella HUP1     ----------------------TRRYMPQ|LLTSFVTQFFQQFTGINAIIF|YVPVIFSSLGSANSAA--
Arabidopsis STP1   ----------------------LRRKYRPH|LTMAVMIPFFQQLTGINVIMF|YAPVLFNTIGFTTDAS--
Synechocystis glcP ----------------------SRRGGLLPI|VWIGMGLSALQQFVGINVIFY|YSSVLWRSVGFTEEKS--
Leishmania Pro-1   ----------------------------PR|LLMGCVMAGTLQLTGINAVMN|YAPTIMGSLGLAP-----
Neurospora qa-y    ----------------------KQRKVQWR|FFLGGMLFFWQNGSGINAINY|YSPTVFRSIGITGTDTGF
Aspergillus qutD   ----------------------WTNKRILYR|LFLGSMLFLWQNGSGINAINY|YSPRVFKSIGVSGGNTSL
E. coli  AraE      ---------------------KINRNVRRA|VFLGMLLQAMQQFTGMNIIMY|YAPRIFKMAGFTTTEQQ-
E. coli  GalP      ---------------------KENSNFRRA|VFLGVLLQVMQQFTGMNVIMY|YAPKIFEIAGYTNTTEQ-
E. coli  XylE      ----------------------RLLMFGVGV|IVIGVMLSIFQQFVGINVVLY|YAPEVFKTLGASTDIA--
Z. mobilis glf     ---------------------AGLFAFGITV|VFAGVSVAAFQQLVGINAVLY|YAPQMFQNLGFGADTA--

K. pneumoniae cit'---------VFATLLANWQ|VVIAGMMMVAMTTTAFYLITV|YAPTFGKKVLMLSASDS-
E. coli  citA      ---------------------IFTTIAKNWR|IITAGTLLVAMTTTTFYFITV|YTPTYGRTVLNLSARDS-
E. coli kgtP       ----------------------KGLWRNRR|AFIMVLGFTAAGSLCFYTFTT|YMQKYLVNTAGMHANVA-
S. hygroscopicus   --------------------AFAALRSHPR|QTLLVVGLTIGGNVAFYTWTT|YLPTYATVSTGADKDSA-
bap3
Tn1696 cmlA        --------------------QWSQLLLPVK|CLNFWLYTLCYAAGMGSFFVF|FSIAPGLMMGR-------
E. coli pBR322     ------------------------------|MTVFFIMQLVGQVPAALWVIF|GEDRFRWSATMIG-----
E. coli  RP1       ------------------------------|MAVFFIMQLVGQVPAALWVIF|GEDRFHWTATTIG-----
E. coli  Tn10      ------------------------------|LIIYFSAQLIGQIPATVWVLF|TENRFGWNSMMVG-----
S. aureus NorA     -----------------------------P|VILTLVLSFGLSAFETLYSLY|TADKVNYSPKDIS-----
B. subtilis Bmr    -----------------------------A|FLIILISSFGLASFESLFALF|VDHKFGFTASDIA-----

Bacillus pTHT15    --------------------PGLGKNIPFM|IGVLCGGIIFGTVAGFVSMVP|YMMKDVHQLSTAEIG---
Bacillus pNS1981   --------------------PGLGKNIPFM|IGVLCGGIIFGTVAGFVSMVP|YMMKDVHQLSTAEIG---
Bacillus BS908     --------------------PELGKNVFFV|IGTLCGGLIFGTVAGFVSMVP|YMMKDVHHLSTAAIG---
S. aureus pT181    --------------------PKLGKNIPFM|LGLFSGGLIFSIVAGFISMVP|YMMKTIYHVNVATIG---
Streptomyces mmr   --------REHRVTNPVMPWQLFRGPGFTG|ANLVGFLFNFALFGSTFMLGL|YFQHARGATPFQA-----
S. aureus qacA     --------RNLSSSDPMLDVRLFKKRSFSA|GTIAAFMTMFAMASVLLLASQ|WLQVVEELSPFKA-----
S. coelicol' ActII --------QERRGGATLIELSLLRRSRYAA|GLAVALVFFTGVSGMSLLLAL|HLQIGLGFSPTRAA----
Yeast    ATR1      --------RFAKTPLLPRAVIKDRH-----|MIQIMLALFFGWGSFGIFTFY|YFQFQLNIRQY-------

E. coli  LacY      ------------------------------V|IGVSCTYDVFDQQFANFFTSF|FATGEQGTRVF-------
K. pneumoniae      ------------------------------V|VGVASVYDVFDQQFANFFKGF|FSSPQRGTEVF-------
E. coli  RafB      -----------------------------T|IGVACVYDVFDQQFAIFFRSF|FDTPQAGIKAF-------
E. coli  MelB      ----------------------------SCL|LGMALAYNVASNIITGFAIYY|FSYVIGDADLF-------
```

```
                                   Helix 8                                  Helix 9
CONSENSUS                          TI  G VN  FT VS     VE  GRR L   L G  GM  C     L
Human   GLUT1 307 |VYATIGSGIVNTAFTVSLFV|VERAGRRTLH|LIGLAGMAGCAILMTIALALL|EQLPWM------ 364
Rabbit    GLUT1   |VYATIGSGIVNTAFTVSLFV|VERAGRRTLH|LIGLAGMAACAVLMTIALALL|EQLPWM------
Rat       GLUT1   |VYATIGSGIVNTAFTVSLFV|VERAGRRTLH|LIGLAGMAGCAVLMTIALALL|EQLPWM------
Pig       GLUT1   |VYATIGSGIVNTAFTVSLFV|VERAGRRTLH|LIGLAGMAGCAVLMTIALALL|EQLPWM------
Mouse     GLUT1   |VYATIGSGIVNTAFTVSLFV|VERAGRRTLH|LIGLAGMAGCAVLMTIALALL|ERLPWM------
Human     GLUT2   |VYATIGVGAVNMIFTAVSVFL|VEKAGRRSLF|LIGMSGMFVCAIFMSVGLVLL|NKFSWM------
Rat       GLUT2   |VYATIGVGAINMIFTAVSVLL|VEKAGRRTLF|LAGMIGMFFCAVFMSLGLVLL|DKFTWM------
Mouse     GLUT2   |VYATIGVGAINMIFTAVSVLL|VEKAGRRTLF|LTGMIGMFFCTIFMSVGLVLL|DKFAWM------
Rat       GLUT7   |AYVTLGSGVNFLTTVSLIV|VERAGRRTLF|LAGMIGMFFCAVFMSLVLVLL|DKFTWM------
Human     GLUT3   |IYATIGAGVNTIFTVSLFL|VERAGRRTLH|MIGLGGMAFCSTLMTVSLLLK|DNYNGM------
Mouse     GLUT3   |IYATIGAGVNTIFTVSLFL|VERAGRRTLH|MIGLGGMAVCSVFMTISLLLK|DDYEAM------
Human     GLUT4   |AYATIGAGVNTVFTVSVLL|VERAGRRTLH|LLGLAGMCGCAILMTVALLLL|ERVPAM------
Rat       GLUT4   |AYATIGAGVNTVFTVSVLL|VERAGCRTLH|LLGLAGMCGCAILMTVALLLL|ERVPSM------
Mouse     GLUT4   |AYATIGAGVNTVFTVSVLL|VERAGRRTLH|LLGLAGMCGCAILMTVALLLL|ERVPAM------
Human     GLUT5   |QYVTAGTGAVNVVMTFCAVFV|VELLGRRLLL|LLGFSICLIACCVLTAALALQ|DTVSWM------
Rat       SV2     |YFVSFLGTLAVLPGNIVSALL|MDKIGRLRML|AGSSVLSCVSCFFLSFGNSES|A----------
Yeast     SNF3    |YLVSFITYAVNVVFNVFGLFF|VEFFGRRFVL|VVGGVIMTIANFIVAIVGCSL|KTVAAA------
Yeast     GAL2    |FETSIVIGVVNFASTFFSLWT|VENLGRRKCL|LLGAATMMACMVIYASVGVTR|LYPHGKSQ----
Yeast     HXT1    |FETSIVFGVVNFFSTCCSLYT|VDRFGRRNCL|MWGAVGMVCCYVVYASVGVTR|LWPNGQDQ----
Yeast     HXT2    |FQTSIVLGIVNFASTFVALYT|VDKFGRRKCL|LGGSASMAICFVIFSTVGVTS|LYPNGKDQ----
Yeast     RAG1    |FETSIVLGIVNFASTFFALYT|VDHFGRRNCL|LYGCVGMVACYVVYASVGVTR|LWPDGPDHP---
Yeast     MAL61   |FTFSIIQYCLGIAATFVSWWA|SKYCGRFDLY|AFGLAFQAIMFFIIGGLGCSD|THGAKM------
Yeast     LAC12   |VLMNGVYSIVTWISSICGAFF|IDKTGRRSGF|-LGSISGAALALTGLSICTAR|YEKTKK------
Yeast     IRT1    |SAVSIIVSGTNFIFTLVAFFS|IDKIGRRTIL|LIGLPGMTMALVVCSIAFHFL|GIKFDGAVAVVV
Yeast     IRT2    |SAVSIIVSGTNFVFTLIAFFC|IDKIGRRYIL|LIGLPGMTVALVICAIAFHFL|GIKFNGADAVVA
Chlorella HUP1    |LLNTVVGAVNVGPTLIAVMF|SDKFGRRFLL|TEGGIQCCLAMLTTGVVLAIE|FAKYGTD-----
Arabidopsis STP1  |LMGAVVTGSVNVGATLVSIYG|VDRWGRRFLF|LEGTTGMLICCAVVAACIGAK|FGVDGTP-----
Synechocystis glcP |LLITVITGFINILTTLVAIAF|VDKFGRRPLL|LMGSIGMTITLGILSVVFGGA|TVVNGQP-----
Leishmania Prc-1  |LVGNFVVMLWNFVTTLASIPL|SYVFTMRHVF|LFGSIFTFCMCLFMCGIPVY-|PGVSKKLE----
Neurospora qa-y   |LTTGIFGVVKMVLTIIWLLWL|VDLVGRRFIL|FIGAAGGSLCMWFIGAYIKIA|DPGSNKAE----
Aspergillus gutD  |LTTGIFGVVTAVITFVWLLYL|IDHFGRRNLL|LVGAAGGSVCLWIVGGYIKIA|KPENNPEG----
E. coli  AraE     |MIATLVVGLTFMFATFIAVFT|VDKAGRRPAL|KIGFSVMALGTLVLGYCLMQF|DNGTASS-----
E. coli  GlcP     |MWGTVIVGLTNVLATFIAIGL|VDRWGRRPTL|TLGFLVMAAGMGVLGTMMHIG|IHSPSA------
E. coli  XylE     |LLQTIIVGVINLTFTVLAIMT|VDKFGRRPLQ|IIGALGMAIGMFSLGTAFYTQ|AP----------
Z. mobilis glf    |LLQTISIGVVNFIFTMIASRV|VDRFGRRPLL|IWGALGMAAMMAVLGCCFWFK|VG----------

K. pneumoniae     |LLVTIAVAISNFFWLPVGGAL|SDRFGRRSV-|LIAMTLLALATAWPALTMLAN|AP----------
E. coli  citA     |LVVTMLVGLSNFIWLPIGGAI|SDRIGRRPV-|LMGITLLALVTTLPVMNWLTA|AP----------
E. coli  kgtP     |SGIMTAALFVFMLIQPLIGAL|SDKIGRRTS-|MLCFGSLAAIFTVPILSALQN|VS----------
S. hygroscopicus  |VLAGTVSLIFFGLIQPLGGLL|LRIGGRAM-|MIGFGVAAAVLTVPLLTAMTG|WF----------
bam3
Tn1696 cmlA       |QGVSQLGFSLLFATVAIAMVF|TARFMGRVIP|KWGSFSVLRMGMGGLIAGAVL|LAIT--------
E. coli pBR322    |LSLAVFGILHALAQAFVIGPA|TKRFGEKQAI|IAGMAADALGYVLLAFATRGW|------------
E. coli  RP1      |ISLAAFGILHSLAQAMITGPV|AARLGERRAL|MLGMIADGTGYILLAFATRGW|------------
E. coli  Tn10     |FSLAGLGLLHSVFQAFVAGRI|ATKWGEKTAV|LLFFIADSSAFAFLAFISEGW|------------
S. aureus NorA    |IAITGGGIFGALFQIYFFDKF|MKYFSELTFI|AWSLLYSVVVLILLVFANGYW|------------
B. subtilis Bmr   |IMITGGAIVGAITQVVLFDRF|TRWFGEIHLI|RYSLILSTSLVFLLTTVHSYV|------------

Bacillus pTHT15   |SVIIFPGTMSVIIPGYIGGIL|VDRRGPLYVL|NIGVTFLSVSFLTASFLLETT|------------
Bacillus pNS1981  |SVIIFPGTMSVIIPGYIGGIL|VDRRGPLYVL|NIGVTFLSVSFLTASFLLETT|------------
Bacillus BS908    |SGIIFPGTMSVIIPGYIGGLL|VDRRGSLYVL|TIGSALLSSGFLIAAFFIDAA|------------
S. aureus pT181   |NSVIFPGTMSVIVPGYFGGFL|VDRRGSLFVF|ILGSLSISISFLTIAFFVEFS|------------
Streptomyces mmr  |GLELLPMTIFFPVANIVYARI|SARFSNGTLL|TAFLLLAGAASLSMVTITAST|PY----------
S. aureus qacA    |GLYLLPMAIGDMVPAPIAFGL|AARFGPKIVL|PSGIGTAAIGMFIMYFFGHPL|SYST--------
S. coelicol' ActII |LTMTPWSVFLVVGAILTGAVL|GSKFGRKAL-|HGGLVVIALGVLIMLLTIGDQ|AGGLT-------
Yeast     ATR1    |TALWAGGTYFMFLIWGIIAAL|LVGFTIKNVS|PSVFLFFSMVAFNVGSIMASV|TPVHETYF----

E. coli  LacY     |GYVTTMGELLNASIMFFAPLI|INRIGGKNAL|LLAGTIMSVRIIGSSFATSAL|E-----------
K. pneumoniae     |GFVTTGGELLNALIMFCAPAI|INRIGAKNAL|LIAGLIMSVRILGSSFATSAV|E-----------
E. coli  RafB     |GFATTAGEICNAIIMFCTPWI|INRIGAKNTL|LVAGGIMTIRITGSAFATTMT|E-----------
E. coli  MelB     |PYYLSYAGAANLVTLVFFPRL|VKSLSRRILW|AGASILPVLSCGVLLLMALMS|YHNVVLI-----
```

```
                                                  Helix 10                                  Helix 11
CONSENSUS                                         IV I  FVAFF  G GPIPW    E F Q  R   A A A   NW  NFI G  F
Human   GLUT1      -------SYLS|IVAIFGFVAFFEVGPGPIPWF|IVAELFSQGPRP|AAIAVAGFSNWTSNFIVGMCF|422
Rabbit    GLUT1      -------SYLS|IVAIFGFVAFFEVGPGPIPWF|IVAELFSQGPRP|AAVAVAGFSNWTSNFIVGMCF
Rat       GLUT1      -------SYLS|IVAIFGFVAFFEVGPGPIPWF|IVAELFSQGPRP|AAVAVAGFSNWTSNFIVGMCF
Pig       GLUT1      -------SYLS|IVAIFGFVAFFEVGPGPIPWF|IVAELFSQGPRP|AAIAVAGFSNWTSNFIVGMCF
Mouse     GLUT1      -------SYLS|IVAIFGFVAFFEVGPGPIPWF|IVAELFSQGPRP|ARIAVAGFSNWTSNFIVGMCF
Human     GLUT2      -------SYVS|MIAIFLFVCFFEIGPGPIPWF|MVAEFFSQGPRP|AALAIAAFSNWTCNFIVALCF
Rat       GLUT2      -------SYVS|MTAIFLFVCFFEIGPGPIPWF|MVAEFFSQGPRP|TALALAAFSNWVCNFIALCF
Mouse     GLUT2      -------SYVS|MTAIFLFVCFFEIGPGPIPWF|MVAEFFSQGPRS|TALALAAFSNWVCNFVIALCF
Rat       GLUT7      -------SYVS|MTAIFLFVCFFEIGPIPIPFF|GVREWFTQIWRP|GAIVCVATLDWVPNFKKGICF
Human     GLUT3      -------SFVC|IGAILVFVAFFEIGPGPIPWF|IVAELFSQGPRP|AAMAVAGCSNWTSNFLVGLLF
Mouse     GLUT3      -------SFVC|IVAILIFVAFFEIGPGPIPWF|IVAELFSQGPRP|AALAVAGCSNWTSNFLVGMLF
Human     GLUT4      -------SYVS|IVAIFGFVAFFEIGPGPIPWF|IVAELFSQGPRP|AAMAVAGFSNWTSNFIIGMGF
Rat       GLUT4      -------SYVS|IVAIFGFVAFFEIGPGPIPWF|IVAELFSQGPRP|AAMAVAGFSNWTCNFIVGMGF
Mouse     GLUT4      -------SYVS|IVAIFGFVAFFEIGPGPIPWF|-VAELFSQGPRP|AAMAVAGFSNWTCNFIVGMGF
Human     GLUT5      -------PYIS|IVCVISYVIGHALGPSPIPAL|LITEIFLQSSRP|SAFMVGGSVHWLSNFTVGLIF
Rat       SV2        -----------|MIALLCLFGGVSIASWNALDV|LTVELYPSDKRT|TAFGFLNALCKLAAVLGISIF
Yeast     SNF3       --------KVM|IAFICLFIAAFSATWGGVVWV|ISAELYPLGVRS|KCTAICAAANWLVNFICALIT
Yeast     GAL2       -PSSKGAGNCM|IVFTCFYIFCYATTWAPVAWV|ITAESFPLRVKS|KCMALASASNWVWGFLIAFFT
Yeast     HXT1       -PSSKGAGNCM|IVFACFYIFGFATTWAPIAYV|VISECFPLRVKS|KCMSIASAANWIWGFLISFFT
Yeast     HXT2       -PSSKAAGNVM|IVFTCLFIFFFAISWAPIAYV|IVAESYPLRVKN|RAMAIAVGANWIWGFLIGFFT
Yeast     RAG1       DISSKGAGNCM|IVFACFYIFCFATTWAPIAYV|VISESYPLRVKG|KAMAIASASNWIWGFLIGFFT
Yeast     MAL61      ---------GS|GALLMVVAFFYNLGIAPVVFC|LVSEMPSSRLRT|KTIIIARNAYNVIQVVVTVLI
Yeast     LAC12      ----KSASNGA|IVFIYLFGGIFSFAFTPMQSM|YSTEVSTNLTRS|KAQLLNFVVSGVAQFVNQFAT
Yeast     IRT1       SSGFSSWGIVI|IVFIIVFAAFYALGIGTVPW-|QQSELFPQNVRG|IGTSYATATNWAGSLVIASTF
Yeast     IRT2       SDGFSSWGIVI|IVFIIVYAAFYALGIGTVPW-|QQSELFPQNVRG|VGTSYATATNWAGSLVIASTF
Chlorella HUP1       -PLPKAVASGI|LAVICIFISGFAWSWGPMGWL|IPSEIFTLETRP|AGTAVAVVGNFLFSFVIGQAF
Arabidopsis STP1     GELPKWYAIVV|VTFICIYVAGFAWSWGPLGWL|VPSEIFPLEIRS|AAQSITVSVNMIFTFIIAQIF
Synechocystis glcP   -TLTGAAGIIA|LVTANLYVFGFGFSWGPIVWV|LLGEMFNNKIRA|AALSVAAGVQWIAMFIISTTF
Leishmania Pro-1     ---AKN--GVA|ITGILLFILGFEVCVGPCYYV|LTQIMFPPSPRP|RGASFTQVAQFIFNIINVCY
Neurospora qa-y      DAKLTSGGIAA|IFFFYLWTAFYTPSWNGTPWV|INSEMFTQNTRS|LGQASAAANNWFWNFIISRFT
Aspergillus qutD     -TQLDSGGIAA|IFFFYLWTAFYTPSWNGTPWV|INSEMFDPTVRS|LAQAAAASNNWLWNFLISRFT
E. coli  AraE        -----GLSWLS|VGMTMMCIAGYAMSAAPVVWI|LCSEIQPLKGRD|FGITCSTTTNWVSNMIIGATF
E. coli  GalP        -------QYFA|IAMLLMFIVGFAMSAGPLIWV|LCSEIQPLKGRD|FGITCSTATNWIANMIVGATF
E. coli  XylE        -------GIVA|LLSMLFYVAFAMSWGPVCWV|LLSEIFPNAIRG|FALAIAVAAQWLANYFVSWTF
Z. mobilis glf       -------GVLP|LASVLLYIAFGMSWGPVCWV|VLSEMFPSSIKG|AAMPIAVTGQWLANILVNFLF

K. pneumoniae SFLM|MLSVLLWLSFIYGMYNGAMIP|ALTEIMPAEVRV|AGFSLAYSLATAVFGGFTPVI|
E. coli  cit         -------DFTR|MTLVLLWFSFFFGMYNGAMVA|ALTEVMPVYVRT|VGFSLAFSLATAIFGGLTPAI
E. coli kgtP         -------SPYA|AFGLVMCALLIVSFYTSISGI|LKAEMFPAQVRA|LGVGLSYAVANAIFGGSAEYV
S. hygroscopicus     ---------WS|VLAVQCAGMLVLTAYTSVSGA|INAELFPQEIRG|RGIGLPYAASVALFGGTAFYV

Tn1696 cmlA          ---EIWALQSV|LGFIAPMWLVGIGVATAVSVA|PNGALRGFDHVA|GTVTAVYFCLGGVLLGSIGTL
E. coli pBR322       ---------MA|FPIMILLASGGIGMPALQAML|S-RQVDDDHQGQ|LQGSLAALTSLTSITGPLIVT
E. coli  RP1         ---------MA|FPIMVLLASGGIGMPALQAML|S-RQVDEERQGQ|LQGSLAALTSLTSIVGPLLFT
E. coli  Tn10        ---------LD|FPVLILLAGGGIALPALQGVM|S-IQTKSHEQGA|LQGLLVSLTNATGVIGPLLFT
S. aureus NorA       --------SIM|LISFVVFIGFDMIRPAITNYF|S--NIAGERQGF|AGGLNSTFTSMGNFIGPLIAG
B. subtilis Bmr      --------AIL|LVTVTVFVGFDLMRPAVTTYL|S--KIAGNEQGF|AGGMNSMFTSIGNVFGPIIGG

Bacillus pTHT15      -------SWFM|TIIIVFVLGGLLFTKTVISTI|VSSSLKQQEAGA|GMSLLNFTSFLSEGTGIAIVG
Bacillus pNS1981     -------SWFM|TIIIVFVLGGLSFTKTVISTI|VSSSLKQQEAGA|GMSLLNFTSFLSEGTGIAIVG
Bacillus BS908       -------PWIM|TIIVIFVFGGLSFTKTVISTV|VSSSLKEKEAGA|GMSLLNFTSFLSEGTGIAIVG
S. aureus pT181      -------MWLT|TFMFIFVMGELSFTKTVISKI|VSSSLSEEEVAS|GMSLLNFTSFLSEGTGIAIVG
Streptomyces mmr     -------WVVA|VAVGVANIGAGIISPGMTAAL|VDAAGPENANV-|AGSVLNANRQIGSLVGIAAMG
S. aureus qacA       -----------|MALALILVGAGMASLAVASAL|IMLETPTSKAGN|AAAVEESMYDLGNVFGVAVLG
S. coelicol' ActII   -------SWEL|VPGIAVAGLGMGIMIGLLFDI|ALADVDRQEAGT|ASGVLTAVQQLGFTVGVAVLG
Yeast    ATR1        --------RTQ|LGTMIILSFGMDLSFPASSII|FSDNLPMEYQGM|AGSLVNTVVNYSMSLCLGMGA

E. coli  LacY        -----------|-VVILKTLHMFEVPFLLVGCF|-KYITSQFEVRF|SATIYLVCFCFFKQLAMIFMS
K. pneumoniae        -----------|-VIILKMLHMFEIPFLLVGTF|-KYISSAFKGKL|SATLFLIGFNLSKQLSSVVLS
E. coli  RafB        -----------|-VVILKMLHALEVPFLLVGAF|-KYITGVFDTRL|SATVYLIGFCFSKQLAAILLS
E. coli  MelB        VIAGILLNVGT|ALFWVLQVIMVADIVDYGEYK|LHVRCESIAYSV|QTMVVKGGSAFAAFFIAVVLG
```

Helix 12

```
CONSENSUS                                  G Y F  FA         F  F   VPETKG T EEI   F
                                           *                *       *
Human     GLUT1 423 QYVEQ-----------LC  GPYVFIIFTVLLVLFFIFTYF  KVPETKGRTFDEIASGF   467
Rabbit    GLUT1     QYVEQ-----------LC  GPYVFIIFTVLLVLFFIFTYF  KVPETKGRTFDEIASGF
Rat       GLUT1     QYVEQ-----------LC  GPYVFIIFTVLLVLFFIFTYF  KVPETKGRTFDEIASGF
Pig       GLUT1     QYVEQ-----------LC  GPYVFIIFTVLLVLFFIFTYF  KVPETKGRTFDEIASGF
Mouse     GLUT1     QYVEQ-----------LC  GPYVFIIFTVLLVLFFIFTYF  KVPETKGRTFDEIAFGF
Human     GLUT2     QYIAD-----------FC  GPYVFFLFAGVLLAFTLFTYF  KVPETKGKSFEEIAAEF
Rat       GLUT2     QYIAD-----------FL  GPYVFFLFAGVVLVFTLFTYF  KVPETKGKSFDEIAAEF
Mouse     GLUT2     QYIAD-----------FL  GPYVFFLFAGVVLVFTLFTYF  KVPETKGKSFEEIAAEF
Rat       GLUT7     QSLRD-----------FK  GPYHFWAFHGVVIVWYGNYWF  KVPETKGKSFDEIAAEF
Human     GLUT3     PSAAH-----------YL  GAYVFIIFTGFLITFLAFTFF  KVPETRGRTFEDITRAF
Mouse     GLUT3     PSAAA-----------YL  GAYVFIIFAAFLIFFLIFTFF  KVPETKGRTFEDIARAF
Human     GLUT4     QYVAE-----------AM  GPYVFLLFAVLLLGFFIFTFL  RVPETRGRTFDQISAAF
Rat       GLUT4     QYVAD-----------AM  GPYVFLLFAVLLLGFFIFTFL  RVPETRGRTFDQISATF
Mouse     GLUT4     QYVAD-----------RM  GPYVFLLFAVLLLGFFIFTFL  KVPETRGRTFDQISAAF
Human     GLUT5     PFIQE-----------GL  GPYSFIVFAVICLLTTIYIFL  IVPETKAKTFEEINQIF
Rat       SV2       TSFV------------GI  TKAAPILFASAALALGSSLAL  KLPETRGQVLQ         742
Yeast     SNF3      PYIVDTGSHTS-----SL  GAKIFFIWGSLNAMGVIVVYL  TVPETKGLTLEEIDELY
Yeast     GAL2      PFITS-----------AI  NFYYGYVFMGCLVAMFFYVFF  FVPETKGLSLEEIQELW
Yeast     HXT1      PFITG-----------AI  NFYYGYVFMGCMVFAYFYVFF  FVPETKGLSLEEVNDMY
Yeast     HXT2      PFITS-----------AI  GFSYGYVFMGCLVFSFFYVFF  FVPETKGLTLEEVNEMY
Yeast     RAG1      PFITS-----------AI  HFYYGYVFMGCMVFAFFYVFF  FVPETKGLTLEEVNEMY
Yeast     MAL61     MYQLNSEKW-------NW  GAKSGFFWGGFCLATLAWAVV  DLPETAGRTFDEINELF
Yeast     LAC12     PKAMK-----------NI  KYWFYVFYVFFDIFEFIVIYF  FFVETKGRSLEEIEVVF
Yeast     IRT1      LTMLQN----------IT  PAGTFAFFAGLSCLSTIFCIF  CYPELSGLELEEVQTIL
Yeast     IRT2      LTMLQN----------IT  PTGTFSFFAGVACLSTIFCIF  CYPELSGLELEEVQTIL
Chlorella HUP1      VSMLC-----------AM  EYGVFLFFAGWLVIMVLCAIF  LLPETKGVPIEEVQALY
Arabidopsis STP1    LTMLC-----------HL  KFGLFLVFAFFVVVMSIFVYI  FLPETKGIPIEEMGQVW
Synechocystis glcP  PPLLDT----------VG  LGPAYGLYAISAAISIFFIWF  DVPETKGRTLEQM        468
Leishmania Pro-1    PIATESISGGPSGNQDKG  QAVAFIFFGGLGLICFVIQVF  FLHPWDEERDGKKVVAP
Neurospora qa-y     PQMFIK----------M  EYGVYFFFAGLMLLSIVFIYF  FLPVTKSIPLEAMDRLF
Aspergillus qutD    PQMFTS----------M  GYGVYFFFAGLMILSIVFVFF  LIPETKGVPIESMETLF
E. coli  AraE       LTLLDS----------IG  AAGIFWLYIALNIAFVGITFW  LIPETKNVTLEHIERKL
E. coli  GalP       LPMLNT----------LG  NANTFWVYAALNVLFILLTLW  LVPETKHVSLEHIERNL
E. coli  XylE       PMMDKNSWLVAH----FH  NGFSYWIYGCMGVLAALFMWK  FVPETKGRTLEELEALW
Z. mobilis glf      KVADGSPALNQT----FN  HGFSYLVFAALSILGGLIVAR  FVPETKGRSLDEIEEMW

K. pneumoniae       STALIDYTGIe---KASP  GYWMSFAAICGLLATCYLYRR  S---AVALQTAR        444
E. coli  citA       STALVQLTGD----KSSP  GWWLMCAALCGLAATTMLFAR  L---SSGRQTVENKL      431
E. coli kgtP        ALSLKSI------GMETA  FFWYVTIMAVVAFLVSLMLHR  K---GKGMRL           432
S. hygroscopicus    GTWLKSM------GLNDF  FPWYVAVLCLLTALTAIGLPR  FVPDTCDGPASPAQHLP
bap3
Tn1696 cmlA         IISLLPRNT-------AW  PVVVYCLTLATVVLGLSCVSR  VKGSRGQGEHDVVALQS
E. coli pBR322      AIYAASAS--------TW  NGLAWIVGAALYLVCLPALRR  GAWSRATST           396
E. coli  RP1        AIYAASIT--------TW  NGWAWIAGAALYLLCLPALRR  GLWSGAGQRADR         399?
E. coli  Tn10       VIYNHSLP--------IW  DGWIWIIGLAFYCIIILLSMT  FMLTPQAQGSKQETSA
S. aureus NorA      ALFDVHIE--------AP  IYMAIGVSLAGVVIVLIEKQH  RAKLKEQNM
B. subtilis Bmr     MLFDIDVN--------YP  FYFATVTLAIGIALTIAWKAP  -AHLKAST

Bacillus pTHT15     GLLSIPLLDQRLLPM--E  VDQSTYLYSNLLLLFSGIIVI  SWLVTLNVYKHSQRDF
Bacillus pNS1981    GLLSIPLLDQRLLPM--E  VDQSTYLYSNLLLLFSGIIVI  SWLVTLNVYKHSQRDF
Bacillus BS908      GLLSIGFLDHRLLPI--D  VDHSTYLYSNMLILFAGIIVI  CWLVILNVYKRSRRHG
S. aureus pT181     GLLSLQLINRKLVLE--F  INYSSGVYSNIL
Streptomyces mmr    VVLHSTSD---------W  DHGAAISFLAVGLAYLLGGLS  AWRLIARPERRSAVTAAT   475
S. aureus qacA      SLSSMLYRV*VTSFNDAF  VATALVGGIIMIIISIVVYLL  IPKSLDITKQK         514
S. coelicol' ActII  TLFFGLLGS*NFSTAMVR  TLWVVIALLA...........
qacA insert (at *)          FLDISSFSSKGIVGDLAHVAEESVVGAVEVAKATGIKQLANEA
ActII insert (at *)QATASVDDGASRARTELAAAGASTTEQDRLLADLRVCLRESASQQDSERTPDSCRNLQQ
                    ARPAVAEATARAWRTAHTE
Yeast     ATR1      --TVETQVNSDGKHLLKG  YRGAQYLGIGLASLACMISGL  YMVESFIKGRRQELLQ

E. coli  LacY       VLAGNMYESIGF-----Q  GAYLVLGLVALGFTLISVFTL  SGPGPLSLLRRQVNEVA   417
K. pneumoniae       AWVGRMYDTVGF-----H  QAYLILGCITLSFTVISLFTL  KGSKTLLPATA         416
E. coli  RafB       TFAGHLYDRMGF-----Q  NTYFVLGMIVLTVTVISAFTL  SSSPGIVHPSVEKAPVA
E. coli  MelB       MIGYVPNVEQSTQALLGM  QFIMIALPTLFFMVTLILYFR  FYRLNGDTLRRIQIHIL
```

```
CONSENSUS
                                  *         *         *
Human     GLUT1 468 RQGGASQSD--KTPEELFHPLGADSQV                                492
Rabbit    GLUT1     RQGGASQSD--KTPEELFHPLGADSQV                                492
Rat       GLUT1     RQGGASQSD--KTPEELFHPLGADSQV                                492
Pig       GLUT1     RQGGASQSD--KTPEELFHPLGADSQV                                492
Mouse     GLUT1     RQGGASQSD--KTPEELFHPLGADSQV                                492
Human     GLUT2     QKKSGSAHR--PKAAVEMKFLGATETV                                524
Rat       GLUT2     RKKSGSAPP--RKATVQMEFLGSSETV                                522
Mouse     GLUT2     RKKSGSAPP--RKAAVQMEFLASSESV                                523
Rat       GLUT7     RKKHGGRPP--KLRWITANFIIASDQVKKMKND                          528
Human     GLUT3     EGQAHGADRSGKDGVMEMNSIEPAKETTTNV                            496
Mouse     GLUT3     EGQAHSGK--GPAGVEL-NSMQPVKETPGNA                            492
Human     GLUT4     HRTPSLLEQEVKPSTEL-EYLGPDEND                                509
Rat       GLUT4     RRTPSLLEQEVKPSTEL-EYLGPDEND                                509
Mouse     GLUT4     RRTPSLLEQEVKPSTEL-EYLGPDEND                                510
Human     GLUT5     TKMNKVSE-VYPEKEELKELPPVTSEQ                                501
Yeast     SNF3      IKSSTGVVSPKFNKDIRERALKFQYDPLQRLEDGKNTFVAKRNNFDDETPRNDFRNTI..614
Yeast     GAL2      EEGVLPWKSEGWIPSSRRGNNYDLEDLQHDDKPWYKAMLE                   574
Yeast     HXT1      AEGVLPWKSASWVPVSFRGADYNADDLMHDDQPFYKSLFSRK                 569
Yeast     HXT2      VEGVKPWKSGSWISKEKRVSEE                                     541
Yeast     RAG1      SEGVLPWKSSSWVPSSRRGAEYDVDALQHDDKPWYKAML                    567
Yeast     MAL61     RLGVPARKFKSTKVDPFAAAKAAAAEINVKDPKEDLETSVVDEGRSTPSVVNK     614
Yeast     LAC12     EAPNPRKASVDQAFLAQVRATLVQRNDVRVANAQNLKEQEPLKSDADHVEKLSEAESV 587
Yeast     IRT1      KDGFNIKASKALAKKRKQQVAR---VHELKYEPTQEIIEDI                  584
Yeast     IRT2      KDGFNIKASKALAKKRKQQVAEGAAHHKLKFEPTQEIVES                   612
Chlorella HUP1      ARHWFWNRVMGPAAAEVIAEDEKRVAAASAIIKEEELSKAMK                 533
Arabidopsis STP1    RSHWYWSRFVEDGEYGNALEMGKNSNQAGTKHV                          522
Synechocystis glcP
Leishmania Pro-1    AIGKKELSEESIGNRAE                                          567
Neurospora qa-y     EIKPVQNANKNLMAELNFDRNPEREESSSLDDKDRVTQTENAV                537
Aspergillus gutD    DKKPVWHAHSQLIRELRENEEAFRADMGASGKGGVTKEYVEEA                533
E. coli  AraE       MAGEKLRNIGV                                                472
E. coli  GalP       MKGRKLREIGAHD                                              464
E. coli  XylE       EPETKKTQQTATL                                              491
Z. mobilis glf      RSQK                                                       473

K. pneumoniae cit'
E. coli  citA
E. coli kgtP
S. hygroscopicus    PDFERHA                                                    448
bap3
Tn1696 cmlA         AGSTSNPNR                                                  419
E. coli pBR322
E. coli  RP1
E. coli  Tn10
S. aureus NorA
B. subtilis Bmr

Bacillus pTHT15
Bacillus pNS1981
Bacillus BS908
S. aureus pT181
Streptomyces mmr
S. aureus qacA
S. coelicol' ActII
Yeast     ATR1      NTIALWLSGKRY                                               547

E. coli  LacY
K. pneumoniae
E. coli  RafB      HSEIN                                                      425
E. coli  MelB      DKYRKVPPEPVHADIPVGAVSDVKA                                  469
```