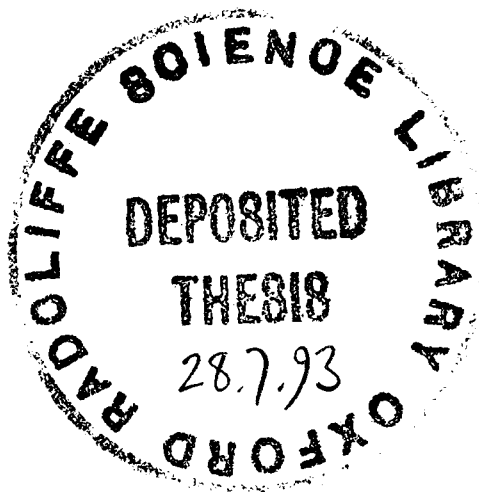


A Fructose-Intolerant Yeast Strain to Select for Sucrose Fructosyl-Transferase Activity



A Thesis submitted for the degree of
Doctor of Philosophy
at the
University of Oxford

Timothy C. Doyle
Magdalen College, Hilary Term 1993

Cum finis est licitus, etiam media sunt licita.

Herman Busenbaum, 1650

Timothy C. Doyle,
Magdalen College,
Oxford.

Hilary Term 1993

Submitted for the degree of
Doctor of Philosophy

A fructose-intolerant yeast strain to select for sucrose fructosyl-transferase activity

A selection system for yeast cells expressing mutated invertase (EC 3.2.1.26, *SUC2*) with altered fructotransferase activity was developed based on the survival of a fructose-intolerant strain in the presence of suitable acceptor substrates and sucrose. Cells of such a strain expressing a wild-type hydrolase activity will not grow due to the release of free fructose from sucrose. Cells expressing an inactive invertase mutant will not grow since they cannot cleave the sucrose, the sole carbon source. Only cells expressing sucrose fructosyl-transferase activity will thrive, growing on the released glucose, the fructosyl moiety not being released.

A strain of *Saccharomyces cerevisiae* was engineered to be intolerant of the presence of fructose in its growth media. This was achieved by inducing a condition in yeast similar to liver cells of humans suffering from hereditary fructose intolerance (MIM 22960). This disorder results from a deficiency of aldolase B (EC 4.1.2.13), and phosphorylation of fructose by ketohexokinase (EC 2.7.1.3) results in an accumulation of fructose 1-phosphate, with a consequent depletion of cytoplasmic phosphate and ATP. Thus, cells in which ketohexokinase phosphorylates fructose, but which lack aldolase B, are intolerant of fructose. Yeast possess neither of these enzymes, and so expression of ketohexokinase in yeast would result in fructose-intolerance. A strain of yeast, for ketohexokinase expression, was initially bred to be unable to metabolize sucrose or fructose, yet remain capable of utilizing glucose, as well as lacking non-specific phosphatases, to prevent remobilization of sequestered fructose 1-phosphate.

Rat liver ketohexokinase was purified to heterogeneity, and the partial amino acid sequence subsequently generated exploited to amplify a region of the ketohexokinase cDNA by PCR. This was used to probe a cDNA library, yielding clones encoding the entire ketohexokinase coding region. This was cloned into *pMA91*, and subsequent expression in yeast resulted in a strain intolerant of fructose in its growth medium, although still capable of growing on glucose. In order to produce a stable fructose-intolerant selection strain, a vector (*pIAD1*) was constructed that allowed multiple integration of an expression cassette containing ketohexokinase cDNA into the rDNA locus of yeast chromosome XII.

Expression of wild type invertase from the episomal plasmid *pIAD3* in this strain resulted in sucrose-intolerance. A preliminary programme of mutagenesis of the *SUC2* gene yielded eight libraries of about one hundred clones each. None of these contained any mutants showing solely sucrose fructosyl-transferase activity, although this system would clearly provide an ideal selection for such mutants from a much larger library.

Table of Contents

Abstract iii

Table of Contents iv

List of Figures xii

List of Tables xiv

Abbreviations xv

Acknowledgements xvi

Chapter 1. A New Strategy of Selection for Fructotransferase Activities

1.1 Naturally Occurring Fructofuranosides 1.1

 1.1.1 Plant and Fungal Fructans 1.3

 1.1.2 Elymoclavine Fructosides 1.4

1.2 Chemical Synthesis of Fructofuranosides 1.6

 1.2.1 The Use of Invertase in Fructofuranosides Synthesis 1.7

1.3 A Selection System to Isolate Sucrose Fructosyl Transferase Activities 1.9

1.4 References 1.12

Chapter 2. Development of the TDS Yeast for Use as a Selection Strain

2.1 Introduction 2.1

 2.1.1 The Life Cycle of *Saccharomyces cerevisiae* 2.2

 2.1.2 Conjugation of Haploids and Dissection of Ascospores 2.5

 2.1.3 Phenotypic Requirement for the TDS Strain 2.7

 2.1.3.1 Sucrose hydrolysing enzymes 2.7

 2.1.3.2 Primary metabolism of fructose and glucose 2.8

 2.1.3.3 Non-specific phosphatase genes 2.9

 2.1.3.4 Auxotrophic markers 2.9

2.2 Materials and Methods 2.11

 2.2.1 Yeast Strains 2.11

 2.2.1.1 TDS parent strains 2.11

 2.2.1.2 Leucine-dependent strains 2.11

 2.2.1.3 Strain storage 2.11

2.2.2 Yeast Growth Media 2.11

2.2.2.1 YEPD 2.11

2.2.2.2 1/4 YEPD 2.11

2.2.2.3 SD 2.12

2.2.2.4 Presporulation medium 2.12

2.2.2.5 Sporulation medium 2.12

2.2.2.6 Burkholder’s defined media 2.12

2.2.3 Determination of Phenotypes and Genotypes 2.13

2.2.4 Haploid Conjugation 2.14

2.2.5 Microforging Loops for the Manipulation of Single Yeast Cells 2.15

2.2.6 Dissection of Asci 2.15

2.3 Results and Discussion 2.16

2.3.1 Mating Strategy 2.16

2.3.2 Phenotype Determination 2.17

2.3.3 TDS II Strain Characteristics 2.18

2.4 References 2.19

Chapter 3. The Purification of Ketohexokinase

3.1 Introduction 3.1

3.1.1 Preparation of Crude Protein Extracts 3.2

3.1.2 Crude Purification by Selective Precipitation 3.3

3.1.2.1 Selective denaturation 3.4

3.1.2.2 Salting-out of proteins 3.4

3.1.3 Chromatographic Separations of Proteins 3.5

3.1.3.1 Gel filtration 3.5

3.1.3.2 Ion-exchange chromatography 3.6

3.1.4 Determination of Protein Molecular Mass 3.8

3.1.5 Protein and Peptide Sequencing 3.10

3.2 Materials and Methods 3.12

3.2.1 General Purification Steps 3.12

3.2.2 Liver Preparation and Enzyme Extraction by Selective Precipitations 3.12

3.2.3 Purification to Homogeneity by Chromatographic Separations 3.13

3.2.4 Assay of Ketohexokinase Activity 3.13

3.2.4.1 Spectrophotometric assay 3.13

3.2.4.2 Colorimetric assay 3.14

3.2.4.3 Determination of protein concentration 3.14

3.2.5 Polyacrylamide Gel Electrophoresis of Protein Samples	3.15
3.2.5.1 Casting and running of SDS discontinuous gels	3.15
3.2.5.2 Fixing and staining of gels	3.16
3.2.6 Molecular Mass Determination of KHK by Mass Spectrometry	3.17
3.2.7 Preparation and Sequencing of Peptides of Ketoheokinase	3.17
3.3 Results and Discussion	3.18
3.4 References	3.23

Chapter 4. The Cloning of a cDNA Probe for the Ketoheokinase Gene

4.1 Introduction	4.1
4.1.1 Screening a cDNA Library	4.1
4.1.1.1 Cloning by complementation	4.1
4.1.1.2 Screening expression libraries with antibody probes	4.2
4.1.1.3 Screening with cDNA probes	4.3
4.1.2 Design of Oligonucleotide Primers	4.4
4.1.3 The Polymerase Chain Reaction (PCR)	4.5
4.2 Materials and Methods	4.7
4.2.1 Bacterial Strain and Plasmid Used	4.7
4.2.1.1 <i>E. coli</i> strain	4.7
4.2.1.2 Cloning vector	4.7
4.2.2 Bacterial Growth Media	4.7
4.2.2.1 LB	4.7
4.2.2.2 Terrific Broth	4.7
4.2.2.3 SOB	4.8
4.2.2.4 SOC	4.8
4.2.2.5 TYM broth	4.8
4.2.3 Strain Storage	4.8
4.2.4 Transformation of <i>E. coli</i> with Plasmid DNA	4.8
4.2.5 Buffered Phenolic Solutions	4.9
4.2.6 Plasmid DNA Purification	4.10
4.2.6.1 Crude small scale preparations (‘minipreps’)	4.10
4.2.6.2 Large scale plasmid purification (‘maxiprep’)	4.10
4.2.6.2.i Preparation of crude plasmid DNA	4.11
4.2.6.2.ii Caesium chloride gradient purification	4.12
4.2.6.2.iii Polyethylene glycol (PEG) precipitation	4.12
4.2.7 DNA Manipulations	4.12

4.2.7.1 DNA precipitation	4.12
4.2.7.2 Restriction enzyme digestion	4.13
4.2.7.3 Removal of terminal phosphate by CIP/proteinase K treatment	4.13
4.2.7.4 Phosphorylation of linear DNA ends with T4 kinase	4.14
4.2.7.5 Blunt-ending of DNA fragments	4.14
4.2.7.6 Ligation of DNA fragments	4.15
4.2.7.7 Horizontal agarose gel electrophoresis of DNA	4.15
4.2.7.8 Purification of DNA from LMP agarose gels	4.16
4.2.8 RNA Purification	4.16
4.2.8.1 Preparation of RNAase-free sterile water	4.17
4.2.8.2 Siliconization of glassware	4.17
4.2.8.3 Isolation of rat liver crude RNA	4.17
4.2.9 cDNA Synthesis, and Second Strand Synthesis by PCR	4.18
4.2.9.1 Preparation of oligonucleotide primers	4.18
4.2.9.2 cDNA first strand synthesis	4.18
4.2.9.3 Second strand cDNA synthesis and amplification by PCR	4.19
4.2.10 Double-Stranded DNA Sequencing	4.19
4.2.10.1 Dideoxy- sequencing reactions	4.19
4.2.10.2 Sequencing gel preparation and electrophoresis	4.20
4.2.10.3 Fixing and exposure of sequencing gel	4.20
4.3 Results and Discussion	4.21
4.3.1 Oligonucleotide Primer Design	4.21
4.3.2 KHK Probe Amplification	4.21
4.3.3 Sequence Analysis of the PCR-generated Fragment	4.22
4.4 References	4.24

Chapter 5. Cloning and Sequencing of the Ketohexokinase cDNA

5.1 Introduction	5.1
5.1.1 cDNA Library Availability and Characteristics	5.1
5.1.2 Sequencing Long Regions of DNA	5.2
5.1.2.1 Generation of a nested set of deletion mutants	5.3
5.1.2.2 Resolution of sequencing gel band-compressions	5.4
5.1.3 Protein Sequence Analysis and Comparison	5.4
5.2 Materials and Methods	5.5
5.2.1 Bacteria, Phage and Plasmids	5.5
5.2.1.1 <i>E. coli</i> strains	5.5

5.2.1.2	Phage library and helper phage	5.6
5.2.1.3	<i>E. coli</i> plasmids	5.6
5.2.1.4	High efficiency electro-transformation of <i>E. coli</i>	5.6
5.2.2	DNA Manipulations	5.7
5.2.2.1	Southern blotting of agarose gels	5.7
5.2.2.2	Synthesis of radioactive probes	5.7
5.2.2.3	Hybridization of radioactive probe to nylon filters	5.8
5.2.3	cDNA Library Screening and Phagemid Rescue	5.9
5.2.3.1	Determination of library titre and phage lysate preparation	5.9
5.2.3.2	cDNA library-plaque colony lifts	5.10
5.2.3.3	Localization of plaques hybridizing probe from the filters	5.11
5.2.3.4	Isolation of the single clones	5.11
5.2.3.5	Determination of the sizes of isolated clonal inserts	5.11
5.2.3.6	Excision of <i>pBluescript</i> plasmids from the phage clones	5.12
5.2.4	DNA Sequencing	5.12
5.2.4.1	Resolution of band compressions on gels	5.12
5.2.4.2	Sequential unidirectional deletion by Exo III/mung bean exonuclease treatment	5.13
5.3	Results and Discussion	5.14
5.3.1	Isolation of the Complete KHK Coding Sequence	5.14
5.3.1.1	Screening of cDNA library	5.14
5.3.1.2	Sequencing the KHK cDNA	5.14
5.3.1.3	Characterization of the KHK cDNA	5.16
5.3.2	Comparison of Ketohexokinase with Other Protein Sequences	5.19
5.3.2.1	Sequence comparison with proteins on the NBRF database	5.19
5.3.2.2	A potential furanose-binding motif	5.21
5.4	References	5.24
5.5	Addendum	5.27
5.5.1	References (continued)	5.28

Chapter 6. Expression of the Ketohexokinase cDNA in the TDS Yeast Strain

6.1	Introduction	6.1
6.1.1	Transformation Systems	6.1
6.1.1.1	Spheroplast transformation	6.1
6.1.1.2	Lithium acetate transformation	6.1
6.1.1.3	Electro-transformation of yeast	6.1
6.1.2	Yeast Plasmid Expression Vectors	6.2

6.1.2.1	Bacterial sequences in yeast vectors	6.2
6.1.2.2	Yeast transformation selection systems	6.3
6.1.2.3	Yeast expression cassettes	6.4
6.1.2.4	Plasmid retention in yeast	6.5
6.1.2.5	Design of a multiple-integration expression vector	6.9
6.1.3	^{31}P -NMR of <i>in vivo</i> Yeast Cells	6.9
6.2	Materials and Methods	6.11
6.2.1	Yeast Plasmids	6.11
6.2.2	Transformation of Yeast with Circular Plasmid DNA	6.11
6.2.2.1	Transformation using lithium chloride	6.11
6.2.2.2	Transformation of spheroplasts	6.12
6.2.2.3	Electro-transformation of yeast	6.13
6.2.3	Transformation of Yeast with Linear DNA for Integration	6.13
6.2.4	Miniprep of Yeast Plasmid DNA	6.13
6.2.5	Pulsed-Field Gel Electrophoresis (PFGE) of Yeast Chromosomal DNA	6.14
6.2.5.1	Preparation of chromosomal DNA in agarose plugs	6.14
6.2.5.2	Equilibration of <i>pH</i> prior to PFGE	6.14
6.2.5.3	Blotting and probing of chromosomal DNA	6.15
6.2.6	Determination of Fructose Intolerance in Transformed Yeast	6.15
6.2.6.1	Preliminary screens on petri dishes	6.15
6.2.6.2	Quantification of intolerance of fructose	6.15
6.2.6.3	Recovery of cell from exposure to fructose	6.15
6.2.6.4	Sensitivity of cells to sucrose	6.15
6.2.7	Partial Purification of Ketohexokinase from Yeast	6.16
6.2.8	Enzymatic Identification of Fructose 1-Phosphate in Transformed Yeast	6.16
6.2.8.1	Perchloric acid extraction of cell metabolites	6.16
6.2.8.2	Enzymatic analysis of fructose 1-phosphate	6.17
6.2.9	Metabolic Studies Using ^{31}P NMR	6.18
6.2.9.1	NMR of intact cells	6.18
6.2.9.2	NMR of perchloric acid extracts from yeast cells	6.18
6.3	Results and Discussion	6.19
6.3.1	Heterologous Expression of Ketohexokinase on Episomally-Borne Plasmids in Yeast	6.19
6.3.1.1	Cloning of the ketohexokinase coding region into <i>pMA91</i>	6.19
6.3.1.2	Transformation of the yeast TDS II	6.20
6.3.1.3	Heterologous expression of ketohexokinase	6.20
6.3.1.4	Determination of fructose intolerance	6.21
6.3.1.5	Enzymatic identification of fructose 1-phosphate in TDB I cells	6.21
6.3.1.6	^{31}P NMR studies of transformed TDB I strains	6.22

6.3.2 Multiple Integration of the Ketohexokinase cDNA into the Yeast Chromosome 6.29

6.3.2.1 Construction of the integration plasmid, *pIAD1* 6.29

6.3.2.2 Transformation and integration of *pIAD1*-KHK 6.30

6.3.2.3 Pulsed-field gel electrophoresis of yeast chromosomes 6.31

6.3.2.4 Tolerance to fructose of integrated plasmid-bearing yeast 6.31

6.3.2.5 Heterologous expression of ketohexokinase from the integrated plasmid 6.33

6.3.3 Genetic Stability of the Integrated *pIAD1* Plasmid 6.34

6.3.4 Sucrose Tolerance of the Fructose Intolerant Yeast Strains 6.34

6.4 References 6.34

Chapter 7. Mutagenesis and Screening of the *SUC2* Gene

7.1 Introduction 7.1

7.1.1 Random Mutagenesis of DNA 7.1

7.1.1.1 Nitrous acid mutagenesis 7.2

7.1.1.2 Hydroxylamine mutagenesis 7.2

7.1.1.3 Bisulphite mutagenesis 7.3

7.1.1.4 Hydrazine mutagenesis 7.4

7.1.1.5 Acid depurination of DNA 7.4

7.1.2 Generation of Single-Stranded DNA 7.4

7.1.2.1 Preparation of single-stranded DNA by filamentous phage 7.5

7.1.2.2 Production of single-stranded DNA by asymmetric PCR 7.5

7.1.3 Design of the Yeast Expression Vector for *SUC2* Expression 7.6

7.2 Materials and Methods 7.7

7.2.1 Yeast Plasmid 7.7

7.2.2 Direct Cloning of PCR Products by dT-Tailing the Plasmid 7.7

7.2.3 Direct Blotting of Bacterial DNA onto Charged Nylon Membranes 7.8

7.2.4 Chemical *in vitro* Mutagenesis of Single-Stranded DNA 7.8

7.2.4.1 Amplification and purification of single-stranded DNA by asymmetric PCR 7.8

7.2.4.2 Preparation of Sephadex G50 spun columns 7.9

7.2.4.3 *In vitro* mutagenesis with formic acid 7.9

7.2.4.4 *In vitro* mutagenesis with hydrazine 7.9

7.2.4.5 *In vitro* mutagenesis with sodium bisulphite 7.9

7.2.4.6 Amplification of mutated DNA for ligation 7.10

7.3 Results and Discussion 7.10

7.3.1 Cloning and Expression of the *SUC2* Gene for Invertase 7.10

7.3.1.1 Construction of the yeast expression vector, *pIAD3* 7.10

7.3.1.2 Cloning of the <i>SUC2</i> coding region	7.10
7.3.1.3 Expression of wild-type invertase	7.13
7.3.2 <i>In Vitro</i> Mutagenesis and Expression of the <i>SUC2</i> Gene	7.15
7.3.2.1 Generation of single-stranded DNA by PCR	7.15
7.3.2.2 Chemical mutagenesis and amplification of mutated DNA	7.15
7.3.2.3 Expression of mutant <i>SUC2</i> genes and screening for fructotransferase activity	7.15
7.4 References	7.16

Chapter 8 General Conclusions: The Future Uses of the Fructose-Intolerant Yeast

8.1 <i>SUC2</i> Screening Strategy	8.1
8.2 A Potential Strategy for the Characterization of <i>SUC2</i> Mutants	8.2
8.2.1 Region-Specific Mutagenesis of 'Hot-Spots'	8.4
8.3 Potential Improvements for Integration of Ketohexokinase	8.5
8.4 Potential for Isolation of Other Fructotransferase Genes	8.6
8.4.1 Expression and Mutation of the <i>Aspergillus niger SUC1</i> Gene	8.6
8.4.2 Screening of an <i>Aspergillus niger</i> Library for a Fructotransferase Gene	8.7
8.4.3 Isolation of the Fructan Synthesizing Proteins of <i>Compositae</i> Plants	8.7
8.5 References	8.8

Appendix. Publications and Presentation of Data

1. Abstract of poster presented at the '16th International Conference of Yeast Genetics and Molecular Biology,' Vienna, Austria, August 1992. (Doyle, T.C., Donaldson, I.A., Matas, N. and Cox, B.S. (1992). Genetic engineering of a fructose-intolerant yeast strain to screen for fructotransferase mutants of invertase. *Yeast*, **8**, S573.)
2. Donaldson, I.A., Doyle, T.C. and Matas, N. (1993). Expression of rat liver ketohexokinase in yeast results in fructose intolerance. *Biochem. J.*, **291**, 179-186.
3. Doyle, T.C., Spickett, C.M., Donaldson, I.A. and Ratcliffe, R.G. (1993). Metabolic Studies of a Fructose-Intolerant Yeast by *in vivo* ³¹P-NMR Spectroscopy. *Yeast*, In press.

List of Figures

<i>Fig. 1.1</i>	Structure of fructose and sucrose	1.2
<i>Fig. 1.2</i>	Structure of fructan isomers	1.2
<i>Fig. 1.3</i>	Structure of elymoclavine fructosides	1.6
<i>Fig. 1.4</i>	Fructose metabolism in liver, muscle and yeast	1.10
<i>Fig. 1.5</i>	Metabolic consequences of mutant invertases in a fructose-intolerant yeast	1.11
<i>Fig. 2.1</i>	The life cycle of <i>Saccharomyces cerevisiae</i>	2.3
<i>Fig. 2.2</i>	Mating and conjugation in yeast	2.4
<i>Fig. 2.3</i>	Tetrad analysis of phenotypic markers from diploid yeast	2.6
<i>Fig. 2.4</i>	Construction of the <i>pho3/pho5</i> double mutant IH530	2.10
<i>Fig. 2.5</i>	Determination of auxotrophic marker genotype	2.14
<i>Fig. 2.6</i>	Development of the TDS II yeast strain	2.17
<i>Fig. 2.7</i>	Results of chromogenic assay for acid and alkaline phosphatases	2.19
<i>Fig. 3.1</i>	Titration curves of three proteins, indicating <i>pH</i> ranges for separation	3.6
<i>Fig. 3.2</i>	Titration curves of Fractogel® TSK CM-650 and DEAE-650	3.7
<i>Fig. 3.3</i>	<i>pH</i> elution gradient of chromatofocusing column	3.8
<i>Fig. 3.4</i>	Schematic explanation of chromatofocusing	3.9
<i>Fig. 3.5</i>	Purification strategy for ketohexokinase	3.19
<i>Fig. 3.6</i>	Chromatographic elution profiles showing KHK activity peaks	3.20
<i>Fig. 3.7</i>	SDS-PAGE of purification fractions for ketohexokinase	3.21
<i>Fig. 3.8</i>	Elution profile of peptides generated from ketohexokinase	3.22
<i>Fig. 4.1</i>	DNA and amino acid sequence of amplified KHK probe region	4.23
<i>Fig. 5.1</i>	The λ ZAP II vector with cDNA insert	5.2
<i>Fig. 5.2</i>	Restriction maps of three rescued clones from the λ ZAP library	5.15
<i>Fig. 5.3</i>	Preliminary sequencing by Exo III deletion of clone 16i	5.16
<i>Fig. 5.4</i>	Strategy for the construction of clone 5i/16i in <i>pUC18</i>	5.17
<i>Fig. 5.5</i>	Construction of deleted clones of the KHK cDNA for sequencing	5.18
<i>Fig. 5.6</i>	Sequencing strategy for the entire KHK cDNA	5.19
<i>Fig. 5.7</i>	DNA and amino acid sequence of rat liver ketohexokinase	5.20
<i>Fig. 5.8</i>	Amino acid comparison of ketohexokinase and other furanose kinases	5.22
<i>Fig. 5.9</i>	Amino acid sequences of putative furanose-binding regions	5.22
<i>Fig. 5.10</i>	Amino acid comparison of phosphofructokinase sequences	5.23
<i>Fig. 5.11</i>	Sequence alignment of KHK and 6-phosphotagatokinase	5.28
<i>Fig. 6.1</i>	Integration of a linearized plasmid into the yeast chromosome	6.8
<i>Fig. 6.2</i>	The rDNA locus of yeast chromosome XII	6.10

<i>Fig. 6.3</i>	Colorimetric assay showing ketohexokinase activity in transformed yeast	6.21
<i>Fig. 6.4</i>	SDS-PAGE of partially purified yeast extracts	6.22
<i>Fig. 6.5</i>	Protein elution profiles from an S100 column, showing KHK activities	6.23
<i>Fig. 6.6</i>	Fructose intolerance of TDB I cells	6.24
<i>Fig. 6.7</i>	Recovery of TDB I cells from fructose stress	6.25
<i>Fig. 6.8</i>	³¹ P NMR spectra of TDB I cells during control experiments	6.26
<i>Fig. 6.9</i>	³¹ P NMR spectra of TDB I cells during 1 and 5mM. fructose loading	6.27
<i>Fig. 6.10</i>	³¹ P NMR spectra of TDB I cells during 25mM. fructose loading	6.28
<i>Fig. 6.11</i>	Construction of <i>pIAD1</i> expression/multiple integration yeast plasmid	6.30
<i>Fig. 6.12</i>	PFGE of chromosomes of several TDB II strains	6.32
<i>Fig. 6.13</i>	Growth profile of TDB II cells upon fructose loading	6.32
<i>Fig. 6.14</i>	Recovery of TDB II cells from fructose stress	6.33
<i>Fig. 7.1</i>	Mechanism of nitrous acid mutagenesis	7.2
<i>Fig. 7.2</i>	Mechanism of hydroxylamine mutagenesis	7.3
<i>Fig. 7.3</i>	Mechanism of bisulphite mutagenesis	7.3
<i>Fig. 7.4</i>	Mechanism of hydrazine mutagenesis	7.4
<i>Fig. 7.5</i>	Strategy for the construction of the <i>pIAD3</i> plasmid	7.11
<i>Fig. 7.6</i>	Construction of the <i>pUC SUC2±</i> plasmid from the <i>pRB58</i> plasmid	7.12
<i>Fig. 7.7</i>	Directional cloning of the <i>SUC2</i> gene into the <i>pIAD3</i> expression cassette	7.14
<i>Fig. 8.1</i>	Strategy for the localization of mutations in the <i>SUC2</i> gene	8.3

List of Tables

<i>Table 1.1</i>	Different types of fructans	1.2
<i>Table 1.2</i>	Some of the reported fructotransferase activities of yeast invertase	1.8
<i>Table 3.1</i>	Methods for preparation of crude protein extracts	3.3
<i>Table 3.2</i>	Separation properties of various gel filtration media	3.5
<i>Table 3.3</i>	Specificities of endoproteases used for peptide generation	3.11
<i>Table 3.4</i>	Recipes for SDS PAGE	3.16
<i>Table 3.5</i>	Purification scheme for rat ketohexokinase	3.21
<i>Table 3.6</i>	Amino acid sequences of KHK-derived peptides	3.22
<i>Table 4.1</i>	Conditions and amount of CIP required for dephosphorylation of linear DNA	4.13
<i>Table 4.2</i>	Separation ranges of DNA fragments by various agarose concentrations .	4.16
<i>Table 4.3</i>	Oligonucleotide sequences of primers used to amplify the KHK probe . .	4.21
<i>Table 5.1</i>	Comparison of proteins using the MULTALIGN program	5.21
<i>Table 6.1</i>	Native <i>S. cerevisiae</i> promoter systems	6.4
<i>Table 6.2</i>	PCR primers for the amplification of the KHK coding region	6.19
<i>Table 7.1</i>	PCR primers for the amplification of the <i>SUC2</i> coding region	7.12
<i>Table 8.1</i>	PCR primers for characterization of <i>SUC2</i> mutations	8.2

Abbreviations

AMP, ADP, ATP	Adenosine 5' mono-, di-, triphosphate
Amp	Ampicillin
ARS	Autonomously replicating sequence
BCIP	5-Bromo-4-chloro-3-indolyl-phosphate
bp, kb	Base pair(s), kilobase pair(s)
BSA	Bovine serum albumin
cDNA	Complementary (to mRNA) DNA
CIP	Calf intestine alkaline phosphatase
Da, kDa	Daltons, kiloDaltons
DEAE	Diethylaminoethyl
DEPC	Diethyl pyrocarbonate
dITP	Deoxyinosine triphosphate
DNA	Deoxyribose nucleic acid
dNTP, ddNTP	Deoxy-, dideoxy-nucleoside triphosphates
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FAD, FADH ₂	Flavin adenine dinucleotide, oxidized and reduced
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GTP	Guanosine triphosphate
HPLC	High pressure liquid chromatography
IPTG	Isopropylthio-β-D-galactoside
KHK	Ketohexokinase
K _m	Michaelis constant
LMP-agarose	Low melting point-agarose
MES	2(<i>N</i> -Morpholino)ethanesulphonic acid
mRNA	Messenger RNA
NAD, NADH	Nicotinamide adenine dinucleotide, oxidized and reduced
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEP	Phospho- <i>enol</i> -pyruvate
PFGE	Pulsed-field gel electrophoresis
P _i	inorganic phosphate
pI	Isoelectric point
POD	Peroxidase
ppm	Parts per million
PTH-amino acid	Phenylthiohydantoin-amino acid
rDNA	Ribosomal RNA-encoding DNA
RNA	Ribose nucleic acid
SDS	Sodium dodecyl sulphate
TEA	Triethanolamine
TEMED	<i>N,N,N',N'</i> -tetramethylethylenediamine
TEMDP	Tetramethyl ester of methylene diphosphonate
TFA	Trifluoroacetic acid
T _m	Melting temperature
TPP	Thiamine pyrophosphate
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
v/v, w/v	Volume/volume, weight/volume
Xgal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside

Acknowledgements

I should firstly like to thank Professors E.M. Southern and G.K. Radda for allowing me to undertake my research in their Department. Thanks also to the Science and Engineering Research Council for 3 years and 8 months funding (thanks Brian).

I am also very grateful to Dr. Brian Cox, formerly of the Department of Plant Sciences, for allowing me to use his laboratory during the early stages of this work, and to him and the rest of his group, especially Drs. Joy Boyce, Shirley McCreedy and Enrico Beradi for all their advice, ideas and assistance. Similarly, I am indebted to Dr. David Bonthon, of the University of Edinburgh, for being a guru on the end of the phone about all aspects of molecular biology.

I should also thank Dr. George Ratcliffe for allowing us NMR time on his magnet, and Dr. Corinne Spickett for 'shimmying' the magnet, obtaining the spectra, and explaining what all the peaks meant. Their enthusiasm for the NMR aspect of this thesis certainly paid off for all concerned.

Thanks must also be expressed to Tony Willis for sequencing the ketohexokinase peptides, Drs. Andy Porter and Uwe Maskos for their assistance with the PCR, Dr. David Harvey for performing the mass spectroscopy of the purified ketohexokinase, and Dr. Geoff Barton for his assistance and advice with the protein sequence comparison work.

In the lab, my thanks must go to Allison Fleetwood for a month's assistance in the calm before the storm, and to Nada Matas, a student from Zagreb University, for giving up her summer vacation to come to work with me on this project. So enjoyable, she came back for more... Thanks also to the Master of the Impossible, Simon, my fellow grad. student, for keeping that certain sanity and insanity alive in the lab' over the last three and a half years.

I am also indebted to Magdalen, for accommodating me for all these years (possibly setting a new record), and all my friends there for making my time out of the lab so varied and exciting.

Finally, and most importantly, I should like to thank my supervisor, Dr. Iain Donaldson, for all his enthusiasm, encouragement and friendship (with the odd beer, here and there) over the last five years. Through his selfless efforts I have been able to undertake this work, and further my scientific career in the States. I hope his advice both scientific and otherwise will not be forgotten or ignored.

Chapter 1

A New Strategy of Selection for Fructotransferase Activities

There has recently been renewed interest in fructofuranosides for both food and non-food uses [1]. Fructans are easily isolated by precipitation and can be hydrolysed as a source of very pure free fructose for use as a sweetener. Alternatively, they may be processed by acid-catalysed hydrolysis and dehydration to form hydroxymethylfurfural as a feedstock for the chemical industry. Furthermore, fructosylation of pharmaceuticals, to decrease their cumulative effects, is of some interest because human tissues do not possess the enzymatic machinery necessary for the hydrolysis of this glycosidic link. The synthesis of fructofuranosides remains most easily accomplished by exploitation of enzymatic fructosyl-transfer from sucrose. The project described in this thesis provides a genetic approach to investigate this activity, through a selection system for genes that encode sucrose fructosyl-transferases.

1.1 Naturally Occurring Fructofuranosides

Fructofuranosides are compounds that occur naturally in plants and lower organisms as derivatives or storage polymers. These compounds contain at least one, terminal, unsubstituted D-fructofuranosyl moiety, usually as the β -anomer, attached via a 2- or 6- linkage (see *Fig. 1.1a*). The most common fructofuranoside is sucrose, which is a double glycosidic disaccharide consisting of an α -D-glucopyranoside joined to a β -D-fructofuranose by a linkage that is α -1,2 with respect to the glucose, and β -2,1 with respect to the fructose (*Figure 1.1b*). This is a transport compound in many plants, allowing the movement of the photoassimilate from source to sink tissues, where it may either be stored or used in other metabolic processes. In most plants, storage of carbohydrate is in the form of the polymer of glucose, starch, but in some families of plants, polymers of fructose comprising a sucrose molecule with additional fructofuranosyl moieties accumulate. Collectively, these fructosides are known as fructans, of which three main forms exist, 1-kestose (inulin), 6-kestose and neokestose (both known as phlein) (see *Table 1.1* and *Figure 1.2*).

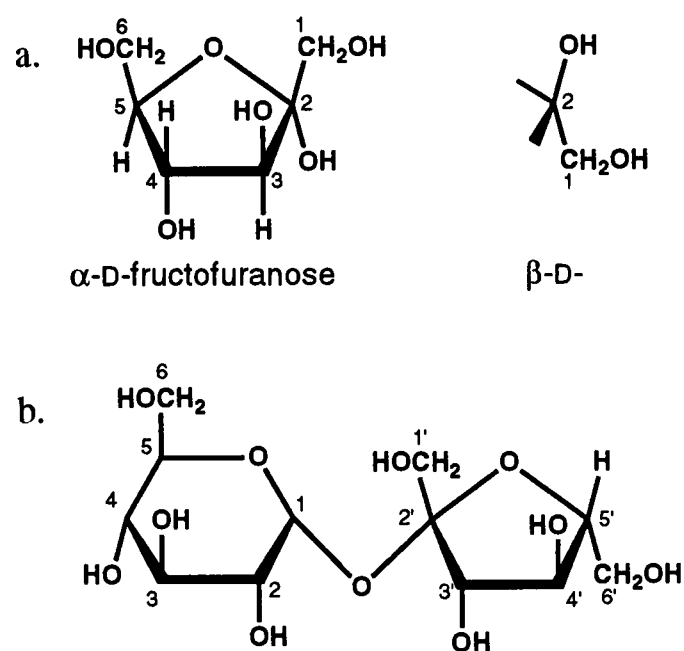


Figure 1.1 a. Structure of the two anomers of the furanose form of D-fructose, indicating the numbering of the carbon atoms.
b. Structure of sucrose (α -D-glucopyranoside-1,2- β -D-fructofuranoside).

Trivial name	Basic trisaccharide	Fructosyl link in elongating chain	Maximum chain length	Distribution
Inulin	1-kestose	β -2,1'	35-50	Compositae Boraginaceae Iridaceae
Phlein	6-kestose Neokestose	β -2,6' β -2,6	250 8-10	Gramineae <i>Asparagus</i>

Table 1.1 Different types of fructan (see also figure 1.2). Taken and corrected from John [2].

Fungi, such as *Aspergillus niger* [4, 5], and some bacteria (e.g. *Bacillus subtilis* [6, 7, 8], *B. amyloliquefaciens* [9] and *B. polymyx*, as well as many oral bacteria [10]) are also capable of synthesizing fructans, although they achieve this through the use of the enzyme

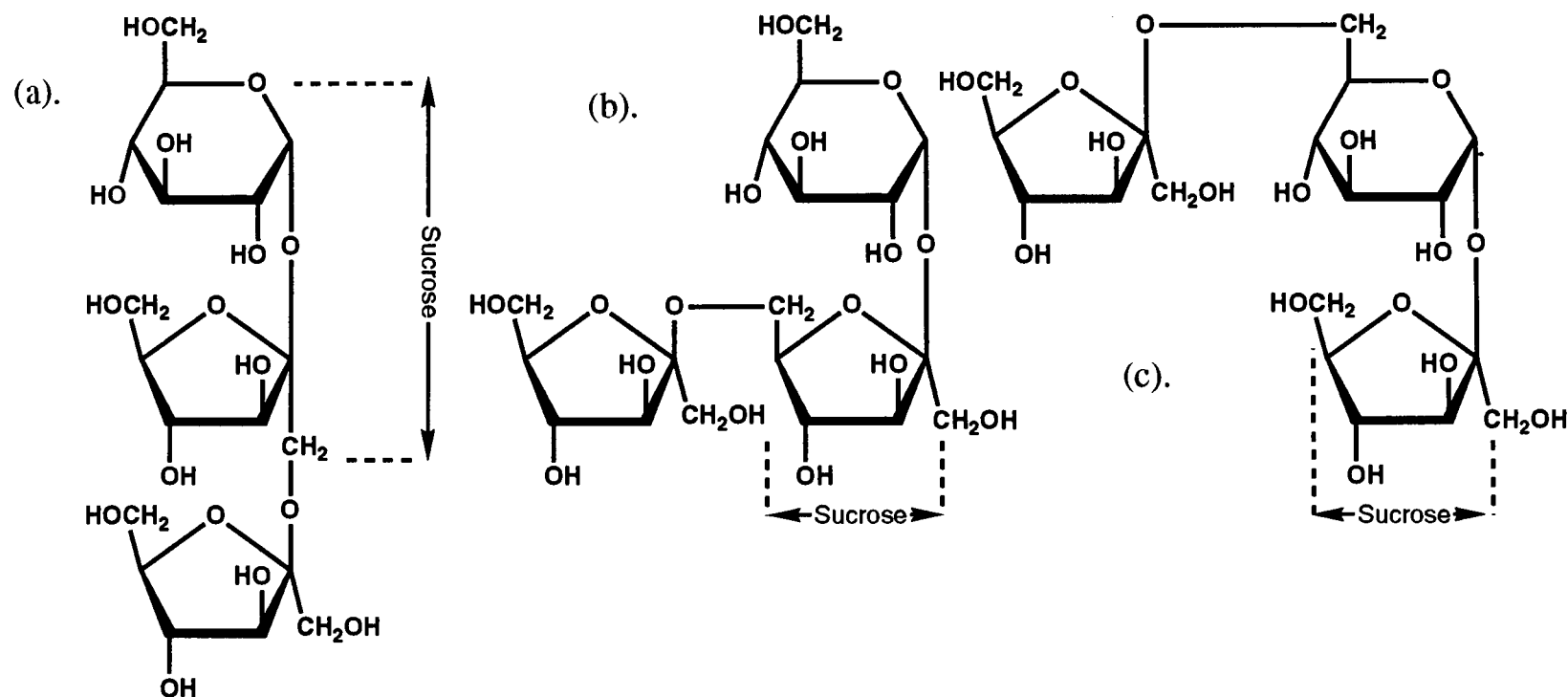


Figure 1.2 The structure of three isomers of trisaccharide fructan found in plants, (a) 1-kestose (isokestose), (b) 6-kestose (kestose), and (c) neokestose [3].

levansucrase. In addition to polymerizing fructose, fungi can also fructosylate other chemicals, such as ergot alkaloids. *Claviceps purpurea* is able to synthesize elymoclavine fructosides [11] with one, two or three *O*- β -D-fructofuranosyl moieties attached. It is probable that these compounds are used by the organism in an antimicrobial rôle [12].

1.1.1 Plant and Fungal Fructans.

Fructose is the sweetest naturally occurring sugar, and can taste 20 to 80% sweeter than sucrose, depending on the test used [2]. In addition, fructose is more soluble than sucrose, less viscous, causes fewer dental caries and can be metabolized at low levels without inducing insulin secretion. For the food industry, fructose offers a number of advantages over sucrose, and upon development of a cheap source of fructose (high-fructose corn syrup, HFCS) in 1975, the sweetener industry was revolutionised. In 1985, 30% of the share of sweeteners used in the USA was HFCS, compared with sucrose with a 43% share. It is expected that by 2005, HFCS and crystalline fructose will together have overtaken sucrose in the USA [1]. This process is, however, confined to the USA, and in the EC, for example, production and utilization of HFCS has been restricted in order to protect the interests of the sugar beet farmers. This policy is unlikely to continue indefinitely, simply because of the technical advantages of fructose over sucrose in the food industry.

Isomerization of glucose from cereal starch is not the only source of fructose, although it may at present be the most economical. Fructan provide a more natural source of fructose, from which the fructose may be more simply obtained. Fructan is accumulated to high levels in storage organs of species belonging to the *Compositae*, notable the Jerusalem artichoke and chicory. Both species accumulate fructan to levels comparable to starch in potato and sugar in beet, and both can be cultivated and harvested. However, Jerusalem artichoke has a very large canopy in proportion to the size of its roots, which makes harvesting inconvenient, and chicory is usually grown under conditions where breakdown of stored fructan is at an advanced stage before the root is available for extraction. There is currently considerable interest in the introduction of genes for fructan synthesis into sugar beets, which it is envisaged would be a much more efficient source of these compounds. Attempts to express the levansucrase from *Bacillus subtilis* in plants have revealed that these long chain, insoluble, fructans appear to be toxic to plants. Accumulation of fructans in plants is normally accomplished by the action of two enzymes (8.4.3) that restrict the products to an average degree of polymerization (DP) of 50-70, which remain soluble. Transfer of this system to species that do not normally accumulate fructans is unlikely to have the same toxicity problems associated with the deposition of insoluble fructans in transformed plants.

The fructose can be released on hydrolysis of the fructan, the half life of β -fructofuranosidic linkages in 0.5*M.* sulphuric acid at 20°C being 370-390 minutes [13], and complete hydrolysis

has been achieved by treatment with 0.01M. hydrochloric acid at 52°C for 3-5 days [14] or with 0.2M. hydrochloric acid at 95°C for 2 minutes [15]. But as well as being a source of fructose, the fructan itself is finding applications in the food industry itself. Fructans have about 10% of the sweetness of a comparable weight of sucrose, and possess many of the technical properties of sucrose required by the food industry. However, when ingested, fructans remain unhydrolyzed until they reach the large intestine where they encourage the selective growth of *Bifidobacteria* which are claimed to benefit health [16, 17]. To these ends, Neosugar, commercially produced utilizing *Aspergillus niger* ATCC 20611, and composed of 1-kestose with 1 or two additional sugars attached, has been implicated in relief of constipation, cholesterol reduction and suppression of putrefactive substances [17]. Because the fructan is less readily absorbed from the digestive tract it is classed as a low-calorie, natural, dietary fibre.

In Mexico, over 20,000ha of agave (*Agave azul tequilana*) is grown for the production of tequila. This alcoholic drink is distilled from the fermented inulin-containing agave juice, ethanolic fermentation from inulin being performed by a yeast species (*Kluyveromyces marxianus*) growing naturally on the agave plant [18]. Although ethanol produced in this way is intended for human consumption, fructan-derived fructose is also suitable as a bulk chemical for the industrial production of ethanol, as well as propanone, butanone, butan-2,3-diol and propan-1,2,3-triol [1]. Although bacterially-produced polyfructoses have not yet been exploited for industrial applications, bacterial fructans have been studied as a potential source of fructans [10]. It has also been suggested that production of fructans from sucrose by *B. polymyxa* might be used for the production of fructose syrups, used directly as a thickener, or to encapsulate flavouring and colouring agents.

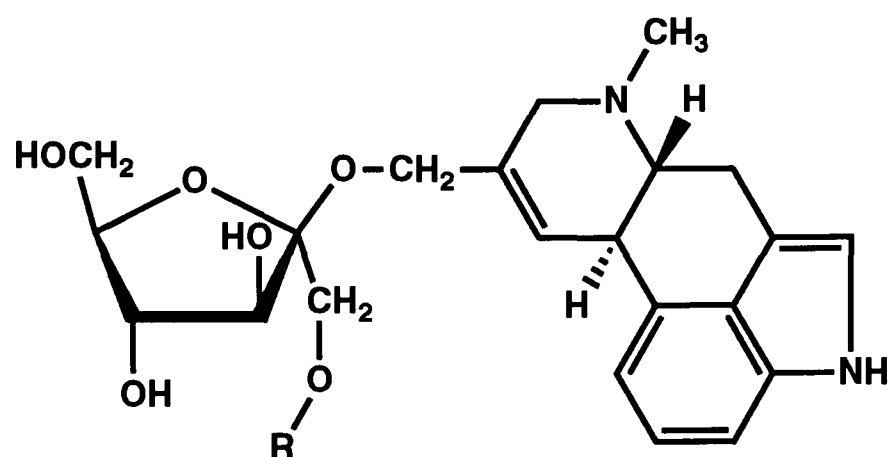
1.1.2 Elymoclavine Fructosides.

Ergot alkaloids have a long history of toxic effects, to animals as well as man, and also of medicinal uses [19]. These tetracyclic indole derivatives are primarily found in the sclerotia of various species of *Claviceps*, which parasitize members of the Gramineae, most prominently, *Claviceps purpurea* on rye. The toxic effects of ergot or ergotism were documented and described long before their cause was recognised. They were especially severe in Central Europe during the Middle Ages, probably due to the consumption of bread made with infected rye. In addition to the 'classical' muscular, vascular and neurological effects of ergot alkaloids, neurohormonal activities have been observed which manifest themselves on various processes that are controlled by the hypothalamic hypophyseal system. The most prominent effects of ergot alkaloids are inhibition of lactation in mammals, first noted as early as 1676 that agalactia could be observed among nursing women suffering gangrenous ergotism [20], and the inhibition of nidation, the implantation of the fertilized egg, in rats [21]. Both effects have

been traced to the inhibition of prolactin release from the anterior pituitary gland by certain ergot alkaloids as the primary mechanism of action [22, 23, 24]. The ability of ergot alkaloids to inhibit prolactin release from the pituitary assumed the additional significance with the discovery of the rôle of prolactin in the development and growth of certain mammary tumors in mammals with a significance in human breast cancer [25]. The available evidence indicated that prolactin seems to be necessary for the induction and growth of dimethylbenzanthracene (DMBA)-induced mammary carcinomas in rats. It was then shown that certain ergot alkaloids significantly inhibit the development of DMBA-induced mammary tumors in rats, and reduced the incidence of spontaneous alveolar hyperplastic nodules, a form of precancerous mammary lesion [26]. Administration of ergot alkaloids to rats bearing DMBA-induced mammary adenocarcinomas over a period of several weeks resulted in regression of tumors [22].

Naturally occurring elymoclavines in *Claviceps purpurea* have been found to fall into three classes. During submerged fermentation, a transient formation of agroclavine is rapidly followed by the generation of elymoclavine, the generation of which peaks after 15-20 days. This is then fructosylated by the action of an invertase (fructosyl transferase) activity present in the fungal mycelium, transferring the fructosyl moiety from sucrose [11]. The natural elymoclavine fructosides were found to contain one, two or three *O*- β -D-fructofuranosyl units (see *Fig. 1.3*). Use of immobilized cells of *Claviceps purpurea* has been reported as a method of production of the elymoclavine fructosides [27], and even yeast invertase has been shown to perform the same glycosylation to produce the monosaccharide derivative [28]. The fructosylation of elymoclavine by the fungus increases the water solubility of the alkaloid, and thereby decreases its inhibitory effect on the growth of the fungus. Thus, levels of production of the alkaloid by cultures grown on sucrose, in which this protective fructosylation takes place, are greater than those using other carbon sources.

Fructosylated elymoclavine has been shown to have a similar activity in inhibiting prolactin release as the unfructosylated compound, but offers an additional advantage as a potential mammary carcinoma treatment. Clavine alkaloids are toxic, and thus prolonged treatment may result in the accumulation of the chemical, with consequential toxic manifestations. It would seem likely that elymoclavine would accumulate in lipids, and its biological half-life would be considerably longer than its aqueous equivalent. The addition of a fructosyl moiety to the compound should render the alkaloid amphipathic, aiding its excretion from the body. Mammals do not produce fructohydrolases, although this enzyme is made by bacteria in the lower gut. Consequently, intravenous administration of the drug will prevent its exposure to the fructohydrolase activity, and consequently, the drug will remain fructosylated, and hence will be relatively quickly excreted. The extent of glycosylation might be used to control the retention time of the compound in the body.



R = H, (2→1)-O-β-D-fructofuranoside, or disaccharide

Figure 1.3 The structure of elymoclavine fructoside, a chemical implicated as a potential mammary carcinoma treatment.

1.2 Chemical Synthesis of Fructofuranosides

Chemical methods for the synthesis of β-D-fructofuranosides are few and limited in scope [29]. The standard method for the synthesis of fructosides is that of Fischer [30]. Free, unprotected fructose is dissolved in a highly polar dry alcohol, which also serves as the aglycone. A strong acid catalyst is usually employed, but more recently, strong cation exchange resins (*e.g.* Dowex 50) have been used [31, 32]. At room temperature, the reaction involves the relatively fast production of α- and β-fructofuranosides with tautomerization to pyranosides proceeding at such a slow rate that their contamination of the synthesis is easily avoided by keeping the temperature below 25°C [33]. In no cases have phenolic fructofuranosides been synthesized by this method, although benzyl-thiofructosides have been formed from fructose and toluene thiol in dimethyl sulfoxide (DMSO) using sulphuric acid as a catalyst [34]. Fischer methods are limited to short chain, highly polar alcohols in which it is possible to dissolve fructose.

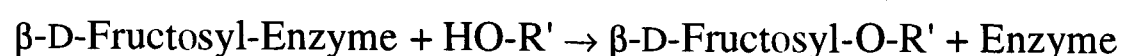
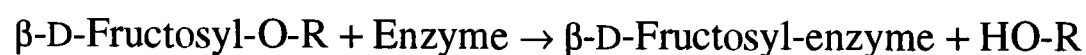
Attempts to generate fructofuranosides by the thermolysis of sucrose in DMSO in the presence of alcohols was later found to have a mechanism similar to the Fischer method [35], whereby traces of sulphuric acid in the DMSO were found to hydrolyse the sucrose generating a fructosyl carbonium ion, with which the alcohol group could react [31]. As a synthetic method, this reaction is fraught with problems, such as the preferential production of the α-fructofuranoside, substantial isomerization of the furanosyl products to the pyranose form due to the high temperature required to cleave the sucrose, and contamination with glucosyl compounds. Due to the difficulties encountered in this method of fructofuranoside synthesis, attention is now turning back to the use of invertase to catalyse transfructosylation reactions, interest in which had waned in the 1960's due to the hydrolytic activity of invertase.

1.2.1 The Use of Invertase in Fructofuranoside Synthesis.

β -D-fructofuranoside fructohydrolase (invertase, EC 3.2.1.26) catalyses the hydrolysis of sucrose to glucose and fructose, and solutions in which this reaction proceeds undergo a net inversion of optical rotation owing to the highly laevorotatory nature of free fructose. This characteristic has led to the assignment of the trivial name 'invertase' to the enzyme, although some α -glucosidases (EC 3.2.1.20) will 'invert' sucrose solutions *in vitro* [36, 37, 38], and mammals hydrolyse sucrose with an intestinal α -glucosidase [39].

The ability of yeast to produce invertase is conferred by the '*SUC*' genes [37, 40], a family of genes responsible for sucrose utilization [41]. Studies of the *SUC2* gene have shown that it encodes two forms of invertase, a secreted, glycosylated form which is regulated by glucose repression, and a non-glycosylated enzyme which is produced constitutively in small amounts and remains within the cell [42, 43].

All of the characteristics of invertase action are consistent with a two-step mechanism in which a fructosyl-enzyme complex is formed [44].



R is normally the glucosyl residue of sucrose, but may be a methyl group [45], and R' is usually an hydrogen of water but may be one of a number of organic compounds which can serve as acceptor-substrates (see *Table 1.2*). Invertase is therefore also a fructotransferase, although this activity is unlikely to be of any physiological significance. With certain acceptor substrates, however, yeast invertase shows a distinct preference for fructosyl transfer rather than hydrolysis, despite the enormous concentration of water [46].

It can be seen from *Table 1.2* that the simplest system for the generation of fructofuranosides involves use of sucrose as both donor and acceptor substrate to produce 6-kestose, 1-ketose or neokestose [48-52]. Synthesis of β -D-fructofuranosides by transfructosylation is currently limited by the acceptor substrate specificity of invertase (*Table 1.2*), and the attrition of yields by the hydrolytic action of the enzyme [46]. It is perhaps for these reasons that interest in the fructotransferase activity of invertase as a catalyst for organic synthesis has waned since extensive investigations of acceptor-substrate specificities during the mid-nineteen sixties [56].

Donor	Acceptor	Product	Reference
Sucrose	Glucose	<i>O</i> - β -D-fructofuranosyl-(2-6)-D-glucose	47
"	Fructose	<i>O</i> - β -D-fructofuranosyl-(2-1)-D-fructose	47
"	"	<i>O</i> - β -D-fructofuranosyl-(2-6)-D-fructose	47
"	Sucrose	6-Kestose	48-52
"	"	1-Kestose	48-52
"	"	Neokestose	48-52
Methyl- β -D-fructofuranoside	Sorbitol	β -D-fructofuranosyl sorbitol	45
"	Mannitol	β -D-fructofuranosyl mannitol	45
"	Glucose	<i>O</i> - β -D-fructofuranosyl-(2-6)-D-glucose	45
"	Mannose	β -D-fructofuranosyl mannose	45
"	Galactose	β -D-fructofuranosyl galactose	45
Sucrose	Methanol	Methyl- β -D-fructofuranoside	46, 53, 54
"	Ethanol	Ethyl- β -D-fructofuranoside	45, 46, 55
"	Propan-1-ol	n-Propyl- β -D-fructofuranoside	45, 46
"	Butan-1-ol	n-Butyl- β -D-fructofuranoside	45
"	Benzyl alcohol	Benzyl- β -D-fructofuranoside	46
"	Glycerol	Monoglyceryl- β -D-fructofuranoside	48, 50
"	Hydroquinone	4-Hydroxy-phenyl- β -D-fructofuranoside	56
"	Resorcinol	2-Hydroxy-phenyl- β -D-fructofuranoside	56

Table 1.2 Some of the reported fructotransferase activities of yeast invertase

The advent of molecular genetic manipulation techniques now opens up new possibilities to produce mutations which yield invertase with increased transferase/hydrolase activity ratios and broaden the enzyme's acceptor-substrate specificity so that other β -D-fructofuranosides (some of pharmacological interest) may be synthesized by transfructosylation from sucrose. The prospect of achieving such mutants are enhanced since the fructotransferase activity of invertase is unlikely to have been subjected to positive selection pressures during yeast evolution, and so there is a greater likelihood that more favourable sequence permutations may be found. From a biotechnological standpoint, yeast invertase is attractive because, in 'over-expressing' strains, the enzyme is secreted into the yeast growth medium [57], and systems for its immobilization in reactors have already been developed [58, 59].

As well as producing novel catalysts for use in organic synthesis, *in vitro* mutagenesis of the *SUC2* gene could provide more fundamental information regarding the active site of invertase. No tertiary structural information is yet available on invertase, and only two residues have been implicated in the active site of the enzyme, an aspartate and histidine residue, the former having been identified as Asp-23 [60]. On the strength of the involvement of these two residues, two mechanisms have been proposed for the mechanism of invertase action, both involving an enzyme-fructosyl intermediate [61, 62].

1.3 A Selection System to Isolate Sucrose Fructosyl-Transferase Activities

A selection system for yeast cells expressing invertase with desirable fructotransferase activity was designed based on the survival of a 'fructose-intolerant' strain in the presence of suitable acceptor substrates and sucrose. Cells of such a strain which express wild-type hydrolase activity will die owing to the release of free fructose from sucrose; cells expressing inactive mutations of *SUC2* will not grow, because they cannot cleave sucrose which will be the sole source of carbon provided; only cells displaying fructotransferase activity will form colonies, because the fructosyl moiety is not freed, but the glucose released is available as a carbon source.

To produce such a strain, a plan was formulated for genetically engineering a yeast so that it would display a condition similar to that observed in liver cells of humans suffering from 'hereditary fructose intolerance' [reviewed in 63]. This disorder is inherited as an autosomal recessive trait and its frequency has been estimated at approximately one in thirty thousand [64]. The condition results from a genetic abnormality in the peculiar metabolism of fructose in liver cells. In most cells (including yeast), fructose is phosphorylated by hexokinase (EC 2.7.1.1) to form fructose 6-phosphate which is subsequently phosphorylated by phosphofructokinase to form fructose 1,6-bisphosphate, which is cleaved by aldolase 'A' to form dihydroxy acetone phosphate and 3-phosphoglyceraldehyde. In liver cells, however, ketohexokinase (EC 2.7.1.3) phosphorylates fructose to fructose 1-phosphate which is directly cleaved to dihydroxyacetone phosphate and glyceraldehyde by aldolase 'B' [65] (see *Fig 1.4*). Patients with 'hereditary fructose intolerance' lack the latter enzyme, and dietary fructose, and compounds from which it can be formed, such as sorbitol and sucrose, are toxic to their liver cells because phosphorylation of fructose results in a potentially lethal sequestration of cellular phosphate as fructose 1-phosphate which cannot be metabolized further. Thus, cells in which ketohexokinase phosphorylates fructose, but which lack aldolase 'B', are intolerant of the hexose. Yeast possesses neither ketohexokinase nor aldolase 'B' and so the expression of the former enzyme in yeast will result in fructose intolerance (*Fig. 1.4*). The gene for ketohexokinase had not been cloned, but purification procedures for the enzyme from rat livers had been developed [66, 67]. Since the enzyme is relatively abundant in this tissue, of low molecular weight (34kDa) and is not glycosylated, it was felt that the cloning of its cDNA by exploiting partial amino acid sequence of the isolated protein was a realistic goal.

In designing the fructose intolerant yeast, capable of expressing the plasmid-borne mutated *SUC2* gene, and selecting fructotransferase mutants, several genetic lesions were thought to be desirable. The rationale for their inclusion, and the basis of the transferase screen are shown in *Figure 1.5*.

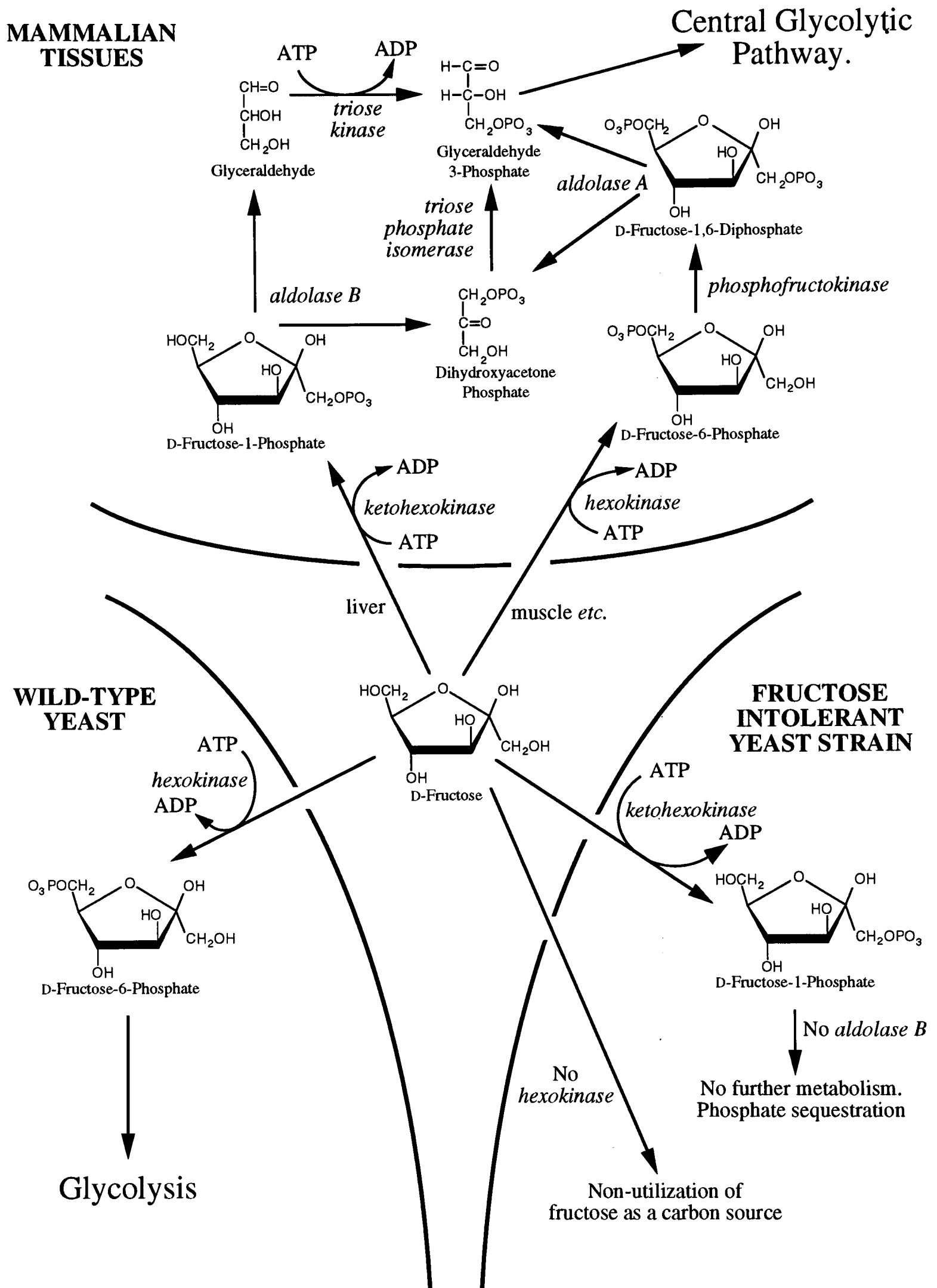


Figure 1.4 Fructose metabolism in mammalian tissues (top), showing the different pathways in liver (top left) and muscle (top right) tissues. A similar metabolic pathway to that of muscle occurs in wild-type yeast (bottom right). Expression of mammalian ketohexokinase in a yeast strain lacking endogenous hexokinases (bottom right) is expected to result in a condition similar to that found in individuals suffering from hereditary fructose intolerance, which is due to a lack of aldolase B in the liver (see top left).

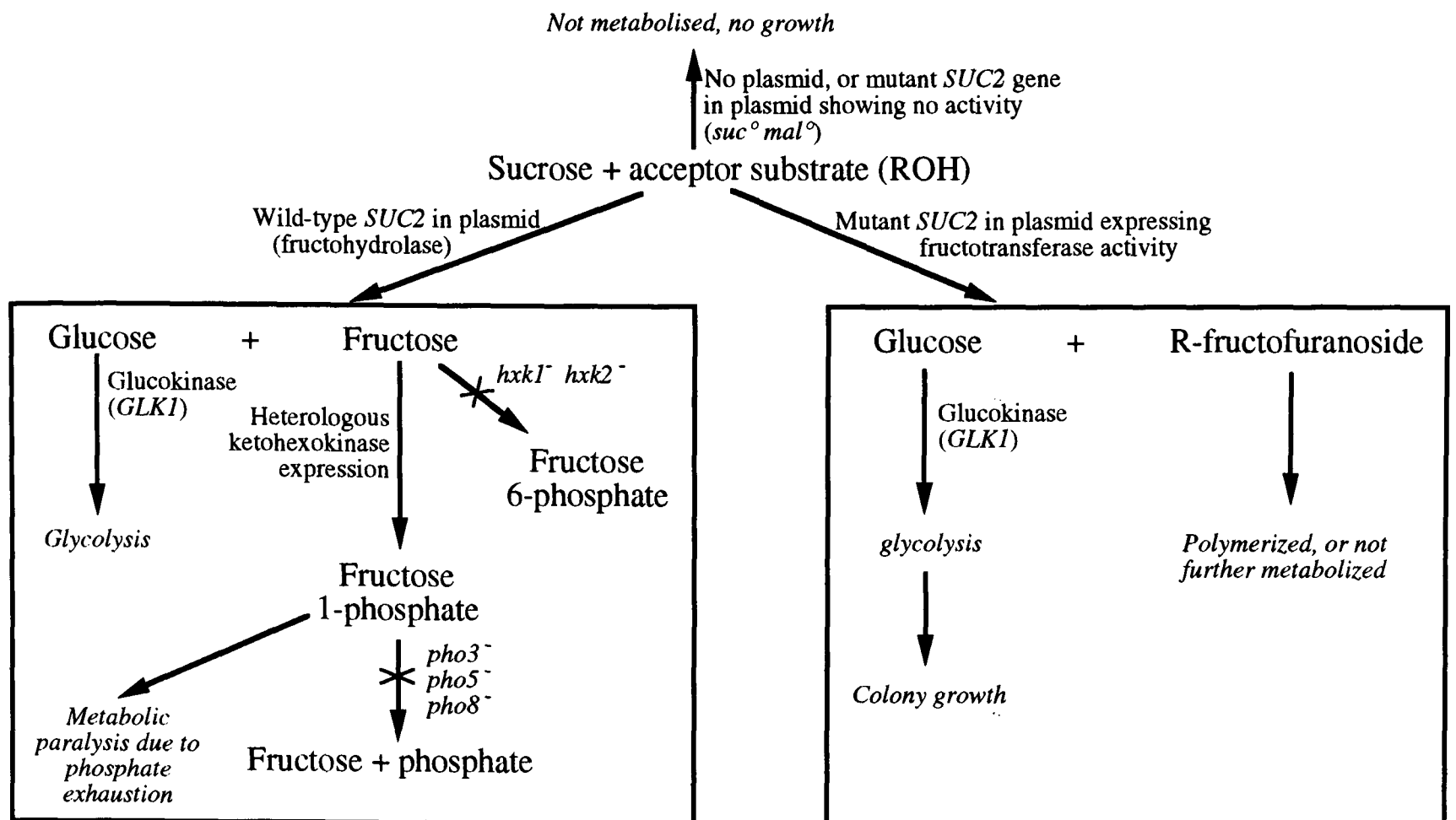


Figure 1.5 Metabolic consequences of the addition expression of various invertase mutations in a fructose-intolerant yeast strain, when supplied with sucrose as the sole carbon source. The absence of the *SUC2* gene, or mutations not showing any activity will not be able to grow, due to the lack of a metabolizable carbon source (top). Wild-type invertases (bottom left) hydrolyse sucrose, releasing free fructose, causing metabolic paralysis. Mutations conferring fructotransferase activity will not release fructose, and hence yeast growth occurs using the released glucose (bottom right).

The production and characterization of the fructose-intolerant yeast, and a preliminary screen of the *SUC2* gene that had been mutagenized are described in this thesis. For convenience, the work has been separated into six chapters reporting the different steps that led to the required strain. These are as follows:

- Chapter 2:** The breeding of a yeast strain with the required metabolic lesions, that would be used for the expression of ketohexokinase and mutated invertase.
- Chapter 3:** Purification of rat liver ketohexokinase to homogeneity, and the generation and sequencing of peptides to yield partial amino acid sequence.
- Chapter 4:** Using degenerate oligonucleotides, complementary to two peptides, as primers for the polymerase chain reaction, a region of the cDNA for ketohexokinase was amplified to generate a probe to screen a cDNA library at high stringency, .
- Chapter 5:** Use of the amplified region of ketohexokinase cDNA to screen a cDNA library and isolate a full length clone encoding the ketohexokinase coding region. This was then sequenced, the predicted amino acid sequence determined, and sequence comparisons performed with those present in sequence data banks.

- Chapter 6:** Expression of the ketohexokinase gene in the yeast strain designed for this purpose, and the characterization of the resultant fructose intolerance. Expression was initially from the episomal multicopy plasmid *pMA91*, and subsequently an integration plasmid was designed and constructed that targetted multiple copy integration into the tandemly repeated rDNA locus on yeast chromosome XII. Comparisons of fructose-intolerance and mitotic stability of the integrated and episomal plasmid were made.
- Chapter 7:** Amplification and cloning of the yeast *SUC2* gene, and its expression in yeast from a single-copy vector designed for this purpose. The *SUC2* gene was then mutagenized and cloned back into the expression vector. Relatively small libraries were transformed into the fructose-intolerant yeast strain and screened for mutants showing fructotransferase activity.
- Chapter 8:** Although no mutants that conferred the desired phenotype were isolated, limitations of time prevented a more exhaustive screening. A strategy is suggested that would allow a rapid characterization of the site of suitable mutations. It is envisaged that this method would allow the identification of mutations that affect activity, and not 'silent' ones. The use of the fructose intolerant yeast in the isolation of native genes that express a fructotransferase activity is also suggested.

1.4 References

1. Fuchs, A. (1987). Potentials for Non-Food Utilization of Fructose and Inulin. *Starch/Stärke*, **39**, 335-343.
2. John, P. (1992). Fructan, in: *Biosynthesis of the major crop products*, Wiley Biotechnology Series (Bryant, J.A. and Kennedy, J.F., Series Eds.), Wiley, Chichester, 55-69.
3. Pollock, C.J. and Chatterton, N.J. (1988). Fructans, in: *The Biochemistry of Plants. Volume 14 Carbohydrate* (Preiss, J., Ed.), Academic Press, New York, 109-140.
4. Hidaka, H., Hirayama, M. and Sumi, N. (1988). A Fructooligosaccharide-producing Enzyme from *Aspergillus niger* ATCC 20611. *Agric. Biol. Chem.*, **52**, 1181-1187.
5. Hirayama, M., Sumi, N. and Hidaka, H. (1989). Purification and Properties of a Fructooligosaccharide-producing β -Fructofuranosidase from *Aspergillus niger* ATCC 20611. *Agric. Biol. Chem.*, **53**, 667-673.
6. Dedonder, R. (1966). Levansucrase from *B. subtilis*. *Meths. Enzymol.*, **8**, 500-505.
7. Chambert, R. and Gonzy-Tréboul, G. (1976). Levansucrase of *B. subtilis*; kinetic and thermodynamic aspects of transfructylation process. *Eur. J. Biochem.*, **62**, 55-64.

8. Steinmetz, M., Le Coq, D., Aymerich, S., Gonzy-Tréboul, G. and Gay, P. (1985). The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. *Mol. Gen. Genet.*, **200**, 220-228.
9. Tang, L.B., Lenstra, R., Borchert, T.V. and Nagarajan, V. (1990). Isolation and characterization of levansucrase-encoding gene from *Bacillus amyloliquefaciens*. *Gene*, **96**, 89-93.
10. Clarke, M.A., Bailey, A.V., Roberts, E.J. and Tsang, W.S. (1991). Polyfructose: a New Microbial Polysaccharide, in: *Carbohydrates as Organic Raw Materials* (Lichtenthaler, F.W., Ed.), VCH, Weinheim, 169-181
11. Fleiger, M., Zelenkova, N.F., Sedmera, P., Kren, V., Novák, J., Rylko, V., Sajdi, P. and Reháček, Z. (1989). Ergot alkaloid glycosides from saprophytic cultures of *Claviceps*, I. Elymoclavine fructosides. *J. Nat. Prod.*, **52**, 506-510.
12. Eich, E., Eichberg, D., Schwarz, G., Clas, F. and Loos, M. (1985). Antimicrobial Activity of Clavines. *Arzneim.-Forsch./Drug Res.*, **35**, 1760-1762.
13. Szejtli, J., Henriques, R.D. and Castimeira, M. (1971). The acid hydrolysis of inulin. *Acta Chim. Acad. Sci. Hung.*, **70**, 379-389.
14. BeMiller J.N. (1972). Isolation and hydrolysis of alkali-stable inulin. *Carbohydr. Res.*, **21**, 154-155.
15. Wolf D.D. and Ellmore, T.L. (1975). Automated hydrolysis of nonreducing sugars and fructans from plant tissues. *Crop Sci.*, **15**, 775-777.
16. Hidaka, H., Eida, T., Takizawa, T., Tokunaga, T. and Tashiro, Y. (1986). Effects of Fructooligosaccharides on Intestinal Flora and Human Health. *Bifidobacteria Microflora*, **5**, 37-50.
17. Hidaka, H. and Hirayama, M. (1991). Useful characteristics and commercial applications of fructo-oligosaccharides. *Biochem. Soc. Trans.*, **19**, 561-565.
18. Fuchs, A. (1991). Current and potential food and non-food applications of fructans. *Biochem. Soc. Trans.*, **19**, 555-560.
19. Cassady, J.M. and Floss, H.G. (1977). Ergolines as Prolactin and Mammary Tumor Inhibitors. *Lloydia*, **40**, 90-106.
20. Dodart, D.J. (1676) des Scavans, Académie des Sciences, de l'imprimerie royale, Paris, 76.
21. Shelensnyak, M.C. (1957). Some experimental studies on the mechanism of ova-implantation in the rat. *Rec. Progr. Hormone Res.*, **13**, 269-322.
22. Floss, H.G, Cassady, J.M. and Robbers, J.E. (1973). Influence of ergot alkaloids on pituitary prolactin and prolactin-dependent processes. *J. Pharm. Sci.*, **62**, 699-715.
23. Cassady, J.M., Li, G.S., Spitzner, E.B., Floss, H.G. and Clemens, J.A. (1974). Ergot Alkaloids. Ergolines and Related Compounds as Inhibitors of Prolactin Release. *J. Med. Chem.*, **17**, 300-307.
24. Barfknecht, C.F. and Rusterholz, D.B. (1974). Inhibition of Prolactin by Ergoline Congeners. *J. Med. Chem.*, **17**, 308-312.

25. Boot, L.M. (1970). Prolactin and mammary gland carcinogenesis— The problem of human prolactin. *Int. J. Cancer*, **5**, 167-175.
26. Sweeney, M.J., Poore, G.A., Kornfeld, E.C., Bach, N.J., Owen, N.V. and Clemens, J.A. (1975). Activity of 6-Methyl-8-substituted Ergolines against the 7,12-Dimethyl-benz[a]anthracene-induced Mammary Carcinoma. *Cancer Res.*, **35**, 106-109.
27. Kren, V., Flieger, M. and Sajdl, P. (1990). Glycosylation of ergot alkaloids by free and immobilized cells of *Claviceps purpurea*. *App. Microbiol. Biotech.*, **32**, 645-650.
28. Kren, V. (1992). Fructosylation of ergot alkaloids by yeast invertase. *Biotech. Lett.*, **14**, 769-772.
29. Green, J.W. (1966). The glycofructosides. *Adv. Carbohydr. Chem.*, **21**, 95-142.
30. Fischer, E. (1895). Ueber die Verbindungen der Zucker mit den Alkoholen und Ketonen. *Berichte der Deutschen Chemischen Gesellschaft* **28**, 1160-1165.
31. Moody, W. and Richards, G.N. (1982). Effect of traces of acids on reaction of sucrose in dimethyl sulphoxide. *Carbohydr. Res.*, **108**, 13-22.
32. Horvard, A.E. and Metzenberg, R.L. (1963). A simple chemical synthesis of β -methyl-fructofuranoside, a β -fructofuranosidase substrate. *Biochim. Biophys. Acta*, **108**, 165-167.
33. Szeja, W. and Kiersznicki, T. (1978). Glycosylation of kestoses. Part II. The study on kinetics of D-fructose glycosidation. *Polish J. Chem.*, **52**, 2097-2105.
34. Moody, W. and Richards, G.N. (1983). Formulation and equilibration of D-fructosides and 2-thio-D-fructosides in acidified dimethyl sulfoxide: synthetic and mechanistic aspects. *Carbohydr. Res.*, **97**, 201-213.
35. Moody, W. and Richards, G.N. (1981). Thermolysis of sucrose in the presence of alcohols. A novel method for the synthesis of D-fructofuranosides. *Carbohydr. Res.*, **97**, 247-255.
36. Donaldson, I.A. and Jørgensen, J.H. (1988). Barley powdery mildew “invertase” is an alpha-glucosidase. *Carlsberg. Res. Commun.*, **53**, 421-430.
37. Khan, N. and Eaton, N.R. (1967). Purification and characterization of maltase and α -methyl glucosidase from yeast. *Biochim. Biophys. Acta*, **146**, 173-180.
38. Urlaub, H. and Wober, G. (1978). α -Glucosidase, a membrane-bound enzyme of glucan metabolism in *Bacillus amyloliquefaciens*. *Biochim. Biophys. Acta*, **522**, 161-173.
39. Kolinska, J. and Semenza, G. (1967). Studies on intestinal sucrase and on sugar transport V. Isolation and properties of sucrase-isomaltase from rabbit small intestine. *Biochim. Biophys. Acta*, **146**, 181-195.
40. Carlson, M. Osmond, B.C. and Botstein, D. (1981). Mutants of yeast defective in sucrose utilization. *Genetics*, **98**, 25-40.
41. Gancedo, A. and Serrano, R. (1989). Energy-Yielding Metabolism, in *The Yeasts* (Rose, A.H. and Harrison, J.S., Eds.), Academic Press, London, vol 3, 205-259.
42. Carlson, M. and Botstein, D. (1982). Two Differentially Regulated mRNAs with Different 5' Ends Encode Secreted and Intracellular Forms of Yeast Invertase. *Cell*, **28**, 145-154.

43. Carlson, M., Taussig, R., Kustu, S. and Botstein, D. (1983). The Secreted Form of Invertase in *Saccharomyces cerevisiae* Is Synthesized from mRNA Encoding a Signal Sequence. *Mol. Cell. Biol.*, **3**, 439-447.
44. Fischer, H., Knots, L. and Fellig, J. (1951). Propriétés de l'invertase purifiée. *Helv. Chim. Acta.*, **137**, 1132-1138.
45. Whelan, W.J. and Jones, D.M. (1953). β -Methyl Fructoside as a Substrate in Transfructosylation. *J. Biochem.*, **54**, xxxiv.
46. Bacon, J.S.D. and Bell, D.J. (1957). A Proof of the Structure of Methyl β -D-Fructofuranoside Prepared Enzymatically from Sucrose and Methanol. *J. Chem. Soc.*, **1957**, 3581-3585.
47. Edelman, J. (1954). Transfer Reactions Catalysed by some Sucrase Preparations. *Biochem. J.*, **57**, 22-33.
48. Bacon, J.S.B. and Edelman, J. (1950). The Action of Invertase Preparations. *Arch. Biochem. Biophys.*, **28**, 467-468.
49. Blanchard, P.H. and Albon, N. (1950). The Inversion of Sucrose; a Complication. *Arch. Biochem. Biophys.*, **29**, 220-222.
50. Bacon, J.S.B. (1953). Transfructosylation. *Ann. Rep. Chem. Soc. London*, **50**, 281-287.
51. Bacon, J.S.D. (1954). The Oligosaccharide Produced by the Action of Yeast Invertase Preparations on Sucrose. *Biochem. J.*, **57**, 320-328.
52. Anderson, B., Thiesen, N. and Broe, P.E. (1969). The Transferring Activity of β -Fructofuranosidase from Yeast: A Quantitative Description. *Acta Chem. Scand.*, **23**, 2367-2374.
53. Bacon, J.S.D. (1952). Transfructidation by a Yeast invertase Preparation. *Biochem. J.*, **50**, xviii.
54. Ishizawa, K., Iriki, Y. and Miwa, T. (1957). Enzymatic Transfructosylation and its applicability to the Preparation of Alkyl β -D-fructofuranosides. *Sci. Rep. Tokyo Kyoiku Daigaku*, **8**, 102-113.
55. Baseer, A. and Shall, S. (1971). The enzymatic synthesis of fructosides using purified external β -fructosidase from *Saccharomyces cerevisiae*. *Int. J. Biochem.*, 503-506.
56. Nakamura, S. (1965). Synthesis of Aryl β -D-Fructofuranoside by Enzymatic Transglycosylation. *Sci. Rept. Tokyo Kyoiku Daigaku*, **12**, 55-72.
57. Neumann, N.P. and Lampen, J.O. (1967). Purification and Properties of Yeast Invertase. *Biochemistry*, **6**, 468-475.
58. Combes, D. and Monsan, P. (1982). Continuous hydrolysis of concentrated sucrose solutions by immobilized invertase. *Enzyme Eng.*, **6**, 305-306.
59. Monsan, P. and Combes, D. (1984). Application of immobilized invertase to continuous hydrolysis of concentrated sucrose solutions. *Biotech. Bioeng.*, **26**, 347-351.
60. Reddy, V.A. and Maley, F. (1990). Identification of an Active-site Residue in Yeast Invertase by Affinity Labeling and Site-directed Mutagenesis. *J. Biol. Chem.*, **265**, 10817-10820.

61. Waheed, A. and Shall, S. (1971). Chemical modification of the active site of yeast invertase. *Biochim. Biophys. Acta*, **242**, 172-186.
62. Straathof, A.J.J., Vrijenhoef, J.P., Sprangers, E.P.A.T., van Bekkum, H. and Kieboom, A.P.G. (1988). Enzymatic formation of β -D-fructofuranosides from fructose: Activity and selectivity of invertase in mixtures of water and alcohol. *J. Carbohydr. Chem.*, **7**, 223-238.
63. Gitzelmann, R., Steinmann, B. and van den Berghe, G. (1989). Disorders of fructose metabolism, in *The Metabolic Basis of Inherited Disease* (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., Eds.), MacGraw-Hill, New York, vol 1, 399-424.
64. Cornblath, M. and Scharz, R. (1976). Disorders of carbohydrate metabolism in infancy, 2nd ed., Saunders, Philadelphia, 322.
65. Van den Berghe, G. (1986). Fructose: Metabolism and Short-Term Effects on Carbohydrate and Purine Metabolic Pathways. *Prog. Biochem. Pharmacol.*, **21**, 1-32.
66. Adelman, R.C., Ballard, F.J. and Weinhouse, S. (1967). Purification and Properties of Rat Liver Fructokinase. *J. Biol. Chem.* **242**, 3360-3365.
67. Sánchez, J.J., González, N.S. and Pontis, H.G. (1971). Fructokinase from rat liver. I. Purification and properties. *Biochim. Biophys. Acta* **227**, 67-78.

Chapter 2

Development of the TDS Yeast for Use as a Selection Strain

2.1 Introduction

The requirement in this work for an expression and selection organism is perfectly fulfilled by the budding yeast *Saccharomyces cerevisiae*. This yeast has been used by man for centuries, both in brewing and baking. However, it is only in this century that a better understanding of its biology been gained, which has enabled genetic manipulation to give many desired characteristics. This was achieved by a combination of chemical and physical mutagenesis of the yeast chromosome, and, more recently, by the use of recombinant DNA technology to undertake precise manipulation of both the cells' own genes, and introduction of heterologous genes.

It was the pioneering work of Winge [1] that first observed homothallism in *S. cerevisiae*, in which diploid cells developed in a culture grown from a single spore. These diploid cells could sporulate but did not have mating ability. Later, Lindegren [2] succeeded in the isolation of mutants that resulted in stable heterothallic haploid (n) cultures from single spores, which were unable to sporulate. Until then, *Saccharomyces* strains existed stably in diploid ($2n$) or polyploid states ($>2n$), with consequent multiple copies of any given gene. Random mutagenesis of these strains would therefore present difficulties in isolation of mutant phenotypes, as the effect of the mutant could be suppressed by the other wild-type gene. In isolating stable haploid cells, Lindegren enabled the development of stable mutations of yeast genes to be generated and propagated. By simple crossing of opposite mating type, hybrid strains could be developed of parent strains of differing phenotypes.

Upon nitrogen or carbon starvation, diploid ($2n$) cells may undergo meiotic division and sporulation to form a tetrad of haploid cells known as an ascospore. This ascospore, or ascus, comprises two pairs of cells of opposite mating types, known as a and α . The ascospore is surrounded by a thick spore wall that enables it to survive the starvation conditions. Should the spore's environment improve, it is able to regenerate into the vegetative form, and the cells mate to form diploids [3]. However, it is possible to dissect the tetrad into its individual spores and allow these to grow in separate haploid colonies [4]. With wild-type yeast, the mating type

of the individual spores is able to switch [5], which allows mating within the colony, and consequently, diploid vegetative growth. Haploids can be stably maintained in their *a* or α states only by a genetic block on this switching [6]. It was the isolation of a mutant yeast strain that was unable to switch mating type that has allowed yeast genetics to develop so rapidly.

2.1.1 The Life Cycle of *Saccharomyces cerevisiae*.

Both haploid and diploid cells will grow in a nutrient rich media with a doubling time of about ninety minutes, although doubling times increase as the nutrient supply is reduced. Cell division is asymmetric, with a daughter cell growing as a small bud on the surface of the original or mother cell [7]. This daughter cell will either remain attached to the mother cell, allowing the cells to flocculate, or can become detached, and so cells remain unicellular. It has recently been reported that, in certain conditions of starvation, *S. cerevisiae* is able to grow in a filamentous fashion [8], and cells at the periphery of a colony are able to grow and divide in a unidirectional fashion, enabling the yeast to ‘explore’ the surrounding medium for nutrients.

The mating type of *S. cerevisiae* is controlled by the *MAT* locus on chromosome III [reviewed in 5], and the mating type is dependent on which of the *MATa* or *MAT α* alleles is expressed, *MATa* expression conditioning mating type *a*. These alleles control the expression of the α and *a* haploid-specific genes, amongst which are the *MF α* and *MFa* genes that encode the α -factor and the *a*-factor respectively. These are small hormone-like peptides that induce cells of the opposite mating type to mate. Upon mating, the presence of both the *MATa* and *MAT α* loci, in the *a*/ α diploid formed, leads to the repression of the α -specific and *a*-specific genes and the induction of sporulation-specific genes. During starvation, these genes encode proteins specific for meiosis and spore formation.

Both of the *MATa* and *MAT α* alleles are present in the yeast genome at loci outside of the *MAT* locus. The loci that these occupy are known as the *HML* and *HMR*, in which the *MAT* genes are silent. The allele at the *MAT* locus can be altered by a gene conversion from either the *HML* or *HMR* loci, resulting in a change of mating type. The gene conversion at the *MAT* locus which results in “mating type switching” is under a complex regulation, involving the *HO* (homothallism) gene. Strains with mutations in the *HO* gene are unable to switch their mating type and are termed heterothallic. The different behaviours of the Winge and Lindegren strains were due to the presence of the *HO* gene in the Winge strains, and the *ho* mutant allele in the Lindegren strains (*Fig. 2.1*). It is the latter strains that are now routinely used in the laboratory.

Upon mixing, haploid cells of opposite mating type form a mating aggregate, and most of the *a* and α cells become agglutinated to neighbours of opposite mating type [10]. This mating

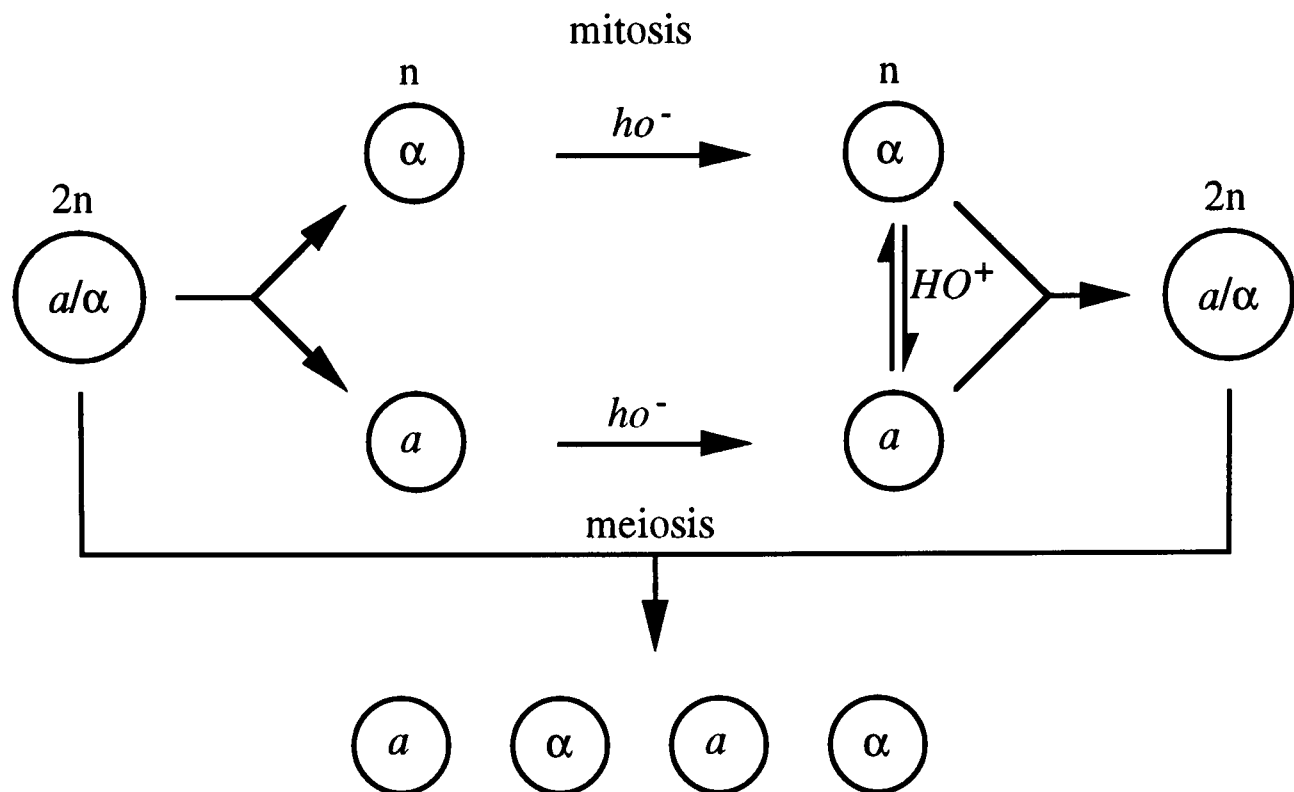


Figure 2.1 A simplified representation of the life cycle of *Saccharomyces cerevisiae*. The two mating type alleles of the *MAT* locus are represented by *a* and α , the diploid ($2n$) strain containing both alleles, and the haploid (n) strains containing one active allele. The wild type yeast strain is homothallic (HO^+) and able to switch mating type in the haploid state, whereas the common laboratory strains are heterothallic (ho^-) and remain either α or *a*. (Taken from Tuite, M. F. (1992) [9].)

response in yeast is initiated by the binding of the mating factor of the opposite haplotype. Each haploid cell type secretes this specific oligopeptide pheromone into the culture medium, which then triggers the biochemical changes needed to allow the mating to proceed. These responses include, in order, cell-surface alterations that enhance the strength and selectivity of cell-cell contacts, a transient arrest of cell growth in the G_1 phase that effectively synchronizes the growth stages of the mating partners, and initiates the new cell wall growth appropriate to cell fusion rather than normal budding. (Diploid cells neither produce, nor respond to, these pheromones.) After about one cell cycle, usually about 90-120 minutes, growth has stopped in the G_1 phase of the cell-division cycle. Sixty to ninety minutes later the agglutinated cells undergo polar elongation towards each other to form either pear-shaped “shmoo” or even longer and more bizarre forms with conjugation-tube projections. Successfully mating cells can then undergo cell-envelope dissolution and resynthesis, in a controlled manner that allows cell fusion but avoids cell lysis. Conjugation begins with cytoplasmic fusion to produce a zygote containing two haploid nuclei. Karyogamy, the fusion of the two nuclei, generally begins soon after cytoplasmic fusion. Two nuclei, whose spindle pole bodies are orientated toward the site of cell fusion, migrate toward each other and fuse to form a zygote with a single diploid nucleus. These zygotes are then able to continue vegetative growth, and consequent production of diploid daughter cells, within an hour of their formation (Fig. 2.2).

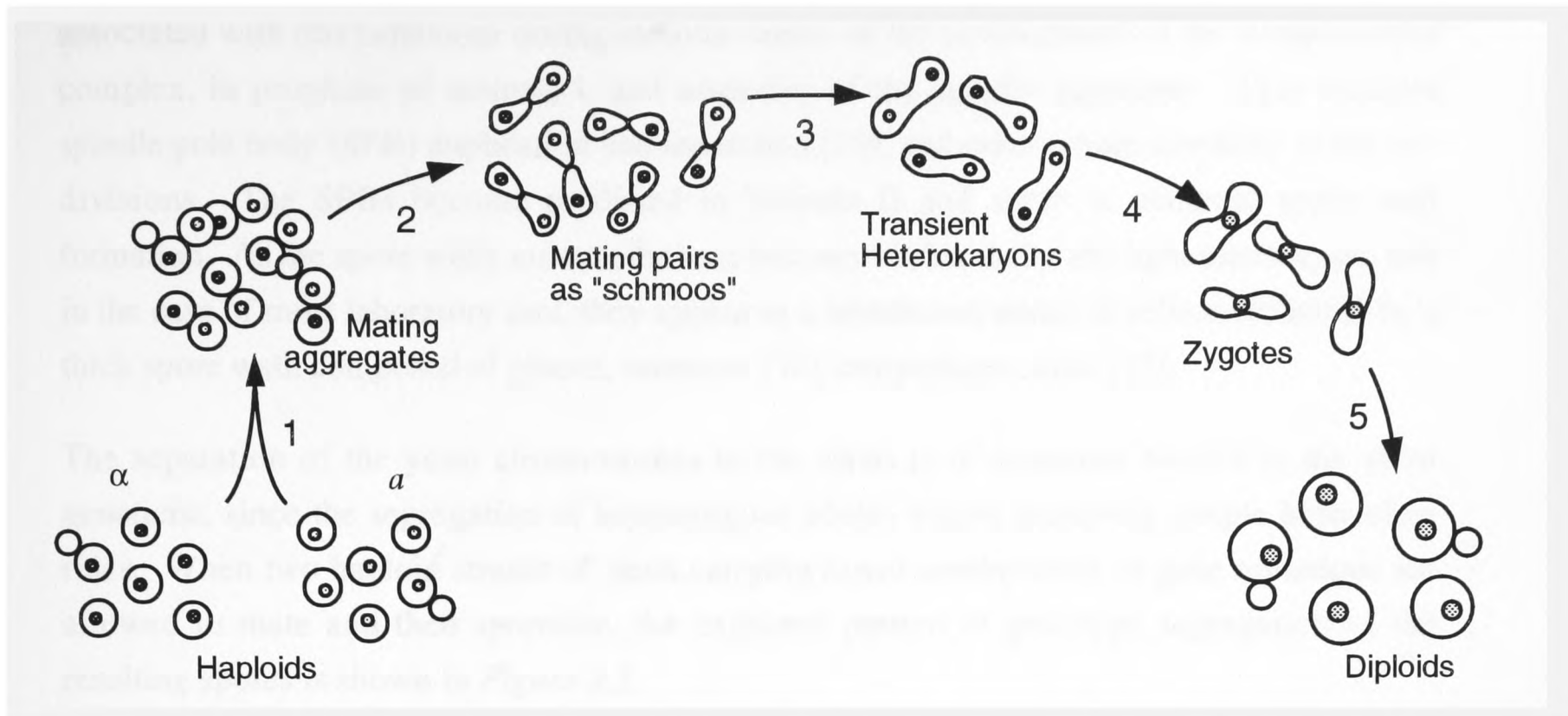


Figure 2.2 Diagram representing the mating and conjugation process in yeast. Cells of opposite mating type come together and form a mating aggregate by cell-cell adhesion (1). The cell cycle is arrested and under the influence of the α - and a -factors the cells grow towards each other as "schmoos" (2). The cytoplasm of the cells fuse (3) to form heterokaryons, and following nuclear fusion (4), zygotes are formed. The cells can then continue vegetative growth (5) and exist as diploids.

Cells of opposite mating types mate efficiently, whereas those of like mating types mate only rarely [5]. In contrast to a/a and α/α cells, which behave as their respective haploids, a/α cells are unable to mate but can be induced to undergo meiosis and sporulation. Sporulation of yeast includes meiosis and ascospore development. When diploid cells undergo starvation conditions, vegetative growth ceases and the cells become committed to meiosis and sporulation [11], in which each diploid cell is transformed into an ascus usually consisting of four haploid spores. The ascospores and haploid cells arising from them after germination constitute the haplophase.

Sporulation can be induced in cells in stationary phase, especially in the presence of a non-fermentable carbon source, but maximal sporulation is achieved when cells are transferred from growth medium (*e.g.* presporulation medium 2.2.2.4) to sporulation medium (2.2.2.5) [12]. Under these conditions, no cell division occurs other than the completion of the cells' current cell cycle to the G_1 phase. In most strains, sporulation is not synchronous, but the time of initiation of spore production is similar, and occurs approximately 10 hours after exposure to sporulation medium [10]. Maximal sporulation is, however, achieved by 36-48 hours in glucose-grown, acetate-induced cells [13].

Once cells have reached the G_1 phase of vegetative growth, they then enter the meiotic cell cycle, during which the chromosomes undergo premeiotic DNA synthesis, pairing, genetic recombination, reduction and equatorial segregation [14]. The morphological changes

associated with this behaviour during meiosis centre on the development of the synaptonemal complex, in prophase of meiosis I, and assembly of the spindle apparatus. This includes spindle pole body (SPB) duplication and separation [15], and microtubule assembly in the two divisions. The SPBs become modified in meiosis II and serve to nucleate spore wall formation. As the spore walls mature, the asci become visible under the light microscope, and in the case of most laboratory asci, they appear as a tetrahedral tetrad of cells surrounded by a thick spore wall, composed of glucan, mannose [16] and perhaps chitin [17].

The separation of the yeast chromosomes in the ascus is of immense benefit to the yeast geneticist, since the segregation of heterozygous alleles follow relatively simple Mendelian rules. When two haploid strains of yeast carrying novel combination of gene mutations are allowed to mate and then sporulate, the expected pattern of genotype segregation of the resulting spores is shown in *Figure 2.3*.

2.1.2 Conjugation of Haploids and Dissection of Ascospores.

When haploid strains of opposite mating types are allowed to mate and sporulate, the resulting spores can be dissected and the individual spores separated. This is achieved by the digestion of the ascus wall with either zymolyase or glusulase to release the spores for dissection. The digested mix can then be placed on an agar plate and the tetrads are dissected with the aid of a glass loop or needle [4]. The individual spores are isolated and allowed to generate into distinct colonies under non-selective conditions, before determination of the phenotypes.

For normal meiosis of a diploid cell with a heterozygous allele, the spores generated will usually give a 2:2 segregation, although gene conversion may lead to 1:3 segregations, albeit at a low frequency [18]. With unlinked genes, the separation of the alleles will generate spores such that 1 in 2^n spores will possess a given phenotype, where n is the number of heterozygous alleles. In the case of *Figure 2.3*, this means that 1 in 2^2 , or 1 in 4, spores will have any given phenotype (*i.e.* AB, Ab, aB or ab). With alleles linked on the same chromosome, their segregation into the spores is more complicated. Since the chromosomes are able to undergo recombination by crossing-over during meiosis [reviewed in 19], the segregation of linked heterozygous alleles is dependent on their distance apart on the chromosome. Consequently, genes lying very close to each other are relatively unlikely to recombine, but those separated by large distances are more likely to undergo this process. It can therefore be useful to consult a genetic map to determine the likelihood of the required allele segregation.

In practical terms, a number of tricks can be used to make this mating and dissection of the resulting tetrads easier. If the haploid parent strains used have different auxotrophic markers, the resulting diploid's growth will be independent of these supplements, provided these

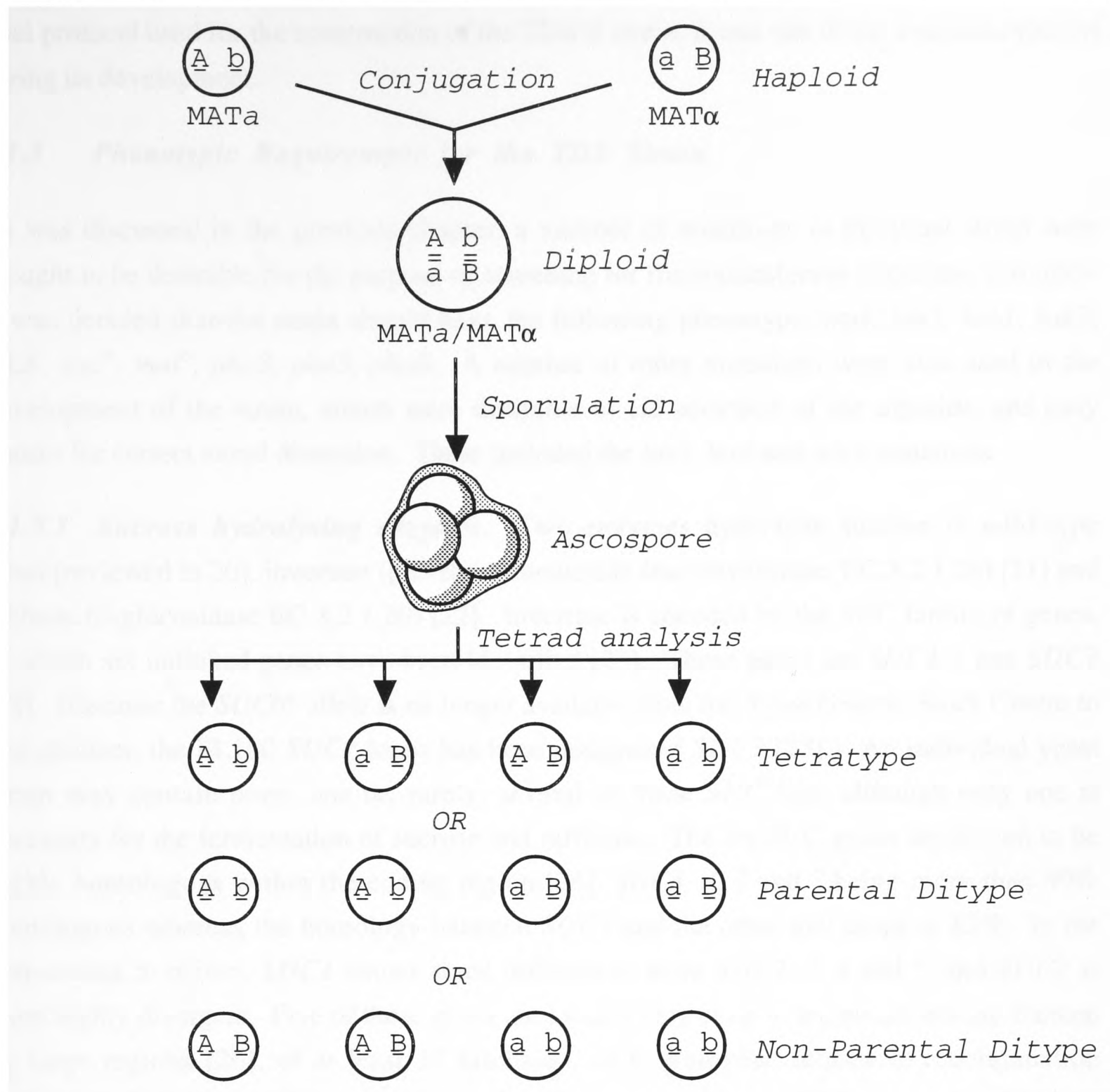


Figure 2.3 Tetrad analysis used to generate novel combinations of gene mutations. The genes *A* and *B* (and their respective mutant alleles, *a* and *b*) are unlinked, and the mating of the *Ab* and *aB* strains followed by the possible ascus tetratypes are shown. In the tetratype, only two recombinant spores occur, and the parental ditype, all the spores are the same as the parent haploids and in the nonparental ditype, all the spores are recombinant. The mating types of the spores is not indicated, but will follow a similar pattern to the other alleles, with a 2:2 segregation. (Taken and adapted from Tuite (1992) [9].)

markers are recessive (e.g. *HIS3*, *leu2* $\times$ *his3*, *LEU2* will give the diploid *HIS3:his3*, *LEU2:leu2*, which will have a *LEU*, *HIS* phenotype). Consequently, plating of a mating mix on to defined medium lacking these supplements will only allow diploid growth. The selection of haploid spores can be aided by employing a diploid strain that is heterozygous for a recessive mutation that confers a resistance to a toxin, such as *can1* that confers canavanine resistance. Plating such a sporulated diploid strain on a canavanine-containing medium will result in growth of only those haploid spores carrying this resistance mutation. These may then be screened for other mutations of interest. Although this strategy does not appear in the

final protocol used for the construction of the TDS II strain, it was one of the avenues explored during its development.

2.1.3 Phenotypic Requirement for the TDS Strain.

As was discussed in the previous chapter, a number of mutations in the yeast strain were thought to be desirable for the purpose of screening for fructotransferase enzymatic activities. It was decided that the strain should have the following phenotype: *trp1*, *leu2*, *hxl1*, *hxl2*, *GLK*, *suc*^o, *mal*^o, *pho3*, *pho5*, *pho8*. A number of other mutations were also used in the development of the strain, which were desirable to aid selection of the diploids, and easy checks for correct tetrad dissection. These included the *his3*, *his4* and *ade2* mutations.

2.1.3.1 Sucrose hydrolysing enzymes. Two enzymes hydrolyze sucrose in wild-type yeast [reviewed in 20], invertase (β -D-fructofuranoside fructohydrolase, EC 3.2.1.26) [21] and maltase (α -glucosidase EC 3.2.1.20) [22]. Invertase is encoded by the *SUC* family of genes, of which six unlinked genes have been identified [23]. These genes are *SUC1-5* and *SUC7* [24]. (Because the *SUC6*⁺ allele is no longer available from the Yeast Genetic Stock Centre to test allelism, the FL100 *SUC*⁺ locus has been designated *SUC7* [23].) An individual yeast strain may contain none, one or, rarely, several of these *SUC* loci, although only one is necessary for the fermentation of sucrose and raffinose. The six *SUC* genes are known to be highly homologous within the coding region [26], *SUC3*, 4, 5 and 7 being more than 99% homologous whereas the homology between *SUC2* and the other loci drops to 82%. In the non-coding 5' region, *SUC1* shows some differences from *SUC2*, 3, 4 and 5, and *SUC2* is more highly divergent. Five of these genes are located very close to telomeres and are flanked by large regions [26], of at least 12 kilobases, of homologous sequence; recombination between telomeres could account for the dispersal of these *SUC* genes to different chromosomes. *SUC2* is not located within a telomere and has no restriction map homology to the regions adjacent to the other *SUC* genes [25]. The relationship between *SUC2* and the other *SUC* genes is therefore not obvious.

The enzymatic properties of the invertases encoded by these six genes are very similar, although the level of activity of strains expressing the different genes varies widely [27]. Strains expressing *SUC4* show the highest level of expression, whereas those expressing *SUC7* are too low even to be able to ferment raffinose. All of the *SUC* genes are transcribed into two mRNAs, with starting points about 60bp apart [25, 28]. The shorter mRNA is constitutively expressed and encodes a minor non-glycosylated cytoplasmic form of invertase of unknown function. The larger 1.9kb mRNA encodes a secreted, glycosylated form of the enzyme, the expression of which is carbon catabolite (most notably glucose) repressed [29].

A naturally silent allele (*suc2^o*) of the *SUC2* locus has been found [30] that does not direct the synthesis of active invertase, either internally or externally, although it has been found to produce both the 1.8kb (internal) mRNA and 1.9kb mRNA under normal regulation by glucose [28]. It is this silent allele that was available for the construction of the TDS strain.

S. cerevisiae is also able to hydrolyse sucrose with α -glucosidases, such as maltase and α -methylglucosidase, the former is also able to hydrolyse maltose (α -1-4-) but not isomaltose (α -1-6-) [20]. Like invertase, the maltase genes exist in a variety of locations on the yeast chromosomes, and five such loci have been identified (*MAL1-4* and *MAL6*) [31]. Unlike invertase, maltase is only present as an intracellular enzyme [32]. In strains carrying any of the *MAL1*, 2, 3 and 6 genes, which are all inducible, the basal level of enzyme activity is low, and insufficient to give a positive fermentation response to the presence of sucrose, and it has been suggested that sucrose is unable to induce these genes [33, 34]. However, if these genes are induced by maltose, the cells will rapidly acquire the ability to ferment sucrose. The *MAL4* gene, which permits constitutive synthesis of maltase, can ferment sucrose in the absence of either maltose or the *SUC* genes.

Like the *SUC* family, there also exist strains of yeast lacking any of the active *MAL* alleles, and these strains are similarly termed *mal^o*. It seems that some of the *MAL* and *SUC* alleles are closely linked on the yeast chromosomes [32], and genetic and physical maps would seem to confirm this (*SUC1* and *MAL1* are on the telomere of 7R, and *SUC3* and *MAL3* are on the telomere of 2R [35]), and like most of the *SUC* alleles, those *MAL* genes that have been mapped occur on telomeres. Evidence suggests that these two alleles (*suc^o*, *mal^o*) will segregate together when tetrads are dissected [36], although there is a small possibility that the alleles could segregate away from each other by recombination.

2.1.3.2 Primary metabolism of fructose and glucose. The primary metabolism of these hexoses by phosphorylation at their C-6 is achieved by the members of the hexokinase family, hexokinases PI, PII and glucokinase [20]. These proteins are coded for by the *HXK1*, *HXK2* and *GLK1* genes, respectively. There is a widely held assumption that glucokinase shows substrate specificity for glucose [37]. However, mammalian hexokinases and glucokinase are capable of phosphorylating both glucose and fructose *in vitro*, and their relative substrate specificities are about the same (each has a ten times lower K_m for glucose than fructose) [38]. The much higher K_m values observed for mammalian glucokinase does, however, effectively preclude it from the primary metabolism of fructose. The observation that the hexokinase double mutants of yeast (*hvk1*, *hvk2*), in the presence of the wild type *GLK1* gene, are unable to utilise fructose as a sole carbon source, but will grow well when supplied with glucose [39] suggests a similar situation occurs in this organism. The rôle of phosphorylating fructose is therefore fulfilled by these two hexokinase isozymes *in vivo*.

Unlike the *SUC* and *MAL* genes, the two hexokinase genes are not tightly linked, but exist on separate chromosomes (*HXK1* on 6R and *HXK2* on 7L) [35].

2.1.3.3 Non-specific phosphatase genes. A large number of genes have been identified in yeast that are responsible for phosphatase activities and the control of their transcription [reviewed in 40]. Although a number of phosphatases have been identified, only three are thought to have non-specific activities [41], and be present at high levels, and it was a lack of these that was identified as being desirable in the final TDS strain.

In yeast there are two non-specific acid phosphatases (orthophosphoric monoester phosphohydrolase; EC 3.1.3.2), which are both highly glycosylated and located in the periplasmic space [42], one of which is produced constitutively (*PHO3*), and the other is repressed by inorganic phosphate (*PHO5*) [43]. Yeast alkaline phosphatases (EC 3.1.3.1), however, are located intracellularly, and cellular extracts were found to contain two activities at alkaline pH. One of these was found to be a nonspecific alkaline phosphatase (*PHO8*) [44], which is also a glycoprotein repressible by inorganic phosphate [45], and the other has restricted specificity, and was thus not important to this study.

The constitutive acid phosphatase is coded for by the *PHO3* gene, and requires two additional complementary genes, *PHO6* and *PHO7*, for its synthesis [46]. *PHO2*, *PHO80*, *PHO81*, *PHO85* and a gene cluster, *pho82-PHO4*, all participate in the expression and response to inorganic phosphate of the repressible acid phosphatase enzyme *PHO5*. The two enzyme structural genes, *PHO5* and *PHO3*, are adjacent, but separable by recombination, on chromosome II. The strain used in this work was developed by Albert Hinnen [46] as described in Figure 2.4. This strain is effectively devoid of acid phosphatase activity, but may still contain small activities derived from other minor acid phosphatases such as *PHO10* and *PHO11* [40]. Due to the nature of this construction, the *pho3*, *pho5*, *URA3* phenotype conferred by it can be considered as a single mutation for tetrad phenotype selection.

The nonspecific alkaline phosphatase, encoded by the *PHO8* gene on chromosome IX, is repressible by inorganic phosphate, although a basal level of activity is retained in a repressed state [45]. It, too, is regulated by the same genes as *PHO5*, except it has no requirement for *PHO2*, and *PHO9* is indispensable for the alkaline, but not the acid, phosphatase.

2.1.3.4 Auxotrophic markers. A number of auxotrophic markers were used in the development of the TDS strain, and some were retained in the final strain to allow their use to select for markers on plasmids. The two required in the TDS strain were *leu2* and *trp1*, the former encoding β -isopropylmalate dehydrogenase [47, 48], and the latter N-C5'phosphoribosyl-anthranilate isomerase [49]. A variant of the first of these two markers (*leu2-d*) is especially useful for selection of high copy number plasmids. This gene is deficient in some of

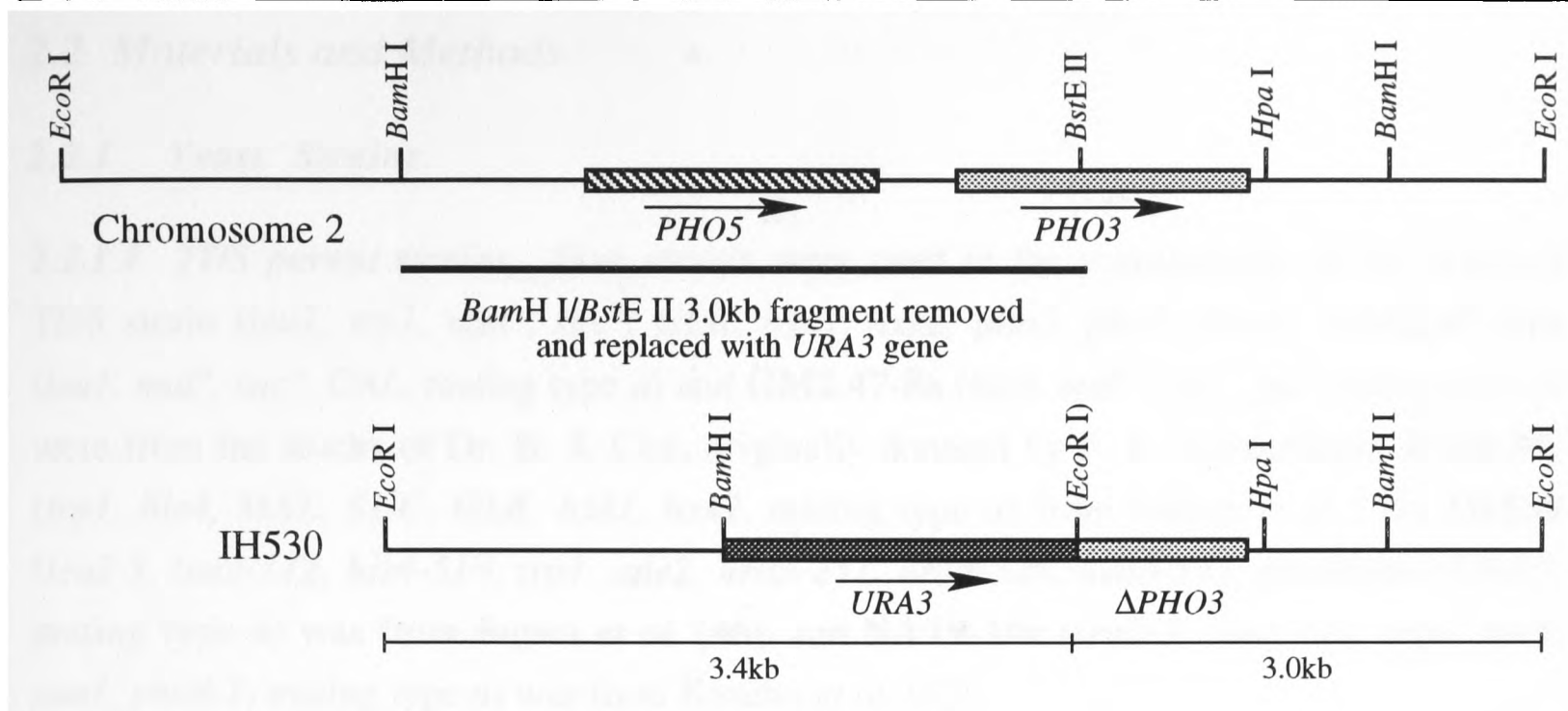


Figure 2.4 Construction of the IH530 strain in the laboratory of A. Hinnen. The *Bam*H I/*Bst*E II fragment containing the *PHO5* gene and the 3' end of the *PHO3* gene was removed and replaced with a *URA3* gene. This fragment was then used to transform a *ura3-251*, *ura3-328*, *ura3-373* strain and the *URA3*⁺ recombinants selected as the IH530 strain. (Not to scale.)

the 5' upstream region, resulting in severely attenuated expression of the protein, so multiple copies are required to complement a *leu2* chromosomal mutation [50]. A number of plasmids have been developed to take advantage of this phenomenon, and it was the aim of this study to use the *leu2-d* mutation to select for multiple copies of the plasmid containing the cloned ketohexokinase gene. The *leu2* mutation used in strains of yeast employed in this study was a double mutant (*leu2-3*, *leu2-112*), which decreases the frequency of natural reversion (see 6.1.2.2).

Several other auxotrophic markers were exploited in the development of the TDS strains. The *ade2.1* allele is a mutation in the gene for phosphoribosylamino-imidazole-carboxylase within the adenine synthesis pathway [51]. This genetic lesion results in an accumulation of an intermediate, which is then stored in the vacuole, where it auto-oxidises to form a red compound, resulting in a pink coloration in colonies [52, 53, 54]. The *leu1* marker is another mutation in the leucine biosynthesis pathway, this gene encoding isopropylmalate isomerase [55]. A mis-labelling of the strain resulted in this marker being incorporated in the first TDS strain (TDS I), and was later removed by back crossing to give the final TDS II strain. Two mutations in the histidine biosynthesis pathway were also used, and these are also represented in some of the TDS II strains, although there is no immediate requirement for them. These markers are *his3* (a double mutation) which encodes a mutant imidazoleglycerolphosphate dehydratase enzyme [56, 57], and *his 4-519*, which encodes a polyfunctional protein which can be divided into 3 domains: HIS4 A (phosphoribosyl-AMP cyclohydrolase), HIS4 B (phosphoribosyl-ATP pyrophosphohydrolase) and HIS4 C (histidinol dehydrogenase) [58].

2.2 Materials and Methods

2.2.1 Yeast Strains.

2.2.1.1 TDS parent strains. Five strains were used in the construction of the required TDS strain (*leu2*, *trp1*, *mal*[°], *suc*[°], *GLK*, *hvk1*, *hvk2*, *pho3*, *pho5*, *pho8*). **GM2.47-21a** (*leu1*, *mal*[°], *suc*[°], *GAL*, mating type *a*) and **GM2.47-8a** (*his4*, *mal*[°], *suc*[°], *gal*, mating type α) were from the stocks of Dr. B. S. Cox, originally donated by F. K. Zimmerman, **hvk0-9C** (*trp1*, *his4*, *MAL*, *SUC*, *GLK*, *hvk1*, *hvk2*, mating type α) from Entian *et al.* [39], **IH530** (*leu2* 3, *leu2*-112, *his4*-519, *trp1*, *ade2*, *ura3*-251, *ura3*-328, *ura3*-373, *pho5/pho3*-URA3, mating type *a*) was from Bajwa *et al.* [46], and **NA79-10c** (*leu2*-3, *leu2*-112, *trp1*, *his3*, *can1*, *pho8*-2, mating type α) was from Kaneko *et al.* [45].

2.2.1.2 Leucine-dependent strains. Due to the presence of two different leucine auxotrophic markers in the parent strains (*leu1* in GM2.47-21a and *leu2* in IH530 and NA79-10c), it was necessary to determine the genotype of strains with a *leu* phenotype. Four strains were used for this: **5548-8A** (*arg4.27*, *leu1*, *ade2.1*, mating type *a*), **5548-1A** (*arg4.27*, *leu1*, *ade2.1*, *lys1.1*, mating type α), **MD40/4c** (*leu2*-3, *leu2*-112, *his3*-11, *his3*-15, *ura2*, *trp1*, 2 μ ⁺, mating type α) and **AH22** (*leu* 2-3, *leu2*-112, *his3*-52, mating type *a*). The first two strains were taken from the stocks of Dr. Brian Cox, and the second pair were gifts of Dr. Melanie Dobson.

2.2.1.3 Strain storage. Due to the probability of reversion or mutation of strains kept at 4°C or continually replated, strains could be cryo-stored for long periods by suspending the strain, either from a plate or liquid culture, in 50% v/v glycerol by freezing in liquid nitrogen and storing at -70°C.

2.2.2 Yeast Growth Media.

Six media were used in the development of the required yeast strain, and in the general propagation of strains.

2.2.2.1 YEPD is a general purpose growth media used when non-selective conditions are required.

2% w/v Glucose
1% w/v Bacto-peptone [Difco]
1% w/v Yeast extract [Oxoid]

2.2.2.2 1/4 YEPD is used to test for the *ade2* mutation.

4% w/v Glucose
1% w/v Bacto-peptone
0.25% w/v Yeast extract

2.2.2.3 SD is a defined synthetic media, used to grow yeast when selection of auxotrophic markers is desired. Glucose can be substituted with other carbon sources, *e.g.* sucrose (SS), fructose (SF), or glycerol (SG).

1% w/v Glucose
0.17% w/v Yeast nitrogen base w/o amino acids [Difco]
0.5% w/v Ammonium sulphate

2.2.2.4 Presporulation medium is a carbon-rich media used to grow diploids prior to carbon and nitrogen starvation on sporulation media (see 2.2.2.5).

10% w/v Glucose
0.3% w/v Bacto-peptone
0.8% w/v Yeast extract

2.2.2.5 Sporulation medium is a carbon and nitrogen starvation media used to induce asci formation in diploid strains. Asci tetrads will usually appear within 3 to 7 days.

0.05% w/v Glucose
0.1% w/v Yeast extract
1% w/v Potassium acetate

2.2.2.6 Burkholder's defined media [59, using the modification of 60] were used when an absolute control of the yeast media was required, especially when defining phosphate concentration, for the induction of acid and alkaline phosphatases.

20 g/l. Glucose
2 g/l. Asparagine
0.5 g/l. Magnesium chloride heptahydrate
0.33g/l. Calcium chloride dihydrate
2 g/l. Ammonium sulphate
0.1 mg/l. Potassium iodide

High phosphate medium contained 1.5g/l. potassium dihydrogen phosphate, low phosphate medium contained 30mg/l. potassium dihydrogen phosphate and 1.5g/l. potassium chloride. The media were adjusted to pH 5.0 with sodium hydroxide and were autoclaved before the addition of filter-sterilized 100× micro-nutrient supplement.

20mg/l. Thiamine
10mg/l. Riboflavin
20mg/l. Pyridoxin
20mg/l. Nicotinic acid
20µg/l. Biotin
20mg/l. Pantothenic acid
100mg/l. Inositol
1mg/l. *Para*-amino benzoic acid (PABA)
10% v/v Trace elements stock solution

The trace elements stock solution consists of the following:

0.1g/l.	Boric acid
40mg/l.	Manganese sulphate tetrahydrate
40mg/l.	Zinc sulphate heptahydrate
45mg/l.	Copper sulphate pentahydrate
20mg/l.	Ammonium molybdate
25mg/l.	Iron sulphate heptahydrate

With both the defined medium and SD medium, filter-sterilized auxotrophic nutritional supplements were added as necessary (0.24% w/v L-tryptophan, 0.24% w/v L-histidine, 0.36% w/v L-leucine, 0.24% w/v uracil as 120× stock solutions). For plates, agar was added to 2% w/v prior to autoclaving, defined medium, low phosphate plates were poured in glass petri dishes, otherwise all petri dishes were polystyrene.

2.2.3 Determination of Phenotypes and Genotypes.

The phenotypes of yeast strains were generally established by determining their nutritional requirements. Ability to metabolize sugars was checked by comparison of the growth of the strain on SD with a replica plate on SD in which the glucose was replaced with the carbohydrate of interest (fructose at 0.2% w/v, sucrose at 0.6% w/v, raffinose at 1.0% w/v). Sugars were filter-sterilized (0.2µm) and added after autoclaving. Ability to metabolize maltose could not be established in this way, due to the contamination of commercial preparations of the sugar with glucose. Consequently *mal* strains were tested by their inability to hydrolyse para-nitrophenol- α -D-glucoside (PNPG) due to the absence of α -glucosidase. The phenotype was checked by growing the cells in 2ml. supplemented SD liquid medium containing 100mg/l. maltose and glucose and 1mM. PNPG overnight at 30°C. The liquid culture was then alkalinized by addition of 2ml. 1M. sodium carbonate. Strains lacking α -glucosidases remain colourless, as compared with strains containing the enzyme which gave intense yellow solutions.

The auxotrophic requirements of strains were evidenced by absence of growth when deprived of nutrient supplements. The determination of the leucine genotype involved the crossing of the unknown strain with each of the four strains described in 2.2.1.2. Each of the leucine auxotroph mating type was streaked out on a YEPD plate in a linear fashion, such that each strain lay parallel with its opposite mating type. The strain of unknown leucine genotype was then streaked through each pair of known genotype in a unidirectional, single stroke fashion, and the plate allowed to grow. The colonies were then replica plated onto a SD leu- plate, allowed to grow, and the genotype determined as shown in *Figure 2.5*.

Strains were checked for the *pho* mutations by a modification of the method of Toh-e *et al.* [44] using 5-bromo-4-chloro-3-indolyl-phosphate (BCIP or X-phos) [61] as the chromogenic

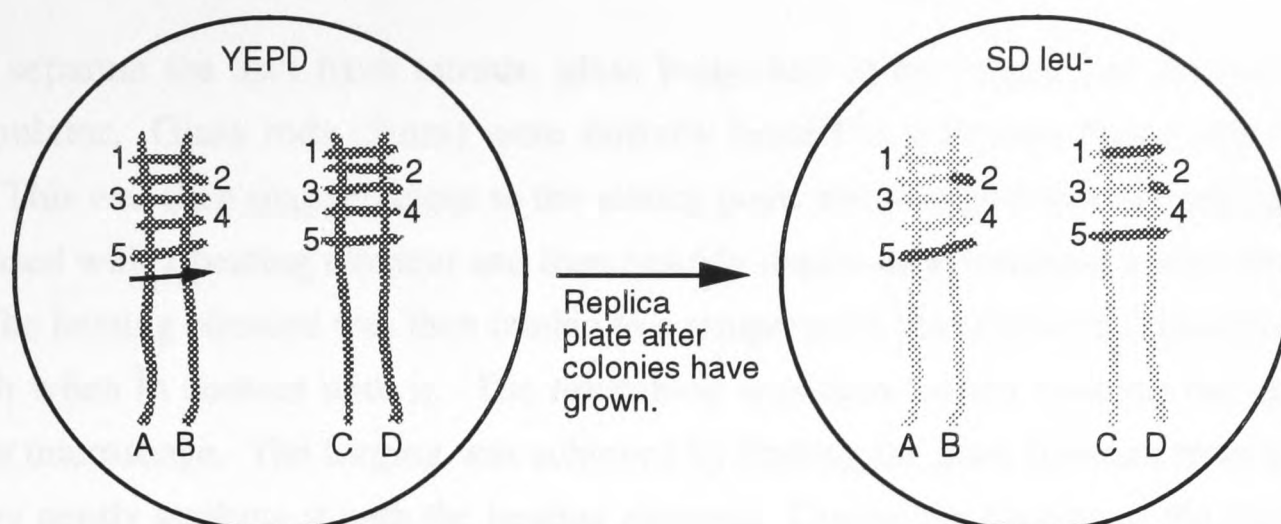


Figure 2.5 Diagram demonstrating the determination of an auxotrophic marker genotype. The left hand petri dish shows the cell strains streaked on YEPD medium. Strain A is *leu1 LEU2 α* , Strain B *leu1 LEU2 α* , Strain C *LEU1 leu2 α* , and Strain D *LEU1 leu2 α* . Strains 1-5 represent the strains of unknown genotype, and the arrow indicates the direction each of these strains is streaked through the known strains. The plates are incubated overnight to allow the cells to mate and grow, before replica plating on to SD *leu-*, and again allowed to grow (shown on the right). On this plate, cells that have grown are shown in heavy shading, and those that remain *leu*, and hence are unable to grow, are lightly shaded. These results would indicate the following genotypes: 1: *leu1 LEU2 α* , 2: *LEU1 leu2 α* , 3: *leu1 LEU2 α* , 4: *leu1 leu2*, 5: *LEU1 LEU2*. The mating types of strains 4 and 5 is undeterminable from this experiment.

substrate. The cells were grown on low phosphate defined medium and then each plate was inverted over 1ml. chloroform for 1 hour. The lysed cells were then overlaid with 5ml. buffered 1% w/v agar containing BCIP at 1mg/ml. For scoring expression of the acid phosphatase, the agar was buffered with 100mM. sodium acetate/acetic acid, pH 4.0, and for the alkaline phosphatase, 100mM. Bis-Tris propane.HCl, pH 9.0 was used. Strains expressing phosphatase activity were able to hydrolyse the dye at their optimum pH to yield an intense dark blue stain.

2.2.4 Haploid Conjugation.

Cells of opposite mating types were mixed on a YEPD plate and left for several hours at room temperature. The cells were periodically checked to see if they had mated by taking a small sample and mixing in 100 μ l. water on a microscope slide and viewing under a light microscope. Cells which had mated take on a dumbbell shape. When a large number of cells had reached this stage, a sample was streaked on to a fresh YEPD plate and single diploid cells were isolated by micromanipulation. Alternatively, if the diploid displayed an auxotrophic independence not shown by either parent haploid, a sample of the mating mix was spread on an SD plate lacking these nutrients, and diploids selected. Diploid cells were grown until colonies became visible, and then were replated on presporulation medium plates and left at 28°C for 2 days. Cells were then transferred to Fogel's sporulation medium plates and left to sporulate at 28°C for about 5 days. Sporulation was checked by observation of asci under a phase-contrast light microscope.

2.2.5 Microforging Loops for the Manipulation of Single Yeast Cells.

In order to separate the asci from tetrads, glass loops had to be forged and mounted on a micromanipulator. Glass rods (3mm) were initially heated in a Bunsen flame and the rod drawn out. This was then snapped close to the stretch point and mounted in a microforge. The glass was fused with a heating element and then quickly drawn away creating a very fine glass filament. The heating element was then cooled to a temperature that allows the glass to soften but not melt when in contact with it. The microloop was then forged viewing the progress under a light microscope. The forging was achieved by heating the glass filament from the side to be bent by gently stroking it with the heating element. During the forging of the loop, care must be taken to ensure it remains in the plain of the microscope depth of field in order to prevent a coil being formed. With practice, loops of 20 - 40 μ m in diameter could be made (an ascospore is approximately 5 μ m in diameter).

2.2.6 Dissection of Asci.

When diploid cells sporulated, small asci appeared under the light microscope as either a diamond shape, or a tetrahedron of four closely packed cells. A sample of the sporulated yeast was diluted into approximately 1ml. water containing 5mg/ml. zymolyase (ICN Flow) and incubated for 30 minutes. The completion of the digestion was checked by examining a sample under phase-contrast at 400 $\times$ magnification. The sample was ready when the asci appeared as loose clusters of four cells in a diamond shape. When at this stage, the digested cell suspension was streaked on one side of a fresh YEPD plate. When most of the liquid had soaked into the agar, the plate was mounted upright on a microscope stage. A microdissection loop was mounted in a de Fonbrune-type manual micromanipulator [4]. The microscope was brought into focus on the cells and the loop brought into the field of view. Well separated asci were found and individually picked up with the loop. This was achieved by lowering the loop over the ascus until it makes contact with the agar. Bringing the loop up again would lift up the ascus by virtue of the surface tension of the liquid trapped in the loop. The plate was then moved to an area free of cells, and the ascus separated into the four individual haploid cells by gently tapping the loop on the surface of the agar. In this way, the individual cells were positioned at predefined positions on the plate. When sufficient asci had been dissected (usually ten per plate) the cells were incubated at 30°C until the colonies of haploid cells were visible. The cells were then replated, grown, and used for replica plating to determine the phenotypes of the asci. The phenotypes of the haploid colonies were established to confirm that the expected genotypic segregation had occurred, and those of the desired phenotype were selected for further crosses as necessary.

The number of tetrads to be dissected to ensure that the desired phenotype was acquired in at least one of the spores was determined using the 2ⁿ calculation. Since there were no cases

when it was necessary to separate two closely linked recessive alleles, this calculation would usually hold. The only closely linked alleles were *pho3/pho5-URA3* (see 2.1.3.3), and the *suc*[°], *mal*[°] mutation (see 2.1.3.1), and in both case, the pairs of mutations are required in the final strain. Therefore, in both of these cases, the pairs of mutant alleles could be considered as a single allele. To ensure that the desired mutant phenotype was obtained, the value of *n* was routinely increased by one, so that the number of tetrads dissected was 2^{n+1} , so that eight colonies of the desired phenotype would be obtained (each tetrad yields four spores). This apparently superfluous number would ensure that losses due to plate infection or failure of dissected spores to divide would not effect the acquisition of the required strain. It would also improve the chances of obtaining the desired strain in either mating type.

2.3 Results and Discussion

2.3.1 Mating Strategy.

The strategy used in the development of the TDS strains is shown in *Figure 2.6*. This involved the development of a strain with the required lesions in sugar metabolism, and a second strain that lacked the genes conditioning non-specific phosphatases. These were then crossed, and 128 asci dissected, and of these seven were determined to have the desired phenotypes. Strains within this group were termed TDS I, representatives of which also possessed the *ade2.1* allele, the *his* allele, both or neither of these phenotypes. The presence of these mutations was not important in this study, since the medium they are grown on can be supplemented to allow their growth, but these characteristics were maintained to allow some flexibility in plasmid selection.

A minor problem was encountered when an unsuccessful transformation of one of the seven TDS I strains with a *leu2-d* bearing plasmid, revealed that the strains had inherited a *leu1* mutation from its parents, and selection of transformants by complementing *leu2* was not possible. At this stage of the study, all the TDS I strains were only available as glycerol stocks, and rejuvenation of all the strains revealed that two of them had lost the *pho8* phenotype, and one of them the *hvk* phenotype. Of the other strains, none contained a *LEU1*, *leu2* phenotype, as determined according to *Figure 2.5*, and so the desired TDS II strain was bred by mating two of the TDS I strains (*ADE2*, *leu1*, *LEU2*, *trp1*, *his4*, *pho3*, 5, 8, *suc*[°], *mal*[°], *hvk1*, 2, mating type *a*), was crossed with *ade2*, *LEU1*, *leu2*, *trp1*, *HIS*, *pho3*, 5, *PHO8*, *suc*[°], *mal*[°], *hvk1*, 2 mating type *α*), selecting for leucine independence in the diploids, sporulating and dissecting tetrads formed. Of the twenty asci dissected, five were found to have the desired phenotype, although one was later determined to be inviable. All four of the remaining strains were all of the same mating type (*α*), two were *ade2*, *his4*, one was *ADE2*,

his4 and the fourth was *ADE2*, *HIS4*, and it was this latter strain that was used in the transformation work. This strain therefore perfectly fulfils the strain requirement, with no additional mutant alleles.

2.3.2 Phenotype Determination.

The determination of the auxotrophic markers was done according to standard procedures, but the other methods for the determination of mutant alleles underwent some modification. The scoring of hexose respiration and disaccharide hydrolysis proved difficult due to the impurities

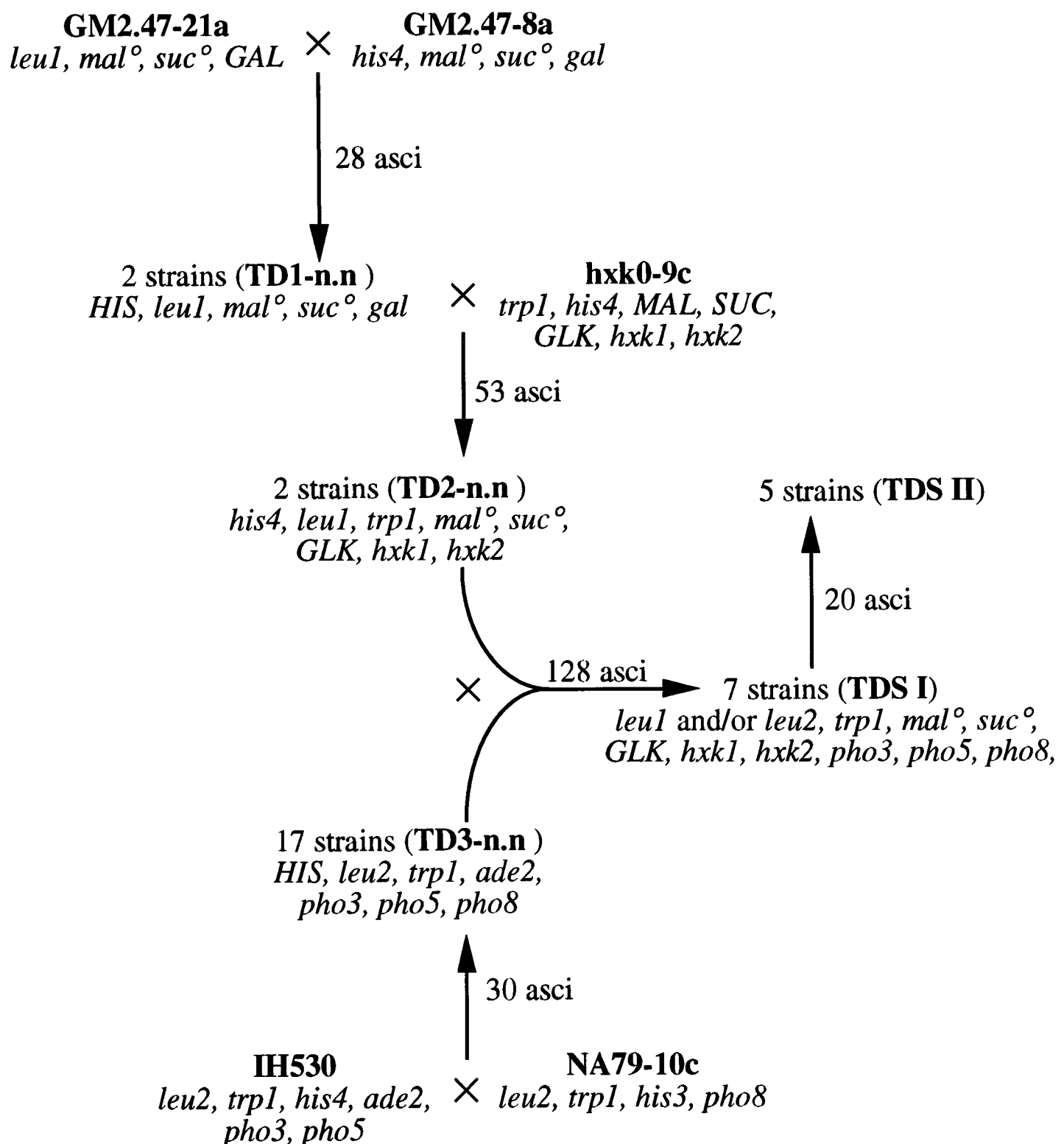


Figure 2.6 Mating strategy for the development of the TDS II strain. In this final strategy, two strains were developed before the initial TDS strain was bred, one being devoid of the sugar metabolizing genes, (TD2-n.n), and the other devoid of the major non-specific phosphatases (TD3-n.n). These development strains were named TD1-n.n etc., where the first n is the number of the tetrad dissected, and the second n is the designation of the spore from each tetrad (labelled a-d).

of the sugars available. Consequently, for all sugars except maltose, a range of sugar concentrations had to be used in the defined medium plates (SS and SF) and a direct comparison made between the growth of strains with known mutations in their metabolism, and that of strains with unknown phenotypes. In the case of sucrose, the sugar had to be made up as a 10% w/v filter-sterilized stock, and added to the autoclaved medium when its temperature was about 60°C to decrease thermal hydrolysis. With sucrose hydrolysis by yeast, 0.6% w/v final concentration was found to give the clearest distinction, and with fructose utilization, this was obtained at 0.2% w/v. Variation of growth temperature (either 28°C or 34.5°C) did not improve discrimination of phenotypes.

The ability of maltase, but not invertase, to hydrolyse *p*-nitrophenol α -glucoside was again used to discriminate between these activities. Cells were used to inoculate a medium effectively free of carbon source containing the chromogenic substrate, and a small amount of maltose added to induce maltase synthesis. The cultures were left overnight with shaking and the medium alkalinized to allow visualization of the released *p*-nitrophenol group. As expected, the *mal*^o phenotype generally segregated with the *suc*^o phenotype (of the ten tested in this way, only one showed segregated alleles). Raffinose (1-6 α -D-galactopyranosyl sucrose) is a substrate for invertase, but not maltase, and growth on this substrate yielded results in determination of phenotype that were consistent with those obtained by the other methods.

The use of 5-bromo-4-chloro-3-indolyl phosphate as a chromogenic substrate for the determination of the phosphatase genes proved a vast improvement on the use of *p*-nitrophenol phosphate as described by Toh-e *et al.* [44]. This substrate gave easily determinable results (see *Fig. 2.7*) without the need to alkalinize the plates to visualize activity, and appears to be stable to acid or alkaline hydrolysis at the *pH*s used for several weeks when kept at 4°C.

2.3.3 TDS II Strain Characteristics.

In spite of the metabolic indignities imposed on the TDS II strain, growth characteristics did not seem to be adversely affected, even when the cells were grown in low phosphate medium. The cells were also later shown to be recipient to plasmid DNA by the lithium acetate transformation method, although spheroplast or electro-transformation methods were unsuccessful (see 6.2.2.2 and .3). Since both these approaches are strain dependent, this was not surprising, and since the lithium acetate method is by far the easiest, this did not represent a problem.

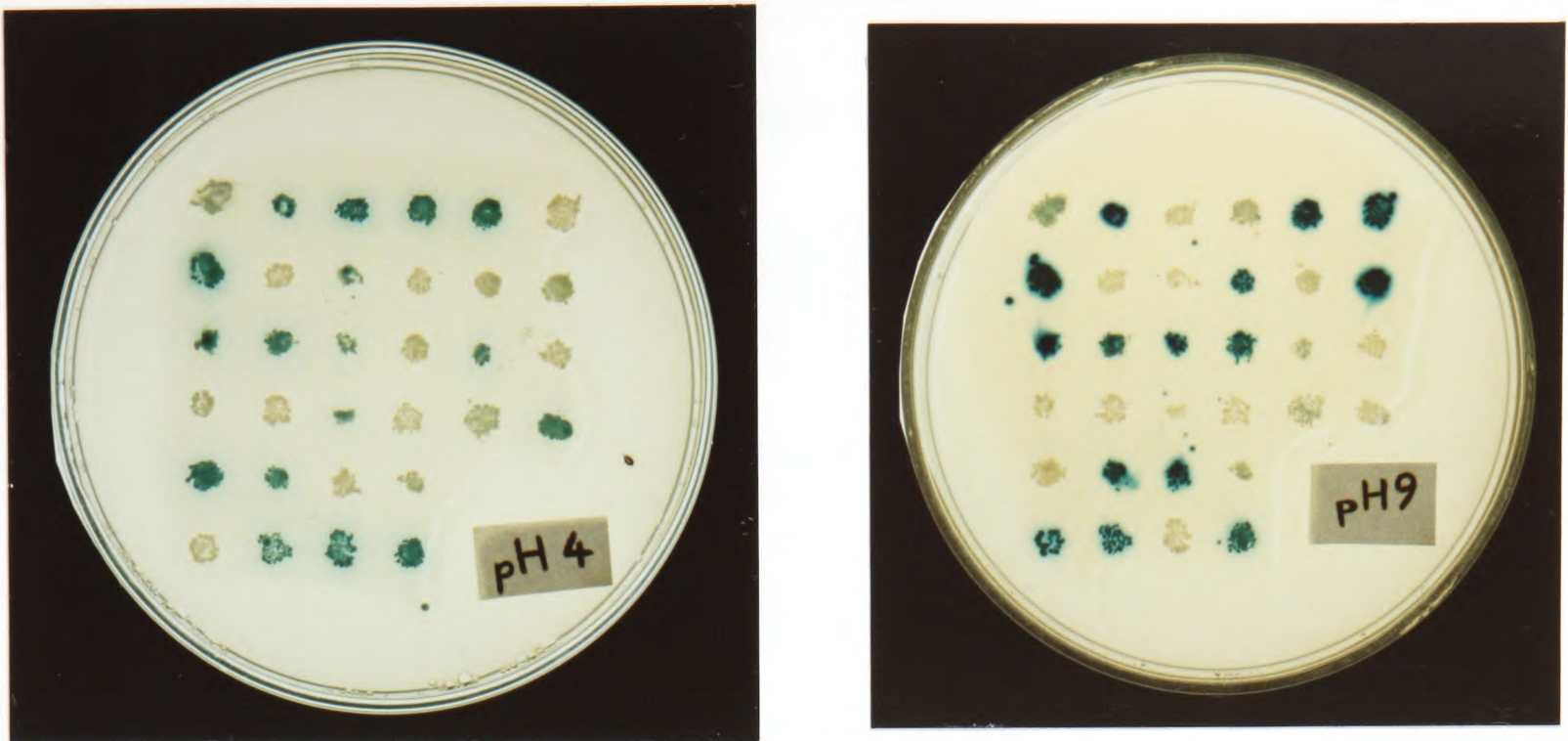


Figure 2.7 Photographs showing the results of a typical assay for the acid (pH4) or alkaline (pH9) phosphatases. The phenotypes of the spores from eight asci are shown, each ascus has been dissected to form a square of colonies. A number of double mutants can clearly be seen.

2.4 References

1. Winge, O. (1935). On haplophase and diplophase in some *Saccharomyces*. *Comptes Rendus du Travail de la Laboratoire Carlsberg, Série de Physiologie*, **21**, 77-111.
2. Lindegren, C.C. and Lindegren, G. (1943). A new method for hybridizing yeast. *Proc. Natl. Acad. Sci. USA*, **29** 306-308.
3. Herskowitz, I. (1988). Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **52**, 538-553.
4. Sherman, F. and Hicks, J. (1991). Getting Started with Micromanipulation and Dissection of Asci. *Methods. Enzymol.*, **194**, 21-37.
5. Herskowitz, I. and Oshima, O. (1982). Control of cell type in *Saccharomyces cerevisiae*: mating type and mating type interconversions, in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. (Strathern, J.N., Jones, E.W. and Broach, J.R., Eds.), Cold Spring Harbor, New York, vol 2, 181-209.
6. Kostriken, R., Strathern, J.N., Klar, A.J.S., Hicks, J.B. and Hefron, F. (1983). A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell*, **35**, 167-174.
7. Pringle, J.R. and Hartwell, L.H. (1982). The *Saccharomyces cerevisiae* Cell Cycle, in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. (Strathern, J.N., Jones, E.W. and Broach, J.R., Eds.), Cold Spring Harbor, New York, vol 2, 97-142.

8. Gimeno, C.J., Ljungdahl, P.O., Styles C.A. and Fink, G.R. (1992). Unipolar Cell Division in the Yeast *S. cerevisiae* Lead to Filamentous Growth: Regulation by Starvation and RAS. *Cell*, **68**, 1077-1090.
9. Tuite, M.F. (1992). Strategies for the Genetic Manipulation of *Saccharomyces cerevisiae*. *Crit. Rev. Biotech.*, **12**, 157-188.
10. Thorner, J. (1982). Pheromone Regulation of Development in *Saccharomyces cerevisiae*, in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. (Strathern, J.N., Jones, E.W. and Broach, J.R., Eds.), Cold Spring Harbor, New York, vol 2, 143-180.
11. Byers, B. (1982). Cytology of the Yeast Cell Cycle, in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. (Strathern, J.N., Jones, E.W. and Broach, J.R., Eds.), Cold Spring Harbor, New York, vol 2, 59-96.
12. Roth, R. and Halvorson, H.O. (1969). Sporulation of yeast harvested during logarithmic growth. *J. Bacteriol.*, **98**, 831-832.
13. Peterson, J.G.L., Olson, L.W. and Zicker, D. (1978). Synchronous sporulation of *Saccharomyces cerevisiae* at high cell concentrations. *Carlsberg Res. Commun.* **43**, 241-254.
14. Esposito, R.E. and Klapholz (1982). Meiosis and Ascospore Development, in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. (Strathern, J.N., Jones, E.W. and Broach, J.R., Eds.), Cold Spring Harbor, New York, vol 2, 211-287.
15. Moens, P.B. and Rapport, E. (1971). Spindles, spindle plaques, and meiosis in the yeast, *Saccharomyces cerevisiae* (Hansen). *J. Cell Biol.*, **50**, 344-361.
16. Kane, S.M. and Roth, R. (1974). Carbohydrate metabolism during ascospore development in yeast. *J. Bacteriol.*, **118**, 8-14.
17. Ballou, C.E., Maitra, S.K., Walker, J.W. and Wheelan, W.L. (1977). Developmental defects associated with glucosamine auxotrophy in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*, **76**, 4351-4355.
18. Fogel, S., Mortimer, R.K., Lusnak, K. and Tavares, F. (1979). Meiotic gene conversion: a signal of the basic recombination event in yeast. *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 1325-1341.
19. Petes, T.D., Malone, R.E. and Symington, L.S. (1991). Recombination in Yeast, in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics*. (Broach, J.R., Pringle, J.R. and Jones, E.W., Eds.), Cold Spring Harbor, New York, vol 1, 407-521.
20. Gancedo, C. and Serrano, R (1989). Energy-Yielding Metabolism, in *The Yeasts*. (Rose, A.H. and Harrison, J.S., Eds.), Academic Press, London, vol 3, 205-259.
21. Goldstain, A. and Lampen, J.O. (1975). β -D-Fructofuranoside Fructohydrolase from Yeast. *Meths. Enzymol.*, **42**, 504-511.
22. Khan, N.A., Zimmermann, F.K. and Eaton, N.R. (1973). Genetic and Biochemical Evidence of Sucrose Fermentation by Maltase in Yeast. *Mol. Gen. Genet.*, **123**, 43-50.
23. Carlson, M., Osmond, B.C. and Botstein, D. (1981). Mutants of yeast defective in sucrose utilization. *Genetics*, **98**, 25-40.

24. Sarokin, L. and Carlson, M. (1985). Comparison of two yeast invertase genes: conversion of the upstream regulatory region. *Nuc. Acids Res.*, **13**, 6089-6103.
25. Hohmann, S. and Gozalbo, D. (1988). Structural analysis of the 5' regions of yeast *SUC* genes revealed analogous palindromes in *SUC*, *MAL* and *GAL*. *Mol. Gen. Genet.*, **211**, 446-454.
26. Carlson, M., Celenza, J. and Eng, F.J. (1985). Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosomal telomeres. *Mol. Cell. Biol.*, **5**, 2894-2902.
27. Grossman, M.K. and Zimmermann, F.K. (1979). The Structural Genes of Internal Invertases in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **175**, 223-229.
28. Carlson, M. and Botstein, D. (1982). Two Differentially Regulated mRNAs with Different 5' Ends Encode Secreted and Intracellular Forms of Yeast Invertase. *Cell*, **28**, 145-154.
29. Carlson, M., Taussig, R., Kustu, S. and Botstein, D. (1983). The Secreted Form of Invertase in *Saccharomyces cerevisiae* Is Synthesised from mRNA Encoding a Signal Sequence. *Mol. Cell. Biol.*, **3**, 439-447.
30. Carlson, M. and Botstein, D. (1983). Organization of the *SUC* Gene Family in *Saccharomyces*. *Mol. Cell. Biol.*, **3**, 351-359.
31. Mortimer, R.K. and Hawthorne, D.C. (1969). Yeast Genetics, in: *The Yeasts*, (Rose, A.H., Harrison, J.S., Eds.), Academic Press, London, vol 1, 385-460.
32. Khan, N.A., Zimmermann, F.K. and Eaton, N.R. (1973). Genetic and Biochemical Evidence of Sucrose Fermentation by Maltase in Yeast. *Mol. Gen. Genet.*, **123**, 43-50.
33. Khan, N.A. and Eaton, N.R. (1967). Purification and characterization of maltase and α -methyl-glucosidase from yeast. *Biochim. Biophys. Acta*, **146**, 173-180.
34. Khan, N.A. and Eaton, N.R. (1971). Genetic control of maltose formation in yeast. I. Strains producing high and low basal levels of enzyme. *Mol. Gen. Genet.* **112**, 317-322.
35. Mortimer, R.K., Contopoulos, C.R. and King, J.S. (1991). Genetic and Physical Maps of *Saccharomyces cerevisiae*. 11th Edition, in: *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics*. (Broach, J.R., Pringle, J.R. and Jones, E.W., Eds.), Cold Spring Harbor, New York, vol 1, 737-752.
36. Hackel, R.A. (1975). Genetic Control of Invertase in *Saccharomyces cerevisiae* I. Isolation and Characterization of Mutants Affecting Sucrose Utilization. *Mol. Gen. Genet.*, **140**, 361-370.
37. Maitra, P.K. (1970). A Glucokinase from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **245**, 2423-2431.
38. Cárdenas, M.L., Rabajille, E. and Niemeyer, H. (1984). Fructose is a good substrate for rat liver 'glucokinase' (hexokinase D). *Biochem. J.*, **222**, 363-370.
39. Entian, K.-D., Kopetzki, E., Fröhlich, K.-U. and Mecke, D. (1984). Cloning of hexokinase isoenzyme PI from *Saccharomyces cerevisiae*: PI transformants confirm the unique role of hexokinase isoenzyme PII for glucose repression in yeasts. *Mol. Gen. Genet.*, **198**, 50-54.

40. Oshima, Y. (1982). Regulatory Circuits for Gene Expression: The Metabolism of Galactose and Phosphate, in *Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. (Strathern, J.N., Jones, E.W. and Broach, J.R., Eds.), Cold Spring Harbor, New York, vol 1, 159-180.
41. Schurr, A. and Yagil, E. (1971). Regulation and Characterization of Acid and Alkaline Phosphatase in Yeast. *J. Gen. Microbiol.*, **65**, 291-303.
42. Toh-E, A., Ueda, Y., Kakimoto, S.-I. and Oshima, Y. (1973). Isolation and Characterization of Acid Phosphatase Mutants in *Saccharomyces cerevisiae*. *J. Bacteriol.*, **113**, 727-738.
43. Bajwa, W., Meyhack, B., Rudolph, H., Schweingruber, A.-M. and Hinnen, A. (1984). Structural analysis of the two tandemly repeated acid phosphatase gene in yeast. *Nuc. Acids Res.*, **12**, 7721-7739.
44. Toh-E, A., Nakamura, H. and Oshima, Y. (1976). A gene controlling the synthesis of non specific alkaline phosphatase in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta*, **428**, 182-192.
45. Kaneko, Y., Tamai, Y., Toh-E, A. and Oshima, Y. (1985). Transcriptional and Post-Transcriptional Control of *PHO8* Expression by *PHO* Regulatory Genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **5**, 248-252.
46. Bajwa, W., Rudolph, H. and Hinnen, A. (1987). *PHO5* Upstream Sequences Confer Phosphate Control on the Constitutive *PHO3* Gene. *Yeast*, **3**, 33-42.
47. Dobson, M.J., Kingsman, S.M. and Kingsman, A.J. (1981). Sequence variation in the *LEU2* region of the *Saccharomyces cerevisiae* genome. *Gene*, **16**, 133-139.
48. Andreadis, A., Hsu, Y.-P., Hermodson, M., Kohlhaw, G. and Schimmel, P. (1984). Yeast *LEU2*. Repression of mRNA Levels by Leucine and Primary Structure of the Gene Product. *J. Biol. Chem.*, **259**, 8059-8062.
49. Tschumper, G. and Carbon, J. (1980). Sequence of a yeast DNA fragment containing a chromosomal replicator and the *trp1* gene. *Gene*, **10**, 157-166.
50. Erhart, E. and Hollenberg, C.P. (1983). The Presence of a Defective *LEU2* Gene on 2 μ DNA Recombinant Plasmids of *Saccharomyces cerevisiae* Is Responsible for Curing and High Copy Number. *J. Bacteriol.*, **156**, 623-635.
51. Stotz, A. and Linder, P. (1990). The *ADE2* gene from *Saccharomyces cerevisiae*: Sequence and new vectors. *Gene*, **95**, 91-98.
52. Ephrussi, B., Hottinguer, H. and Tavlitzki, J. (1949). Action de l'acriflavine sur les levures. *Ann. Inst. Past.*, **76**, 419-442.
53. Reaume, S.E. and Tatum, E.L. (1949). Spontaneous and nitrogen mustard-induced nutritional deficiencies in *Saccharomyces cerevisiae*. *Arch. Biochem.*, **22**, 331-338.
54. Roman, H. (1956). Studies of gene mutation in *Saccharomyces*. *Cold Spring Harbor Symp. Quant. Biol.*, **21**, 175-185.
55. Hsu, Y.-P. and Schimmel, P. (1984). Yeast *LEU1*. Repression of mRNA levels by leucine and relationship of 5'-noncoding region to that of *LEU2*. *J. Biol. Chem.*, **259**, 3714-3719.

56. Struhl, K. and Davis, R.W. (1981). Promoter mutants of the yeast *his3* gene. *J. Mol. Biol.*, **152**, 553-568.
57. Struhl, K. (1985). Nucleotide sequence and transcriptional mapping of the yeast *pet56-his3-ded1* gene region. *Nuc. Acids Res.*, **13**, 8587-8601.
58. Donahue T.F., Farabaugh, P.J. and Fink, G.R. (1982). The nucleotide sequence of the *HIS4* region of yeast. *Gene*, **18**, 47-59.
59. Burkholder, P.R. (1943). Vitamin deficiencies in yeasts. *Amer. J. Bot.*, **30**, 206-211.
60. Cutts, N.S. and Rainbow, C. (1949). Nutrition of a strain of brewer's yeast exacting to *p*-aminobenzoic acid. *Nature*, **164**, 234-235.
61. Gibson, D.M., Christen, A.A. and Mullaney, E.J. (1988). Direct screening for acid phosphatase production on BCIP-agar plates. *Biotech. Technia.*, **2**, 63-68.

Chapter 3

The Purification of Kethexokinese

3.1 Introduction

Kethexokinese (ATP; D-fructose 1-phosphotransferase; EC 2.7.1.3) is the enzyme responsible for the primary metabolism of dietary fructose in mammals [1, 2]. The enzyme is most abundant in the liver, but is also present in the kidney, pancreas and small intestine [3]. All of these tissues possess a specialized pathway employing kethexokinese, aldolase B (EC 4.1.2.13) and triokinese (EC 2.7.1.28) that together convert fructose into intermediates for glycolysis and gluconeogenesis [4]. Two inherited abnormalities of this pathway have been recognised in humans. Hereditary fructose intolerance is a potentially lethal condition caused by a deficiency of aldolase B [5]. As a result of the activity of kethexokinese, ingestion of fructose by individuals with this disorder is followed by accumulation of fructose 1-phosphate and depletion of ATP in tissues possessing this pathway (liver, kidneys, small intestine and pancreas). Essential fructosuria, in contrast, is a benign defect caused by the absence of kethexokinese, which results in an inability to synthesize fructose 1-phosphate [6]. Under these circumstances, uptake of dietary fructose leads to a transient increase in the concentration of blood fructose. This gradually decreases and approximately 80-90% of the ingested fructose is metabolized, the rest being excreted in urine, indicating that, under these circumstances, fructose is being utilised by other pathways [2].

Mammalian hexokinese (EC 2.7.1.1) are also able to phosphorylate fructose [7]. In rats there are four isozymes of this enzyme, belonging to a family of kinases that display antigenic cross-reactivity and regions of amino acid homology [8]. Brain contains hexokinese type I only, skeletal muscle, type II, and adipose tissue and heart contain both of these. Kidney contains types I and III, whereas the intestine and liver contains all three isozymes. Isozymes I, II and III have affinities for glucose roughly ten times that for fructose, phosphorylate a range of other hexoses *in vitro*, and are strongly inhibited by glucose 6-phosphate [9]. Hexokinese type IV (glucokinese), the predominant isozyme in liver, is not inhibited by its product, and is regulated by insulin and glucagon [7, 9]. It shows similar substrate specificity to the other hexokinese, but has a much lower affinity for its substrates, and is not responsible for the primary metabolism of fructose *in vivo*. Nonetheless, this enzyme has recently been shown to

have substantial regions of amino acid homology with isozymes I to III, and so was confirmed to be a member of the hexokinase family, which also includes the yeast isozymes PI and PII [8].

The properties of ketohexokinase show substantial differences from those of enzymes belonging to the hexokinase family. It phosphorylates its substrate at the C1 position rather than the C6 position [10, 11], has an absolute requirement for monovalent cations (physiologically potassium) [12], will only phosphorylate furanose substrates, and unlike glucokinase, is not induced hormonally or by its substrate. In common with hexokinases, it does show marked preference for Mg^{2+} ATP as a nucleotide, and is inhibited by ADP.

Liver ketohexokinase has been isolated from a number of sources [11, 13-15]. Extensive studies of the enzyme's kinetic properties and substrate specificity have been performed. However, a primary structure of ketohexokinase has not been reported. It was the intention of this study to isolate the enzyme using improved techniques and generate peptides of the homogeneous protein. Since the published purification of the rat enzyme were achieved in the late sixties and early seventies, it was thought that purification would be improved by the use of improved chromatographic techniques and matrices. The enzyme was purified from liver of the Sprague-Dawley strain of rat (*Rattus norvegicus*), since this tissue was easily available fresh, and cDNA libraries of this strain of rat were in our possession.

3.1.1 Preparation of Crude Protein Extracts.

Prudent selection of an extraction technique can aid the purification of enzymes, and a number of extraction techniques are available (see *Table 3.1*) [16]. A rapid protocol that limits degradation by endogenous protease activities is desirable. To these ends, the extraction buffer can have inhibitors to these proteases included, such as leupeptin, pepstatin and benzamidine. An extraction technique that yields a crude extract of as high a purity as possible is also preferred. This can be assisted, in the case of differentiated organisms, by isolating the protein from an organ in which it is present in greatest abundance; in the case of ketohexokinase, this is the liver. This protein is cytoplasmic, and so unnecessary contamination of the extract by proteins from organelles can be diminished by avoiding their disruption during breakage of the cell, and their removal from the extract intact. Ketohexokinase was previously extracted from liver by blade homogenization with a Waring blender [11, 13-15], which breaks open organelles. However, use of gentler techniques, such as a Potter-Elvehjem homogenizer or a Seward "stomacher", greatly decreases contamination of the cytoplasmic fraction by proteins of organellar origin.

Technique	Example	Principle
<i>Gentle</i>		
Cell lysis	Erythrocytes	Osmotic disruption of cell membrane
Enzyme digestion	Lysozyme treatment of bacteria	Cell wall digested, leading to osmotic disruption of cell membrane
Chemical solubilization / autolysis	Toluene extraction of yeast	Cell wall (membrane) partially solubilized chemically; lytic enzymes released complete the process
Stomacher	Brain and liver tissue	Cells disrupted by gentle shear forces generated by rhythmic beating of paddles against wall
Potter-Elvehjem homogenizer	Liver tissue	Cells forced through narrow gap, rips off cell membrane
Mincing (grinding)	Muscle <i>etc.</i>	Cells disrupted during mincing process by shear force
<i>Moderate</i>		
Blade homogenizer (Waring blender)	Muscle tissue, most animal tissues, plant tissues	Chopping action breaks up large cells, shears apart smaller ones
Grinding with abrasive (<i>e.g.</i> sand, alumina)	Plant tissues, bacteria	Abrasive surface removes cell walls
<i>Vigorous</i>		
French press	Bacteria, plant cells	Cells forced through small orifice at high pressure; shear forces disrupt cells
Ultrasonication	Cell suspensions	Micro-scale high-pressure sound waves cause disruption by shear forces and cavitation
Bead mill	Cell suspensions	Rapid vibration with glass beads breaks cell walls open
Manton-Gaulin homogenizer	Cell suspensions	As for French press above, but on a larger scale

Table 3.1 Cell disintegration techniques for the generation of cell or tissue homogenates. (Table taken and adapted from Scopes [16].)

3.1.2 Crude Purification by Selective Precipitation.

Purification of proteins by selective precipitation is one of the oldest tools used by the protein biochemist, and few advances have been made in this field over recent years. Consequently, no improvements were made on the published protocols for the preparation of ketohexokinase by selective precipitation, and the basis of these methods will only be briefly commented upon.

These methods rely on either solubility of proteins at different salt concentrations, or selective denaturation.

3.1.2.1 Selective denaturation was used in the first two steps of the purification procedure, and involved isoelectric precipitation and temperature denaturation of contaminating proteins, and their consequent precipitation. Some enzymes are very robust, and can stand quite extreme conditions. By exposing the impure preparation to a suitable extreme, unwanted proteins denature, and precipitate out of solution. The further away from its physiological *pH* a protein is taken, the less stable it becomes as internal electrostatic attractions become disrupted. The objective of a selective denaturation step is to create conditions where the enzyme in question will not quite denature - or only 10-20% is lost, whereas many other components are denatured, and preferentially precipitated by the treatment [16].

Denaturation of protein by heat is a first order process, obeying normal thermodynamic principles. The initiation of denaturation involves the rupturing of internal non-covalent bonds, a marked conformational change in the protein, and release of associated solvent molecules, resulting in an increase in entropy. Since different proteins have differing energies required to allow this process, temperatures can be selected that will completely disorder one protein, and yet leave another virtually unaffected. The denaturation process is highly *pH* sensitive, and so this should be defined and controlled.

The further away from a physiological *pH* a protein is taken, the less stable it becomes, a protein generally being most stable at its physiological *pH*. Extremes of *pH* cause denaturation because sensitive areas of the protein acquire more like-charges, causing internal repulsion, or perhaps loss of charges which were previously involved in attractive forces holding the protein together. A shift in the *pH* may therefore result in denaturation of the protein and the consequent aggregation will cause its precipitation. Since the denaturation works on a similar mechanism as thermal precipitation, the temperature used whilst subjecting the protein to the extreme *pH* conditions should be carefully controlled, usually at 4-10°C.

3.1.2.2 Salting-out of proteins. Proteins can be selectively precipitated by increasing the salt concentration of the solution, and this property can be exploited in a two-step precipitation, the first leaving the protein of interest in solution, and the second precipitating it. Salting out is largely dependent on the hydrophobic nature of the surface of the protein. As the salt concentration is increased, the removal of water molecules from the hydrophobic regions of the protein, as the salt ions are solvated, causes the exposed hydrophobic regions to attract each other. Those proteins with a larger number, or bigger clusters of such patches, will aggregate sooner, whereas proteins with few nonpolar surface residues will remain in solution even at the highest salt concentrations. The precipitated proteins, not denatured, can then be redissolved by decreasing the salt concentration. Ammonium sulphate is preferred for salting

out, since it is highly soluble, has a low energy of solvation (*i.e.* not endo- or exothermic), and is also available at very high purity, free of heavy metals which poison proteins [16].

3.1.3 Chromatographic Separations of Proteins.

Four chromatographic techniques were applied to the purification of ketohexokinase, although one of these was found not to lead to an increase in purity, and reduced yield. Improvement of chromatographic matrices allowed the optimization of existing purification procedures.

3.1.3.1 *Gel filtration* involves the separation of proteins by size. The protein mixture is applied to a porous gel matrix that retards the flow of solutes and macromolecules, such that smaller molecules are retarded more than larger ones [16]. Consequently, proteins will elute with the largest first. The degree of retardation of macromolecules can be varied by altering the pore size of the matrix. Two matrices are routinely used in this procedure, Sephacryl and Sephadex (Pharmacia LKB), which are both available in differing pore sizes and consequent different protein retention characteristics (see Table 3.2). Sephacryl offers a number of advantages over Sephadex, being both mechanically and chemically more stable, and having a smaller bead size, possesses smaller dead volumes, although it is more expensive. Previous purifications of ketohexokinase have utilized a variety of the matrices available (Sephadex G-200; rat [13], Sephadex G-100; rat [14] and bovine [11] and Sephacryl S-200; human [15]). In the light of the reported sizes of the enzyme from different sources (rat; monomer of 28,000 [14], bovine; dimer of 56,000 [11], and human; dimer of 75,000 [15]), the best reported separation of ketohexokinase using G100 by Sánchez *et al.* [14], and the recent introduction of Sephacryl S-100, it was decided to use this matrix to separate the protein.

Gel type	Fractionation range for globular proteins
Sephadex G-75	3 000 - 80 000
Sephadex G-75 Superfine	3 000 - 70 000
Sephadex G-100	4 000 - 150 000
Sephadex G-100 Superfine	4 000 - 100 000
Sephadex G-150	5 000 - 300 000
Sephadex G-150 Superfine	5 000 - 150 000
Sephadex G-200	5 000 - 600 000
Sephadex G-200 Superfine	5 000 - 250 000
Sephacryl S-100 HR	1 000 - 100 000
Sephacryl S-200 HR	5 000 - 250 000
Sephacryl S-300 HR	10 000 - 1 500 000
Sephacryl S-400 HR	20 000 - 8 000 000

Table 3.2 Separation properties of different gel filtration media available from Pharmacia LKB. (Information from Pharmacia [17].)

3.1.3.2 Ion-exchange chromatography. Proteins bind to ion exchangers by electrostatic forces between the proteins' surface charges (mainly) and the dense clusters of charged groups on the exchangers [16]. Different proteins exhibit different titration curves (Fig. 3.1), and since the net charge of protein is determined by the pH of its solvent buffer, it is possible to use either an anion or cation exchanger to bind the protein, by selecting the appropriate pH . In practice, the choice is based on which exchanger type and pH give the best separation of the molecules of interest (see Fig 3.1), within the constraint of their pH stability. Therefore, if a protein is more stable below its pI , a cation exchanger is used, and *vice versa*. To obtain optimal binding of the protein to be purified, a pH is selected one unit from its pI .

The individual ionic groups of ion exchangers (see Fig. 3.2) have steep titration curves, and thus do not have any buffer activity [18]. Consequently, proteins may be eluted according to their pI if a pH gradient is applied to the proteins bound to the column matrix, such that the gradient covers the pI values of the proteins to be eluted. Alternatively, by gradually increasing the salt concentration of the buffer, and maintaining the pH , proteins are displaced as the counter ions interact with the matrix. Since the strength of the interaction between protein and matrix is dependent on the nature of the protein, different proteins will elute at different salt concentrations, and hence aid purification. By selecting a pH at which the protein of interest

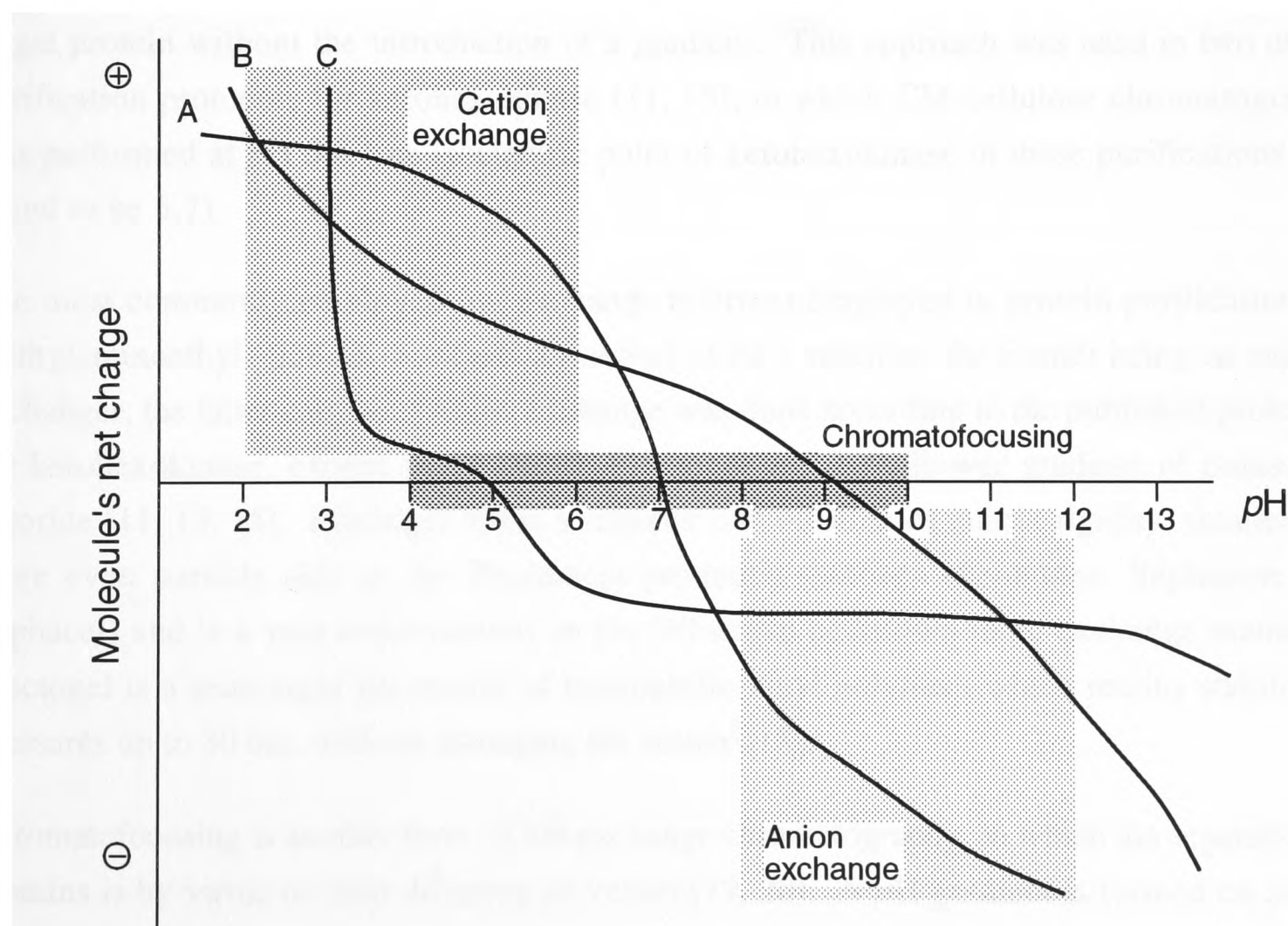


Figure 3.1 The titration curve of three theoretical proteins, showing the pH range of various ion exchange types. Each protein has a different isoelectric points (A; $pI = 7$, B; $pI = 9$, and C; $pI = 5$), and these characteristics may be used to separate the proteins by chromatofocusing. At a pH range below 5, all the proteins are positively charged, allowing separation on a cationic exchanger, and above pH 9, the proteins are negatively charged, allowing separation on an anionic exchanger.

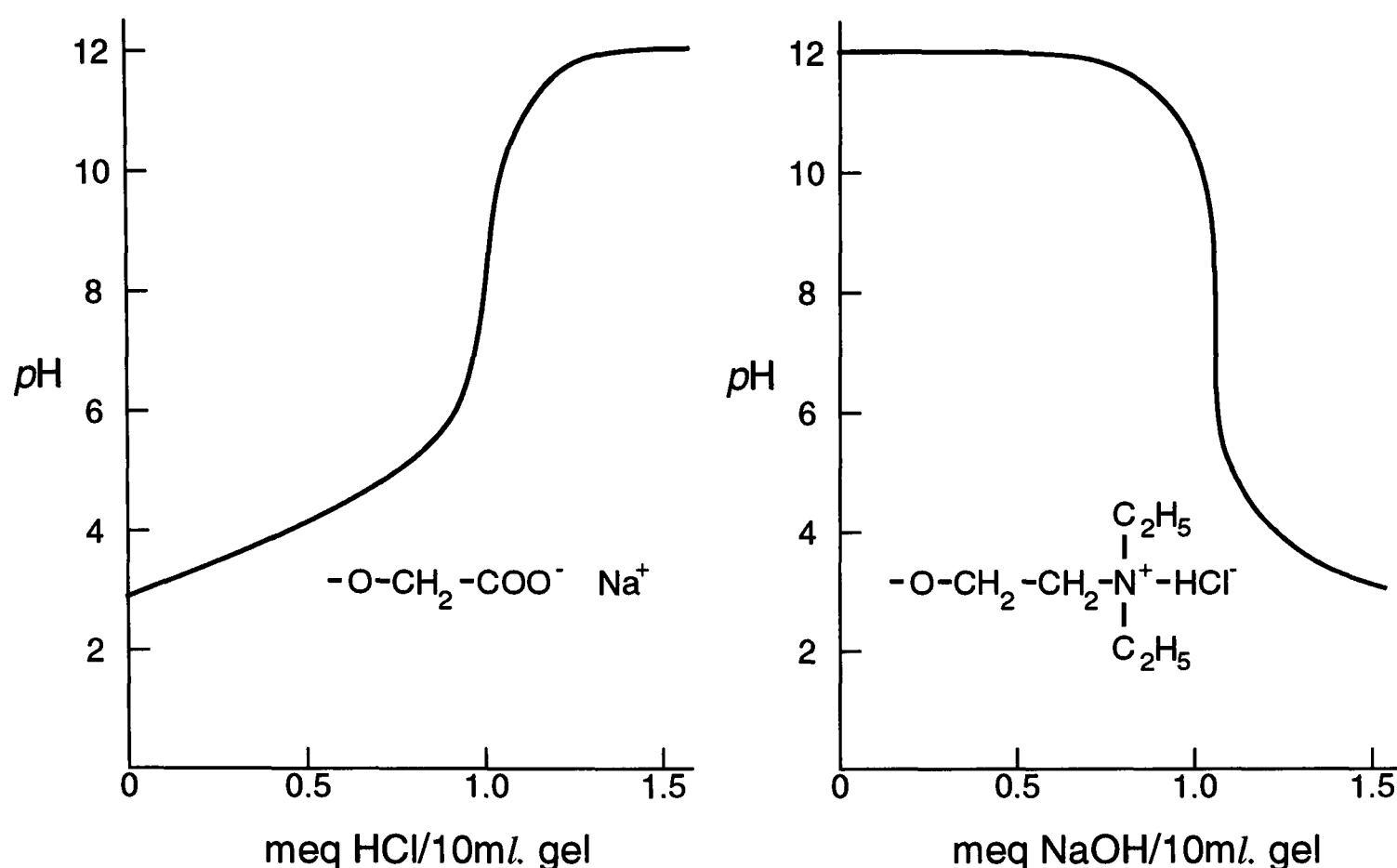


Figure 3.2 Titration curves of Fractogel® TSK CM-650 with hydrochloric acid in 0.5M sodium chloride (left) and Fractogel® TSK DEAE-650 with sodium hydroxide in 0.5M sodium chloride [18].

will not bind, an ion-exchange column can be used to bind contaminating proteins, eluting the target protein without the introduction of a gradient. This approach was used in two of the purification protocols for ketohexokinase [11, 15], in which CM-cellulose chromatography was performed at pH 5.8 (the isoelectric point of ketohexokinase in these purifications was found to be 5.7).

The most commonly employed ion exchange matrixes employed in protein purification are diethylaminoethyl- (DEAE-) or carboxymethyl- (CM-) matrices, the former being an anionic exchanger, the latter cationic. Anion exchange was done according to the published protocols for ketohexokinase, except using DEAE-Fractogel and a shallower gradient of potassium chloride [11, 13, 15]. Fractogel offers a superior combination of matrix rigidity, smaller and more even particle size to the Pharmacia produced matrices (Sephadex, Sepharose and Sephacel) and is a vast improvement on the Whatman celluloses ionic exchange materials. Fractogel is a semi-rigid gel matrix of hydrophilic vinyl polymers which retains stability at pressures up to 30 bar, without damaging the matrix [18].

Chromatofocusing is another form of ion exchange chromatography, in which the separation of proteins is by virtue of their differing *pI* values [19, 20]. A pH gradient is formed on an ion exchanger by taking advantage of the buffering action of charged groups of an ion exchanger (polyethyleimine-agarose), which, though similar to DEAE-absorbents, has a continuous titration of its groups through a large pH range (*Fig. 3.3*). The pH gradient is not formed before loading the buffer on to the column, but generated continuously within the column by

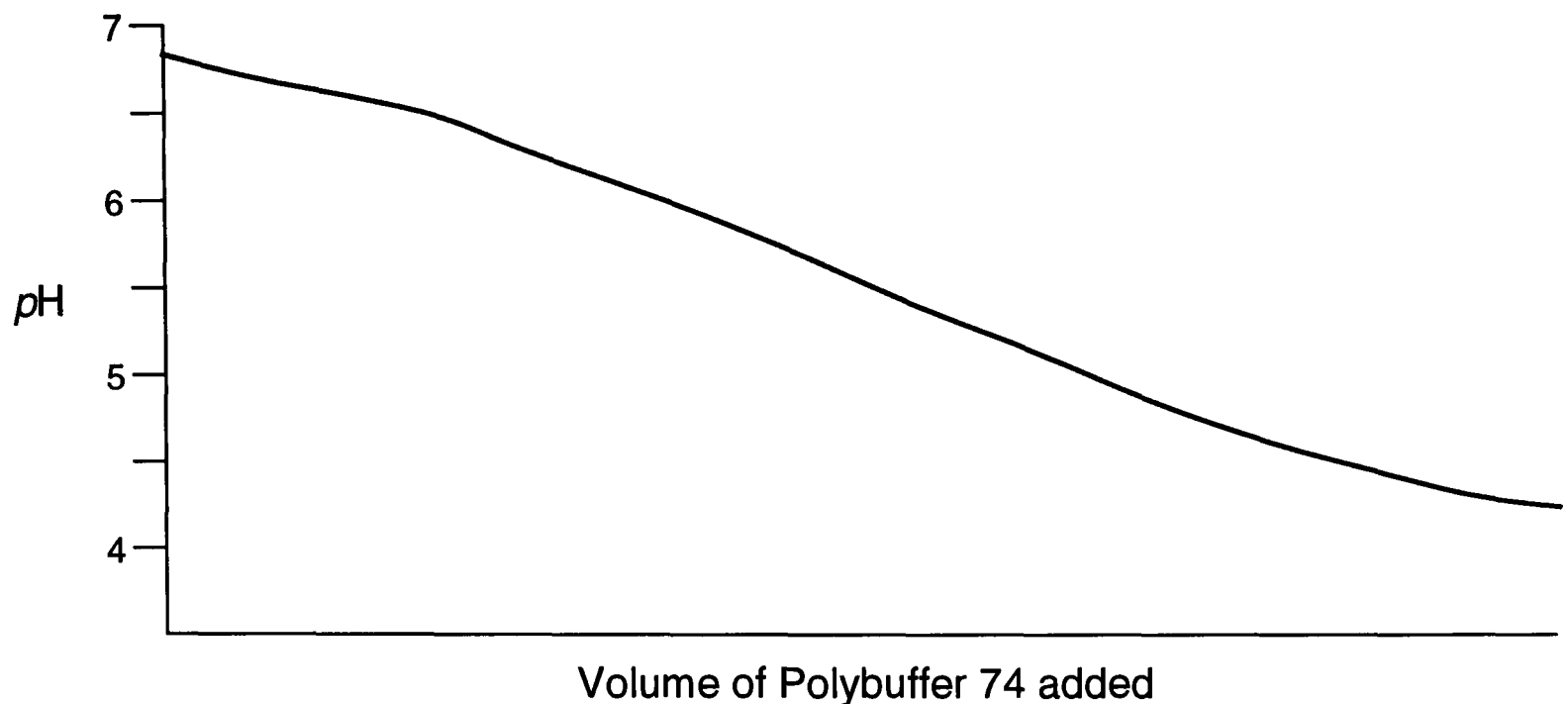


Figure 3.3 The pH elution gradient generated in a column of Polybuffer exchanger PBE 94 eluted with Polybuffer 74, the column initially equilibrated at pH 7, and the Polybuffer added at pH 7.4. (Taken and adapted from Pharmacia technical manual [21].)

addition of the acid form of the polybuffer, just as if two buffers at different pH were gradually mixed in the mixing chamber of a gradient maker.

If such a pH gradient is used to elute proteins bound to the ion exchanger, the proteins elute in order of their isoelectric points (*Fig. 3.4a*). Furthermore, focusing effects takes place, resulting in band sharpening, sample concentration and very high resolution. Consequently, if a second sample of protein is added during a run, it will migrate down the column at the same speed as the eluent until it reaches its *pI*, assuming this pH has not already been eluted, from where it migrates more slowly until it elutes (*Fig. 3.4b*). The power of this technique is outweighed by its expense, and since problems may occur if a protein is unstable at its isoelectric point, it is not practical to use it with crude protein preparations. The use of this method was reported in the purification of the human ketohexokinase [15], the technique was unavailable for the earlier preparations.

The use of an affinity column was also investigated during the development of the purification strategy, the protein was specifically bound to an ATP-AH-Sepharose 4B matrix and eluted with an ATP gradient [22]. However, since no noticeable improvement of purification was detected, this step was dropped from later enzyme preparations, and will not be detailed.

3.1.4 Determination of Protein Molecular Mass.

Three methods were used for the determination of the protein's molecular mass, one in its native form and the other two of the denatured protein. Of the four chromatographic separations detailed above, only one of them separates the protein on the basis of size, and this can be used to determine the molecular mass of the protein in its native state. To achieve this, the gel filtration column must be calibrated with proteins of known molecular weight. Proteins

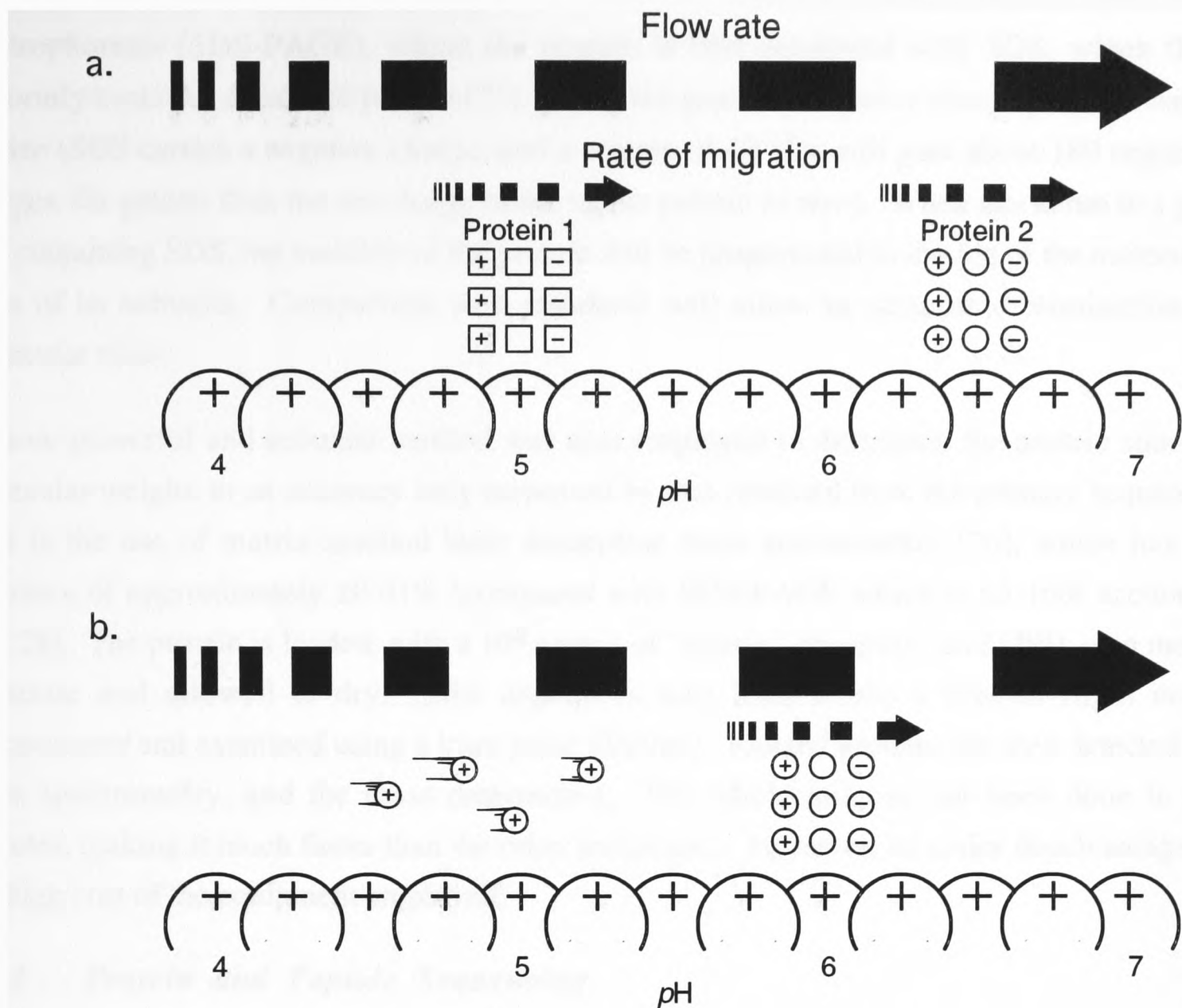


Figure 3.4 a: Behaviour of protein molecules in chromatofocusing. The pH increases down the length of the column. Protein molecules above their isoelectric point in the gradient become negatively charged and bind to the gel. Molecules above their isoelectric point on the gradient become negatively charged and bind to the gel. Molecules below their isoelectric point are repelled from the gel. Molecules at their isoelectric point are neither bound nor repelled. Protein 1 has a hypothetical $pI = 5$, protein 2, $pI = 6.5$.
 b: Focusing effect in chromatofocusing. If a second sample of protein already loaded is applied to the column after the first, it will catch up and co-elute with the first. If a second sample of a protein with a higher pI is applied, it will pass through the first zone and elute before it. (Taken and adapted from a Pharmacia technical manual [21].)

are eluted proportionally to the log of their molecular weight, the largest proteins eluting first. By comparing the elution profile of the standards to that of the protein of interest, the approximate molecular mass can be determined.

Electrophoretic analysis of proteins on polyacrylamide gels [16, 23, 24] also reveals useful information on protein size, either in its native state or as subunits. To achieve the former, proteins are generally run on an gel in an alkaline (pH range 8-9) buffer, in which most proteins are negatively charged, and so move towards the anode. By adding urea to the gel to a concentration of 6-8M., the protein will be denatured and will move proportionally to the logarithm of their subunit size. With both of these methods, however, protein mobility is also affected by the net charge of the protein, and determination of size by comparison to standards can give misleading results. This can be overcome by the use of SDS polyacrylamide gel

electrophoresis (SDS-PAGE), where the protein is first denatured with SDS, which then uniformly coats the denatured protein [25], giving the protein a negative charge proportional to its size (SDS carries a negative charge, and a protein of 40 kDa will gain about 180 negative charges, far greater than the net charge of the native protein *in vivo*). When this is run in a gel, also containing SDS, the mobility of the protein will be proportional to the log of the molecular mass of its subunits. Comparison with standards will allow an accurate determination of molecular mass.

A more powerful and accurate method was also employed to determine the protein subunit molecular weight, to an accuracy only surpassed by that obtained from the primary sequence. This is the use of matrix-assisted laser desorption mass spectrometry [26], which has an accuracy of approximately $\pm 0.01\%$ (compared with SDS-PAGE which is $\pm 5\text{--}10\%$ accurate) [27, 28]. The protein is loaded, with a 10^4 excess of “matrix” (sinapinic acid [29]), on a metal substrate and allowed to dry. This deposit is then loaded into a time-of-flight mass spectrometer and examined using a laser pulse (355nm). Ionized proteins are then detected by mass spectrometry, and the mass determined. The whole process can be done in 15 minutes, making it much faster than the other techniques. However, its major disadvantage is the high cost of the equipment employed.

3.1.5 Protein and Peptide Sequencing.

The sequencing of proteins and peptides has now become a routine process, with almost all the steps involved having been automated. The determination of the sequence of the N-terminus of a protein by the Edman degradation reaction [30-33] may present a problem if the N-terminal residue has undergone modification, rendering it blocked to this reaction. Internal sequence information may then be obtained by internal cleavage of the protein sequence to generate a number of peptides, some or most of which may then be sequenced. Protein cleavage may be performed by enzymatic or chemical methods, the sites of cleavage being dependent on the method used. Traditionally, complete cleavage with proteases was performed, but the advent of automated sequencing equipment has led to the tendency to limit digestion conditions to yield fewer and larger protein fragments. Several proteases are routinely used to digest proteins, and are described in *Table 3.3*. A number of residue-specific chemical reactions have also been elucidated to aid the generation of peptides, although most are not very convenient (*e.g.* cleavage at proline by sodium metal in anhydrous liquid ammonia at -33°C [35]) and only a handful are in routine use. Most notable are the cleavages at methionine using cyanogen bromide [36], which is highly specific and efficient, generating a C-terminal homoserine or homoserine lactone residue, and at tryptophan with *o*-iodosobenzoic acid [37], which yields 70-90% cleavage for most tryptophanyl bonds.

Enzyme	Enzyme commission number (if available)	Specificity
<i>Acidic residues</i>		
<i>Staphylococcus aureus</i> (V8) protease	EC 3.4.21.19	C-terminal to glutamate residues, and some asptyl bonds
Endoprotease Asp-N	-	N-terminal to aspartic and cyteic residues
<i>Basic residues</i>		
Trypsin	EC 3.4.21.4	C-terminal to arginine, lysine and S-aminoethyl-cysteine residues
Clostripain	EC 3.4.22.8	C-terminal to arginine residues
Endoproteinase Lys-C	-	C-terminal to lysine residues
<i>Hydrophobic residues</i>		
Chymotrypsin	EC 3.4.21.1	C-terminal to hydrophobic residues (Phe > Trp > Tyr > Leu)
Thermolysin	EC 3.4.24.4	N-terminal to hydrophobic residues (Leu > Ile > Phe > Val)
Proline-specific endopeptidase	EC 3.4.21.76	C-terminal to proline residues
<i>Less specific proteases</i>		
Elastase	EC 3.4.23.36	Non-polar residues such as alanine, serine, glycine, valine and leucine
Pepsin	EC 3.4.23.36	Not arginine, lysine, proline or isoleucine-containing bonds
Papain and subtilisin	EC 3.4.22.1, EC 3.4.21.14	Non-specific

Table 3.3 Specificities of a variety of endoproteases available for the cleavage of proteins for peptide generation [34].

High performance liquid chromatography (HPLC) and gel electrophoresis are two techniques which have found extensive use in the separation of the peptides generated. Peptides may be separated by their hydrophobic differences on reverse phase columns, or on the basis of charge and size as described above for protein separation. Preparation of microgramme, or smaller, quantities of peptides may then be used to determine extended (20-40 residues) amino acid sequences [38]. The purified peptides are now sequenced by automated solid-phase microsequencing, where the Edman degradation sequencing chemistry centres around the delivery of reagents under the most appropriate conditions and the subsequent removal of the reaction products in such a way as to allow recovery of the released amino acid derivatives, whilst leaving the remainder of the peptide in the reaction chamber. The solid-phase approach

[39] involves the covalent attachment of the protein or peptide to an insoluble support, the washing and recovery processes thereby being rendered both straightforward and efficient. A second approach utilizes the spinning cup technology, which relies on differential solubility in the washing solutions of peptide versus derivatized amino acid, and on the volatility of certain of the reagents and biproducts. This has been developed into gas-liquid (*e.g.* Applied Biosystems 470A) and pulsed liquid-phase (*e.g.* Applied Biosystems 477A) sequencers, in which the sample is physically retained in a glassfibre filter. In the former case, ionic and hydrogen bonding interactions also help to maintain the polypeptide *in situ*. Washout of protein by polar reactants is prevented by supplying these either in the vapour phase or as metered pulses of liquid, which are subsequently removed from the filter by evaporation. The increased yields of these sequencers has led to the development of microbore HPLC systems for identification and quantification of the phenylthiohydantoin (PTH)-amino acids. With such equipment, routine sequencing of 100 pmol amounts is possible, with detection limits down to 1 pmol or less [40].

3.2 Materials and Methods

3.2.1 General Purification Steps.

The procedure used to purify ketohexokinase was a modification of the methods of Adelman *et al.* [13], Sánchez *et al.* [14], Bais *et al.* [15] and Raushel and Cleland [11], using improved column materials and conditions. All steps were done at 4°C, and centrifugations were performed at 27,750g, except where stated to the contrary. Filtrations were made through miracloth [Calbiochem]. Enzyme activity in fractions of column eluent was assayed as described below, and the activity of pooled samples by the method of Adelman *et al.* [13].

3.2.2 Liver Preparation and Enzyme Extraction by Selective Precipitations.

Eight male Sprague Dawley rats (250-300g), fed on a diet of standard rat chow, were killed by cervical dislocation and exsanguinated by severing the carotid arteries. Their livers were used immediately for enzyme purification. Excised livers (c.135g) were twice washed in ice cold 0.25M. mannitol before being homogenized for 10 minutes in 1.5 volumes of the same using a “stomacher”. The homogenate was filtered and the filtrate (Fraction A) was centrifuged for 5 minutes. The pH of the filtered supernatant (Fraction B) was adjusted to 5.0 by addition of 1M. HCl with constant stirring, and was then centrifuged (30 minutes) and the supernatant filtered to yield Fraction C. The pH of this fraction was adjusted to 7.3 with 1M. KOH, heated to 65°C for 2 minutes, all with stirring, quickly cooled to 4°C, filtered and then centrifuged for 30 minutes. The supernatant, Fraction D, was slowly (10 minutes) taken to 40% saturation

with solid ammonium sulphate and stirred for 30 minutes. This was centrifuged for 10 minutes and solid ammonium sulphate added in the same manner to the supernatant (Fraction E) to bring the saturation to 50%. Centrifugation was repeated and the pellet was dissolved in 10ml. 10mM. triethanolamine (TEA) buffer, pH 7.3, 1mM. dithiothreitol to yield Fraction F.

3.2.3 *Purification to Homogeneity by Chromatographic Separations.*

Fraction F was applied to a 108cm × 4cm² Sephacryl S100 [Pharmacia LKB Biotechnology] gel filtration column previously equilibrated with at least 3 volumes TEA buffer. This, and other columns used during the purification of ketohexokinase, were custom-made with dimensions calculated to give optimal separation according to Scopes [16]. Fractions (3ml.) of eluent were collected (flow rate = 3ml./min.) and assayed for both protein concentration (A₂₈₀) and activity (see below). Fractions showing activity were pooled, concentrated [Amicon Ultrafiltration Cell, model M3 IG 2364, Diaflo[®] membrane YM10] to yield Fraction G which was loaded on to a 19cm × 5cm² Fractogel[®] TSK DEAE - 650(S) [Merck] ion exchange column pre-equilibrated with the same buffer. A 20ml. volume of buffer was eluted before applying a 0-200mM. potassium chloride gradient in TEA buffer over 400ml. at 2ml./min. Fractions (2ml.) were collected and assayed as above and their conductance measured to estimate potassium chloride concentration. Samples showing activity were pooled and concentrated by ultrafiltration (as above) before dialysis against 25mM. histidine-HCl, pH 6.2 (Fraction I). This was loaded on to a PBE94 chromatofocusing column (15cm × 0.75cm²) of pre-equilibrated with the same buffer. The column was then washed with 1.5 column volumes before eluting at a flow rate of 0.33ml./min. with Polybuffer 74-HCl, pH 4.5 [Pharmacia LKB Biotechnology], thus generating a pH gradient (pH 6.2-4.5). Fractions (0.5ml.) were assayed as above, and their pH determined using a pH meter. Those showing enzyme activity were pooled, concentrated and dialysed against 100mM. ammonium carbonate / acetic acid, pH 8.1 (Fraction Chr) in preparation for treatment with proteases.

3.2.4 *Assay of Ketohexokinase Activity.*

3.2.4.1 Spectrophotometric assay. This was performed according to Adelman *et al.* [13]. The formation of ADP was coupled to PEP, pyruvate kinase, NADH and lactate dehydrogenase and the diminution in A₃₄₀ was continuously recorded. Each cuvette contains 3ml. reaction mixture

33mM.	TEA, pH 7.3
100mM.	Potassium chloride
20mM.	Magnesium chloride
0.33mM.	NADH
15mM.	ATP
3.3mM.	PEP
5mM.	Fructose

containing 10U. each of pyruvate kinase and lactate dehydrogenase, and a 100 μ l. sample of enzyme. The reaction was measured at 25°C against a reference of the same lacking fructose. For samples A-C, a control was also conducted in the absence of ATP to account for the magnitude of sorbitol dehydrogenase activity. Under these conditions described above, KHK activity was linear with time for at least 10 minutes and was proportional to the amount of enzyme added [13].

3.2.4.2 Colorimetric assay. To allow the rapid detection of fractions eluted from chromatographic columns, a simple colorimetric assay was developed. In this method, the phosphorylation of fructose was linked to the generation of pyruvate. Detection of this product was accomplished using pyruvate oxidase, coupled to the POD / 4-aminoantipyrene/phenol indicator system [41, 42]. A 5 μ l. aliquot of sample was added to 1ml. reaction mixture

75mM. Potassium phosphate, pH 7.2
 300mM. Potassium chloride
 20mM. Magnesium chloride
 15mM. Adenosine 5'-triphosphate
 5mM. Phospho-*enol*-pyruvate
 0.2mM. Thiamine pyrophosphate chloride (TPP)
 8 μ M. Flavin adenine dinucleotide (FAD)
 10mM. Phenol
 8mM. 4-Aminoantipyrene
 5mM. Fructose

containing 2U. peroxidase, 2U. pyruvate oxidase and 3U. pyruvate kinase¹. The reaction was incubated at 37°C for 60 minutes prior to quenching with 100 μ l. 110mM. EDTA. The absorbance at 510nm was then read spectrophotometrically.

3.2.4.3 Determination of protein concentration. Protein concentrations of fractions were determined approximately by spectrophotometric measurement at 280nm. The concentration in Fractions A-Chr was determined by the method of Lowry *et al.* [43] using the

¹To simplify the assay, stock solutions, sufficient for the performance of fifty assays, were prepared as follows:

- | | |
|-------------|---|
| Solution 1: | 413.25mg ATP, disodium salt, trihydrate,
58.5mg phospho- <i>enol</i> -pyruvate
80mg aminoantipyrene
5ml. 250mM. potassium phosphate, pH 7.2, 2mM. TPP, 80 μ M. FAD
10ml. 250mM. potassium phosphate, pH 7.2
13.5ml. water. |
| Solution 2: | 1.118g potassium chloride
200mg magnesium chloride
47mg phenol
45mg fructose
15ml. water. |
| Solution 3: | 100U pyruvate oxidase (lyophilized, other enzymes added directly to bottle)
40 μ l. peroxidase (2500U/ml. suspension)
75 μ l. pyruvate kinase (2000U/ml. suspension)
882 μ l. 250mM. potassium phosphate, pH 7.2, 80 μ M. FAD. |

10 μ l. sample, 670 μ l. Solution 1, 300 μ l. Solution 2 and 20 μ l. Solution 3 are mixed for each reaction. Stock solution could be stored at -20°C for several months without noticable effect upon the assay.

modification of Bensadoun and Weinstein [44] that first precipitates the protein to remove impurities that might affect the assay. Samples containing 5-50 μ g of protein in solution were taken to 3ml. with water and 25 μ l. 2% w/v sodium deoxycholate was added and allowed to stand for 15 minutes. 1ml. 24% w/v TCA was added, and this was vortexed and centrifuged, in flat bottomed tubes, at 3300g for 10 minutes. The supernatant was aspirated away and the pellet dissolved in 1.5ml. Lowry C, which was freshly made:

50ml. Lowry A	0.1M. Sodium hydroxide
	2% w/v Sodium carbonate
1ml. Lowry B	1% w/v Sodium potassium tartrate
	0.5% w/v Copper sulphate.

Finally 150 μ l. Lowry D (Folin Coicalteau phenol [45]:water, 1:1) was added and the reaction was left for 45 minutes before spectrophotometric measurement at 660nm against a reference lacking protein. Protein concentration was determined by comparison with a set of BSA standards.

3.2.5 Polyacrylamide Gel Electrophoresis of Protein Samples.

3.2.5.1 Casting and running of SDS discontinuous gels. SDS discontinuous gel electrophoresis [46] was performed according to Hames [24]. The resolving gel was made according to *Table 3.4*, and the mixture is degassed under reduced pressure before the addition of the polymerization agents (ammonium persulphate and TEMED [BDH Electran]) and casting between the glass plates in a Biorad PROTEAN I gel casting cassette. Gradient gels were prepared using a gradient former containing each of the two extreme acrylamide concentrations of the gel in each side of the gradient maker, the constant flow through the apparatus and to the plates being achieved using a peristaltic pump.

When a gel (20 $\times$ 17 $\times$ 0.15cm) had been poured, it was immediately overlayed with 50% v/v aqueous isopropanol and left to polymerize overnight. This isopropanol was then removed, the gel surface washed with 1 $\times$ resolving buffer and a sample well comb placed in position. A stacking gel mixture is made as in *Table 3.4*, again degassed before the addition of the polymerization agents, and was then poured on the resolving gel. When the stacking gel had set (2-3 hours), the sample comb was slowly removed, allowing the wells to fill with 1 $\times$ reservoir buffer (10 $\times$ reservoir stock is 0.25M. Tris, 1.92M. glycine, pH 8.3, 1% w/v SDS).

Protein samples to be loaded were first precipitated and washed with acetone, dissolved in 50 μ l. loading buffer

62.5mM.	Tris.HCl pH 6.8
2% w/v	SDS
5% w/v	Dithiothreitol
10% v/v	Glycerol
0.002% w/v	Bromophenol blue

Stock solution	Stacking gel	Final acrylamide concentration in resolving gel (%) ^a						
		20.0	17.5	15.0	12.5	10.0	7.5	5.0
Acrylamide-bisacrylamide ^b	2.5	20.0	17.5	15.0	12.5	10.0	7.5	5.0
Stacking gel buffer stock ^c	5.0	-	-	-	-	-	-	-
Resolving gel buffer stock ^c	-	3.75	3.75	3.75	3.75	3.75	3.75	3.75
10% SDS	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3
1.5% ammonium persulphate	1.0	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Water	11.3	4.45	6.95	9.45	11.95	14.45	16.95	19.45
TEMED	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015

Table 3.4 Recipe for Gel Preparation using the SDS-discontinuous Buffer System ([24]).
Final concentrations of buffers: stacking gel; 0.125*M*. Tris.HCl, pH 6.8
resolving gel; 0.375*M*. Tris.HCl, pH 8.8
^aThe columns represent volumes (ml.) of the various reagents required to make 30ml. of gel mixture.
^b30g acrylamide and 0.8g bisacrylamide [BDH Electran] was dissolved in water to a total volume of 100ml. The solution was filtered through Whatman No. 1 and stored in the dark at 4°C. Prolonged storage results in the hydrolysis of acrylamide, and so solutions must be replaced after about two months.
^cStacking gel buffer stock (0.5*M*. Tris.HCl, pH 6.8) and Resolving gel buffer stock (3.0*M*. Tris.HCl, pH 8.8) were filtered through Whatman No. 1 filter paper and stored at 4°C.

and boiled for 3 minutes before loading a 25µl. aliquot. SDS Molecular Weight Markers MW-SDS-70L [Sigma Chemical Company] were also loaded according to the manufacturer's instructions. Gels were run at 100mA constant current with water cooling to 12°C in a Biorad PROTEAN I Vertical gel electrophoresis cell until the bromophenol blue dye front had reach the end of the gel (c.18 hrs.). The upper reservoir had thioglycollic acid added to 1mM. to prevent oxidation of proteins by residual hydrogen peroxide produced during the polymerization of acrylamide [47, 48].

3.2.5.2 Fixing and staining of gels. Gels were either stained with Coomassie Blue and/or silver. All steps in both staining procedures used gentle agitation on a rotary shaker for each step. Coomassie Blue staining was performed as described in Hames [24], the gel was first fixed and stained overnight with Coomassie Blue (40% v/v methanol, 7% v/v acetic acid 0.5% w/v Coomassie Blue) and then destained over 2 days with several washes with 30% v/v methanol, 10% v/v acetic acid containing Dowex 1 (chloride form) ion exchange resin held in dialysis tubing. The silver staining protocol was a modification of that of Morrissey [49]. The gel was first fixed in 25% v/v propanol, 10% v/v acetic acid for 30 minutes before soaking first in 5% v/v propanol, 7% v/v acetic acid, then 10% v/v glutaraldehyde, each for 30 minutes. It was then washed at least four times over 2 hours in double distilled water, and then soaked in fresh aqueous dithiothreitol (5µg/ml.) for thirty minutes. This was poured away, and without

rinsing, 0.1% w/v silver nitrate was added and left for 30 minutes. The gel was rinsed once, rapidly, in distilled water, and then twice in developer (3% w/v sodium carbonate with 500 μ L/l. 37% w/v formaldehyde freshly added). It was then soaked in developer until the desired level of exposure was achieved on the protein bands. The developer was then poured away, and without rinsing, background remover (2.1% w/v sodium thiosulphate, 3.7% w/v Tris base) was added and left to soak until the desired effect was again achieved. The gel was then washed in water for 30 minutes and soaked in 0.03% w/v sodium carbonate for 10 minutes before storage in 3% v/v aqueous glycerol.

3.2.6 Molecular Mass Determination of KHK by Mass Spectrometry.

The molecular mass of the purified protein was accurately determined by matrix assisted laser adsorption on a Finnigan MAT LASERMAT "Time of flight mass spectrometer". A 0.2 μ L or 1.0 μ L sample of the purified protein in PBE94 polybuffer (c. 0.76mg/ml.) was mixed with 0.2 μ L sinapinic acid matrix on a gold plated, stainless steel target and allowed to dry under an infra-red lamp. This was then put in the path of the ionizing laser of the mass spectrometer and the relative molecular masses of ions generated during irradiation were measured.

3.2.7 Preparation and Sequencing of Peptides of ketohexokinase.

A sample (c. 30 μ g) of purified ketohexokinase in 45 μ L ammonium carbonate buffer, pH 8.1 was diluted with 50 μ L 100mM ammonium bicarbonate pH 8.3 / 5% v/v acetonitrile. To this, 2 μ g each of Endoproteinase Glu-C and Trypsin [both Boehringer "Sequencing Grade" enzymes] were added and the reaction incubated for 60 minutes at 37°C. A control reaction containing protease only was also performed. After digestion, 100 μ L 8M Guanidine HCl / 0.5 M Tris / 0.35mM EDTA was added to each. The samples were then reduced by the addition of 2 μ L of a 100mg/mL solution of dithiothreitol and incubated at 37°C for 60 minutes. In order to alkylate the cysteines, a 2 μ L volume of 4-vinyl pyridine was then added to each and the mixtures were incubated at 37°C for 30 minutes. Both samples were then chromatographed by reverse-phase HPLC on a Brownlee Aquapore RP300 (C8) 220 $\times$ 2mm column loading with 0.1% v/v trifluoroacetic acid (TFA) in water and applying a 5-50% v/v gradient of 0.09% v/v TFA in 80% v/v acetonitrile over 45 minutes, then 50-90% v/v over 15 minutes. The sample and control were then compared to eliminate peptides derived from the protease themselves, and the seven major peaks were sequenced on an Applied Biosystems 470A protein sequencer with on-line phenylthiohydantoinyl (PTH) analysis.

3.3 Results and Discussion

The protein purification scheme is summarized in *Fig. 3.5*, and shows the code for each of the fractions along the process. Rat liver ketohexokinase was purified to homogeneity, using silver-stained SDS-PAGE as the criterion for purity. The degree of purification (166-fold) (*Table 3.5*), though low, when compared with previous reports [11, 13-15], results from the use of a purer starting homogenate than those methods employing bladed blenders that break open organelles. The mild homogenization employed allows the sedimentation of organelles intact, producing a cytoplasmic fraction that is minimally contaminated with proteins released from organelles. This is evinced by the considerably higher specific activity of our starting homogenate when compared with previous reports. The 166-fold purification suggests that ketohexokinase constitutes approximately 0.6% of the cytoplasmic proteins of rat livers. Comparisons of elution profiles (*Fig. 3.6 a-c*) reveal that all of our modifications to the chromatographic methods resulted in considerably improved resolution compared with those previously used for purification of this protein.

The colorimetric assay for ketohexokinase developed to locate activity in chromatographic fractions allows easy detection of the enzyme and obviates the requirement for controls to account for sorbitol dehydrogenase activity. However, the increase in absorbance at 510nm was found to be insufficiently linear to allow precise quantitative analysis, and so the spectrophotometric assay described by Adelman *et al.* [13] was employed for this purpose.

The size of the native protein was estimated to be about 130 kDa (see *Fig. 6.5*), comparing the elution profile of the Sephacryl S-100 gel filtration column with that of the protein standards γ -globulin (155 kDa), bovine serum albumin (67 kDa), ovalbumin (38 kDa) and cytochrome c (14.5kDa). Comparing this value to the separation efficiencies of the column matrices (see *Table 3.2*) would suggest that the wrong matrix had been used. However, comparison of the S-100 profile with that from a S-300 column of the same dimensions (I. A. Donaldson, personal communication) showed that resolution of proteins eluting with the ketohexokinase activity was not improved on the S-300 column, although improved resolution was observed with proteins of higher molecular weights than ketohexokinase. The purified protein displayed a relative mobility on SDS/PAGE (*Fig. 3.7*) consistent with a molecular mass of 31 kDa, but a more accurate value of 32.8 kDa was obtained by mass spectrometry. This value is in close agreement with that later calculated from the primary sequence (32, 728 Da, see 5.3.1.3). The mass spectrum also revealed a potassium ion peak, which, since the buffer does not contain potassium, we assume was bound to the protein. This is consistent with the requirement for potassium by the enzyme [12]. These results indicate that ketohexokinase exists as a tetramer in its native form. This is in contrast to previous reports of the quaternary homodimer

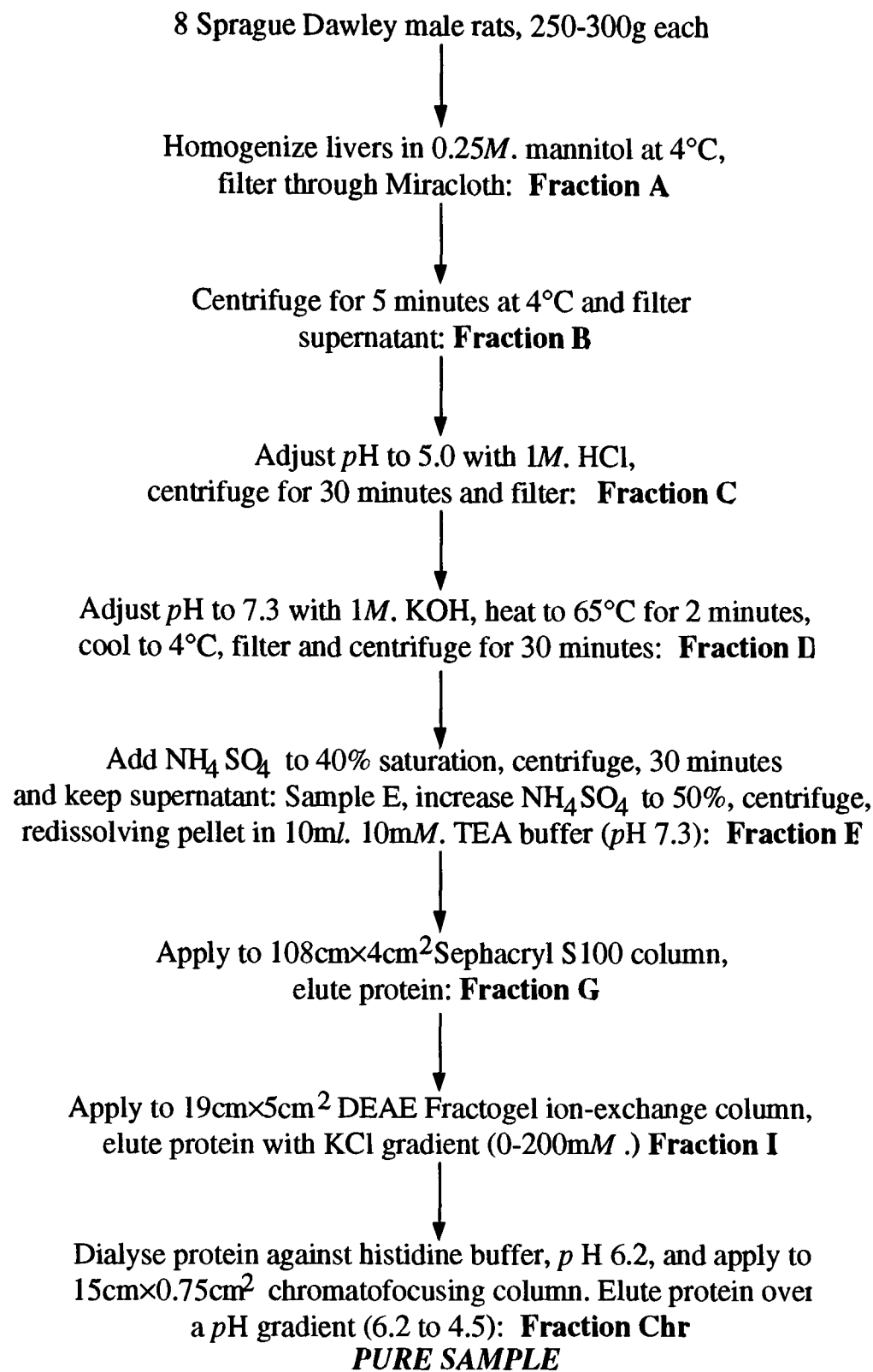


Figure 3.5 Diagrammatic representation of the purification of ketohexokinase from a crude rat liver homogenate. Fractions along the purification are shown as described in sections 3.2.2-3, and the elution profiles of the chromatographic separations are shown in **Fig. 3.6**.

structure of the native protein from bovine [11] and human [15] livers, and the monomeric structure previously reported for the enzyme from rat liver [13, 14].

The N-terminus of ketohexokinase was not accessible to Edman degradation, and was therefore assumed to be blocked. The protein proved to be considerably resistant to enzymatic cleavage. Combined treatment with trypsin (EC 3.4.21.4) and Endoproteinase Glu-C (EC 3.4.21.19) eventually yielded peptides of suitable size for amino acid sequencing. Fractions from the seven largest peptide peaks resolved by reverse phase HPLC (**Fig. 3.8**) were sequenced. Two of these, KHK 38 and KHK 43, were found to contain two peptides,

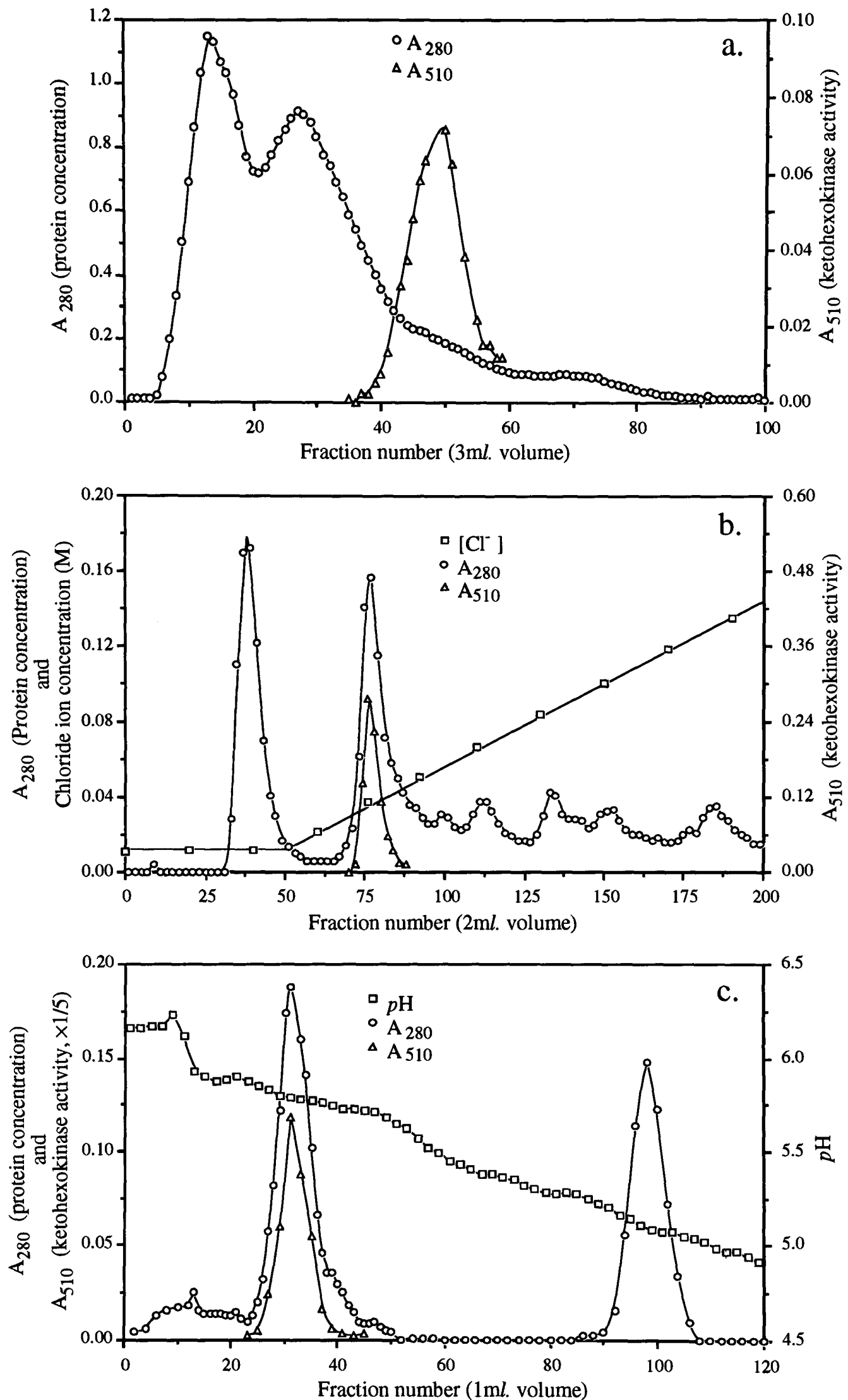


Figure 3.6 Chromatographic elution profiles showing the ketohexokinase activity peaks. In each case, the plot symbols are shown on the graph, the ketohexokinase activity was assayed by the colorimetric assay described in the materials and methods (3.2.4.1). Graph a shows the gel filtration profile, graph b, the anion exchange profile with potassium chloride gradient, and graph c shows the chromatofocusing profile of the two remaining proteins eluting at different pH values.

Fraction	Volume (ml.)	Total protein (mg)	Activity ($\times 10^2$) ($\mu\text{mol/min/mg}$)	Purification	Total activity ($\mu\text{mol/min}$)	% Total
B	120	4080	7.0	1.0	285	100
C	109	2450	12.6	1.8	309	108
D	83	500	26.0	3.7	130	45.4
E	90	450	28.0	4.0	126	44.0
F	12	100	91	13.0	93	32.5
G	11.7	11.7	309	44.3	46	16.1
I	7.8	2.9	800	115	23.4	8.2
Chr.	11	1.7	1160	166	19.8	6.9

Table 3.5 Purification scheme for rat hepatic ketohexokinase, starting with the livers from eight rats. Details are given in the text.

yielding a pair of PTH-amino acid products for two of the four (KHK 38), and seven of the twenty one (KHK 43) cycles of Edman degradation. This ambiguity was later easily resolved from the cDNA, which allowed distinction of peptides KHK 43a and KHK 43b, and KHK 38a and KHK 38b from one another. The sequence of all the peptides is shown in *Table 3.6*.

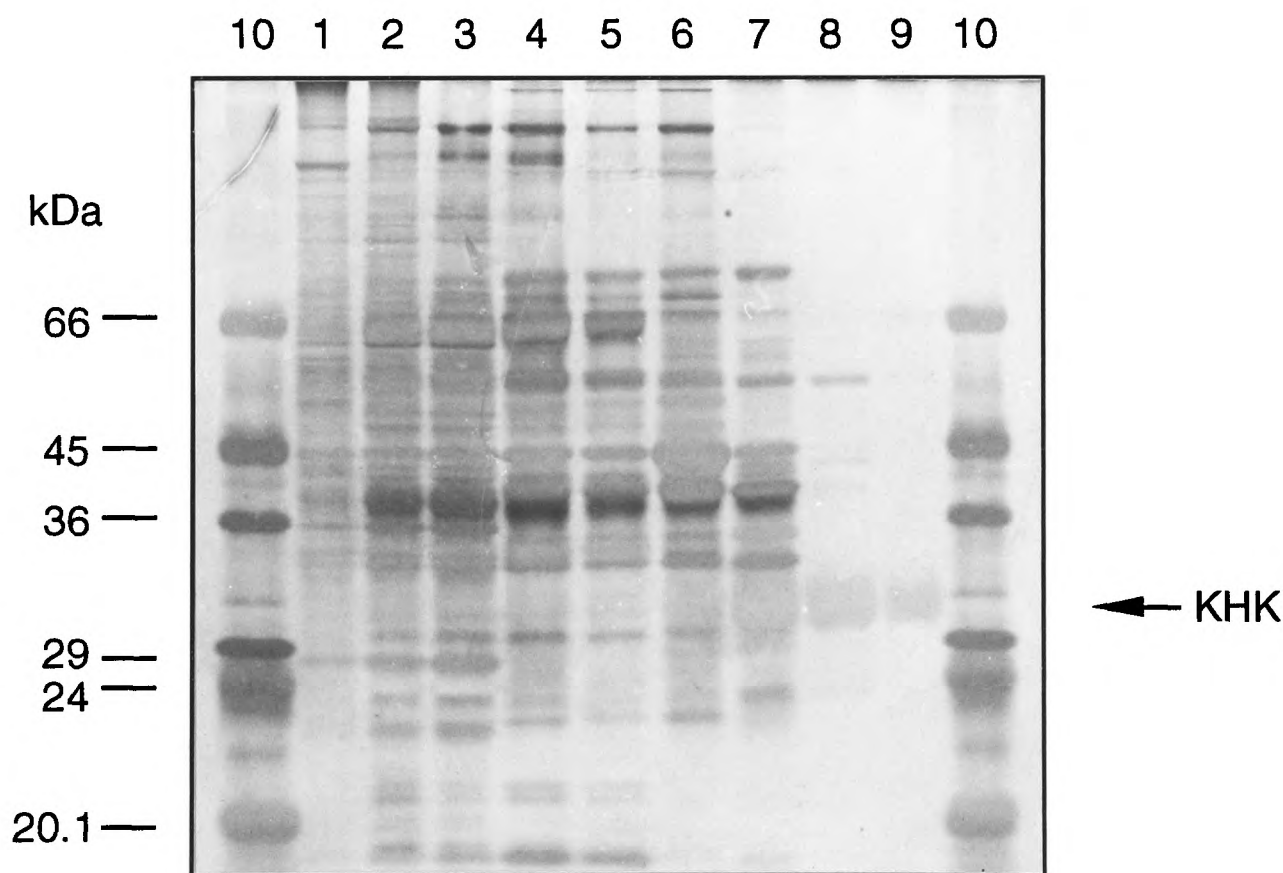


Figure 3.7 Silver stained SDS-polyacrylamide gel electrophoresis showing the purification of KHK from the crude homogenate to the purified sample. From left, the samples are: 1; crude homogenate, 2; crude supernatant (Fraction B), 3; acid treatment (C), 4, heat treatment (D), 5; 40% saturation ammonium sulphate supernatant (E), 6; 50% saturation ammonium sulphate precipitate (F), 7; S-100 size exclusion elution (G), 8; anion exchange elution (I), 9; Chromatofocused pure protein (Chr.), and 10; protein standards, with sizes indicated.

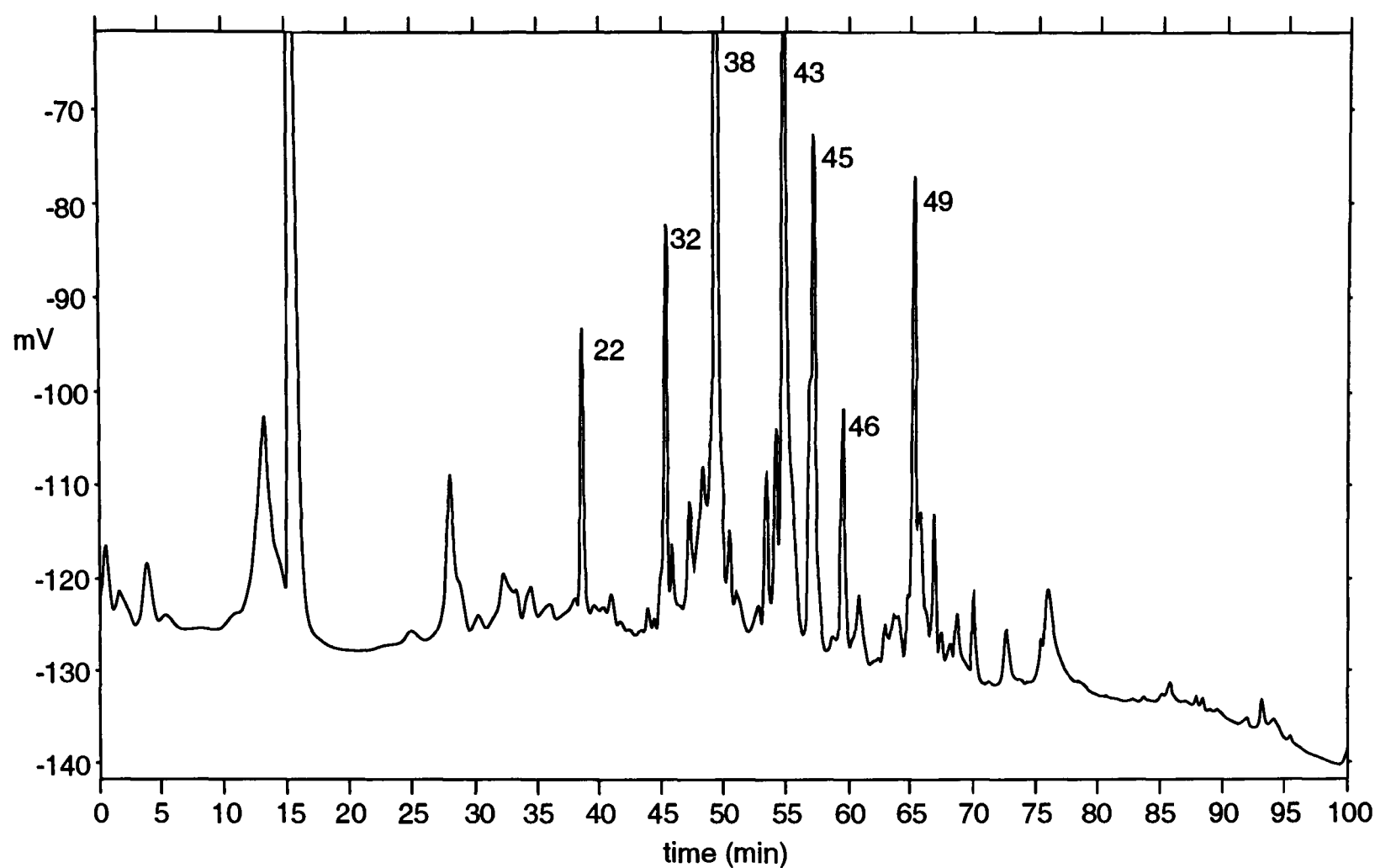


Figure 3.8 Peptide elution profile of ketohexokinase digested with endoproteinase Glu-C and trypsin, separated on a Brownlee Aquapore RP300 (C8) HPLC column. The major peaks that were sequenced are indicated, and their sequences reported in Table 3.6.

Peptide name	Peptide sequence
KHK 22	I E Q Y N A G E
KHK 32	I I N V V D D Y P E
KHK 38a	G V D
38b	I H I Q
KHK 43a	G A D A L G P D G Q L L H S D A F P P P R
43b	T L I C A W A E
KHK 45	T I I L Y D T N L P D V S A K
KHK 46	G G N A S N S - T E
KHK 49	L F Q L F G Y G E

Table 3.6 Peptide sequences of peptides generated by a trypsin / Endoproteinase Glu-C double digestion. Amino acids in italics are those that did not match sequence derived from the cDNA sequence. Where two sequences are given (termed a and b), two peptides were subjected to sequencing in the same reaction. The sequences were resolved from the cDNA derived sequence. All peptides sequenced were confirmed on the derived sequence.

3.4 References

1. Cori, G.T., Ochoa, S., Slein, M.W. and Cori, C.F. (1951). The metabolism of fructose in liver. Isolation of fructose-1-phosphate and inorganic pyrophosphate. *Biochim. Biophys. Acta* **7**, 304-317.
2. Van den Berghe, G. (1986). Fructose: Metabolism and Short-Term Effects on Carbohydrate and Purine Metabolic Pathways. *Prog. Biochem. Pharmacol.* **21**, 1-32.
3. Malaisse, W.J., Malaisse-Lagae, F., Davies, D.R. and Van Schaftingen, E. (1989). Presence of fructokinase in pancreas islets. *FEBS Letts.* **255**, 175-178.
4. Gitzelmann, R., Steinmann, B. and Van den Berghe, G. (1989). Disorders of fructose metabolism, in *The metabolic basis of inherited disease*. (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., Eds.), McGraw-Hill, New York, 6th edn. vol. 1, 399-424.
5. Cross, N.C.P., Tolan, D.R. and Cox, T.M. (1988). Catalytic Deficiency of Human Aldolase B in Hereditary Fructose Intolerance Caused by a Common Missense Mutation. *Cell* **52**, 881-885.
6. Schapira, F., Schapira, G. and Dreyfus, J.C. (1961/1962). La lésion enzymatic de la fructosurie bénigne. *Enzym. Biol. Clin.* **1**, 170-175.
7. Càrdenas, M.L., Rabajille, E. and Niemeyer, H. (1984). Fructose is a good substrate for rat liver 'glucokinase' (hexokinase D). *Biochem. J.* **222**, 363-370.
8. Griffin, L.D., Gelb, B.D., Wheeler, D.A., Davison, D., Adams, V. and McCabe, E.R.B. (1991). Mammalian Hexokinase 1: Evolution Conservation and Structure to Function Analysis. *Genomics*, **11**, 1014-1024.
9. Walker, D.G. (1968). The Nature and Function of Hexokinases in Animal Tissues. *Essays Biochem.* **2**, 33-67.
10. Raushel, F.M. and Cleland, W.W. (1973). The Substrate and Anomeric Specificity of Fructokinase. *J. Biol. Chem.* **248**, 8174-8177.
11. Raushel, F.M. and Cleland, W.W. (1977). Bovine Liver Fructokinase: Purification and Kinetic Properties. *Biochemistry* **16**, 2169-2175.
12. Sánchez, J.J., González, N.S. and Pontis, H.G. (1971). Fructokinase from rat liver. II. The role of K⁺ on the enzyme activity. *Biochim. Biophys. Acta* **227**, 79-85.
13. Adelman, R.C., Ballard, F.J. and Weinhouse, S. (1967). Purification and Properties of Rat Liver Fructokinase. *J. Biol. Chem.* **242**, 3360-3365.
14. Sánchez, J.J., González, N.S. and Pontis, H.G. (1971). Fructokinase form rat liver. I. Purification and properties. *Biochim. Biophys. Acta* **227**, 67-78.
15. Bais, R., James, H.M., Rofo, A.M. and Conyers, R.A.J. (1985). The purification and properties of human liver ketohexokinase. *Biochem. J.* **230**, 53-60.
16. Scopes, R.K. (1987). Protein Purification: Principles and Practice. 2nd Edn., Springer-Verlag, New York.
17. Pharmacia LKB Biotechnology technical bulletin (1991). Gel Filtration; Principles and Methods, 5th Edn.

18. Merck technical bulletin (1985). Fractogel® TSK: Polymeric Media for Biochromatography.
19. Sluyterman, L.A.Æ. and Elgersma, O. (1978). Chromatofocusing: isoelectric focusing on ion exchange columns. I. General principles. *J. Chromatogr.*, **150**, 17-30.
20. Sluyterman, L.A.Æ. and Wijdenes, J. (1978). Chromatofocusing: isoelectric focusing on ion exchange columns. II. Experimental verification. *J. Chromatogr.*, **150**, 31-44.
21. Pharmacia LKB Biotechnology technical bulletin (1988). Chromatofocusing with Polybuffer™ and PBE™.
22. Gilham, P.T. (1971). The covalent binding of nucleotides, polynucleotides and nucleic acids to cellulose. *Methods in Enzymol.*, **21**, 191-197.
23. Ornstein, L. (1964). Disc electrophoresis. I. Background and Theory. *Ann. N.Y. Acad. Sci.*, **121**, 321-349.
24. Hames, B.D. (1981). An introduction to Polyacrylamide Gel Electrophoresis, in *Gel electrophoresis of proteins; a practical approach* (Hames, B.D. and Rickwood, D., Eds.), IRL Press, Oxford, 1-91.
25. Reynolds, J.A. and Tanford, C. (1970). Binding of dodecyl sulphate to proteins at high binding ratios: Possible implications for the state of proteins in biological membranes. *Proc. Nat. Acad. Sci. USA.*, **66**, 1002-1007.
26. Geisow, M.J. (1992). Mass measurement at high molecular weight-new tools for biotechnologists. *Trends Biotech.*, **10**, 432-441.
27. Beavis, R.C. and Chait, B.T. (1990). Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc. Natl. Acad. Sci. USA*, **87**, 6873-6877.
28. Beavis, R.C. and Chait, B.T. (1990). High-Accuracy Molecular Mass Determination of Proteins Using Matrix-Assisted Laser Desorption Mass Spectrometry. *Anal. Chem.*, **62**, 1836-1840.
29. Beavis, R.C. and Chait, B.T. (1989). Cinnamic Acid Derivatives as Marices for Ultraviolet Laser Desorption Mass Spectrometry of Proteins. *Rapid Commun. Mass Spec.*, **3**, 432-439.
30. Edman, P. (1950). Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chem. Scand.*, **4**, 283-293.
31. Edman, P. (1960). Phenylthiohydantoins in Protein Analysis. *Ann. N. Y. Acad. Sci.*, **88**, 602-610.
32. Tarr, G.E. (1977). Improved Manual Sequencing Methods. *Meths. Enzymol.*, **47**, 335-357.
33. Levy, W.P. (1981). Manual Edman Sequencing Techniques for Proteins and Peptides at the Nanomole Level. *Meths. Enzymol.*, **79**, 27-31.
34. Aitkin, A., Geisow, M.J. Findlay, J.B.C., Holmes, C. and Yarwood, A. (1989). Peptide preparation and characterization, in *Protein sequencing; a practical approach* (Findlay, J.B.C. and Geisow, M.J., Eds.), IRL Press, Oxford, 43-68.

35. Hempel, J. and Jornvall, H. (1985). Cleavage at Acyl-Proline Bonds with Sodium in Liquid Ammonia: Application with Nanomolar Amounts of Products by High-Performance Liquid Chromatography for Structural Analysis. *Anal. Biochem.*, **151**, 225-230.
36. Gross, E. and Witkop, B. (1961). Selective cleavage of the methionyl peptide bonds in ribonuclease with cyanogen bromide. *J. Am. Chem. Soc.* **83**, 1510-1511.
37. Mohoney, W.C., Smith, P.K. and Hermondson, M.A. (1981). Fragmentation of Proteins with *o*-Iodosobenzoic Acid: Chemical Mechanism and Identification of *o*-Iodoxybenzoic Acid as a Reactive Contaminant that Modifies Tyrosyl Residues. *Biochemistry*, **20**, 443-448.
38. Wilson, K.J. and Yuan, P.M. (1989). Protein and peptide purification, in *Protein sequencing; a practical approach* (Findlay, J.B.C. and Geisow, M.J., Eds.), IRL Press, Oxford, 1-41.
39. Laursen, R.A. and Bonner, A.G. (1970). An automated solid-phase peptide sequenater. *Fed. Proc.*, **29**, 727.
40. Geisow, M.J. and Aitken, A. (1989). Gas- or pulsed liquid-phase sequence analysis, in *Protein sequencing; a practical approach* (Findlay, J.B.C. and Geisow, M.J., Eds.), IRL Press, Oxford, 85-98.
41. Bergmeyer, H.U., Gassl, M. and Walter, H.-E. (1983). Pyruvate Oxidase, in *Methods in Enzymatic Analysis* (Bergmeyer, H.U. Ed.) 2nd Edn. Verlag Chemie, Weinheim, Vol. 2, 304-305.
42. Bauminger, B. B. (1974). Micro method for manual analysis of true glucose in plasma without deproteinization. *J. Clin. Pathol.*, **27**, 1015-1017.
43. 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.
44. Bensadoun, A. and Weinstein, D. (1976). Assay of Proteins in the Presence of Interfering Materials. *Anal. Biochem.*, **70**, 241-250.
45. Folin, O. and Ciocalteau, V. (1927). Tyrosine and tryptophan determination in proteins. *J. Biol. Chem.*, **73**, 627-650.
46. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.
47. Brewer, J.M. (1967). Artifact Produced in Disc Electrophoresis by Ammonium Persulphate. *Science*, **156**, 256-257.
48. Woods, E.F. (1967). Molecular Weight and Subunit Structure of Tropomyosin B. *J. Biol. Chem.*, **242**, 2859-2871.
49. Morrissey, J.H. (1981). Silver Stain for Proteins in Polyacrylamide Gels: A Modified Procedure with Enhanced Uniform Sensitivity. *Anal. Bioch.*, **117**, 307-310.

The Cloning of a cDNA Probe for the Kethexokinase Gene

4.1 Introduction

4.1.1 Screening a cDNA Library.

The availability of several microgrammes of purified kethexokinase, and the partial amino acid sequence described in the previous chapter (see *Table 3.5*) allowed progression to the screening of a cDNA library, in order to obtain the complete coding sequence for the protein.

cDNA libraries can be constructed in a number of forms, which can be categorized as either expression or non-expression libraries. All these libraries exist as cDNA sequences cloned into specific regions of a phage or plasmid. It is the plasmid or phage used that determines whether the inserted cDNA is expressed. Expression libraries are screened either by complementing deficiencies in mutants, or by probing with antibodies generated against the protein of interest. Libraries that do not express cloned cDNA can be screened by using cDNA probes representing differentially expressed mRNAs [1], or by hybridizing DNA probes. Both these methods, however, may also be used to screen expression libraries. In most cases, the number of independent clones of a cDNA in a library is related to the abundance of its species of mRNA in the mRNA population, which broadly relates to the concentration of its protein in the source tissue. Consequently, a very rare protein would require the screening of a large number of clones to yield the desired sequence. The size of the genome of the organism from which a cDNA library has been constructed also has an important bearing on the number of clones that need to be screened [2].

4.1.1.1 Cloning by complementation. Although this is a very powerful technique, it is the least frequently accessible of the methods for screening cDNA libraries, because it requires a suitable mutant for selection. Its main advantage, however, is the ability to select genes, without having isolated the proteins which they encode. This technique is not restricted to screening in bacteria, but can be extended to eukaryotic organisms such as yeast, provided suitable mutants are available. The cDNA library is cloned into an expression cassette in a plasmid, and this is used to transform the selection mutant. The transformants are then screened for the desired phenotype conferred by the cDNA. As well as requiring a suitable

selection organism, this method is further limited by the requirement for full length clones to produce active proteins, and the potential problem of the heterologous protein being toxic to the selection organism. Furthermore, the tendency for false-positives to arise from either reversion of the mutation, or the ability of more than one gene to complement, may also restrict the application of this approach. As it was, the lack of a suitable mutant for selection rendered this approach inapplicable for the cloning of ketohexokinase.

4.1.1.2 Screening expression libraries with antibody probes. If a suitable quantity of purified protein is available, it is possible to raise polyclonal antibodies against it by injecting a small mammal, such as a rabbit, with the protein, and after several weeks isolating the antibody from a sample of the serum [3]. As long as antigen boosts are given every 6 - 8 weeks, and the rabbit remains healthy, antibody may be collected weekly. The cDNA library, in an expression vector, such as λ gt11 [4], is used to infect a lawn of *E. coli*. Since constitutive expression of the library might be toxic to the *E. coli* host, and might result in cell death before sufficient amounts of antigen is produced, the cDNA sequences are cloned into a *lacZ* site, under the tight control the *lac* operon [5]. Once the number of infected cells around the plaques is sufficiently large, the *lacZ*-directed gene expression is induced by the addition of IPTG [6]. This is often added by impregnating a nitrocellulose filter with the IPTG and placing this over the lawn for several hours. When the filter is lifted, proteins released from the lysed cells will be bound to it, and the filter can be washed and incubated with the antibody solution. The antibody binding can then be visualized by binding with ^{125}I -labelled *Staphylococcus aureus* protein A and exposure to x-ray film.

Polyclonal antibodies have been used successfully for the isolation of genes from many diverse sources [5], and may offer an advantage over monoclonal antibodies, since they are not restricted to one epitope on the antigen. The quality of the antibody is clearly very important, and steps are often taken to purify away components that may bind *E. coli* proteins. This is achieved by exposing the antibody mix to immobilized *E. coli* proteins. Also, as the antibody may be reused (the same antibody mix may be used up to ten times), these contaminating components are again diluted.

There are, however, a number of disadvantages to screening libraries with antibody probes. Probably the most important is the potential danger of ^{125}I -label, which has a half-life (60.4 days) over four times longer than ^{32}P , and the ability to be absorbed and concentrated very efficiently by the thyroid, against huge concentration gradients. Other, non-radioactive, luminescent labelling systems have appeared on the market recently, but these were in their infancy at the time of the development of the strategy for this work. On the practical level, a number of problems can arise. Firstly, the library must be of very good quality, since ideally full length transcripts are needed for the production of proteins which will have a native

conformation. In practice, false positives are abundant and therefore some amino acid sequence of the target protein is required to confirm the isolation of the correct clone. Given that the raising of antibodies requires protein of a similar purity to that required for amino acid sequencing, and that some amino acid sequence is necessary to identify the isolated clones as the desired cDNA, one may as well obtain some amino acid sequence in the first place and use it to make a DNA probe. If this is sufficiently long, the nucleic acid probe can then be used to screen the library at high stringency. The recent development of sequence amplification by PCR has enabled the generation of such long regions of DNA from a gene, even if relatively small segments of peptide sequence are available.

4.1.1.3 Screening with DNA probes. This is the most commonly used and reliable method of screening libraries for clones of interest, and was the method of choice for the cloning of the cDNA for ketohexokinase. Any type of cDNA library may be screened by this method, whether they are in an expression or non-expression host vectors [5]. As a first step, a DNA probe which is complementary to, or is thought to show significant homology with, the required cDNA is generated. Such a probe can be derived from the amino acid sequence of the protein, or if this is not available, a sequence from a homologous gene may be used. Oligonucleotides derived from amino acid sequences are, necessarily, 'degenerate', that is, they represent a pool of molecules comprising a large number of sequences, all of which could encode the peptide.

cDNA libraries are screened by plating at a density at which clones are confluent, and each clone is about 0.5 - 1mm in diameter. A nitrocellulose or positively charged nylon membrane is then laid over the clones, which soak on to the filter. The DNA is then denatured and immobilised, and unoccupied regions on the membrane are 'blocked'. The DNA probe, labelled either chemically or radioactively, is then allowed to bind to the target DNA. By washing the filters under appropriate conditions, the DNA probe that remains specifically bound to the target sequence can be located. The degree of homology of the probe to the template to be used is of prime importance, since, the higher the fidelity, the stronger the hybridization to the cDNA. Weaker low-specificity interactions between probes and cDNA templates can be discriminated from accurately hybridized probe by identifying conditions under which only the latter remains bound. The more 'stringent' the conditions used (*i.e.* lower salt and higher temperature), the higher the degree of complementation required for the probe to remain hybridized to the cDNA. Probing libraries with highly degenerate probes requires laborious incremental increases in stringency in order to eliminate all but the most strongly hybridized clones, without removing the probe completely. Use of a high fidelity probe, on the other hand, allows the use of highly stringent conditions from the outset.

Although screening with DNA probes does not require a library containing full length clones, this is obviously an advantage. If a clone is obtained that is found not to be a full length clone, the library can be rescreened with a second probe obtained from the sequence of the truncated clone, and the new clones consequently isolated may encode the missing region of the cDNA.

The work in this chapter describes the design of a series of degenerate oligonucleotides complementary to the peptide sequences, which were then used as primers in a PCR to generate a long sequence of DNA that would be highly complementary to the cDNA for ketohexokinase. This was then used as a high fidelity probe to screen a cDNA library at high stringency.

4.1.2 *Design of Oligonucleotide Primers.*

There are two problems associated with the design of oligonucleotides complementary to amino acid sequences, the first being the acquisition of a sequence which has the highest probability of matching the gene of interest, and the second is finding conditions under which both oligonucleotides will specifically anneal to their target sequences. There are no rules that guarantee the selection of the correct codon at an ambiguous position, but it has been estimated that a random choice of codon preference would result in at least 76% homology [7]. Of the nine peptides that were sequenced by N-terminal gas-phase Edman degradation, two (KHK 45 and 49) were found to have regions of sufficiently low degeneracy to allow the design and synthesis of small pools of oligonucleotides complementary to their amino acid sequence. The orientation of the peptides KHK 45 and 49 relative to one another within the protein was not known, and this necessitated the design of two pairs of primers, in order to account for both possibilities.

The design of the four primers was formulated using the following principles,

- a) primers should not complement within themselves or to each other,
- b) CpG intercodons are under-represented in the coding regions of mammalian DNA [7, 8],
- c) codon preferences are apparent in genes from rat liver tissue [9]. The following codons were found to be under-represented and therefore excluded: CCG; proline, CUU, UUA and UUG; leucine, GGU; glycine, and ACG; threonine,
- d) the omission of the third base of the 3' codons ensures perfect matches at the transcribing termini of primers 2 and 4.

By applying these principles, the number of primers in the degenerate pool should be vastly decreased from the total sequences possible.

4.1.3 *The Polymerase Chain Reaction (PCR).*

The development of the polymerase chain reaction (PCR) has been one of the biggest breakthroughs in molecular biology in recent years. It allows the amplification of a given sequence of DNA from a few molecules to several microgrammes, and single copy sequences can be routinely extracted from a complex mix of genomic sequences and visualized on an agarose gel. The technique relies on the repeated denaturation, primer annealment, and primer extension of the target DNA sequence, with two primers at either end of the amplified sequence, hybridized to opposite strands. Denaturation is achieved by heating the sample to 90°C, and the mixture is then cooled to allow the primers to anneal to the template. Since the primers are in great excess over the target sequence, these will bind in preference to template-template annealment. After extension of the primers by a DNA polymerase, the cycle is repeated, usually thirty times. As each cycle should result in a doubling of the number of template sequences, thirty cycles of PCR will result in a 2^{30} -fold increase (just over a billion) in the number of the desired sequence.

Although the PCR was first unveiled as a method of DNA amplification at the American Society of Human Genetics Conference in 1985, this early method relied on the repeated addition of DNA polymerase after each denaturation step, and the automation of this process involved its repeated injection. The method underwent several developments, the most notable of which was the application of a thermostable DNA polymerase isolated from *Thermophilus aquaticus* (*Taq*) [10], which overcame this requirement for the continual addition of a thermolabile enzyme after each cycle [11, 12, 13]. The main disadvantage of *Taq* DNA polymerase is its lack of a proof reading activity [14], and its use for the preparation of DNA therefore carries the risk of introducing point mutations at a frequency of approximately one error per kilobase. If such a misincorporation occurred during the first cycle of the PCR, then it would be present in 25% of molecules resulting from the amplification. Although this does not represent a problem for some applications, *e.g.* direct use of the product as a sequencing template or probe, cloning of the PCR product can result in perpetuation of the error. This risk can now be minimised by use of one of the recently isolated enzymes possessing proof-reading activity. Subsequent to the use of PCR in this project to amplify the KHK probe, a number of other thermostable enzymes have been isolated from thermophilic organisms (*e.g.* VentTM [New England Biologicals], *Pfu* [Stratagene] and *Tth* [USB] [15]), and recombinant forms of these enzymes are now marketed, all of which offer greater thermostability and copying fidelity [16].

The PCR is a very sensitive method, capable of amplifying a single template molecule from a large pool of a complex mixture of sequences, but where this is required it is especially important to optimize the conditions used in order to prevent the amplification of non-specific

templates. To these ends, it is very important to use oligonucleotide primers that are unique, or have a low degeneracy, to the sequence of interest, combined with a template preparation that is as free of competing sequences as possible. The selection of the primer sequences is obviously dependent on the nature of the experiment undertaken, and the sequence data available. In the case of the amplification of a region of DNA from peptide sequence data, the use of as small a pool of degenerate sequences as possible will potentially lead to fewer non-specific sequence amplifications. If the primer sequences to be used are highly represented in the DNA used by molecular biologists (such as the M13 forward and reverse sequencing primers, which are present in many cloning vectors), care must be taken to avoid contamination (*e.g.* from aerosols from pipette barrels) of the amplification mixtures with competing DNA sequences. It is possible to perform PCR with an RNA template [15], however, the amplification of a sequence from a crude RNA mix could yield products of unprocessed mRNA sequences, that might contain non-transcribed intron sequences. Enrichment of processed mRNA sequences by poly-A tailed mRNA selection could help remove these products, however, generation of a single stranded cDNA template using oligo-dT primer effectively accomplishes the same thing with greater ease. The partial purification of a complex mix of DNA sequences before attempting amplification can decrease unwanted amplification (*e.g.* purification of plasmid-borne sequences to be amplified).

The PCR can also be optimized by altering reaction conditions [17], and variation of magnesium chloride concentration, annealing temperature, cycle number and primer extension times can affect the efficiency of the reaction. Primer annealment to the template sequence is temperature sensitive, and the lower the annealing temperature used, the more likely that incorrect annealing can occur. Consequently, melting temperatures (T_m) of the two primers should be similar (by having similar G/C : A/T ratios in both primers; the higher the G/C content, the higher the T_m), and an annealing temperature 5°C below this T_m should be used. Therefore, stringent annealing temperatures, especially in the first few cycles, will increase specificity. Magnesium ion concentration should also be optimized, since this can affect primer annealment, strand dissociation of both original template and PCR product, formation of primer-dimer artifacts, and polymerase activity and specificity. Consequently, a range of magnesium ion concentrations should be tried to determine the optimum concentration (usually 0.5 to 2.5mM. over total dNTP concentration). The optimum number of cycles will depend on the number of target sequences available, and the target length. The more abundant the target sequence, the fewer the number of cycles required to achieve a useful amplification, since too many cycles will result in a reduction of the dNTP and primer concentrations, leading to an increase in mis-incorporation and decrease in fidelity. This depletion of dNTP also occurs with increasing length of the sequence to be amplified, so this also has an important bearing on the number of cycles chosen. Furthermore, the target sequence length should be considered when determining the extension time used (estimates of nucleotide incorporation vary from 35 to 100

nucleotides per second, depending on conditions used). Control experiments using each of the two amplification primers alone should also be performed, to ensure artifacts, caused by self priming, do not lead to a misidentification of the amplified region. Since the determination of the optimum conditions for specific template amplification depends on so many parameters, they are best determined experimentally.

4.2 Materials and Methods

4.2.1 Bacterial Strains and Plasmids.

4.2.2.1 *E. coli* strain. The strain of *E. coli* used in this work was DH5 α , (*supE44* Δ *lacU169* (ϕ 80 *lacZ* Δ *M15*) *hsdR17* *recA1* *endA1* *gyrA96* *thi-1* *relA1*) [18]. This was selected as a general cloning strain because, being *lacZ*⁻, DH5 α can be used for blue/white selection with plasmids containing a cloning site within the *lacZ* gene. It is also *recA1* and so will not be able to integrate plasmid DNA into the chromosome.

4.2.1.2 Cloning vector. The plasmid used for cloning the PCR product was *pUC18* [19] which confers ampicillin resistance to transformed *E. coli* strains. *pUC18* has the M13mp18 multiple cloning site located with a *lacZ'* gene, recombinant plasmids become unable to produce β -lactamase, and hence can be blue/white selected on plates containing Xgal and IPTG (see 4.2.4). The replication of this plasmid is 'relaxed' within *E. coli*, and is present at high copy number in transformed cells.

4.2.2 Bacterial Growth Media.

Five media were used to culture bacteria [20].

4.2.2.1 LB, a general purpose growth medium for *E. coli*, is adjusted to pH 7.0 with sodium hydroxide.

1% w/v	Bacto-tryptone
0.5% w/v	Bacto-yeast extract
1% w/v	Sodium chloride

4.2.2.2 Terrific Broth is a nutrient-rich liquid medium which was used to grow DH5 α in preparation for plasmid minipreps (4.2.6.1). The following mixture was made up in 900ml. water and autoclaved. 2.31g potassium dihydrogen phosphate and 12.54g dipotassium hydrogen phosphate in 100ml. water were autoclaved separately and then added.

12g	Bacto-tryptone
24g	Bacto-yeast extract
4ml.	Glycerol

4.2.2.3 SOB was used for plating newly transformed cells. The pH of this was adjusted to 7.0 with sodium hydroxide.

2% w/v	Bacto-tryptone
0.5% w/v	Bacto-yeast extract
0.05% w/v	Sodium chloride
10ml./l.	250mM. Potassium chloride
5ml./l.	2.0M. Magnesium chloride, added after autoclaving

4.2.2.4 SOC was the same as SOB with the addition of 20ml./l. filter-sterilized 1M. glucose after autoclaving. It was used to allow transformed cells to recover prior to plating on SOB plates.

4.2.2.5 TYM broth was used in the preparation of competent *E. coli* cells.

2% w/v	Bacto-tryptone
0.5% w/v	Bacto-yeast extract
0.1M.	Sodium chloride
5ml./l.	2.0M. Magnesium chloride, added after autoclaving

Agar was added to the above media before autoclaving, as necessary, to 1.5% w/v for plates and 0.7% w/v for 'top agar', used when measuring the titre of 'phage. When required, ampicillin was added to the media to a final concentration of 50µg/ml. (agar plates) or 100µg/ml. (liquid medium). This concentration was used for selection of relaxed plasmids such as *pUC*, one third this concentration was used for stringent plasmids. Stock ampicillin (50mg/ml. in water) was filter sterilized (0.2µm) and stored at -20°C until use. It was diluted (×1000 for agar or ×500 for liquid media) after the media were autoclaved and had cooled to below 60°C.

4.2.3 Strain Storage.

Long term storage of *E. coli*, either with or without plasmids, was done by cryostorage at -70°C. Cells were either grown in liquid phase or taken directly from colonies on petri dishes, suspended in 20% v/v glycerol, transferred into 1.8ml. cryotubes [Nunc] and immediately frozen in liquid nitrogen. The tubes were then transferred to a -70°C freezer for long-term storage. Cells could be rejuvenated by taking a small aliquot of the frozen suspension with a sterile tooth pick and streaking directly on to nutrient agar, containing an antibiotic as necessary.

4.2.4 Transformation of *E. coli* with Plasmid DNA.

Bacteria were transformed using calcium chloride-treated competent cells using an adapted method of Sambrook *et al.* [20]. A single fresh colony of the desired strain (DH5α) was

grown for 2 hours in 5ml. TYM broth (4.2.2.5) with vigorous shaking at 37°C. This was transferred to 100ml. fresh, pre-warmed TYM broth and incubated under the same conditions for a further 2 - 3 hours until the A₅₅₀ was approximately 0.5 and the cells were growing with a doubling time of not longer than 20 minutes. The cells were immediately cooled and then sedimented by centrifugation (2500 rpm, 5 minutes, 4°C), drained and resuspended in 30ml. ice-cold filter-sterilized TfbI

30mM. Potassium acetate, pH 5.8
50mM. Manganese chloride
100mM. Potassium chloride
10mM. Calcium chloride
15% v/v Glycerol

and left on ice for 15 minutes. The cells were gently centrifuged (2000 rpm, 5 minutes, 4°C) and the pellet resuspended in 4ml. ice cold filter-sterilized TfbII

10mM. MOPS/sodium hydroxide pH 7.0
75mM. Calcium chloride
10mM. Potassium chloride
15% v/v Glycerol

and 400μl. aliquots were frozen in liquid nitrogen. Cells could be stored for at least six months at -70°C.

The cells were transformed using the following procedure. After thawing on ice, 100μl. aliquots of competent cells were mixed with a sample of plasmid DNA (maximum volume 10μl., containing up to 10ng DNA) in 15ml. 'snap-cap' Falcon tubes and left on ice for 30 minutes. The cells were then heat-shocked at 42°C for 45 seconds and then placed on ice for a further 2 minutes. SOC media (900μl.) was added and the cells incubated at 37°C with vigorous shaking (250 rpm) for 1 hour before plating 100μl. aliquots on SOB plates containing 50μg/ml. ampicillin. For blue/white selection, plates were spread with 40μl. 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal, 20mg/ml. in dimethylformamide) and 4μl. isopropylthio-β-D-galactoside (IPTG, 200mg/ml., filter sterilized) and were allowed to dry in a flow hood for one hour, before inoculation. Plated cells were incubated overnight at 37°C, and colonies were sub-cultured on to fresh LB/ampicillin plates the following day.

4.2.5 Buffered Phenolic Solutions.

Phenolic solutions are used in the separation of DNA from proteins. However, DNA is soluble in phenol unless its pH is greater than 7.5. Consequently, all phenol must be buffered to pH 8.0. Crystalline high-quality phenol, (chromatography grade, Merck Ltd.) was saturated with double distilled water and an equal volume added to 1M. Tris base and mixed. The two phases were separated by centrifugation and the pH determined using pH paper. This was repeated with 1M. Tris.HCl pH 8.0 until the phenolic phase has a pH of 8.0. The Tris phase

was then totally removed and replaced with a small volume of a mixture of 100mM. Tris.HCl pH 8.0, 1mM. β -mercaptoethanol, and then 8-hydroxyquinoline was added to the phenolic phase to a concentration of 0.1% w/v. The latter two chemicals help prevent oxidation of the phenol, which is unstable under alkaline conditions. Phenol is stable for up to two months in this state if stored at 4°C in the dark. This phenol was also used to make phenol/chloroform/isoamyl alcohol (25:24:1, v/v/v) (henceforth referred to as 'phenol/chloroform'). Similarly, to reduce frothing, chloroform was mixed with isoamyl alcohol (50:1, v/v) and will be referred to as 'chloroform' when it is used after phenol/chloroform treatment.

4.2.6 Plasmid DNA Purification.

4.2.6.1 Crude small scale preparations ('minipreps'). A simple miniprep method which involves partial permeabilization of the cells with phenol-chloroform was used for the isolation of plasmid DNA from DH5 α . Although this technique works well with DH5 α , with some strains (*e.g.* HB101) the high carbohydrate content of the cells interferes with later purification steps. Individual colonies were used to inoculate 5ml. Terrific Broth containing 100 μ g/ml. ampicillin in 50ml. Falcon tubes, and were grown overnight with shaking at 37°C. The cells in a 1.5ml. aliquot were then sedimented in a 1.5ml. Eppendorf tube by centrifugation for 1 minute and resuspended in 100 μ l. TNE.

10mM. Tris.HCl pH 8.0
1mM. EDTA
100mM. Sodium chloride

An equal volume of phenol/chloroform was added, the mixture vortexed for 20 seconds, and then centrifuged at 13,000 rpm for 5 minutes. The upper aqueous phase was then removed from the phenol interphase and was mixed with an equal volume of chloroform. After a 1 minute centrifugation the aqueous phase was removed and incubated with DNAase-free RNAase [Boehringer Mannheim] (20 μ g/ml.) for 30 minutes at 37°C. The sample was then washed with an equal volume of phenol/chloroform and then similarly with chloroform. 25 μ l. of 10M. ammonium acetate was added to give a final concentration of 2M. and then 2 volumes (250 μ l.) absolute ethanol were added. The mixture was vortexed and the sample centrifuged for 10 minutes at 13,000 rpm. The DNA pellet was then washed with 70% v/v aqueous ethanol and allowed to dry before being dissolved in 50 μ l. water or TE (10mM. Tris.HCl pH 8.0, 1mM. EDTA). Typical yields for a *pUC* based plasmid is about 10 μ g, and usually a detectable, but insignificant, amount of chromosomal DNA contaminates these preps.

4.2.6.2 Large scale plasmid purification ('maxiprep'). Large scale purification of plasmid DNA was performed by one of two methods, the first using a caesium chloride gradient to purify a crude plasmid preparation, the other using polyethylene glycol precipitation

of the plasmid DNA. Both methods can yield several hundred microgrammes of very pure super-coiled DNA, and use the same initial purification of the plasmid DNA, which is an adaptation of Birnboim and Doly's miniprep procedure [21]. The caesium chloride method yields plasmid DNA of very high quality, which, with care, can be free of nicked or linear DNA [20]. It is, however, more time consuming, utilizing hazardous reagents, and is expensive. Polyethylene glycol precipitation yields slightly lower quality DNA, which is nonetheless suitable for most purposes.

4.2.6.2.i. Preparation of crude plasmid DNA. A 30ml. culture of plasmid-bearing *E. coli* was grown in LB/amp to late exponential growth phase ($OD_{600} \approx 0.6$), and this was used to inoculate 500ml. pre-warmed LB/amp. This was grown with shaking at 37°C until the $OD_{600} \approx 0.4$, when 2.5ml. chloramphenicol (34mg/ml. in ethanol) was added and the culture left overnight with shaking at 37°C. Chloramphenicol stops cell growth, but allows plasmid replication in the cells. Although its inclusion is only essential for stringent plasmids, more consistent yields were obtained by including it, and the cell pellet produced was of a more manageable size. It also leads to a higher plasmid copy number than would otherwise be obtained.

The cells were then harvested by centrifugation (4000 rpm, 15 minutes, 4°C) and then resuspended in 100ml. TNE (**4.2.6.I**), and reharvested as before. These washed cells were then resuspended in 20ml. ice cold Solution I

50mM. Glucose
25mM. Tris.HCl, pH 8.0
10mM. EDTA

after which 40ml. freshly made Solution II

200mM. Sodium hydroxide
1% w/v SDS

were added, and the contents mixed well. This was left at room temperature for 10 minutes before the addition of 30ml. ice cold Solution III.

60ml. 5M. Potassium acetate
11.5ml. Glacial acetic acid
28.5ml. Water

This was mixed by inversion, and after incubation on ice for 10 minutes, this was centrifuged (6000 rpm, 4°C, 15 min, no brake) and the supernatant filtered through miracloth. Isopropanol (0.6 volumes, approx 60ml.) was added, mixed, and left at room temperature for 10 minutes. The DNA was then sedimented by centrifugation (6000 rpm, 15 min, 25°C), the pellet washed with 70% v/v ethanol, and air-dried.

4.2.6.2.ii. Caesium chloride gradient purification. The DNA pellet was taken up in 20ml. TE (10mM. Tris.HCl, pH 8.0, 1mM. EDTA), and 22g of caesium chloride added, and dissolved. 400µl. ethidium bromide (10mg/ml.) was added, mixed and the DNA solution was transferred to a Beckman 27ml. 25×64mm Bell-top Quick-Seal™ centrifuge tube (Part No. 343665). The tube was topped up with paraffin oil, balanced with a second tube, and sealed. These were then ultra-centrifuged in a VTi 50 rotor (45,000 rpm, 20°C, overnight). The tubes were then carefully removed, and the DNA visualized with long-wave (302nm) UV light. Two bands were usually seen, a thick super-coiled plasmid layer, and above this, a fainter band of relaxed plasmid and degraded chromosomal DNA. The RNA and protein were coated on the outer wall of the centrifuge tube. Holding the tube in a clamp, two syringe needles were inserted, one at the top of the tube to act as a vent, and a second just below the super-coiled DNA. Using a syringe, the flat bevel of the needle was positioned just below the DNA, which was then slowly drawn off. The DNA solution was then transferred to a polystyrene tube, and the ethidium bromide removed by five washes with 2ml. of sodium chloride-saturated isopropanol, leaving the DNA solution colourless. The DNA was then precipitated by the addition of four volumes of water and ten volumes of ethanol, followed by centrifugation at 15,000 rpm for 15 minutes. The DNA pellet was washed in 70% v/v ethanol, dried, dissolved in 500µl. TE and the concentration determined by A₂₈₀ measurement (a 50µg/ml. DNA solution gives a A₂₈₀=1).

4.2.6.2.iii. Polyethylene glycol (PEG) precipitation. The crude DNA pellet was dissolved in 10ml. TE (10mM. Tris.HCl, pH 8.0, 1mM. EDTA) and RNaseA added to 10µg/ml., which was then incubated at 37°C for 30 minutes. Solid sodium chloride (0.88g) and 2.5ml. 30% w/v PEG 8000, 1.5M. sodium chloride were added and this was mixed and left on ice for 30 minutes. The DNA was then sedimented by centrifugation (10,000 rpm, 15 minutes, 4°C), and the pellet redissolved in 500µl. TE, before two phenol-, four phenol/chloroform-, and two chloroform-extractions. The DNA was reprecipitated by the addition of 125µl. 10M. ammonium acetate and 625µl. ethanol, and sedimented by centrifugation (13,000rpm, 15 min, 4°C). The pellet was washed in 70% v/v ethanol, air dried, dissolved in 500µl. TE and the DNA concentration determined as before.

4.2.7 DNA Manipulations [20].

4.2.7.1 DNA precipitation. Samples of DNA were precipitated after phenol/chloroform extraction of proteins by the addition of one tenth volume 3M. sodium acetate, pH 5.2, and 2.5 volumes absolute ethanol. If small quantities of DNA were being used, 0.5µl. yeast tRNA (25mg/ml.) [Boehringer Mannheim], as a DNA carrier, was added (this could be removed later, if necessary, by the addition of 1 unit DNAase-free RNAase [Boehringer Mannheim], and incubating for 30 minutes at 37°C). This was vortexed, and the DNA was then sedimented

by centrifugation (13,000 rpm, 10 minutes). This pellet was washed with 70% v/v ethanol and allowed to dry. The dry pellet was then taken up in a suitable volume of water or TE.

4.2.7.2 Restriction enzyme digestion. Restriction enzyme digestion was used for the analysis of DNA (in conjunction with agarose gel electrophoresis, 4.2.7.7) or for preparation of DNA for further manipulations, such as sticky-end ligation (4.2.7.8). Restriction enzyme digests were performed according to manufacturers instructions, using the buffers provided. Digestions were performed either with 10 units of enzyme for 1 hour, or 1 unit overnight, and double digests (*i.e.* using two different enzymes) were either done in the same reaction, if both enzymes use the same buffer, or sequentially, changing the buffers with the enzymes. If necessary, the DNA fragments were then purified by LMP agarose gel electrophoresis (4.2.7.8 and 4.2.7.9).

4.2.7.3 Removal of terminal phosphate by CIP/proteinase K treatment. DNA cut with a restriction enzyme will be able to ligate either to itself, or to a second identical piece of DNA with the same end (either sticky or blunt-ended). To prevent self-ligation, especially of plasmids, the terminal phosphate groups of the DNA can be removed. This prevents its ligation to another dephosphorylated end, but still allows ligation to a phosphorylated end.
10× Calf intestine alkaline phosphatase (CIP) buffer

- 100mM. Tris.HCl, pH 8.3
- 3.0mM. Zinc chloride
- 3.0mM. Magnesium chloride

was added to the cut DNA, and CIP [Boehringer Mannheim] added and incubated under the appropriate conditions (see Table 4.1). The stock CIP (20 units/μl.) was first diluted ten-fold with CIP dilution buffer.

- 30mM. Triethanolamine.HCl pH 7.6
- 3M. Sodium chloride
- 1mM. Magnesium chloride
- 100μM. Zinc chloride

	Protruding 5' termini	Blunt or recessed termini
Amount of CIP needed	1 unit / 100 pmoles.	1 unit / 2 pmoles.
Incubation conditions	30 minutes, 37°C.	15 minutes, 37°C, then add a second aliquot of CIP and incubate for a further 45 minutes at 55°C.

Table 4.1 Conditions and amount of CIP required for dephosphorylation of linear DNA. 1μg linear DNA, 1kb long, contains approximately 3.5 pmoles phosphorylated ends. (Taken from Sambrook *et al.* [20])

The CIP must be removed before proceeding, as residual enzyme would also dephosphorylate other added DNA, thus preventing ligation. Unlike most proteins, CIP is not adequately extracted with phenol, and so the enzyme must first be inactivated with proteinase K, which itself must be removed. Proteinase K [Boehringer Mannheim] was therefore added to a final concentration of 100µg/ml. with 0.5% w/v SDS, 5mM. EDTA, and the mixture incubated for a further 30 minutes at 56°C, before four phenol/chloroform extractions and a chloroform extraction. DNA was then precipitated (see 4.2.7.1).

4.2.7.4 Phosphorylation of linear DNA ends with T4 kinase. To phosphorylate the 5' ends of a DNA fragment, especially those generated by PCR [22], ATP to 1mM. and 10 units of T4 polynucleotide kinase [Gibco BRL] were added directly to the reaction mixture in which the unphosphorylated fragments were generated. This was incubated at 37°C for 15 minutes.

4.2.7.5 Blunt-ending of DNA fragments. Two methods were used to blunt-end fragments of DNA, depending on whether the overhangs to be removed were 3' or 5', or the source of the DNA. Filling-in of overhangs generated by restriction enzyme digests were blunt-ended using the Klenow fragment of *E. coli* DNA polymerase I. This enzyme exhibits a 5'→3' polymerase activity as well as a much smaller 3'→5' exonuclease activity. Since this enzyme is active in the pH range of restriction endonucleases, it is the method of choice for blunt-ending, since no pH change need be performed and thus can be done in the restriction digest mix. The blunt-ending reaction is done in two steps, the first in the presence of the enzyme alone. This will cut back the DNA 3'→5', removing a 3' overhang if present, and generating a short 5' overhang. dNTPs are then added, and the polymerase activity fills in the 5' overhang. By adding the dNTPs first, a 5' overhang generated by a restriction enzyme cleavage may be filled-in with no cutting back. In either case, dNTPs are added to 1mM. each, and 1 unit of Klenow fragment added. Each reaction is incubated at 37°C for 15-30 minutes, before terminating by phenol/chloroform extraction.

Since PCR reaction mixtures contain dNTPs, it is not possible to employ Klenow to blunt-end the 3' overhangs, generated during amplification. The overhanging DNA was therefore removed by cutting back with the single strand-specific DNA exonuclease, mung bean nuclease. Since this enzyme does not have a polymerase activity, the presence of dNTPs will not effect its activity. It is, however, inactive at most pHs used in other reactions (e.g. PCR) and so the pH of the reaction mixture must first be altered. The pH of the DNA solution or reaction mixture was changed to 4.5 by the addition of 1/10 volume 10× mung bean buffer [23]

300mM. Sodium acetate/acetic acid, pH 4.5
500mM. Sodium chloride
10mM. Zinc chloride
50% w/v Glycerol.

Ten units of mung bean nuclease [Gibco BRL] were added, and the mixture was incubated for 15 minutes at 37°C.

4.2.7.6 Ligation of DNA fragments. DNA fragments were ligated into an appropriately cut plasmid with a range of insert to vector concentrations. Ligations of blunt-ended fragments were usually performed at higher ratios (10:1 and 20:1) than those of 'sticky-ended' inserts (2:1 and 5:1). Insert and plasmid DNA were incubated with the appropriate volume of 5× BRL ligation buffer

250mM. Tris.HCl, pH 7.6
50mM. Magnesium chloride
5mM. ATP
5mM. DTT
25% w/v PEG-8000

with 1 unit T4 ligase and water to volume (either to 10µl. or 20µl.). This was incubated overnight at room temperature (21°C) before transformation of *E. coli* as described (see 4.2.4 or 5.2.1.4).

4.2.7.7 Horizontal agarose gel electrophoresis of DNA. Analytical and preparative separations of DNA fragments were achieved using horizontal gel electrophoresis. For analytical separation, DNA samples were run in Ultra-Pure Agarose [Gibco BRL] using 0.5× TBE buffer. 5× TBE buffer was stored at room temperature and replaced when a white precipitate is seen.

54g/l. Tris base
27.5g/l. Boric acid
20ml./l. 500mM. EDTA pH 8.0

For preparative work, DNA samples were run in Ultra-Pure low melting point (LMP) Agarose [Gibco BRL] with 1× TAE buffer at 4°C. 50× TAE buffer was stored at 4°C.

242g/l. Tris base
57.1ml./l. Glacial acetic acid
100ml./l. 500mM. EDTA pH 8.0

For analytical electrophoresis, the agarose was melted by micro-waving in 0.5× TBE buffer to give a final volume of 50ml. (10cm×10cm×0.5cm gel), with the agarose concentration dependent on the size of DNA to be separated, and determined using *Table 4.2*. When all the agarose had melted, the gel was allowed to cool to about 60°C when 2.5µl. 10mg/ml. ethidium bromide (0.5µl./10ml. gel mix) were added and the agarose was poured into a 10×10cm cast with a sample comb in position. The gel was allowed to set for 30-45 minutes before mounting it in the gel apparatus and adding just sufficient fresh 0.5× TBE buffer to submerge the gel. DNA samples were prepared by adding gel loading buffer (6× gel loading buffer is 0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol FF, 30% v/v aqueous glycerol) and

Amount of agarose in gel (% w/v)	Efficient range of separation of linear DNA molecules (kb)
0.3	5 - 60
0.6	1 - 20
0.7	0.8 - 10
0.9	0.5 - 7
1.2	0.4 - 6
1.5	0.2 - 3
2.0	0.1 - 2

Table 4.2 Effective separation ranges of linear DNA fragments in TBE buffered agarose gels at various agarose concentrations. (Taken from Sambrook *et al.* [20])

then introduced into the sample wells. A 1kb DNA ladder [Gibco BRL] was run in parallel tracks to enable the size of sample DNA fragments to be determined. The gel was then run at a maximum of 5V/cm constant voltage for 3 hours or until the DNA sample was sufficiently resolved. The progress of the separation could be continually determined by observation of the ethidium bromide-stained DNA fragments under incident long-wavelength (302nm) UV light. A Polaroid photograph, or video image, of the final gel was taken while trans-illuminating the gel with short-wavelength (254nm) UV light.

Preparative LMP agarose gels were prepared similarly, except they were run in 1× TAE buffer at 4°C and at a maximum of 1.5V/cm to prevent the gel melting during the run. DNA fragments were excised from the gel under long-wavelength UV light, keeping the the UV irradiation time as short as possible to prevent UV-induced nicking in the DNA.

4.2.7.8 Purification of DNA from LMP agarose gels. The excised agarose fragment was trimmed of non-UV-fluorescent agarose with a blade, put into a pre-weighed Eppendorf and the volume of the fragment determined by weight (assuming a density of 1g/ml.). Five volumes of water was added and the gel melted at 65°C for 10 minutes, vortexed briefly and then heated at 65°C for a further minute. One volume of buffer phenol was added and mixed by vortexing for 60 seconds. After centrifugation at top speed for 5 minutes, the aqueous phase was removed leaving the scum at the interface. The extraction was repeated first using phenol/chloroform and then chloroform.

4.2.8 RNA Purification.

Due to the abundance and resistance of RNases, a number of precautions must be taken in the preparation of RNA to prevent its degradation.

4.2.8.1 Preparation of RNAase-free sterile water. The water used in all work was first treated with diethyl pyrocarbonate, which inactivates any RNAase present prior to autoclaving. This is essential when working with RNA, and water prepared in this way is also suitable for work with DNA. Diethyl pyrocarbonate (DEPC) was added to 0.1% v/v to double distilled water, and this was left overnight at 37°C to inactivate the RNAase [24, 25]. The water was then autoclaved at 121°C for 20 minutes to sterilize it and destroy the residual DEPC present.

4.2.8.2 Siliconization of glassware. Glassware to be used for work with RNA must first be siliconized, since RNA adheres to untreated glassware. The glass was first thoroughly cleaned by soaking over-night in concentrated chromic acid. This was prepared by carefully adding nine volumes of Analar sulphuric acid [Merck Ltd.] to one volume potassium dichromate-saturated water. The glassware was then thoroughly rinsed with distilled water, dried in an oven and then siliconized by immersion in 2% v/v dimethyldichlorosilane in 1,1,1-trichloroethane [Merck Ltd.] for thirty minutes. The glassware was then rinsed first with propanone and then with distilled water before baking at 180°C overnight.

The glass plates used in DNA sequencing must also be siliconized, to aid their removal from the run sequencing gel. The plates were prepared slightly differently than above for this, since the requirement for them to be RNAase-free was not required. Only one side of each plate was siliconized (the side in contact with the gel), and this was first cleaned by wiping with a 5% w/v methanolic potassium hydroxide solution. After rinsing the plates with distilled water and drying at 80°C, the cleaned surface was wiped with 2% v/v dimethyldichlorosilane solution. This was rinsed first with propanone and then distilled water before drying. The plates could be used for two gels before they needed to be retreated.

4.2.8.3 Isolation of rat liver crude RNA. Isolation of RNA from rat liver tissue was performed by the acid guanidinium thiocyanate-phenol-chloroform (AGPC) extraction method [26] which provided high yield and purity of undegraded RNA. The rat was killed and exsanguinated as described in 3.2.1. The liver was excised and placed on an ice cold petri dish. The tissue was chopped into small pieces, and a 5g sample was immediately homogenized at room temperature in a Potter-Elvehjem homogenizer containing 50ml. 'Solution D'

4M. Guanidinium thiocyanate
25mM. Sodium citrate, pH 7.0
0.5% w/v Sarcosyl
0.1M. 2-Mercaptoethanol.

Guanidinium thiocyanate is a strong inhibitor of ribonucleases [27], and until the last step of the extraction the RNA is protected from degradation. The following components were then

sequentially added to the homogenate, with brief shaking between each addition; 5ml. 2M. sodium acetate, pH 4.0, 50ml. water saturated phenol and 10ml. chloroform-isoamyl alcohol (49:1 v/v). The mixture was then divided between two siliconized Corex centrifuge tubes, cooled on ice for 15 minutes and centrifuged at 10,000g at 4°C for 20 minutes. The aqueous phase was then transferred to a fresh pair of these tubes and the RNA was precipitated by adding 1 volume of isopropanol and leaving at -20°C for 1 hour. The centrifugation was repeated, the pellet was dissolved in 15ml. of 'Solution D' and the RNA reprecipitated, in a chloroform-treated 15ml. snap-cap polypropylene Falcon tube, in the same manner. After centrifugation at 12,000g at 4°C for 20 minutes, the pellet was washed in 75% v/v ethanol, sedimented and vacuum dried, ready for dissolving in RNAase-free DEPC-treated water.

4.2.9 *cDNA Synthesis, and Second Strand Synthesis by PCR.*

4.2.9.1 *Preparation of oligonucleotide primers.* Oligonucleotides were synthesized on an 'Applied Biosystems' DNA synthesizer with approximately 88% efficiency (130 nmoles of each oligonucleotide). The oligonucleotides were deprotected, by treatment with 3ml. 88% ammonia for 1 hour, followed by overnight incubation at 55°C, and purification (23mers from ≤22mers) to give a yield of between 25-33%. The latter was accomplished using Oligonucleotide Purification Cartridges [Applied Biosystems 400771] which were first treated with 5ml. acetonitrile (HPLC grade) followed by 5ml. 2M. triethylamine acetate. All solutions used were in polypropylene syringes [Aldrich]. The crude, deprotected, tritylated oligonucleotide samples containing ~20 OD units were diluted by adding 0.5ml. water to 1.5ml. sample prior to twice passing through the purification cartridge. The cartridge was then washed with 15ml. 1.5M. ammonia, and then with 10ml. water. The bound oligonucleotides were finally detritylated by washing the cartridge with 5ml. 2% v/v trifluoroacetic acid (TFA). The cartridge was then washed with 10ml. water, and the purified oligonucleotides were eluted with 1ml. 20% v/v acetonitrile, vacuum dried, dissolved in 100μl. water and stored at -20°C.

The concentrations of the oligonucleotides was determined spectrophotometrically by measuring the absorbance of the solutions at 260nm. The molar extinction coefficient of each oligonucleotide was calculated by summing the contribution of every nucleotide component (contributions are; G: 12 010, A: 15 200, T: 8400, C: 7050).

4.2.9.2 *cDNA first strand synthesis.* First strand cDNA synthesis (reverse transcription of mRNA) was performed by adding 10× amplification buffer,

500mM. Potassium chloride
100mM. Tris.HCl, pH 8.3
0.1% w/v Gelatin

with the four dNTPs each to a final concentration of 0.25mM., RNAase inhibitor (1U/ μ l. final concentration), crude RNA (20 μ g), primer {d(T)₁₂₋₁₈, 50ng/ μ l. final concentration} or synthetic oligonucleotides 1 or 3 (50pmoles) and Moloney murine leukemia virus reverse transcriptase (200-400U). The magnesium chloride concentration used was 5mM. The reaction was incubated at 42°C for 45 minutes before inactivating the reverse transcriptase by heating to 95°C for 5 minutes.

4.2.9.3 Second strand cDNA synthesis and amplification by PCR. To 20 μ l. of the first strand cDNA mix (4.2.9.2), oligonucleotide primers ('upstream' and 'downstream', 50pmoles each, see 4.2.9.1) were added, followed by 6 μ l. 10 $\times$ amplification buffer, 53.5 μ l. water and 2.5 Units (0.5 μ l.) of *Taq* polymerase [*AmpliTaq*®, Perker-Elmer], and the reaction mix was overlayed with 100 μ l. paraffin oil. A set of amplification reactions was performed with template from reverse transcription reactions which had been primed with different primers (d(T)₁₂₋₁₈, oligonucleotide 1, 2, 3, 4 or combinations of these). Each of these were amplified at a range of concentrations of magnesium chloride (1.25mM., 3.0mM. and 5.0mM.) to test which combinations would give the best amplification efficiency and no single primer products. Thirty cycles of PCR were carried out on a 'Perkin-Elmer Cetus DNA Thermal Cycler' beginning with a template melting step (94°C, 5 minutes for the first cycle and 1.5 minutes for subsequent cycles), followed by an annealing step (2 minutes at 50°C) and a polymerization step (72°C for 2 minutes and 10 minutes for the last cycle).

4.2.10 Double-Stranded DNA Sequencing.

4.2.10.1 Dideoxy- sequencing reactions. DNA sequencing was performed by the dideoxy- method [28, 29] using the Pharmacia T⁷Sequencing Kit (27-1682-01). For sequencing, the template plasmid was denatured by adding an equal volume of 0.4M. sodium hydroxide, 0.4mM. EDTA, left for 1 minute and then neutralizing with 1/5 volume of 2M. ammonium acetate, pH 4.5 before precipitation of the DNA with ethanol. The template DNA (1.5-2 μ g) was dissolved in 10 μ l. water and 2 μ l. M13 universal primer (optimum primer:template ratio is 8.88:3-5) and 2 μ l. annealing buffer were added. The mixture was then heated to 60°C for 10 minutes and the primer allowed to anneal to the template by allowing the mixture to cool to 37°C. 'Enzyme premix' was prepared by adding the following in order: 1 volume water, 3 volumes Pharmacia labelling mix, 2 volumes T7 DNA polymerase [30] (1.5 units/ μ l.) and 1 volume [α -³⁵S] dATP α S (specific activity >1000Ci/mmol) [Amersham]. A 'labelling reaction' was prepared by adding 6 μ l. 'enzyme premix' to the annealed primer and template. This was then incubated at room temperature for 5 minutes. Samples (4.5 μ l.) of this were added to each of four, pre-warmed (37°C) 'mix short' dNTP and ddNTP sequencing mixes (A, C, T and G) and incubated for 5 minutes at 37°C. The reaction was terminated by

adding 5 μ l. 'stop solution' containing bromophenol blue and xlenecyanol FF electrophoretic marker dyes.

4.2.10.2 Sequencing gel preparation and electrophoresis. The DNA sequence was obtained from a 6% w/v acrylamide, 50% w/v urea gel in 1 $\times$ MRC buffer

50mM. Tris.HCl pH 8.8
2.75g/l. Boric acid
0.95g/l. EDTA.

The gel solution was degassed by filtering with a 0.45 μ m hydrophilic membrane [Millipore Durapore] and then 1ml. 10% w/v ammonium peroxodisulphate (APS) and 15 μ l. *N,N,N',N'*-tetramethylethylene-diamine (TEMED) were added immediately before the gel was poured (400 $\times$ 310 $\times$ 0.4mm) between two siliconized glass plates separated by 0.4mm teflon spacers. The gel was allowed to polymerize for at least two hours, and preferably overnight, before being mounted in the BRL S2 Sequencing Gel Electrophoresis Apparatus.

1 $\times$ MRC buffer was used as the running buffer, the lower reservoir containing 3% w/v potassium acetate. (This forms a concentration gradient through the gel which lowers the electrical resistance, and thus decreases the voltage down the gel. The DNA therefore slows down as the potassium acetate concentration increases, thus compressing the spaces between adjacent bands of DNA at the lower end of the gel [31, 32].). All electrophoresis was performed at 60W constant power. After pre-electrophoresis of the gel for 30 minutes, long run samples were loaded in the wells of 'shark-tooth combs' and the gel electrophoresed until the xlenecyanol FF (the slower of the two dyes in the stop solution) was about 2cm from the bottom of the gel. The power was disconnected, short run samples were then loaded, and the electrophoresis was continued until the bromophenol blue dye-front of this sample had reached the bottom of the gel.

4.2.10.3 Fixing and exposure of sequencing gel. The gel was then removed from the sequencing apparatus and one of the siliconized plates removed from the gel. The gel, still attached to the other plate, was then soaked for 20 minutes in 10% v/v acetic acid, 10% v/v methanol, to remove the urea from the gel, before aspirating away the soaking solution. A sheet of Whatman 3MM was placed carefully on the gel, and appressed by placing the other glass plate on top. The gel became attached to the filter paper, which was then removed and a piece of Saran Wrap was placed over the gel before drying it on a Biorad Model 583 Gel Dryer at 80°C for 2 hours. The Saran Wrap was removed and the dried gel was clamped in an exposure cassette with GRI-AX 'Blue' sensitive X-ray film for 16 hours at room temperature. The film was then developed in an automatic developing machine, revealing the DNA sequence. The sequence was read and then analysed on an Apple Macintosh computer using the 'DNA Strider™' version 1.0 program [33].

4.3 Results and Discussion

4.3.1 Oligonucleotide Primer Design.

Two peptides, KHK 45 and KHK 49, were found to have regions of sufficiently low degeneracy to allow the design and synthesis of small pools of oligonucleotides complementary to their amino acid sequence (Table 4.3). Two pairs of complementary oligonucleotides, 23 base pairs long, were used. These comprise all probable codons for seven amino acids according to the rules in section 4.1.2, along with the first two conserved nucleotides of the eighth codon. The degeneracy of both pairs of 23mers was forty eight. The GC/AT ratio was about one, which ensures that the annealing temperature is almost constant (when the oligonucleotide with the lowest possible GC content was compared with that of the highest, the difference would result in only a 6°C change of annealing temperature).

4.3.2 KHK Probe Amplification.

Rather than employ these degenerate pools of oligonucleotides to screen a cDNA library directly, it was decided to use them as primers for the amplification of the region of cDNA that they flank. This strategy allows the generation of a PCR product with a substantial tract of DNA that is perfectly matched to the cDNA template. This fragment can therefore be used as a probe to screen the cDNA library at very high stringency from the outset.

A 5g sample of rat liver tissue yielded 18mg of mixed RNA, of which 100µg was used as template for an oligo-dT-primed synthesis of single stranded cDNA, and this cDNA mixture was employed as the template for PCR. Primers 2 and 3 yielded a 224 base pair PCR product, 3mM magnesium chloride giving the optimal yield. This DNA fragment was blunt ended by mung bean exonuclease treatment (4.2.7.5) and ligated into the CIP-treated *Sma* I site of *pUC18*. Blue/white selection of DH5α transformants of this ligation mix, and minipreps of

Peptide	KHK45	KHK49
amino acid sequence	Tyr Asp Thr Asn Leu Pro Asp Val	Phe Gln Leu Phe Gly Tyr Gly Glu
Derived mRNA	UAU GAY ACH AAY CUS CCW GAU GU	UUY CAR CUS UUU GGV UAU GGW GA
cDNA 'first strand'	ATA CTR TGD TTR GAS GGW CTA CA (Primer 1)	AAR GTY GAS AAA CCB ATA CCW CT (Primer 2)
cDNA 'second strand'	TAT GAY ACH AAY CTS CCW GAT GT (Primer 3)	TTY CAR CTS TTT GGV TAT GGW GA (Primer 4)

Table 4.3 Amino acid sequence of peptides KHK 45 and KHK 49, the derived nucleotide sequences of the mRNAs, and the first and second strand cDNAs, used for amplification primers. The primer names are indicated after their sequences. The degenerate bases have been assigned the letter codes recommended by the IUPAC-IUB Biochemical Nomenclature Commission.

selected white clones, yielded plasmids with PCR-generated inserts. Restriction enzyme analysis of these revealed the presence of single *Bam*H I and *Pst* I sites, and this information was used to determine orientation of inserts within the plasmid, and the presence of multiple copies of the PCR generated fragment (of the ten transformants analysed, one was found to contain a triple insert). The sequence of both strands was determined using the M13 forward primer and clones of opposite orientation.

4.3.3 *Sequence Analysis of the PCR-generated Fragment.*

The sequence of the PCR-generated fragment is shown in *Fig. 4.1*. The amino acid translation of the sequence showed that both peptides from which the oligonucleotide primers were designed are in the same reading frame, and that those portions of the peptide sequences not represented in the primer design are also present in the amplified region. This amplified region was also found to contain a section complementary to the amino acid sequence of a further peptide, KHK 22. Sequence comparison of the derived peptide sequence with those in the PIR databases using the TFASTA program of the GCG/Staden sequence analysis package showed that this sequence was unique, and consequently represents a region of a new protein. It was believed that these characteristics adequately confirm the identity of the PCR product as being a fragment of the coding region of ketohexokinase cDNA.

Sequences showed several mismatches between the primers and the template. This was expected due to the degenerate primers used, and comparison with the final sequence obtained from the cDNA library (see *Fig. 5.7*) showed that there were several more such mismatches. Comparison of the amplified sequence with the sequence subsequently obtained from the library revealed that there were no mis-matches, which might have been generated by the infidelity of *Taq* polymerase, in this region.

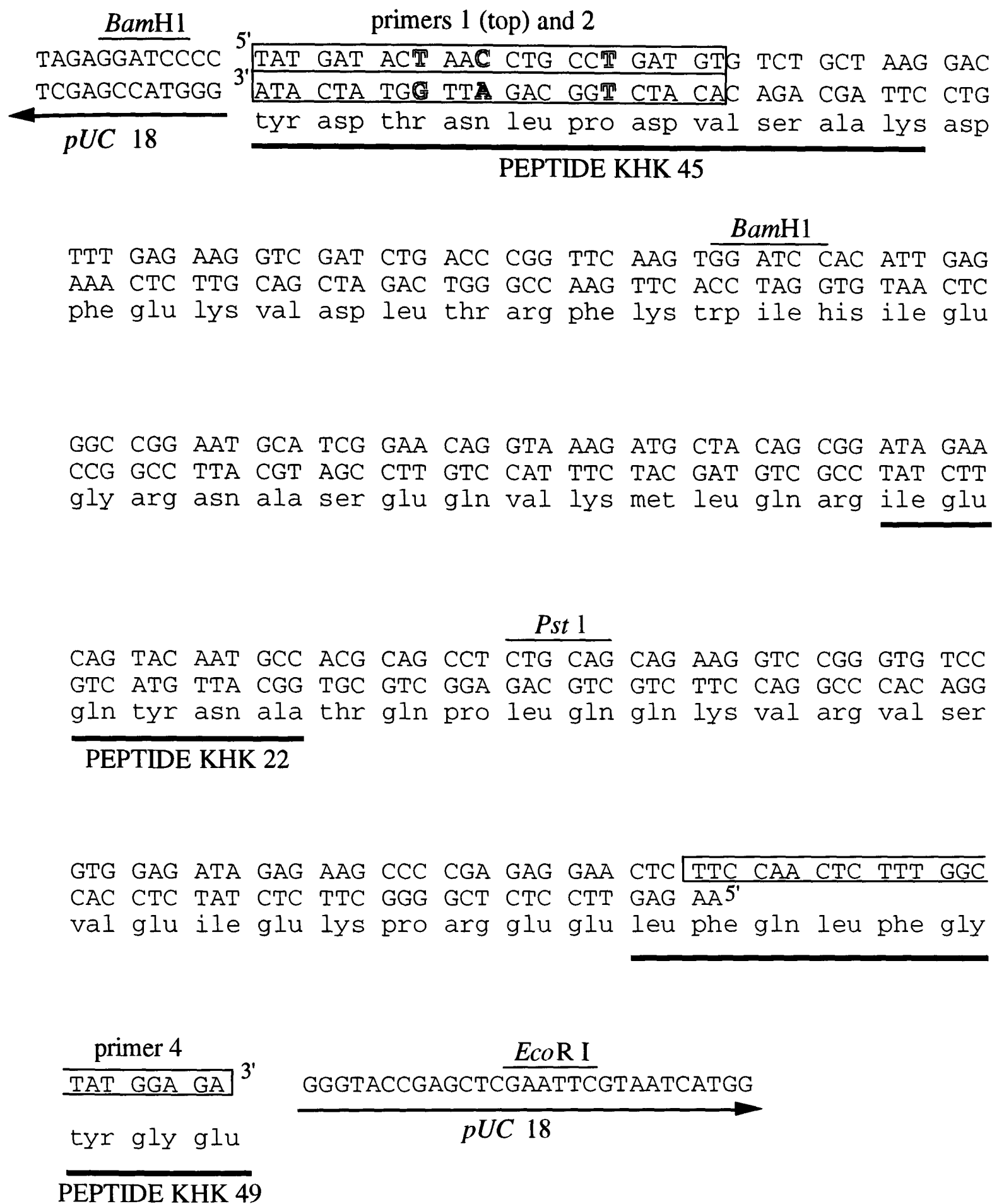


Figure 4.1 Sequence of the amplified region of the KHK cDNA, used to screen a cDNA library. The derived amino acid sequence is shown under the DNA sequence, and the regions that had been identified by peptide sequencing are shown with thick underlining. Sequences derived, or complementary to, the oligonucleotide primers that were revealed from the sequencing gel are boxed, and mis-matched base pairs are shown outlined (primers 1 and 4 were not used in the final amplification, the regions boxed as such are those complementary to their respective pair). Restriction enzyme recognition sequences are indicated with thin lines, and the pUC18 sequences flanking the insert are also shown, as sequenced.

4.4 References

1. Zimmerman, C.R., Orr, W.C., Leclerc, R.F., Barnard, E.C. and Timberlake, W.E. (1980). Molecular cloning and selection of genes regulated in *Aspergillus* development. *Cell*, **21**, 709-715.
2. Williams, .G. (1981). The preparation and screening of a cDNA clone bank, in *Genetic Engineering* (Williamson, R., Ed.), Academic Press, London, vol. 1, 1-59.
3. Carroll, S.B. and Laughon, A. (1985). Production and purification of polyclonal antibodies to the foreign segment of β -galactosidase fusion proteins, in *DNA Cloning; A Practical Approach*, (Glover, D., Ed.), IRL Press, Oxford, vol 1, 89-111.
4. Young, R.A. and Davis, R.W. (1983). Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA*, **80**, 1194-1198.
5. Huynh, T.V., Young, R.A. and Davis, R.W. (1985). Construction and Screening cDNA Libraries in λ gt10 and λ gt11, in *DNA Cloning; A Practical Approach*, (Glover, D., Ed.), IRL Press, Oxford, vol 1, 89-111.
6. Young, R.A. and Davis, R.W. (1983). Yeast RNA Polymerase II Genes: Isolation with Antibody Probes. *Science*, **222**, 778-782.
7. Lathe, R. (1985). Synthetic oligonucleotide probes deduced from amino acid sequence data: Theoretical and practical considerations. *J. Mol. Biol.*, **183**, 1-12.
8. Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nuc. Acids Res.*, **8**, 1499-1504.
9. Aota, S., Golobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1989). Codon usage tabulated from the GenBank Genetic Sequence data. *Nuc. Acids Res.* (suppliment), **16**, r315-r402.
10. Saiki, R.K. and Gelfand, D.H. (1989). Introducing AmpliTaq DNA polymerase. *Amplifications*, **1**, 4-6.
11. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985). Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350-1354.
12. Mullis, K.B., Faloona, F., Scharf, S., Saiki, R.K., Horn, G.T., and Erlich, H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 263-273.
13. Mullis, K.B. and Faloona, F. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Meths. Enzymol.*, **155**, 335-350.
14. Tindall, K.R. and Kunkel, T.A. (1988). Fidelity of DNA Synthesis by the *Thermus aquaticus* DNA Polymerase. *Biochemistry*, **27**, 6008-6013.
15. Myers, T.W. and Gelfand, D.H. (1991). Reverse Transcription and DNA Amplification by a *Thermus thermophilus* DNA Polymerase. *Biochemistry*, **30**, 7661-7666.
16. Mattila, P., Korpela, J., Tenkanen, T. and Pitkänen, K. (1991). Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase-an extremely heat stable enzyme with proofreading activity. *Nuc. Acids Res.*, **19**, 4967-4973.

17. Innis, M.A. and Gelfand, D.H. (1990). Optimization of PCRs, in *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., Eds.), Academic Press, Inc., San Diego, 3-12.
18. Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.*, **166**, 557-580.
19. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **133**, 103-119.
20. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual*, 2nd Edn, Cold Spring Harbor Laboratory Press.
21. Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.* **7**, 1513-1523.
22. Bothron, D.T. (1990). L-Asparaginase II of *Escherichia coli* K-12: cloning, mapping and sequencing of the *ansB* gene. *Gene*, **91**, 101-105.
23. Gubler, U. (1987). Second-strand cDNA synthesis: mRNA fragments as primers. *Meths. Enzymol.*, **152**, 330-335.
24. Fedorcsák, I. and Ehrenberg, L. (1966). Effects of diethyl pyrocarbonate and methyl methanesulfonate on nucleic acids. *Acta Chem. Scand.*, **20**, 107-112.
25. Kumar, A. and Lindberg, U. (1972). Characterization of messenger ribonucleoprotein and messenger RNA from KB cells. *Proc. Natl. Acad. Sci. USA*, **69**, 681-685.
26. Chomczynski P. and Sacchi N. (1987). Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Anal. Bioch.* **162**, 156-159.
27. Cox, R.A. (1968). The use of guanidinium chloride in the isolation of nucleic acids. *Meths. Enzymol.*, **12B**, 120-129.
28. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
29. Sanger, F. (1981). Determination of Nucleotide Sequences in DNA. *Science*, **214**, 1205-1210.
30. Tabor, S. and Richardson, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA*, **84**, 4767-4711.
31. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983). Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.
32. Hong, G.F. (1987). A systematic DNA sequencing strategy. *J. Mol. Biol.*, **158**, 539-549.
33. Marck, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nuc. Acids Res.*, **16**, 1829-1836.

Cloning and Sequencing the Ketohexokinase cDNA

5.1 Introduction

The 224 base pair ketohexokinase sequence cloned in the previous chapter offers a high fidelity probe with which to screen a cDNA library. This sequence represents a region encoding a peptide of 74 amino acids and a molecular mass of 8.8 kDa. Since the size of the purified protein was determined to be 32.8 kDa by mass spectrometry, the probe represents about a quarter the length of the total ketohexokinase coding sequence. It was therefore seemed reasonable that this would hybridize to any clones in the cDNA library that contained substantial portions of the complete ketohexokinase cDNA.

5.1.1 cDNA Library Availability and Characteristics.

The first requirement of a cDNA library was that it had been prepared from mRNA from the same tissue of the same strain of rat from which ketohexokinase had been purified. This would ensure the highest possible complementarity between the probe and the target sequence. The vector in which the library was constructed was of secondary importance, since screening a library by hybridization can be achieved on any library. Two libraries were made available for this work, the first in λ gt11 [1] and later one in λ ZAP [2]. When screened, the first of these did not yield any positive clones. The cDNA was therefore isolated from the λ ZAP library, which also provided the added convenience of direct rescue of isolated clones into a phagemid without the requirement to make large scale preparations of bacteriophage.

The λ ZAP library is made of a *pBluescript SK(-)* plasmid linked between lambda phage arms (see *Fig. 5.1*). With the aid of an M13 helper phage, the *pBluescript* plasmid can be excised from the lambda arms [3] and packaged in the helper phage coat proteins. The plasmid, within the helper phage, can then easily transform bacteria and replicate. The M13 helper phage supplies trans-acting proteins that recognize two separate signals positioned within the Lambda ZAP arms. The two signals are derived from two overlapping halves of an f1 (M13) origin of replication. They are termed initiator, which denotes the signal for initiation of DNA synthesis, and terminator, which denotes the signal for termination of DNA synthesis [4]. Both these signals are recognized by the M13 helper phage gene II protein. Following synthesis of a new

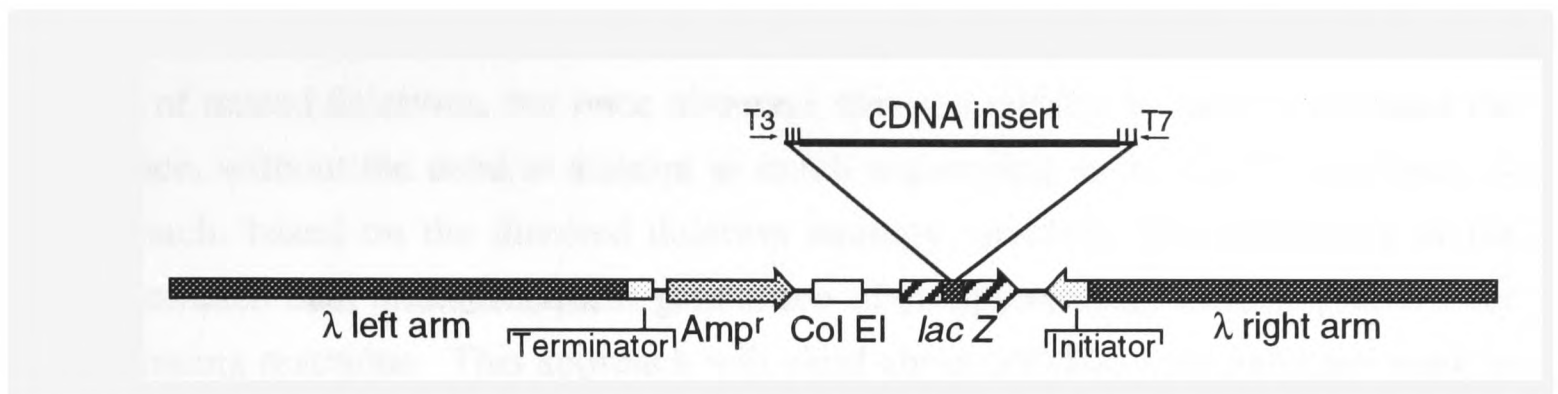


Figure 5.1 The λ ZAP II vector with cDNA insert (not to scale). A linearised *pBluescript SK (-)* plasmid is inserted between the two lambda phage arms, and the cDNA library is cloned into the multiple cloning site of the plasmid (at the *EcoR* I site in λ ZAP II libraries, and directionally at the *Xho* I and *EcoR* I sites in λ UniZAP II). The initiator and terminator sequences flanking the plasmid sequence are recognised by the M13 helper phage gene II protein facilitating excision of the plasmid from the lambda arms, during which they become the f1 origin in the excised plasmid. T3 and T7 are RNA polymerase promoters [2].

DNA strand, the displaced strand is circularized to form the *pBluescript SK (-)* plasmid containing the cloned insert, and the second strand is then synthesized by the host proteins [5]. Since *pBluescript SK (-)* contains the signals for packaging as a M13 filamentous phage, the circularised plasmid is packaged and secreted from the cell [4]. *pBluescript* plasmids are recovered by infecting an F' bacterial strain and selecting transformants on ampicillin plates. The rescued clones can then be used to analyse the insert sequence.

5.1.2 Sequencing Long Regions of DNA.

The dideoxy- method used for sequencing DNA in this project is only suitable for the direct sequencing of regions of DNA up to about 400 base pairs for a single set of sequencing reactions. For DNA sequences longer than this, a strategy must be undertaken to allow the sequencing of a number of overlapping regions, which, when viewed together, will yield the entire sequence of interest. This was not necessary for the PCR-amplified region of the KHK cDNA, since the fragment was only 224 base pairs. Whatever the length of the DNA to be sequenced, it is also vital to obtain the sequence of both strands, since sequencing ambiguities can arise from imperceptible band-compressions on the sequencing gel.

Two general strategies can be used to obtain the entire sequence of a DNA fragment of greater than one kilobase, a random (or shotgun) approach, or a directed approach. The former relies on the generation of a number of subclones of the target sequence, each of which is cloned and sequenced [6]. These data can then be analysed (*e.g.* by the Staden package [7] on the VAX), and the entire sequence obtained. This method is heavily dependent on access to the appropriate computer facilities, and requires the sequencing of five to seven times more nucleotides than the sequence is long. A contiguous sequence usually only becomes available when 90% of the sequence of both strands is obtained, and problems can arise when the target contains regions of repeated sequences [8].

The directed deletion approach necessitates a considerable investment of time to generate a complete set of nested deletions, but once obtained, this can rapidly be used to obtain the entire sequence, without the need to attempt as much sequencing as the random method. A slower approach, based on the directed deletion strategy, involves the sequencing of the fragment from each end, and subsequent generation of complementary custom primers, for further sequencing reactions. This approach will yield about 600-800 base pairs per week in each direction of sequencing, and requires easy access to an oligonucleotide synthesizer. It was decided to use the directed deletion approach to sequence the KHK cDNA.

5.1.2.1 Generation of a nested set of deletion mutants. Several methods have been developed that generate a set of deletion mutants that lack progressively more nucleotides from one end or other of the target sequence. The method used in this study uses exonuclease III (Exo III), which catalyses the stepwise removal of 5' mononucleotides from the recessed or blunt 3'-hydroxyl termini of double stranded DNA [9]. Most protruding 3' termini are completely resistant to the activity of the enzyme [10] (exceptions being those generated by *Apa* I, *Pst* I and *Sac* II). The rate of removal of nucleotides is uniform and predictable, so that by simply sampling and quenching a reaction at incremental times, a range of deletions of one strand can be generated [11, 12]. The exposed single-stranded DNA can then be removed by digestion with either S1 or mung-bean nuclease, and the plasmid DNA recircularized. In order to allow digestion in one direction, deleting the target DNA and not the plasmid, a double digest is performed such that both enzymes cut at one side of the target, and that only the cut end that is adjacent to the insert is susceptible to Exo III digestion (*i.e.* generates a blunt- or recessed 3'-end). If suitably positioned sites are unavailable in the construct, the DNA can be cut first with one enzyme and filled in with α -thiophosphate dNTPs [13, 14], which are also resistant to Exo III digestion, and then cut again, closer to the target, to generate a susceptible end. The pool of recircularized DNA is then analysed to find a series of clones of appropriate sizes for sequencing. If this is performed at both ends of the target DNA, the sequence can be determined on both strands.

If a full restriction map is available for the target DNA, a set of deletions can be generated by a series of restriction digests, blunt-ending if necessary, and religation. In the sequencing of the KHK cDNA, primary sequence information was obtained on rescued clones by restriction mapping and sequencing of the flanking regions of the inserts. One of the clones was then subjected to Exo III/mung-bean digestion, and from the pooled information from these experiments, a nested set of deletion clones was generated by restriction enzyme digestion/religation of another rescued clone. This allowed the entire cDNA to be sequenced, on both strands, in thirteen sets of sequencing reactions.

5.1.2.2 Resolution of sequencing gel band-compressions. Problems can arise in the determination of the sequence from a polyacrylamide gel due to the formation of short stretches of dyad symmetry. These result in DNA secondary structure within the gel, leading to a compression of the bands and their unreadability [8]. These ambiguities can sometimes be resolved from the sequence from the opposite strand, but the possibility of compressions within the sequence on this strand cannot be ruled out, even if they do not appear to be present. Several solutions to this problem are available, depending on the equipment and sequencing protocol used. One is to run the gels at a higher temperature that causes these secondary structures to denature, although to be successful, this relies on the use of special equipment, available from, for example, Stratagene. A second approach is the use of denaturing formaldehyde gels that prevent the formation of secondary structure within the DNA, allowing the ambiguities to be resolved. However, formaldehyde is toxic, and the gels must be run in fume cupboards. Alternatively, the structures formed by internal Hoogsteen G-G [15] or Watson-Crick G-C pairing can be eliminated by substituting the dGTP in the reaction mixture with 7-deaza dGTP, 7-deaza dITP, or dITP [16]. Although the Klenow fragment of *E. coli* DNA polymerase I will accept these analogues, Sequenase™ (T7 DNA Polymerase) incorporates these analogues more efficiently (as well as being more tolerant of thionucleotides used to label the DNA), and so was the enzyme of choice for our sequencing reactions.

5.1.3 Protein Sequence Analysis and Comparison.

The most successful approach to prediction of the structure and function of a protein is to identify similarities between the sequence of the molecule and other well characterized proteins. To these ends, the limiting factor in performing these comparisons will be access to a computer capable of undertaking such comparisons, and the availability of databases and suitable software.

The simplest analysis available for attempting to characterise a protein from its sequence is the scanning of the target sequence with motifs that have been determined from other protein sequences. Location of such motifs (such as N-glycosylation sites) in a target sequence, however, does not confirm that these are physiologically relevant in the native protein, and often further biochemical analysis will have to be performed to confirm the presence of such features. Similarly, the presence of multiple cysteine residues in the primary structure of a protein is a good indication of the presence of disulphide bonds involved in the maintenance of tertiary structure, but again these must be confirmed by biochemical or crystallographic analysis.

Automatic sequence comparisons are used to search for homology between a newly determined sequence, or one of interest, and those present in various protein sequence data banks. The

alignment of two or more protein sequences can provide a wealth of information to guide further investigations, especially if one of the aligned proteins has been biochemically or crystallographically well defined [17]. Although rigorous methods for the alignment of proteins are available, the calculation of an optimal alignment of four or more sequences is currently beyond the capabilities of all but the most powerful computers [18, 19, 20, 21].

Broadly, there are three categories of methods for the comparison of protein sequences [21]. Segment methods compare all overlapping segments of a pre-determined length from one protein with all segments from the others. The distribution of scores obtained for all segment pairs can be used directly to infer homology. Segment methods have the advantage of simplicity, but they cannot allow insertions or deletions (indels) within a sequence. The optimal global alignment methods allow the best overall score for the comparison of two sequences, which allow indels. The Needleman and Wunsch algorithm [22] was the first description of global alignments of sequences, it has undergone many developments and gives the best overall comparison of sequences, including a consideration of gaps [*e.g.* 23]. Indeed, one or more alignments can be produced consistent with a best score. All the methods rely on a scoring scheme for the matching of each of the 210 possible pairs of amino acids, which in the more sophisticated programs, incorporate knowledge about the physical properties of the amino acids. Needleman and Wunsch suggested that their algorithm could be extended to the simultaneous comparison of many sequences, and this has been achieved in the MULTALIGN program developed by Geoff Barton and Michael Sternberg [18-20]. The successful application of this program to the alignment of sequences to the ketohexokinase sequence was performed by its author, to whom I express my gratitude.

5.2 *Materials and Methods*

5.2.1 *Bacteria, Phage and Plasmids.*

5.2.1.1 *E. coli strains.* Two additional *E. coli* strains were used in addition to the DH5 α described in the previous section, XL1 Blue and BB4.

XL1 Blue (*supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac* F'[*proAB⁺ lacI^q lacZ Δ M15 Tn10(*tet^r)*]]) is a recombinant-deficient strain that will support the growth of vectors carrying some amber mutations, but not those with the *Sam100* mutation (*e.g.* λ ZAP) [2]. Transfected DNA is modified, but not restricted. XL1 Blue was used to propagate λ ZAPII recombinants, which are unstable in BB4. The F' in this strain allows blue/white screening on X-gal and permits phage M13 superinfection. Stratagene, who markets the strain, state that it carries an*

‘uncharacterized’ mutation which makes plaques and colonies more intensely blue when plated on X-gal [24].

BB4 (*supF58 supE44 hsdR514 gal K2 galT22 trpR55 metB1 tonA Δ(lac)U169 F'[proAB⁺ lacI^q lacZΔM15 Tn10(tet^r)]*) is a *recA*⁺ strain used for the growth of λZAP and other λ bacteriophages [2]. The F' in this strain carries *lacZΔM15*, which permits α-complementation with the β galactosidase encoded in λZAP. The F' allows superinfection with M13 helper phage, a step required for converting a recombinant λZAP to a *pBluescript* phagemid [25].

5.2.1.2 Phage library and helper phage. The rat liver cDNA library used for the isolation of the KHK cDNA clone was kindly provided by Professor Brian Burchell of The University of Dundee. The Sprague Dawley rat cDNA library was in λZAPII with a titre of approximately 10⁶ pfu/μl. The λZAPII plasmid vectors are a derivative of the lambda phage containing the *pBluescript SK*- plasmid into which the cDNA (up to 10kb.) is cloned. In the case of this library, the cDNA was cloned into the *EcoR* I site of the plasmid. The use of a helper phage enables the plasmid to be excised from the phagemid and transformed directly into an *E. coli* host strain. The f1 helper phage used in this work was R408, and this along with the *E. coli* strains XL1 Blue and BB4 were gifts from Dr. Tamar Enoch.

5.2.1.3 *E. coli* plasmids. The excised *pBluescript-II SK*- plasmid [2] was used for much of the work in cloning the gene, although because of a more convenient distribution of restriction sites within its multiple cloning site, *pUC18* was also used in the generation of deleted constructs for sequencing. The *pBluescript-II SK*- plasmid is similar to *pUC18* (see 4.2.1.2), containing both a similar multiple cloning site in the *lacZ'* gene and a gene conferring ampicillin resistance, but unlike *pUC18*, it also contains the f1(-) phage origin allowing it to be made into a single stranded form and packaged into the lambda coat proteins.

5.2.1.4 High efficiency electro-transformation of *E. coli*. For this work, *E. coli* strains were transformed by electro-transformation which yields up to 10¹² transformants per μg DNA [26, 27, 28]. *E. coli* cells were prepared by growing to mid-log phase and washing in water to decrease the ionic strength of their environment. 500ml. LB broth was inoculated with 5ml. fresh over night culture and incubated at 37°C with vigorous shaking until the A₆₀₀ lay between 0.5 and 1. The cells were then chilled on ice for 15 minutes before sedimenting by centrifugation (4000g, 15 minutes, 4°C). The cells were resuspended in 500ml. ice-cold water, sedimented, resuspended in 250ml. of the same and again sedimented. The cells were then resuspended in 10ml. 10% v/v glycerol, sedimented and resuspended to a total volume of 1.5-2.0ml. 10% v/v glycerol before freezing 200μl. aliquots in liquid nitrogen. Cells prepared in this way remained transformation competent for at least 6 months when stored at -70°C.

For transformation, the cells were thawed on ice and 40 μ l. transferred to a cold Eppendorf tube. 1-2 μ l. DNA and 40 μ l. water were added and the mixture was allowed to stand on ice for 60 seconds. This was then transferred to a cold 0.2cm Gene Pulser Cuvette [Biorad], tapped to the bottom of the cuvette and electroporated in a Gene Pulser Apparatus [Biorad] with a 25 μ F, 2.48kV pulse (giving a pulse time of 4.5-5 msec) with the pulse controller set at 200 Ω . The cuvette was removed from the Gene Pulser and 1ml. SOC was added immediately, and this was incubated at 37°C with shaking for one hour in a snap cap Falcon tube. Ten and one hundred fold dilutions were prepared and plating 100 μ l. aliquots of the undiluted and diluted samples used to inoculate LB plates containing 50 μ g/ml. ampicillin. Colonies containing the plasmid were visible within 24 hours.

5.2.2 DNA Manipulations [8].

5.2.2.1 Southern blotting of agarose gels. Gels to be Southern blotted were first trimmed of excess agarose and the bottom left hand corner was removed to allow for orientation. The DNA in the gel was then denatured by soaking in several volumes of Denaturation solution

1.5M. Sodium chloride
0.5M. Sodium hydroxide

for 45 minutes with gentle agitation. The gel was then rinsed in deionized water and neutralized in

1M. Tris.HCl pH 7.0
1.5M. Sodium chloride

with two washes of 30 and 15 minutes. The gel was then placed on a 3MM filter paper wick and then on top of this was placed, in order, a piece of Hybond N+ nylon membrane [Amersham] cut to the size of the gel, Saran Wrap [GRI] around the edges of the nylon membrane, a second piece of 3MM paper, a 5cm wad of paper towels and a 500g weight. The wick was then placed in a 100mM. sodium phosphate pH 7.2, 1.4M. sodium chloride reservoir, the gel being supported above this reservoir, and left for 24 hours. The nylon membrane was then removed, washed in Denaturation solution for 30 seconds and then neutralized in fresh reservoir solution for 2 minutes. The filter was then dried in a flow hood.

5.2.2.2 Synthesis of radioactive probes. DNA probes were made by the method of Feinberg and Vogelstein [29]. The fragment of DNA was removed from the plasmid by restriction enzyme digestion (4.2.7.2) and an aliquot of this was analysed on a 1% w/v agarose gel with pUC18 standards to determine the amount of DNA and that the digestion had gone to completion. The remaining digest was loaded onto a LMP agarose gel (4.2.7.7) and the insert fragment was isolated and purified (4.2.7.8). 62ng DNA was used for each

labelling reaction. A 25 μ l. reaction was prepared in an Eppendorf tube and put on ice as follows:

11.5 μ l. Solution LS
7 μ l. Solution OL
1 μ l. Nucleic acid enzyme grade BSA (10mg/ml.) [BRL Gibco]
3 μ l. Water

where solution LS is

500mM. Hepes pH 6.6
50mM. Each of dTTP, dGTP and dATP
125mM. Tris.HCl pH 8.0
12.5mM. Magnesium chloride
25mM. Mercaptoethanol

and Solution OL is

1mM. Tris.HCl pH 6.6
1mM. EDTA
90 A₂₆₀ units pdN6/ml. [Pharmacia LKB Biotechnology].

About 62ng of the solution containing the recovered DNA fragment in 1-5 μ l. was heated to 95-100°C for 2 minutes in a sealed capillary tube, immediately cooled to 0°C in an ice bath and added to the reaction mixture. 5 μ l. [³²P]dCTP (3000Ci/mmol) [Amersham] and 2.5 units (0.5 μ l.) Klenow fragment [BRL Gibco] were then added, mixed and the reaction left at room temperature overnight.

The DNA probe was purified from the reaction mixture and the unreacted label using Nensorb®20 columns [NEN Research Products] [30]. The columns were initially pre-washed with 2ml. 100% methanol, eluting at 1 drop per 3-4 seconds with positive pressure before priming the column with 2ml. filter sterilized Reagent A

100mM. Tris.HCl pH 7.7
1mM. EDTA
10mM. Triethylamine.

After this had eluted, the reaction mixture was added to the column in 200ml. Reagent A, and the column was washed with 3ml. Reagent A and 3ml. water. During all these steps, the column was not allowed to run dry. The DNA probe was then eluted by applying 1ml. Reagent B (50% v/v aqueous methanol) collecting 10 drop fractions (c. 300 μ l.). 1 μ l. samples were diluted in 10ml. water (in duplicate) and the amount of labelled DNA probe made determined by Cerekov counting.

5.2.2.3 Hybridization of radioactive probe to nylon filters. The dried filters were layered between monofilament nylon mesh spacers to give a pile of up to five filters thick.

These were then rolled up and put into hybridization tubes [Hybaid] to which 30ml. pre-warmed (65°C) prehybridization solution

100mM. Sodium phosphate pH 7.2
1% w/v SDS
1M. Sodium chloride
200µg/ml. Salmon sperm DNA
5× Denhart's solution

were added. Salmon sperm DNA was made up as a 4mg/ml. stock, dissolving by microwaving, and shearing the DNA by sonication for 2 minutes. The DNA solution was then boiled for 10 minutes. Denhart's solution was made as a fifty times stock, containing

1% w/v Ficoll Type 400 [Pharmacia LKB Biotechnology]
1% w/v polyvinylpyrrolodone
1% w/v BSA fraction V.

The filters were then prehybridized in a Hybaid Hybridization Oven at 65°C for 1 hour. During this time, the ³²P-labelled DNA probe was added to 4ml. prehybridization solution and heated to 100°C for 5 minutes and then cooled to 65°C. After prehybridization, the probe DNA was added to each tube to give a final concentration of 5×10⁵ cpm/ml. and this was allowed to hybridize for 2.5 hours at 65°C. The filters were then washed at high stringency five times with 50ml. 10mM. sodium phosphate pH 7.2, 1% w/v SDS after which the activity of the wash solution was less than 1000 cpm/min/ml. Each wash was at least 2 hours long, and one of the middle washes was overnight. The filters were then removed from the hybridization tubes, allowed to dry, and the hybridized probe located by autoradiography.

5.2.3 cDNA Library Screening and Phagemid Rescue.

5.2.3.1 Determination of library titre and phage lysate preparation. To screen the rat liver cDNA library successfully, the titre of the bacteriophage has to be sufficient for the whole library to be represented. The titre required depends on the size of the genome the library represents, and for rat liver should be 10⁷pfu/ml. To determine the titre, LB (containing 10mM. magnesium chloride) medium was inoculated with *E. coli* strain BB4. After reaching log growth phase (A₆₀₀ = 0.5), the cells were sedimented and resuspended in half the original volume of SM medium

50mM. Tris.HCl, pH 7.5
100mM. Sodium chloride
10mM. Magnesium sulphate
5ml. /l. 2% w/v gelatin

and 1.7µl. phage library in SM (10⁻² to 10⁻⁷, in ten-fold dilutions) were added to 100µl. of the cell suspension. The mixture was incubated at 37°C for 30 minutes without shaking and 3ml. of molten (42°C) soft top agar (LB as above, 0.7% w/v bacto-agar) added. The mixture was

quickly mixed by inversion and evenly distributed by a swirling action over the plate. Plaques were counted after over-night incubation at 37°C.

A preparation of phage was prepared as a plate lysate from a lawn of BB4 infected with phage with a titre of 10^5 pfu/ml. (confluent plaques) by washing with 5ml. SM, incubated for several hours at 4°C. This was recovered, and the plate washed with a further 1ml. SM, which was left for 15 minutes and pooled with the first 5ml. After adding 100 μ l. chloroform to the pooled SM, the cell debris was centrifuged at 4000g for 10 minutes at 4°C. To the supernatant, RNAase (1 μ g/ μ l.) and DNAase (1 μ g/ μ l.) were added and incubated at 37°C for 15 minutes. 1 volume 20% w/v PEG 8000, 2M. sodium chloride in λ diluent

100mM. Tris.HCl, pH 7.7
10mM. Magnesium sulphate

was added and incubated for 1 hour on ice. The precipitated phage particles were recovered by centrifugation at 10,000g, 4°C, for 10 minutes, resuspended in λ diluent and stored, with a drop of chloroform, at 4°C.

5.2.3.2 cDNA library-plaque colony lifts. From the results of the titre assay of the phage (c. 10^9 pfu/ml.), the phage were diluted to give confluent plates (by a 7×10^{-4} fold dilution of the stock, using 5 μ l. phage with 300 μ l. *E. coli*), using 14cm polystyrene petri dishes with BB4 cells as hosts. In order to ensure that a clone containing the required sequence was obtained, twenty such plates were prepared. After overnight growth at 37°C, the plates were cooled to 4°C prior to plaque lifting. Lifts were done using Hybond N+ nylon filters [Amersham] in duplicate for each plate. The circular filters were positioned on the plates, allowed to wet completely, and left for 1 minute. During this time, three holes were made asymmetrically through the filters with a fine syringe needle containing indelible ink to aid alignment of the filters, plates and autoradiograms. The membranes were then lifted and immediately transferred to denaturation solution

0.5M. Sodium hydroxide
1.0M. Sodium chloride

for 1 minute and then placed for 2 minutes in neutralization solution

100mM. Sodium phosphate pH 7.2
1.4M. Sodium chloride

the pH of which was maintained at 7.2 by the addition of 100mM. phosphoric acid, 1.4M. sodium chloride using an automatic pH-stat [Radiometer ABU80 Autoburette, TTT80 Titrator, PHM 84 Research pH Meter]). The filter was then rinsed in fresh neutralization solution for 2 minutes, blotted between filter papers and then allowed to dry on the bench.

5.2.3.3 Localization of plaques hybridizing probe from the filters. The dried filters were placed on a sheet of Saran Wrap [GRI] covered 3MM paper [Whatman] and held down with sticky labels marked with a small spot of fluorescent paint. The filters were then covered with a second sheet of Saran Wrap and after exposing to white light, to allow the paint to glow, was placed in an X-ray cassette, with an intensifying screen, and a sheet of Blue-sensitive X-ray film [GRI]. The autoradiogram was left for 24 hours at -70°C before developing. The filters, still attached to the Saran Wrap/3MM, were located on the autoradiogram by the position of the fluorescent paint, and were aligned with the ink-stained holes in the original plates. Clones giving positive spots on each duplicate filter were then removed by punching out a plug of agar corresponding to the spot on the autoradiogram, using an inverted 1ml. Gilson tip. This plug was then placed in 1ml. SM, containing 1 drop of chloroform, and left 24 hours at 4°C to allow the phage to leach out.

5.2.3.4 Isolation of the single clones. The titre of each phage agar plug supernatant was determined and the phage were then replated for a secondary screening at 500pfu/14cm plate. The phage were then blotted and probed as above. Well-isolated positive plaques were excised from each plate with 1ml. disposable pipettes and subjected to tertiary screening to ensure single clones within each plug.

5.2.3.5 Determination of the sizes of isolated clonal inserts. The presence of M13 forward and reverse primers flanking the cloning site of λ ZAP II allowed the sizing of the clonal inserts by PCR. The method used was that of Friedman *et al* [31]. A 10 μ l. sample of each phage clone was added to 64 μ l. water and heated to 70°C for 5 minutes to denature the phage. 100 μ l. paraffin oil was added to each sample and this was heated to 94°C for 3.5 minutes before the addition of 26 μ l. PCR master mix

10 μ l.	10 $\times$ Amplification buffer (see 4.2.9.2)
9 μ l.	40mM. magnesium chloride (to give a final concentration of 3mM.)
4 μ l.	dNTP mix (5mM. each)
30pmol	M13 forward and reverse primer
2.5 units	<i>Ampli</i> Taq [®] [Perkin-Elmer]
To volume	Water.

30 cycles of PCR were carried out on a 'Perkin-Elmer Cetus DNA Thermal Cycler' (each cycle was: 1.5 minutes at 94°C to denature, 2 minutes at 55°C for primer annealing and then 3 minutes at 74°C for primer extension). For the thirtieth cycle, primer extension was increased to 10 minutes. The samples were then cooled to 4°C. 10 μ l. samples from the PCR were loaded onto a 1% w/v agarose gel and run for 3 hours at 50 volts. A Polaroid photograph was taken of the gel illuminated by short wavelength UV light. The gel was Southern blotted and the membrane probed with the ³²P-labelled DNA probe. An autoradiograph of this, with comparison to the Polaroid photograph, revealed the size of the clonal inserts.

5.2.3.6 Excision of *pBluescript* plasmids from the phage clones. XL1 Blue *E. coli* cells were grown to an A_{600} of approximately 0.5 in 50ml. LB containing 10mM. magnesium chloride and 10mM. maltose and centrifuged at 10,000g for 10 minutes. The cells were then resuspended in 25ml. SM. 200 μ l. of these cells were then mixed with 200 μ l. λ ZAP phagemid clone stock (containing approximately 1×10^5 particles) and 20 μ l. R408 helper phage (1×10^{11} pfu/ml.) in a 50ml. polypropylene tube and incubated for 15 minutes at 37°C. To this, 5ml. 2 $\times$ YT

1.6% w/v Bacto-tryptone
1% w/v Yeast extract
1% w/v Sodium chloride

was added and the mixture was incubated at 37°C for 2-6 hours with shaking. The mixture was then heated to 70°C to inactivate the parent (λ ZAP) phage and to kill the bacteria. This was centrifuged at 10,000g for 5 minutes, and the supernatant, containing the *pBluescript* plasmid packaged in the f1 phage particles, as well as the R408 helper phage, was transferred to a fresh Eppendorf. To recover the excised plasmid, 200 μ l. XL1 Blue host cells, prepared as above, were mixed with 200 μ l. of this mixture and incubated at 37°C for 5 minutes. Aliquots of this were then diluted 10-, 100- and 1000-fold and plated on LB/ampicillin plates. Transformants were selected and replated.

5.2.4 DNA Sequencing.

5.2.4.1 Resolution of band compressions on gels. DNA sequencing was done as in section 4.2.10 except that 5% w/v acrylamide gels were run, and several other sequencing mixes were used. The presence of band compressions were resolved by replacing the 'T⁷Sequencing mixes' with Pharmacia Deaza-T⁷Sequencing mixes [27-1683-01] which substitute dGTP with either 7-deaza-dGTP (c⁷dGTP) or 7-deaza dITP (c⁷dITP). Sequencing was also done using the following sequence mixes substituted in the Pharmacia protocol: 'Labelling mix A, long run';

50 mM. Sodium chloride
7.5 μ M. c⁷dGTP or dGTP
7.5 μ M. dTTP
7.5 μ M. dCTP
45mM. dithiothreitol
10mM. Tris HCl pH 8.0
1mM. EDTA

or 'Labelling mix A, short run';

50 mM. Sodium chloride
1.5 μ M. c⁷dGTP or dGTP
1.5 μ M. dTTP
1.5 μ M. dCTP
45mM. Dithiothreitol
10mM. Tris HCl pH 8.0
1mM. EDTA

were used in place of the 'Labelling Mix-dATP' for either long or short sequencing runs. The 'A'-, 'C'-, 'G'-, 'T'-Mixes-Short and Long were replaced as follows:

	G mix	A mix	T mix	C mix
ddGTP	8 μ M.	-	-	-
ddATP	-	6 μ M.	-	-
ddTTP	-	-	9 μ M.	-
ddCTP	-	-	-	8 μ M.
dGTP or c ⁷ dGTP	85 μ M.	85 μ M.	85 μ M.	85 μ M.
dATP	85 μ M.	85 μ M.	85 μ M.	85 μ M.
dTTP	85 μ M.	85 μ M.	85 μ M.	85 μ M.
dCTP	85 μ M.	85 μ M.	85 μ M.	85 μ M.
Sodium chloride	50 μ M.	50 μ M.	50 μ M.	50 μ M.
Tris.HCl pH 8.0	10 μ M.	10 μ M.	10 μ M.	10 μ M.
EDTA	1 μ M.	1 μ M.	1 μ M.	1 μ M.

using the deaza-mixes to resolve band compressions.

5.2.4.2 Sequential unidirectional deletion by Exo III/mung bean exonuclease treatment. This is a modified method of Hoheisel and Pohl [32] and Henikoff [11, 12] substituting mung bean exonuclease for S1 nuclease. The clone was digested with two restriction enzymes that cut uniquely in the multiple cloning site of the *pBluescript* on the same side of the insert to give opposing 5' and 3' recessed ends. The clone was therefore first digested with *Xho* I to give a 3' recessed end and then *Kpn* I to give the 5' recessed end. This was then phenol/chloroform extracted, ethanol precipitated and taken up in 21 μ l. TE. The DNA concentration was then determined by measuring the A₂₆₀ of diluted samples. To 17.5 μ l. (2.65 pmoles) DNA solution, 5 μ l. water and 7 μ l. 5 $\times$ Exo-buffer

75mM. Tris.HCl pH 8.0
3.3mM. Magnesium chloride

were added and incubated at 37°C for 2 minutes. Exo III was then added to 135U/pmole (5.5 μ l. , 65U/ μ l.) susceptible 3' ends and the 37°C incubation was continued. Every minute for 11 minutes, 3 μ l. was removed from the reaction mixture and added to an equal volume of water, heated to 100°C for 30 seconds and incubated for a further 7 minutes at 70°C before cooling on ice. On ice, 2.5 μ l. 10 $\times$ mung bean buffer (see 4.2.7.5), 20U mung bean nuclease and water to 25 μ l. were added. This was then incubated at room temperature for 20 minutes before adding 5 μ l. stop solution

800mM. Tris.HCl pH 8.0
20mM. EDTA
80mM. Magnesium chloride.

An 8 μ l. sample was then taken for gel analysis, and to the remaining 22 μ l., 2U Klenow fragment was added and incubated for 2 minutes at 37°C. 2 μ l. dNTP mix (0.25mM. each) was then added and incubated for a further 10 minutes before adding 7.5U T4 DNA ligase and

incubating overnight at room temperature. The ligated mixes were then used to electro-transform XL1 Blue cells. Ten minipreps were prepared for each ligation mix and analysed by *Xba* I digestion and fragment separation on 0.6% w/v agarose gels. On the basis of the gel results, twelve clones of increasing deletion size (each deletion being about 200bp) were selected and sequenced from the deleted end.

5.3 Results and Discussion.

5.3.1 Isolation of the Complete KHK Coding Sequence.

5.3.1.1 Screening of cDNA library. Twelve independent isolates were obtained, after three sequential screens, by plaque hybridization of about one million clones. Since the cDNA cloning site in the library is flanked by M13 forward and reverse primers, these were used to amplify the cDNA inserts. These were analysed by agarose gel electrophoresis, Southern blotting and probe hybridization to reveal the size of the inserts range in size from 1kb to 2.5kb. The three largest of these (2.3kb, 2.4kb and 2.5kb) were rescued from the phage to yield the clones in the *pBluescript* SK(-) phagemid (termed clones 5i, 5ii and 16i, see *Fig. 5.2*). Restriction mapping of the clones and preliminary sequencing of the flanking c.300 base pairs of the inserts using the M13 forward and reverse primers revealed that clone 16i had a poly-A tail at one end, clone 5ii had a region encoding one of the sequenced peptides at one end (KHK 32), although no poly-A tail was observed at the other end of this insert, and that clone 5i contained a region encoding one of the sequenced peptides (KHK 49), but did not contain the 3' end of the gene. It was therefore clear that clone 5i would not contain the complete coding region for the KHK protein, and that insufficient information was available to indicate whether clone 5ii contained the 5' end of the coding sequence. It was therefore concluded that clone 16i offered the greatest chance of obtaining the complete coding region.

5.3.1.2 Sequencing the KHK cDNA. A number of 3' deletion clones of clone 16i were made by Exo III exonuclease / mung bean exonuclease digestion that would allow the sequencing of about two kilobases upstream of the poly-A tail (see *Fig. 5.3*). The sequence obtained from these deleted clones revealed that this clone also was incomplete, and was missing the 5' end of the gene, encoding the N-terminus region of the gene. It did, however, reveal that all the clones contained substantial regions of unassociated, coligated DNA sequences, flanked with *EcoR* I linkers. Re-examination of the restriction maps then showed that clone 5i, although lacking the 3' end of the coding region, did contain the 5' sequence that clone 16i did not possess, along with a substantial section of 5' upstream, untranslated sequence (see *Fig. 5.2*).

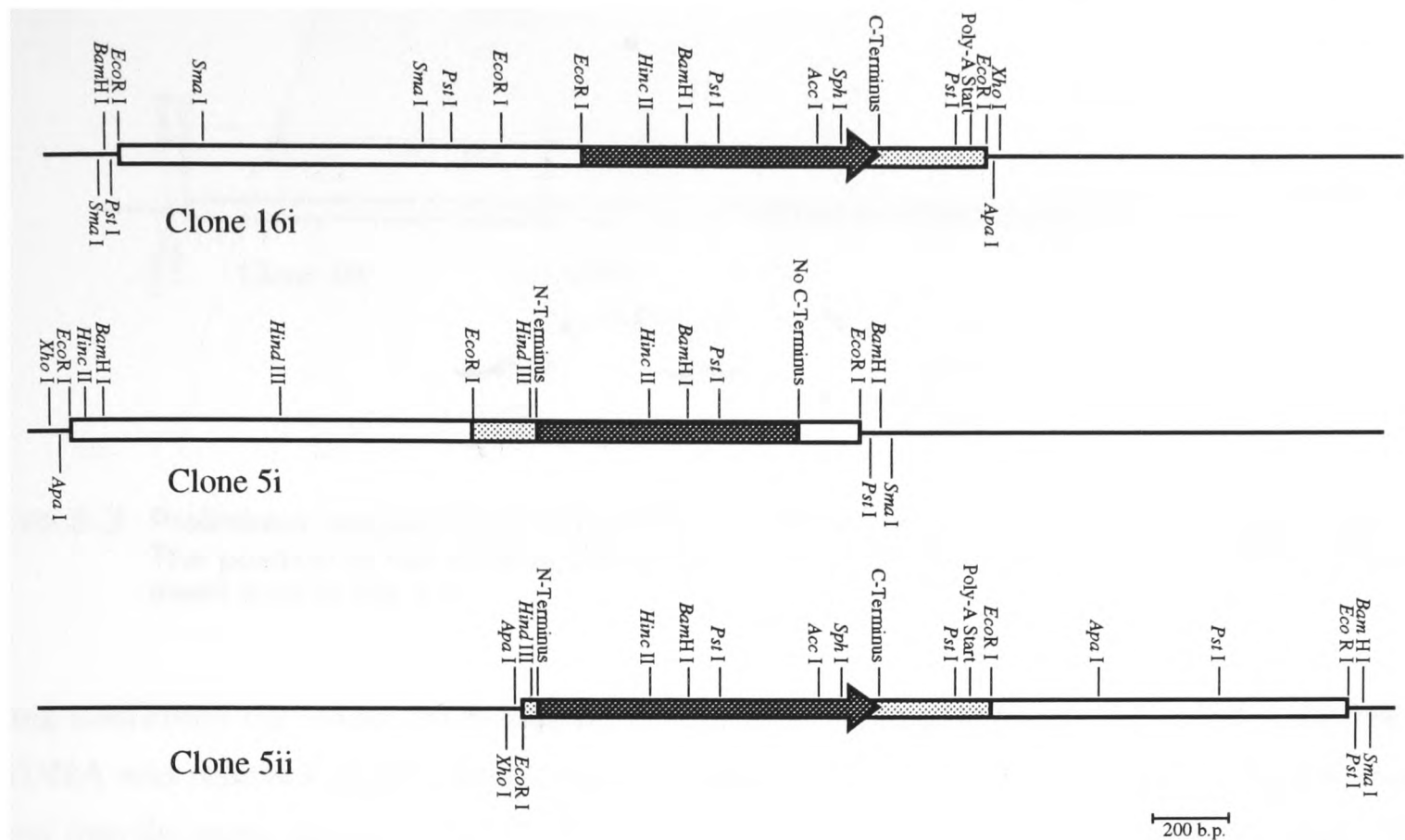


Figure 5.2 Restriction maps of the three largest rescued clones from the λ ZAP II library. The heavily shaded boxes represent the ketohexokinase coding region, the lightly shaded boxes represent 5'- or 3'-untranslated regions, the open boxes represent unassociated, coligated cDNA, and the thin lines, plasmid DNA. All three clones are aligned with respect to the KHK regions.

A hybrid was constructed by splicing the segment downstream of the *Bam*H I site of clone 16i with the segment upstream of the same site from clone 5i (see Fig. 5.4). This construct (clone 5i/16i) therefore contained a section of cDNA containing both the 3'- and 5'-untranslated regions, free of co-ligated cDNA. Using the sequencing information obtained from clone 16i, and the restriction enzyme mapping data, a series of deleted clones was constructed by restriction digestion and religation that allowed the complete sequence of the hybrid clone 5i/16i to be determined on both strands using the M13 forward and reverse primers. This revealed a point mutation in the parent clone 5i (AAT to GAT) resulting in a change of Asn18 to Asp18, but more importantly, it showed that clone 5ii contained the complete coding region for the ketohexokinase gene, along with about fifty base pairs of upstream, untranslated sequence. Two lines of evidence supported the identification of this as a mutation. Firstly, the sequences of clone 5ii and the Exo III deleted constructs of 16i encoded Asn18, and secondly, the amino acid sequence of peptide KHK 32 revealed an Asn at this position. The probability of this Asp to Asn change being a polymorphism is very slight, since the cDNA library and protein sequence were both obtained from the Sprague Dawley strain of rat.

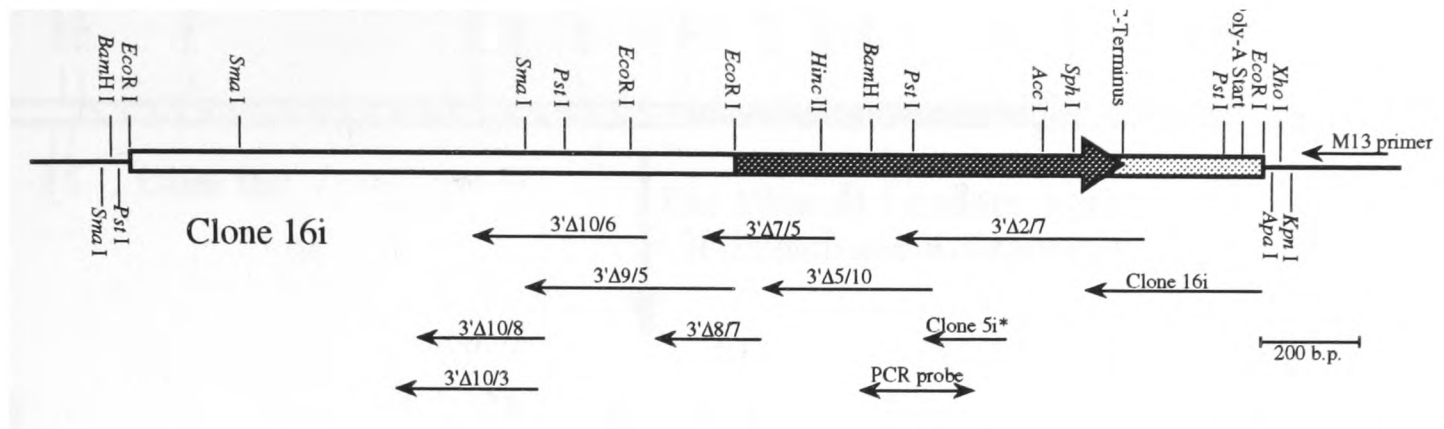


Figure 5.3 Preliminary sequencing of the of clone 16i showing the exonuclease III-deleted clones.. The position of the PCR-amplified probe region is also shown. Shading of the DNA insert is as in Fig. 5.2.

Having confirmed that clone 5ii contained the complete coding sequence for ketohexokinase, the cDNA was relieved of the unassociated, coligated DNA by an *EcoR* I digestion, and was ligated into the same site in *pUC18*. Two clones were obtained from this, with the insert in either direction, and were termed clones *pUC KHK* (+) or (-), depending on the orientation of the KHK gene relative to the β -lactamase gene (see Fig. 5.5). Deletions of these clones were made by restriction enzyme cleavage and religation, as indicated in Figure 5.5, to enable the determination of the sequence of both strands of the entire insert using the M13 forward primer. This sequencing strategy is shown in Figure 5.6.

5.3.1.3 Characterization of the KHK cDNA. The entire sequence of the KHK cDNA, obtained from the three isolated clones from the library, along with the derived amino acid sequence is shown in Figure 5.7. Both clones 16i and 5ii contained poly-A tails at their 3'-termini, that identified the cDNA as being a copy of fully processed mRNA. The consensus polyadenylation signal AATAAA [33] was identified 20 nucleotides upstream of each poly-A tract. The coding region has been verified using data obtained from both strands of at least two independent clones. The translation initiation codon was assigned using several criteria. Firstly, its position, relative to the 3'-termination codon, would generate a protein, the molecular mass of which (32,728 Da) is in close agreement with that determined by mass spectrometry (32,800 Da). Secondly, all nine peptide sequences from the purified ketohexokinase lie in the open reading frame created by this start codon. Thirdly, it is the only in-frame ATG between the region encoding peptide KHK 32 and the in-frame stop codon 198 nucleotides upstream of this initiation site. Fourthly, the overwhelming predominance of C at the -1 and -4 positions of the initiation site sequences from rat cDNAs is conserved within the ketohexokinase cDNA when it is assigned this start codon [34, 35].

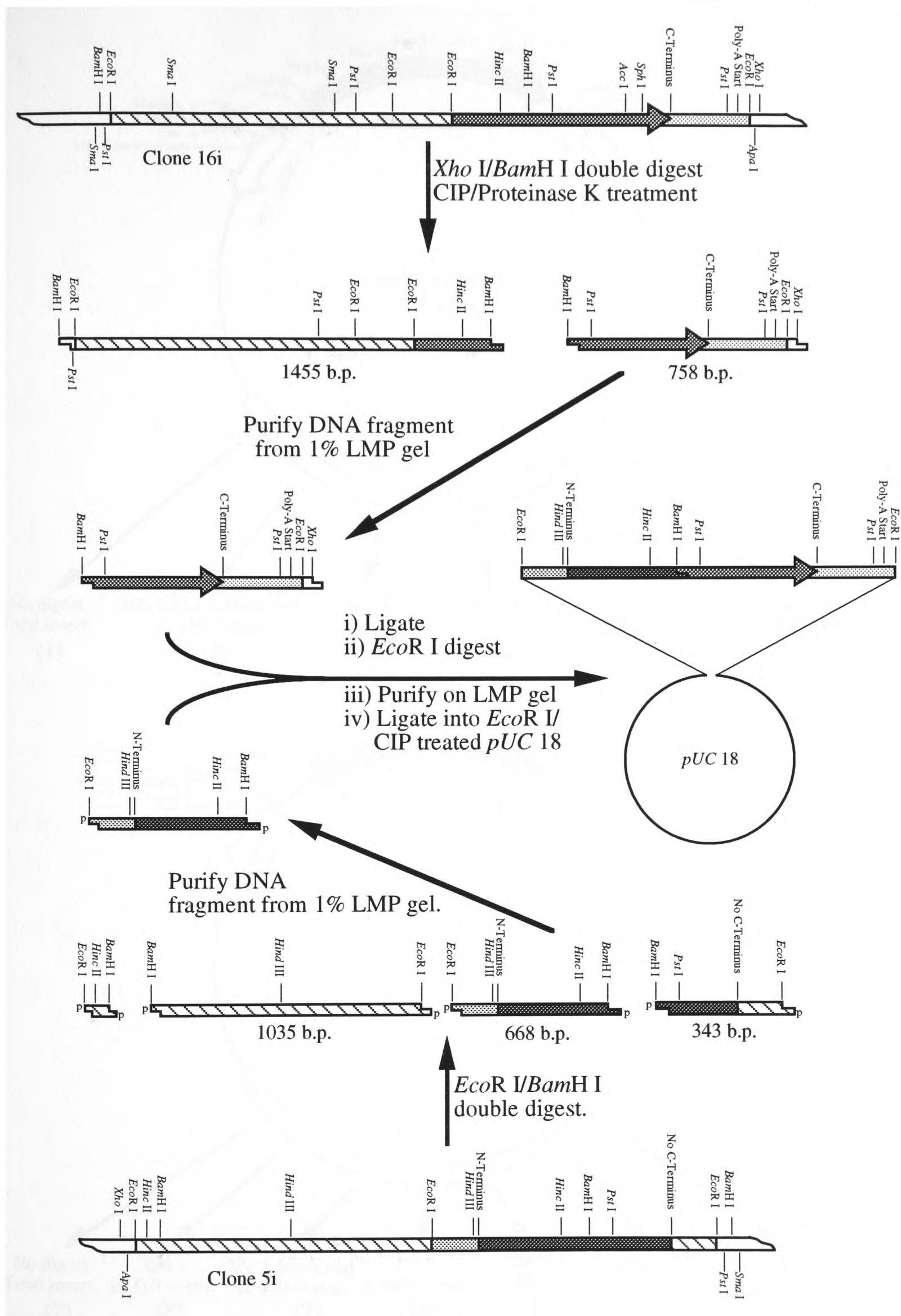


Figure 5.4 Strategy for the construction of clone 5i/16i in pUC18. Inserts of each direction in pUC18 were identified as either + or -, relative to the β -lactamase gene. The insert labelling is as fig. 5.2.

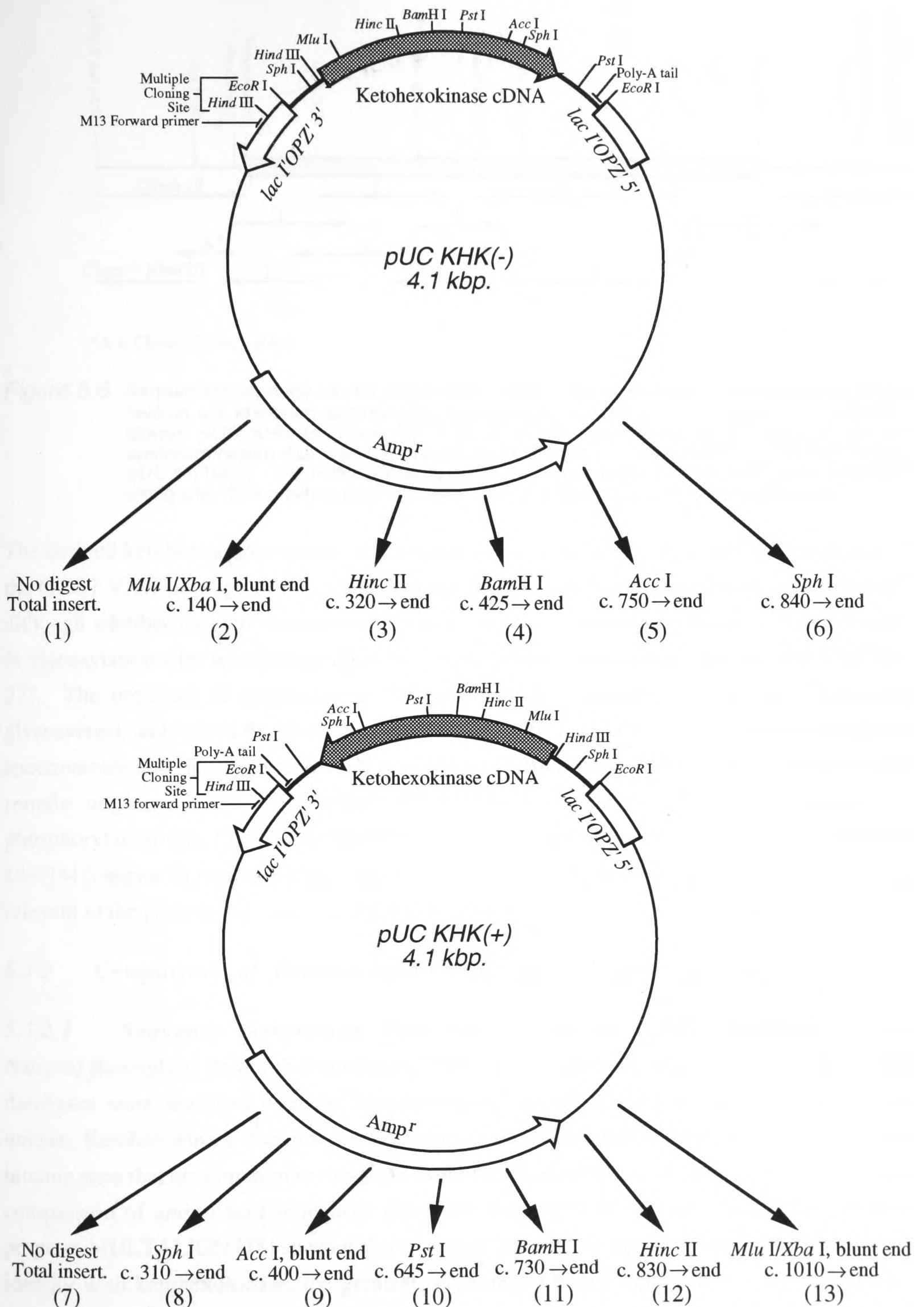
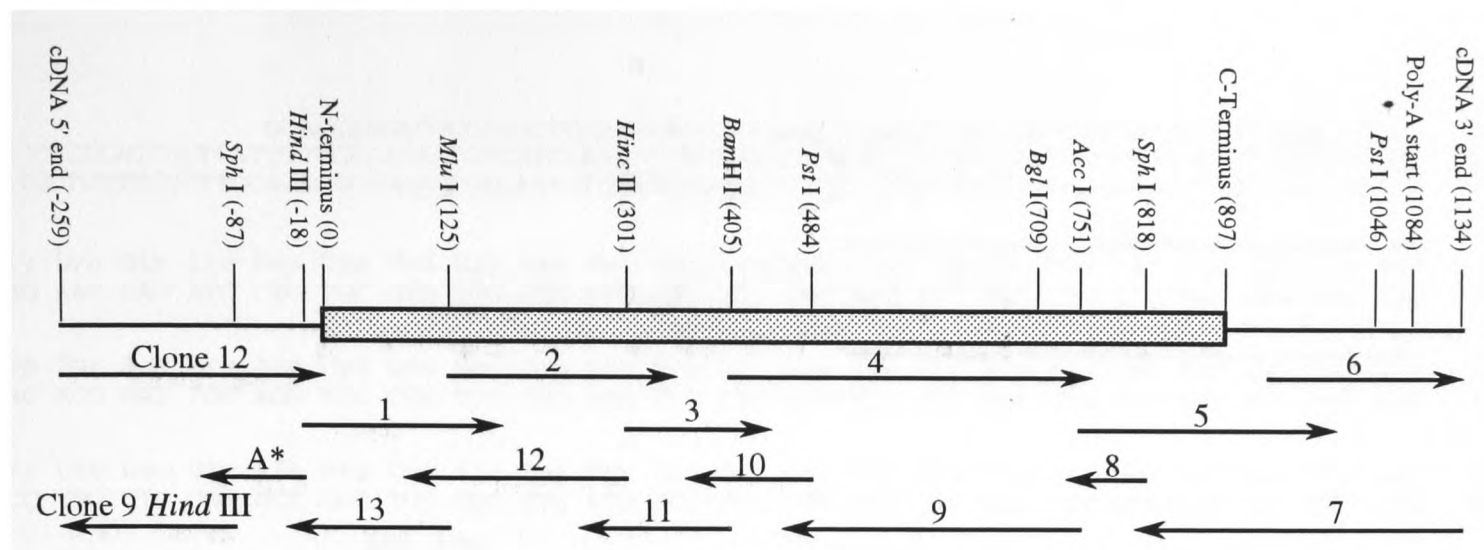


Figure 5.5 Construction of deletion clones of the KHK cDNA for sequencing of both strands. The sequencing strategy for the insert is shown in Fig 5.6.



*A is Clone 12 *Xba* I-*Mlu* I

Figure 5.6 Sequencing strategy for the entire KHK cDNA. The clones used for sequencing each section are identified by an arrow with a number in the case of the deleted clones *pUC* KHK ($\pm$) (see Fig. 5.5), or in the case of the three clones in the 5' upstream region that was not present in *pUC* KHK ($\pm$), were derived from the clones *pUC* 51/16i($\pm$). The restriction enzyme sites of importance for sequencing are marked, along with their position (in base pairs) from the first nucleotide in the start codon.

The derived ketohexokinase amino acid sequence was analysed by the MOTIFS program on the GCG7 VAX package of molecular biology software. A number of features were found, although whether they are present in the native protein remains unconfirmed. Six potential N-glycosylation sites were found, these having the consensus sequence Asn-Xaa-Ser/Thr [36, 37]. The presence of these sites is not sufficient to conclude that the Asn residue is glycosylated, and indeed the close agreement between the molecular mass determined by mass spectrometry and that derived from the amino acid sequence would suggest that these sites remain unglycosylated. The other motifs found (amidation [38], casein kinase II phosphorylation sites [39], N-myristoylation sites [40], and protein kinase C phosphorylation sites [41]) appear in regions of the protein that would not undergo such modification, or are not relevant to the protein, and were assumed to be absent.

5.3.2 Comparison of Ketohexokinase with Other Protein Sequences.

5.3.2.1 Sequence comparison with proteins on the NBRF database. The National Biomedical Research Foundation (NBRF) Protein Identification Resource (PIR) V.29 databases were searched with the ketohexokinase protein sequence, which proved to be unique. Ketohexokinase does not contain sequence complementary to either the ATP or hexose binding sites that are common to members of the hexokinase family of proteins [42]. Pairwise comparison of amino acid sequences using the Alignment of Multiple Sequence (AMPS) program MULTALIGN [18] showed that no single protein had a prominently similar sequence identity with ketohexokinase, the greatest percentage identity being ~20%. Of the protein sequences showing this degree of identity, three furanose sugar kinases were selected for detailed comparison, because of their similar size, function and substrate specificity. These

GCAGCCAGACGCCAGGCTGGATGAAATGAAGAAGGCAATGGCAGGCAGTGGTGCCAGCCTCTGAAGGTGT																							69	
CGAGCTGCTACCTCCCACCGCTCATCTCTCACACAATGCCATCAAACCCCTGCAAGCCACACTTCTTGCCAGCCACCTGGCACCCGAATCTCAGG																							164	
GCCTCCTGCATGCTGTTTCCCTTGAGAGGGCTGAAGAGACAGAGTCTAGCAGGAATCCCCTCCGCTTGCGGGTAGGAAGCTTGGGGAGCAGCCTC																							259	
KHK 32																								
Met	Glu	Glu	Lys	Gln	Ile	Leu	Cys	Val	Gly	Leu	Val	Val	Leu	Asp	Ile	Ile	Asn	Val	Val	Asp	Lys	Tyr	Pro	24
ATG	GAA	GAG	AAG	CAG	ATC	CTG	TGC	GTG	GGG	CTG	GTG	GTG	CTG	GAC	ATC	ATC	AAT	GTG	GTG	GAC	AAA	TAC	CCA	331
															↑	KHK 46								
Glu	Glu	Asp	Thr	Asp	Arg	Arg	Cys	Leu	Ser	Gln	Arg	Trp	Gln	Arg	Gly	Gly	Asn	Ala	Ser	Asn	Ser	Cys	Thr	48
GAG	GAA	GAC	ACG	GAT	CGC	AGG	TGC	CTA	TCC	CAG	AGA	TGG	CAG	CGT	GGA	GGC	AAC	GCG	TCC	AAC	TCC	TGC	ACT	403
Val Leu Ser Leu Leu Gly Ala Arg Cys Ala Phe Met Gly Ser Leu Ala His Gly His Val Ala Asp Phe Leu																							72	
GTG CTT TCC TTG CTC GGA GCC CGC TGT GCC TTC ATG GGC TCG CTG GCC CAT GGC CAT GTT GCC GAC TTC CTG																							475	
KHK 38a																								
Val	Ala	Asp	Phe	Arg	Arg	Arg	Gly	Val	Asp	Val	Ser	Gln	Val	Ala	Trp	Gln	Ser	Gln	Gly	Asp	Thr	Pro	Cys	96
GTG	GCC	GAC	TTC	AGG	CGG	AGG	GGT	GTG	GAT	GTG	TCT	CAA	GTG	GCC	TGG	CAG	AGC	CAG	GGA	GAT	ACC	CCT	TGC	547
KHK 45																								
Ser	Cys	Cys	Ile	Val	Asn	Asn	Ser	Asn	Gly	Ser	Arg	Thr	Ile	Ile	Leu	Tyr	Asp	Thr	Asn	Leu	Pro	Asp	Val	120
TCC	TGC	TGC	ATC	GTG	AAC	AAC	TCC	AAT	GGC	TCC	CGT	ACC	ATT	ATT	CTC	TAC	GAC	ACG	AAC	CTG	CCA	GAT	GTG	619
KHK 38b																								
Ser	Ala	Lys	Asp	Phe	Glu	Lys	Val	Asp	Leu	Thr	Arg	Phe	Lys	Trp	Ile	His	Ile	Glu	Gly	Arg	Asn	Ala	Ser	144
TCT	GCT	AAG	GAC	TTT	GAG	AAG	GTC	GAT	CTG	ACC	CGG	TTC	AAG	TGG	ATC	CAC	ATT	GAG	GGC	CGG	AAT	GCA	TCG	691
KHK 22																								
Glu	Gln	Val	Lys	Met	Leu	Gln	Arg	Ile	Glu	Gln	Tyr	Asn	Ala	Thr	Gln	Pro	Leu	Gln	Gln	Lys	Val	Arg	Val	168
GAA	CAG	GTA	AAG	ATG	CTA	CAG	CGG	ATA	GAA	CAG	TAC	AAT	GCC	ACG	CAG	CCT	CTG	CAG	CAG	AAG	GTC	CGG	GTG	763
KHK 49																								
Ser	Val	Glu	Ile	Glu	Lys	Pro	Arg	Glu	Glu	Leu	Phe	Gln	Leu	Phe	Gly	Tyr	Gly	Glu	Val	Val	Phe	Val	Ser	192
TCC	GTG	GAG	ATA	GAG	AAG	CCC	CGA	GAG	GAA	CTC	TTC	CAG	CTG	TTC	GGC	TAT	GGA	GAG	GTG	GTG	TTT	GTC	AGC	835
Lys	Asp	Val	Ala	Lys	His	Leu	Gly	Phe	Arg	Ser	Ala	Gly	Glu	Ala	Leu	Lys	Gly	Leu	Tyr	Ser	Arg	Val	Lys	216
AAA	GAT	GTG	GCC	AAG	CAC	CTG	GGG	TTC	CGG	TCA	GCA	GGG	GAG	GCC	CTG	AAG	GGC	TTG	TAC	AGT	CGT	GTG	AAG	907
KHK 43b											KHK 43a													
Lys	Gly	Ala	Thr	Leu	Ile	Cys	Ala	Trp	Ala	Glu	Glu	Gly	Ala	Asp	Ala	Leu	Gly	Pro	Asp	Gly	Gln	Leu	Leu	240
AAA	GGG	GCT	ACG	CTC	ATC	TGT	GCC	TGG	GCT	GAG	GAG	GGA	GCC	GAT	GCC	CTG	GGC	CCC	GAC	GGC	CAG	CTG	CTC	979
His Ser Asp Ala Phe Pro Pro Pro Arg Val Val Asp Thr Leu Gly Ala Gly Asp Thr Phe Asn Ala Ser Val																							264	
CAC TCA GAT GCC TTC CCA CCA CCC CGA GTA GTA GAC ACT CTC GGG GCT GGA GAC ACC TTC AAT GCC TCT GTC																							1051	
Ile Phe Ser Leu Ser Lys Gly Asn Ser Met Gln Glu Ala Leu Arg Phe Gly Cys Gln Val Ala Gly Lys Lys																							288	
ATC TTC AGC CTC TCC AAG GGA AAC AGC ATG CAG GAG GCC CTG AGA TTC GGG TGC CAG GTG GCT GGC AAG AAG																							1123	
Cys Gly Leu Gln Gly Phe Asp Gly Ile Val ***																							298	
TGT GGC TTG CAG GGG TTT GAT GGC ATT GTG TGA GAGATGAGCGGTGGGAGGTAGCAGCTCGACACCTCAGAGGCTGGCACCCT																							1207	
GCCTGCCATTGCCTTCTTCATTTTCATCCAGCCTGGCGTCTGGCTGCCAGTTCCTTGGGCCAGTGTAGGCTGTGGAACGGGTCTTTCTGTCTCTT																							1302	
CTCTGCAGACACCTGGAGCAATAAATCTTCCCCTGAGCCAA																							1393	

Figure 5.7 DNA and amino acid sequence of rat liver ketohexokinase. Bases are numbered from the 5'-end of the cDNA sequence in clone 5i. The in-phase termination codon is designated by ***, the in-frame, upstream stop codon in bold underlined letters, and the putative polyadenylation signal is underlined. The region of the structural cDNA that was used to probe the cDNA library is in italics, and the site of the point mutation (A to G) in clone 5i is arrowed. Amino acids are numbered starting with the initiator Met. Amino acids identified by peptide sequencing are indicated by over-lines and the peptide name is shown above each line.

proteins were the *E. coli* minor phosphofructokinase (pfk2, EC 2.7.1.11) [43], *E. coli* ribokinase (EC 2.7.1.15) [44] and *Vibrio alginolyticus* fructokinase (EC 2.7.1.4) [45].

Multiple alignment of the ketohexokinase sequence with those of these three proteins was performed using the MULTALIGN program. This program uses the Needleman and Wunsch algorithm, modified to allow quick comparison of more than two sequences [21]. Rat

	KHK	pfk2	FK	RK	GK	HK
KHK	-	21.4	18.24	19.66	20.27	21.21
pfk2	4.48	-	20.41	22.67	20.78	21.17
FK	3.24	6.17	-	23.21	16.99	19.61
RK	5.82	9.00	8.15	-	22.08	18.45
GK	0.03	-0.68	-0.33	-0.49	-	52.11
HK	1.24	-1.94	1.26	-0.50	59.23	-

Table 5.1 Comparison of proteins using the MULTALIGN program. The percentage identity between the proteins is indicated above the diagonal, and the similarity score is shown below the diagonal. The similarity scores were determined by randomizing the sequence 100 times, and comparing each randomized sequence. The means and the standard deviations of the distribution of comparisons are calculated, and the similarity expressed in terms of the number of units the standard deviation is away from the mean of the random distribution. (Abbreviations: KHK; ketohexokinase, pfk2; *E. coli* minor phosphofructokinase, FK; *V. alginolyticus* fructokinase, RK; *E. coli* ribokinase, GK and HK; rat liver glucokinase and hexokinase.)

hexokinases [46, 47, 48] and glucokinase [49] were also compared using this method, and the results are summarized in *Table 5.1*.

The amino acid sequence of ketohexokinase reveals that it is much more closely related to the prokaryotic furanose kinases examined than it is to the eukaryotic hexokinase family. *Table 5.1* predictably shows a high percentage identity when a direct comparison is made between glucokinase and hexokinase. The ‘percentage identities’ obtained with all other pairwise comparisons are ~20%. However, the ‘similarity scores’ clearly show that, although glucokinase and hexokinase are very closely related to each other, they show no similarity with the other proteins. The ‘similarity scores’ obtained for all possible comparisons between the other proteins show that they are related to one another, although, unsurprisingly, the mammalian ketohexokinase appears to be the most evolutionarily divergent protein within the group.

The multiple alignment of the four proteins revealed a highly conserved motif (*Fig 5.8*), TX(A/G)AGDX₉L, where the amino acids X are nearly always conservative amino acid replacements. This consensus was used to search the PIR databases and was also found in a fifth protein, *E. coli* DNA topoisomerase I [50] (*Fig 5.9*).

5.3.2.2 A potential furanose-binding motif. The most striking similarity between them is the conserved region depicted in *Figures 5.8* and *5.9*. Although the proteins in *Fig. 5.8* are all ATP-dependent kinases, the possibility that this region encodes a common binding site for ATP would seem to be discounted by its conservation in the structural gene for DNA topoisomerase I, that does not employ ATP or NAD as a substrate. The proposed mechanism

Ketohexokinase	MEEKQILCVG LVVLDIINVVDK YPEEDTDRRCLSQRWQR	GGNASNSCTVLSLLGARCAPMGSLAHGHVADFLVADFRRRGV
E. coli ribokinase	MQN AGSLVVLGSINADHILNLQSFPTPGETVTGNHYQVAF	GGKGANQAVAAGRSGANIAFI ACTGDDSIGESVRQQLATDNI
E. coli pfk2	MVRIYTLTLAP SLDSATITPQIYPEENCAVPHRCNP	GGGINVARAIAHLGGSATAIFPAGGAT GEHLVSLADENV
Fructokinase	M NQVWVTGDAVVLDI PESETSLKKC	PGGAPANVAVAIARLSGKSAFFGRVGGDDPFGRFMQSILDQEGV
Ketohexokinase	DVSQVAMQSQGDTPCSCCIWNNSNGSRTI ILYDTNLPDVSAKDFE	KVLDLTRFKWTH IEGRNASEQVKMLQR IEQYNATQPLQQKVRV
E. coli ribokinase	DITPVSVIKGESTGV ALIFVNGEGENVIG IHAGANAALSPA LVEAQRER	IANASALLMQLESPLSVMAAAKIAHQNKTI VAL
E. coli pfk2	PVATVE AKDWTRQNLHVHVEASGEQ YRF VMPG AALNEDEFRQ	LEEQVLEIESGAILVISGSLPPGVKLEKLTQLISLRKNK
Fructokinase	CTEFLIKDPEQRTSTTVVVDLDD QGER SFTFMVKPSADQFMSVED	DMGNFKQGDWLHVCSISLANEPSRSSTFEAIKRAKAAGGFISF
Ketohexokinase	SVEIEKPR EELFQLFGYGEWVFVSKDVAKHLGFR	SAGEALKGLYSRVKKGATLICAWAEEGADALGPDGQLLHSDAFPP
E. coli ribokinase	NPAPARELP DELLALVDIITPNETEAE KLTGIRVEND	EDAAKAAQVLHEKGIRT VLITLGSRGVWASVNGEQRVPGFR
E. coli pfk2	GSAASSTVLGQGLSAAALATGN IELVKPNQKELSALVNRELTPDD	VVRKAAQEI VNSGKAKRVVSLGPGQALGVDSEN CIQVVP
Fructokinase	DPNLRDEVW QDQSEIQAWVMKAVAMADVVKFSEEBELLFLTDST	SMAQGLQQIAAMNIAL VLVTOGAKGVWRVPESQSELITQOV
Ketohexokinase	PRVVDTLGAGDTFNASVIFSLS	KGNSMQEALRFGCQVAGKKCGLQG FDGIV
E. coli ribokinase	VQAVDTIAAGDTFNGALITALL	EEKPLPEAIRFAHAAAIAVTRKGAQPSVPWREE IDAFLDRQR
E. coli pfk2	LKSQSTVGAGDRLVGAMTLKLA	ENASLEEMVRFGVAAGSAATLNQGTRLCSHDDTQKIYAYLSR
Fructokinase	VSPIDTTGAGDAFVGGLLACLS	RHADWKNHPVSSAIQWANGCGALATTQKGAMTALPTQTE LLRFIGQ

Figure 5.8 Amino acid comparison of ketohexokinase and other furanose kinases. The sequences were compared using the MULTALIGN program. The amino acid homology of ribokinase, phosphofructokinase 2 (pfk2) and fructokinase are compared with those of ketohexokinase, and those found to be identical are in bold, and those with conservative replacements are italicised.

for *E. coli* topoisomerase I [51] involves the transient breakage of a DNA backbone bond and simultaneous formation of a phosphotyrosine bond between the protein and the cut 5' phosphoryl group. The cleavage specificity suggests that the enzyme is interacting with both sides of the point it cleaves, and therefore bridges the breakpoint, binding non-covalently with the 3' deoxyribose sugar. The conserved motif TX(A/G)AGDX₉L may be involved in the non-covalent binding of either the 5' deoxyribose phosphate group or the 3' sugar. The conservation of this region may therefore reflect a common function in the binding of furanose sugars.

E. coli pfk2 is only distantly related to *E. coli* pfk1, the crystal structure and substrate binding sites of which have been determined [52] and similar structural data have also been obtained for the phosphofructokinase of *Bacillus stearothermophilus* [53, 54]. *E. coli* pfk1 and pfk2 do not show immunological cross-reactivity [55], and, unlike pfk1, pfk2 is not under allosteric

Ketohexokinase	(248)	PRVVDTLGAGDTFNASVIFSLS
E. coli ribokinase	(250)	VQAVDTIAAGDTFNGALITALL
E. coli pfk2	(250)	LKSQSTVGAGDRLVGAMTLKLA
Fructokinase	(244)	VSPIDTTGAGDAFVGGLLACLS
DNA topoisomerase I	(414)	DSTTLTVGAGDFRLKARGRILR

Figure 5.9 Amino acid sequence comparison for ketohexokinase, ribokinase, pfk2, fructokinase and DNA topoisomerase I, at the proposed region involved in furanose binding (see text). The boxed region indicates the region used to search the PIR databases, which lead to the location of DNA topoisomerase. Labelling of amino acids is as in Figure 5.8.

control by ADP and PEP [56], however, the proteins are of similar size and possess a region of homology that corresponds with the consensus sequence in Fig. 5.9 that pfk2 shares with the other furanose kinases examined. Furthermore, this region of homology is also present in the amino acid sequence of *Bacillus stearothermophilus* phosphofructokinase (Fig. 5.10). The structural data for both *E.coli* pfk1 and *Bacillus stearothermophilus* phosphofructokinase reveal that this region contains residues involved in binding fructose 6-phosphate and is remote from those that bind ATP. This further suggests that the TX(A/G)AGDX₉L region of homology is unlikely to be involved in the binding of ATP, but may be involved in the recognition of furanose sugars.

The quaternary structure of the major phosphofructokinase consists of a tetramer of peptides, arranged as two dimers, with the active sites being present at the junction of the two peptides in each dimer [53]. Size exclusion chromatography reported in this work (see section 3.3) revealed that the quaternary structure of ketohexokinase is also a tetramer, although previous studies have reported ketohexokinase as a dimer. Our results would support the theory that ketohexokinase belongs to an ancient family of phosphofructokinase-like proteins. The report that KHK exists as a dimer may reflect the conditions used to determine the quaternary structure, and may in fact represent the dissociated dimers from a tetramer.

Ketohexokinase has recently been reported in the halophilic archaebacterium *Haloarcula vallismortis* [57]. This is the first report of an ATP-dependent D-fructose 1-phosphotransferase in prokaryotes. It would be interesting if future studies show whether this protein belongs within the same group of furanose kinases revealed in this report, because this would lend support to my belief that this represents an ancient family of proteins.

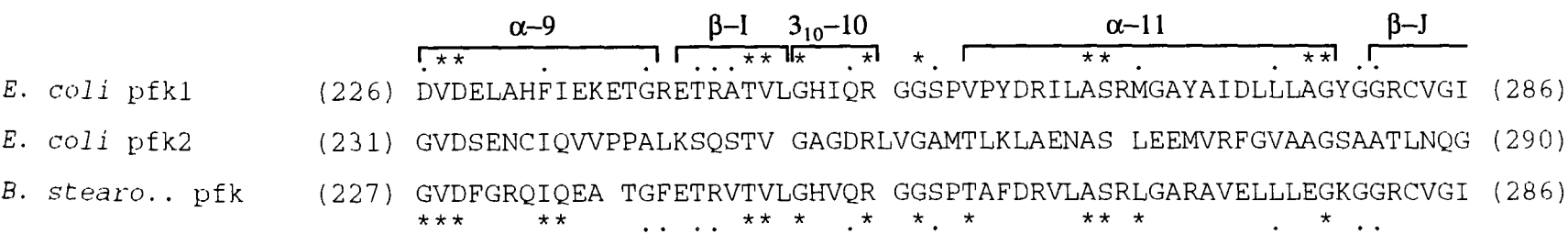


Figure 5.10 Amino acid comparison of the *E. coli* minor phosphofructokinase (pfk2) with the *E. coli* major phosphofructokinase (pfk) and the *B. stearothermophilus* pfk sequences. Sequence identity is shown with a *, and conserved replacements with a •. The structural motifs of the *E. coli* pfk protein are shown above its amino acid sequence. The positions of the first and last amino acid of each sequence is shown in brackets.

5.4 References

1. Young, R.A. and Davis, R.W. (1983). Efficient isolation of genes using antibody probes. *Proc. Natl. Acad. Sci. USA*, **80**, 1194-1198.
2. Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988). λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. *Nuc. Acids Res.*, **16**, 7583-7600.
3. Brenner, S., Cesareni, G. and Karn, J. (1982). Phasmids: Hybrids between ColE1 plasmids and *E. coli* bacteriophage lambda. *Gene*, **17**, 27-44.
4. Dotto, G.P., Horiuchi, K. and Zinder, N.D. (1984). The Functional Origin of Bacteriophage ϕ 1 DNA Replication: Its Signal and Domains. *J. Mol. Biol.*, **172**, 507-521.
5. Geider, K., Beck, E. and Schaller, H. (1978). An RNA transcribed from DNA at the origin of phage ϕ d single strand to replicative form conversion. *Proc. Natl. Acad. Sci. USA*, **75**, 645-649.
6. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.A. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457-465.
7. Staden, R. (1986). The current status and portability of our sequence handling software. *Nuc. Acids Res.*, **14**, 217-231.
8. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual*, 2nd Edn, Cold Spring Harbor Laboratory Press.
9. Weiss, B. (1976). Endonuclease II of *Escherichia coli* is exonuclease III. *J. Biol. Chem.*, **251**, 1896-1901.
10. Rogers, S.G. and Weiss, B. (1980). Exonuclease III of *Escherichia coli* K12, an AP endonuclease. *Meths. Enzymol.*, **65**, 201-211.
11. Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*, **28**, 351-359.
12. Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Meths. Enzymol.*, **155**, 156-165.
13. Guo, L.-H. and Wu, R. (1983). Exonuclease III: use for DNA sequence analysis and in specific deletions of nucleotides. *Meths. Enzymol.*, **100**, 60-96.
14. Jasin, M., Regan, L. and Schimmel, P.R. (1983). Modular arrangement of functional domains along the sequence of an aminoacyl tRNA synthetase. *Nature*, **306**, 441-447.
15. Mizusawa, S., Nishimura, S. and Seela, F. (1986). Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nuc. Acids Res.* **14**, 1319-1324.
16. Barr, P.J., Thayer, R.M., Laybourn, P., Najarian, R.C., Seela, F. and Tolan, D.R. (1986). 7-Deaza-2'-deoxyguanosine-5'-triphosphate: Enhanced resolution in M13 dideoxy sequencing. *BioTechniques*, **4**, 428-432.

17. Russell, R.B. and Barton, G.J. (1992). Multiple Protein Sequence Alignment From Tertiary Structure Comparison: Assignment of Global and Residue Confidence Levels. *Proteins*, **14**, 309-323.
18. Barton, G.J. and Sternberg, M.J.E. (1987). A Strategy for the Rapid Multiple Alignment of Protein Sequences. *J. Mol. Biol.*, **198**, 327-337.
19. Barton, G.J. and Sternberg, M.J.E. (1987). Evaluation and improvements in the automatic alignment of protein sequences. *Prot. Eng.*, **1**, 89-94.
20. Barton, G.J. and Sternberg, M.J.E. (1990). Flexible Protein Sequence Patterns; A Sensitive Method to Detect Weak Structural Similarities. *J. Mol. Biol.*, **212**, 389-402.
21. Barton, G.J. (1990). Protein Multiple Sequence Alignment and Flexible Pattern Matching. *Methods. Enzymol.*, **48**, 403-428.
22. Needleman, S.B. and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 433-453.
23. Smith, T.F. and Waterman, M.S. (1981). Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195-197.
24. Bullock, W.O., Fernandez, J.M. and Short, J.M. (1987). XL1-Blue: A high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *BioTechniques*, **5**, 376-379.
25. Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984). Experiments with Gene Fusions. Cold Spring Harbor Laboratory Press, New York.
26. Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nuc. Acids Res.*, **16**, 6127-6145.
27. Taketo, A. (1988). DNA transformation of *Escherichia coli* by electroporation. *Biochim. Biophys. Acta*, **949**, 318-324.
28. Tsong, T.Y. (1991). Electroporation of cell membranes. *Biophys. J.*, **60**, 297-306.
29. Feinberg and Vogelstein (1983). A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity. *Anal. Biochem.*, **132**, 6-13 (Addendum: *Anal. Biochem.* (1984), **137**, 266-267).
30. Johnson, M.T., Read, B.A., Monko, A.M., Pappas, G. and Johnson, B.A. (1986). A Convenient, New Method for Desalting, Deproteinizing, and Concentrating DNA or RNA. *BioTechniques*, **4**, 64-70.
31. Friedman, K.D., Rosen, N.L., Newman, P.J. and Montgomery, R.R. (1990). Screening of λ gt11 libraries, in *PCR protocols, A guide to methods and Applications*. (Innis M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. Eds.), Academic Press Inc, San Diego, 253-257.
32. Hoheisel, J. and Pohl, F.M. (1986). Simplified preparation of unidirectional deletion clones. *Nuc. Acids Res.* **14**, 3605.
33. Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic mRNA. *Nature*, **263**, 211-214.
34. Kozak, M. (1984). Comparison and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nuc. Acids Res.*, **12**, 857-872.

35. Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283-292.
36. Sox, H.C. and Hood, L. (1970). Attachment of Carbohydrate to the Variable Region of Myeloma Immunoglobulins Light Chains. *Proc. Natl. Acad. Sci. USA*, **66**, 975-982.
37. Marshall, R.D. (1972). Glycoproteins. *Ann. Rev. Biochem.*, **41**, 673-702.
38. Kreil, G. (1984). Occurrence, Detection and Biosynthesis of Carboxy-Terminal Amides. *Methods. Enzymol.*, **106**, 218-223.
39. Pinna, L.A. (1990). Casein Kinase 2: an 'eminence grise' in cellular regulation? *Biochim. Biophys. Acta*, **1054**, 267-284.
40. Towler, D.A., Gordon, J.I., Adams, S.P. and Glaser, L. (1988). The biology and enzymology of eukaryotic protein acetylation. *Ann. Rev. Biochem.*, **57**, 69-99.
41. Woodget, J.R., Gould, K.L. and Hunter, T. (1986). Substrate specificity of protein kinase C. *Eur. J. Biochem.*, **161**, 177-184.
42. Griffin, L.D., Gelb, B.D., Wheeler, D.A., Davidson, D., Adams, V. and McCabe, E.R.B. (1991). Mammalian Hexokinase 1: Evolutionary Conservation and Structure to Function Analysis. *Genomics*, **11**, 1014-1024.
43. Daldal, F. (1984). Nucleotide sequence of gene *pfkB* encoding the minor phosphofructokinase of *Escherichia coli* K12. *Gene*, **28**, 337-342.
44. Hope, J.N., Bell, A.W., Hermodson, M.A. and Groarke, J.M. (1986). Ribokinase from *Escherichia coli* K12. Nucleotide Sequence and Overexpression of the *rbsK* Gene and Purification of Ribokinase. *J. Biol. Chem.*, **261**, 7663-7668.
45. Blatch, G.L., Scholle, R.R. and Woods, D.R. (1990). Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose uptake-encoding region. *Gene*, **95**, 17-23.
46. Schwab, D.A. and Wilson, J.E. (1989). Complete amino acid sequence of the rat brain hexokinase, deduced from the cloned cDNA, and proposed structure of a mammalian hexokinase. *Proc. Natl. Acad. Sci. USA*, **86**, 2563-2567.
47. Thelen, A.P. and Wilson, J.E. (1991). Complete amino acid sequence of the type II isozyme of rat hexokinase, deduced from a cloned cDNA: Comparison with a hexokinase from Novikoff ascites tumor. *Arch. Biochem. Biophys.*, **286**, 645-651.
48. Schwab, D.A. and Wilson, J.E. (1991). Complete amino acid sequence of the type III isozyme of rat hexokinase, deduced from cloned cDNA. *Arch. Biochem. Biophys.*, **285**, 365-370.
49. Andreone, T.L., Printz, R.L., Pilkis, S.J., Magnuson, M.A. and Granner, D.K. (1989). The Amino Acid Sequence of Rat Liver Glucokinase Deduced from Cloned cDNA. *J. Biol. Chem.*, **264**, 363-369.
50. Tse-Dinh, Y.-C. and Wang, J.C. (1986). Complete Nucleotide Sequence of the *topA* Gene Encoding *Escherichia coli* DNA Topoisomerase I. *J. Mol. Biol.*, **191**, 321-331.
51. Wang, J.C. and Liu, L.F. (1979). DNA Topoisomerases: Enzymes That Catalyze the Concerted Breaking and Rejoining of DNA Backbone Bonds, in: *Molecular Genetics*, (Taylor, J.H. Ed.), Academic Press, Inc., New York, vol. III, 65-88.

52. Shirakihara, Y. and Evans, P.R. (1988). Crystal Structure of the Complex of Phosphofructokinase from *Escherichia coli* With its Reaction Products. *J. Mol. Biol.*, **204**, 973-994.
53. Evans, P.R. and Hudson, P.J. (1979). Structure and control of phosphofructokinase from *Bacillus stearothermophilus*. *Nature*, **279**, 500-504.
54. Kolb, E., Hudson, P.J. and Harris, J.I. (1980). Phosphofructokinase: complete amino acid sequence of the enzyme from *Bacillus stearothermophilus*. *Eur. J. Biochem.*, **108**, 587-597.
55. Kotlarz, D. and Buc, H. (1977). Two *Escherichia coli* fructose-6-phosphate kinases. Preparative purification, oligomeric structure and immunological studies. *Biochim. Biophys. Acta*. **481**, 35-48.
56. Kotlarz, D. and Buc, H. (1981). Regulatory Properties of Phosphofructokinase 2 from *Escherichia coli*. *Eur. J. Biochem.*, **117**, 569-574.
57. Altekar, W. and Rangaswamy, V. (1991). Ketohexokinase (ATP: D-fructose 1-phosphotransferase) initiates fructose breakdown via the modified EMP pathway in halophilic archaebacteria. *FEMS Microbiol. Lett.*, **81**, 241-246.

5.5 Addendum

During the final drafting of this chapter, Bork *et al.* [58] reported the convergent evolution of a number of families of sugar kinases, resulting in three families of proteins catalyzing the phosphorylation of hexoses and pentoses. Of the three families reported (hexokinase, ribokinase and galactokinase), the ribokinase family showed a greater number of proteins showing a similarity to rat ketohexokinase, although ketohexokinase was not included in the family, probably due to its recent release on the EMBL databases. All of the proteins did, however, show sequence 'boxes' of similarity, the third of which was identified in this report. Dr. Peer Bork showed great interest in the ketohexokinase sequence, following his receipt of a preprint of our paper reporting it [59], commenting that it was the first mammalian protein member of the family (although the second eukaryotic protein, yeast ribokinase [60] was the first identified), and was closer to the other members than inosine-guanosine kinase from *E. coli*.

Further to these comments, he also performed an alignment of the ketohexokinase sequence with the *Lactococcus lactis* 6-phosphotagatokinase [61] protein (see *Fig. 5.11*), indicating the regions of the three sequence similarity 'boxes'. He was also able to add further evidence for the existence of an ancient family of proteins containing both the phosphofructokinase and ribokinase families of proteins [62].

Chapter 6

Expression of the Kethexokinase cDNA in the TDS Yeast Strain

6.1 Introduction

Saccharomyces cerevisiae was the first eukaryote to be successfully transformed in a way that allowed precise manipulation of the phenotype and genotype of the organism [1]. Application of these techniques may be used to introduce genes into a yeast strain, and the genes may then be maintained episomally or chromosomally.

6.1.1 Transformation Systems.

Since the first reports of the ability to introduce DNA into yeast in 1978 [2, 3], this and other techniques have been developed to obtain high transformation efficiencies.

6.1.1.1 Spheroplast transformation. Hinnen *et al.* [2] and Beggs [3] reported the first successful transformation of cloned DNA in plasmids into *S. cerevisiae*. In both cases, the cell wall was removed from the yeast cells with lytic enzymes to generate spheroplasts. These were maintained in an isotonic solution, the plasmid DNA was added, and the spheroplasts embedded in regeneration agar that only allowed the transformed cells to regenerate viable colonies. Although this method can produce up to 10^5 transformants per μg plasmid DNA, the method is complex and time consuming, is often strain-dependent, and can produce transformants of higher ploidy than the parental strain due to spheroplast fusion [4].

6.1.1.2 Lithium acetate transformation. Due to the problems associated with the spheroplast transformation method, Ito *et al.* [5] developed a method that induced a competent state for DNA uptake in yeast. This method uses a prolonged exposure of yeast cells to monovalent metal ions, especially lithium ions, before the addition of the DNA to be transformed and subsequent heat shock. By the refinement of this method, transformation efficiencies in excess of 10^5 per μg DNA have been achieved [6]. Since this method can be applied to most laboratory strains, and is relatively simple to perform, it has become the choice for transformation of many laboratories.

6.1.1.3 Electro-transformation of yeast. Recently, electro-transformation of yeast has become a realistic method for transforming yeast cells [7, 8], and transformation

efficiencies in excess of 10^7 per μg DNA have been obtained at low DNA concentrations [9, 10]. In a fashion similar to electro-transformation of *E. coli*, the cells must first be resuspended in a low ionic strength, and hence low-conductance, medium. However, it was found that efficiencies were improved immensely if the resuspension medium was osmotically supported, usually using sorbitol, and by the addition of dithiothreitol [9]. Since this method is still in its infancy, and protocols have not been optimised, but transformation efficiencies may be obtained far in excess of those for spheroplast or lithium acetate methods. Therefore, due to the rapid preparation of cells and transformation protocol, coupled with high transformation efficiencies, this method may become the routine method of yeast transformation.

6.1.2 Yeast Plasmid Expression Vectors.

For ease of manipulation, most expression plasmids are shuttle vectors, capable of replication in either *E. coli* or yeast, chromosomally or episomally in the latter organism. There are three basic criteria needed for such a vector:

- i) the ability to replicate in *E. coli* (due to the *ori* sequence from the *colE1* plasmid) to allow the manipulation/construction of the plasmid before transformation into the yeast host,
- ii) the presence of marker genes to allow selection of transformants in either host, and
- iii) a suitable expression cassette into which the gene for expression might be inserted.

The presence of the latter may be obviated if the DNA sequence to be expressed has the necessary expression sequences within it. Plasmids may be maintained episomally or chromosomally integrated in yeast, and may be present either in single or multiple copy. If the plasmid is to be maintained episomally in the yeast host cell, it should also contain a suitable yeast origin of replication. This is not required if the plasmid is to be integrated and maintained chromosomally, in which case, an integration sequence is required.

6.1.2.1 Bacterial sequences in yeast vectors. As there is no rapid and simple method at present for the isolation of large quantities of plasmid DNA directly from yeast, most manipulations and constructions are therefore carried out in *E. coli*. Consequently, most yeast plasmids are constructed around a bacterial plasmid, such as *pBR322* or the *pUC* family, and will therefore contain the necessary bacterial replication and selection sequences. The usual selection system used in bacteria is the ampicillin resistance gene, although it is possible to complement some bacterial auxotrophic mutations with yeast genes, for example the bacterial *leuB* mutation can be complemented with the yeast *LEU2* gene [11, 12]. It is sometimes desirable to introduce DNA sequences into yeast that do not contain bacterial sequences, *e.g.* in the genetic manipulation of yeast to be used in the manufacture of food and drink [1]. Since it is difficult to manipulate such plasmids directly, a new class of vectors has been designed that removes the bacterial sequence from the plasmid only after it has been introduced into the yeast

host [13]. This system exploits the 2 μ site-specific recombination system, whereby specific inverted repeats within the plasmid are recognised by the product of the 2 μ *FLP* gene [14], a recombinase, and a recombination event results in the excision of the sequence between the repeats [15]. If the bacterial sequence is situated between the repeats, this will be looped out of the plasmid, and subsequently lost from the yeast host. For the simple expression of a protein in yeast, such an elaborate excision system is not required.

6.1.2.2 Yeast transformation selection systems. To select transformed yeast cells, it is necessary that the plasmid confers a gene that permits the transformed cell to grow in conditions that the untransformed cells would not be able to do. This can be performed by the selection gene either by overcoming a genetic defect in an essential pathway of the host (auxotrophic selection), or by conferring the resistance to a toxic compound present in the growth media. The former method is most applicable to laboratory strains, where the host strain can be chosen, bred, or engineered to have an appropriate genetic lesion. The use of toxin resistance genes is more applicable for use in ill-defined strains such as polyploid brewing strains.

A variety of single gene genetic defects in yeast have been isolated to which genes have been cloned to complement the lesion. The lesions used for selection systems are generally involved in the biosynthesis of an essential metabolite, which can be supplemented by its addition to the growth medium. If that strain is then transformed with a plasmid bearing a wild-type copy of the defective gene, transformants can be selected simply by the ability to grow in the absence of that metabolite. DNA sequences that lack such a component can also be co-transformed with a plasmid bearing such a gene. The transformants are selected in the same way, and it is assumed that some of these transformants will contain both sequences. This can be subsequently shown by demonstrating the presence of the non-selectable sequence in the yeast by, for example, Southern blotting or PCR. Four genes are routinely exploited in the selection of plasmid transformation, *TRP1*, which complement the tryptophan-requiring *trp1* mutants, *HIS3*, complementing histidine-requiring *his3* mutants (e.g. *his3* Δ 200 deletion), *URA3*, complementing uracil-requiring *ura3* mutants (e.g. *ura3*-52 transposon insertion), and *LEU2*, complementing the leucine-requiring *leu2* mutants (e.g. *leu2*-3,-112 double point mutations). By using genes with deletions within their promoters, such as *leu2-d* [16], *ura3-d* and *trp1-d* [17], selection systems can be developed to select for multiple copies of the gene, and hence the plasmid, since the deletion results in low efficiencies of expression, and hence high copy number is required to complement the host gene mutation.

Selection of transformants in wild-type or undefined strains requires a plasmid-borne selectable gene that confers resistance to a toxin. In a similar fashion to bacterial plasmid selection, resistance to the aminoglycoside antibiotics hygromycin B and gentamicin can be conferred by

the bacterial hygromycin B phosphoribosyl transferase [18] and Km^r (Tn601) [19, 20] genes, respectively. An alternative to the use of antibiotic resistance is the use of copper resistance conferred by multiple copies of the *CUP1* gene [21, 22] of *S. cerevisiae*.

6.1.2.3 Yeast expression cassettes. Following the development of successful yeast transformation systems, expression of foreign genes in yeast began to be investigated [23]. Initial studies used foreign promoter sequences, and it became apparent that either the foreign transcripts were not being properly processed, *e.g.* introns were not spliced [24], or the promoter gave aberrant initiation [25] or was totally inactive [26]. Consequently, for efficient expression of foreign genes it was necessary to use native yeast promoters with cDNA sequences inserted [27]. Yeast promoter elements consist of at least three elements, which regulate the efficiency and accuracy of initiation [reviewed in 28]: upstream activation sequences (UASs), TATA elements and initiator elements. UASs [29], which show some similarities to mammalian enhancers [30], determine the activity and regulation of the promoter through the binding of transcriptional enhancers, and act at variable distances upstream from the transcriptional initiation site. UASs of some constitutively-expressed genes contain poly(dA-dT) tracts which probably activate transcription by affecting nucleosome structure. TATA elements (which have the consensus sequence TATAA) are found 40 to 120 base pairs upstream of the translational initiation site, and provide a window within which initiation can occur [31]. Most regulators are regulated to some extent, but the most powerful, glycolytic promoters are poorly regulated. However, in some cases it is desirable to separate the growth stage from the expression stage, and this can be accomplished using a tightly regulated promoter, such as the *GAL1* promoter, which require the presence of an inducer, in this case, galactose. Some regularly used yeast promoters are shown in *Table 6.1*.

The first promoters used were those of the glycolytic enzymes alcohol dehydrogenase I

Promoter	Host genotype	Strength	Regulation ^a	Reference
<i>PGK^b</i> , <i>GAP</i> , <i>TP1</i>	wt	++++	≤ 20-fold induction (<i>PGK</i>)	32, 33
<i>GAL1</i>	wt	+++ ^c	1000-fold by induction with galactose	34
<i>ADH2</i>	wt	++ ^d	100-fold repression by glucose	35
<i>PHO5</i>	wt	+ / ++	200-fold repression by P_i	36
<i>PHO5</i>	<i>pho4^{ts} pho80</i>	+	50-fold induction by shift 37°C to 24°C	37
<i>CUP1</i>	wt	+	20-fold induction by Cu^{2+}	38
<i>MFa1</i>	wt	+	Constitutive in a cells	39
<i>MFa1</i>	<i>sir3^{ts}</i>	+	10 ⁵ -induction by shift 37°C to 24°C	39

Table 6.1 Native *S. cerevisiae* promoter systems. ^aInduction ratios are reporter-dependent. ^b5% of mRNA in single copy. ^c1% of mRNA in single copy but severely limited in multiple copies. Improved by *GAL4* over-expression. ^dImproved by *ADR1* over-expression. Table taken from Romanos *et al.* [23].

(*ADH1*) [27], phosphoglycerate kinase (*PGK*) [33, 40], or glyceraldehyde 3-phosphate dehydrogenase (*GAP*) [41]. Initially, these were thought to be constitutive, but it was later found that they were induced by glucose (e.g. *PGK* is induced 20- to 30-fold by glucose, when the culture is grown with acetate as a carbon source [33]). These promoters are amongst the most powerful of *S. cerevisiae*, the *PGK* mRNA accumulates to approximately 5% of total, and although they possess relatively poor induction ratios, are extensively used for constitutive promotion. The most powerful tightly-regulated promoters of *S. cerevisiae* are those of the galactose-regulated genes *GAL1*, *GAL7* and *GAL10*, involved in the metabolism of galactose. The regulation of these genes has been well studied, and has become a key model system of eukaryotic transcriptional regulation [reviewed in 42]. Many genes are involved in the regulation of transcription, but the central interaction is between the activating *GAL4* product with the UAS, and its repression by *GAL80*, the protein it encodes binding the *GAL4* unless galactose is added. The promoters are strongly repressed by glucose, but upon removal of glucose and addition of galactose, *GAL1*, *GAL7* and *GAL10* mRNAs are rapidly induced >1000-fold to approximately 1% of the total mRNA [43]. For the purpose of expressing ketohexokinase and its consequent use in the selection of fructotransferase mutants, the *GAL* promoters are of little use since the induction of the enzyme by galactose would initially provide the cell with a carbon source, and the subsequent release of glucose from sucrose by invertase could potentially repress ketohexokinase expression. The *PGK* promoter does not suffer from this drawback, and so was selected as the expression promoter.

Yeast transcriptional terminators are usually present in expression vectors for efficient mRNA 3' end formation. Efficient termination is probably required for maximal expression, although the process is not well understood in yeast. It is probable that the mechanism is similar to the that in higher eukaryotes, where cleavage and polyadenylation occurs downstream of the signal AAUAAA [44]. A number of consensus sequences have been implicated as part of the mRNA 'terminator', especially the tripartite sequence TAG..(T-rich)..TA(T)GT..(AT-rich)..TTT [45] and TTTTATA [46]. The commonly found tripartite motif shows poor conservation and is tolerant of large sequence alterations, suggesting that a general feature such as high AT-content may be critical [47, 48].

6.1.2.4 Plasmid retention in yeast. Plasmids may be maintained in yeast cells either episomally or integrated into the chromosome. In both cases, the plasmid sequence may be present in single or multiple copy, copy-number being dependent on plasmid type. The mitotic stability of plasmid is also dependent on the sequences introduced, but in many cases, a selective pressure must be maintained to ensure plasmid retention.

Two sequences have been identified that function as replication origins in yeast, thereby maintaining plasmids that contain them episomally. The first is a 599 basepair fragment

isolated from the natural yeast plasmid, the 2 μ circle [49, 50, 51], coding inverted repeats from the *REP3* sequence [15]. This sequence has been identified as an important part of the 2 μ plasmid's partitioning and mitotic segregation machinery, allowing plasmids bearing this sequence to replicate autonomously from the yeast chromosomes [1]. Such plasmids are thought to be retained in the nucleus, and with normal selection markers (*e.g.* *URA3*, *LEU2*), exist at a copy number of 10 to 30 per cell. Copy number can be increased by use of the *leu2-d* selection marker, which requires the plasmid to be present at 100-200 copies per cell, or by using the entire 2 μ plasmid coupled to the *leu2-d* marker, which can achieve 200-400 copies per haploid cell [3]. The mitotic stability of the plasmids containing only the short 2 μ sequence is, however, relatively low in the absence of the selective pressure, with the plasmid being retained in the mother cell on division, and on non-selective medium the plasmid may be lost within several generations, the plasmid being lost in 1 to 3% of cells per generation [23]. Plasmids containing the entire 2 μ sequence, especially those capable of losing bacterial sequences by recombination, have a very high mitotic stability, with plasmid-free cells arising at a rate of 1 in 10⁴ per generation [52].

The second sequence that has been identified as being capable of functioning as a yeast origin of replication is the yeast chromosomal autonomously replicating sequence (*ARS*). This was originally isolated due to its proximity to the *TRP1* selectable marker [53], although *ARS*s have been identified on other yeast chromosomal loci, and are probably involved in initiation of chromosomal replication. *ARS* vectors are present at high copy number (2-20), but the mitotic stability of plasmids bearing this sequence are very low, even under selective conditions, and often less than 20% of a population of cells will bear the plasmid [54, 55]. Mitotic stability can be obtained by the addition of a centromeric fragment from *S. cerevisiae* chromosome IV (*CEN4*), making the plasmid mitotically stable under non-selective conditions [56]. However, mitotic stability is gained at the expense of copy number, which is reduced to 1-2 per cell [57].

A further development of such vectors is the introduction of two telomeric sequences [58], separated by a removable, second selection marker. Before transformation, the plasmid is digested to remove this second marker and leave the telomeres at the cut ends, and transformants are selected for by independence of a first, but requirement for the second marker. Such plasmids then behave essentially as normal yeast chromosomes (and thus are referred to as yeast artificial chromosomes, *YACs*) [59, 60], showing mitotic stability and single copy number, providing large enough DNA inserts are cloned into them (*YACs* as small as 55kb are mitotically stable under selective conditions) [61]. Genomic sequences over 1000 kb have been successfully cloned in this way, although fragments containing tandem repeats are highly unstable in yeast. In all cases, the copy number of the plasmids can be increased by placing the origin of replication under the control of an inducible promoter [62].

This is especially useful with YACs, facilitating the purification of the YAC DNA directly from the yeast.

Transformed sequences may also be retained within the yeast host by integrating the plasmid chromosomally [63]. Such plasmids (*Fig. 6.1a*) or plasmid fragments (*Fig. 6.1b*) do not require origins of replication, since they essentially become part of the yeast chromosome. Plasmid integration arises by recombination between a chromosomal sequence and a homologous plasmid-borne sequence. Usually, this sequence is a selectable marker gene, with the wild type allele being the plasmid-borne copy. The frequency of such integration can be increased up to 100-fold by first linearizing the plasmid within this yeast sequence [63]. This is often referred to as transformation by targetted integration (see *Fig. 6.1a*). Since the site of targetted integration is usually the auxotrophic marker, only one integration event is required to complement the allele. Indeed, it is possible to integrate the *leu2-d* auxotrophic marker into a mutant allele at single copy-number, with a resulting wild-type phenotype, as the chromosomal *LEU2* promoter becomes recombined with the wild type gene. Since the integration site remains following the recombination event (indeed, there are now two near-identical sites), it is possible for the integration event to occur again, resulting in multiple integration (2-5 copies per cell) [64]. The stability of this multiple integration may not be very high, as there is no selective pressure to maintain the extra gene copies, and these copies may be lost by a second, excision recombination [2, 64]. The integration event results in the integration of the plasmid DNA as well as the wild-type allele. Consequently, if this plasmid contains a cDNA sequence in an expression cassette, this too will be integrated.

Two approaches have been demonstrated to achieve stable multiple integration in yeast. The first relies on the presence of multiple copies of the same sequence on the yeast chromosome, and the other involves reverse-transcriptase mediated transposition into yeast chromosomes of an expression cassette, inserted within modified *Ty1* elements located on a 2 μ plasmid [65, 66]. The first of these utilizes either the δ -sequence of the yeast transposable element [67], a sequence that is randomly distributed through the yeast genome, or the rDNA locus [68], consisting of 100-200 tandem repeats existing near the centromere of chromosome XII.

The δ -sequence is present at more than 100 copies per cell, although the distribution within the chromosomes varies with strain, and integration appeared to be limited to a single chromosome, which varied with different transformants [67]. Three to five copies of the integrated sequence were detected with each integrated strain, although the *URA3* selection marker was used, requiring only a single copy to complement the chromosomal allele. Integration at the rDNA locus, selecting with the *leu2-d* marker resulted in more the 100 integrated copies per cell [68]. In this case, the rDNA sequence used to target integration contains a putative *ARS* sequence, and it is possible that upon integration, the integrated

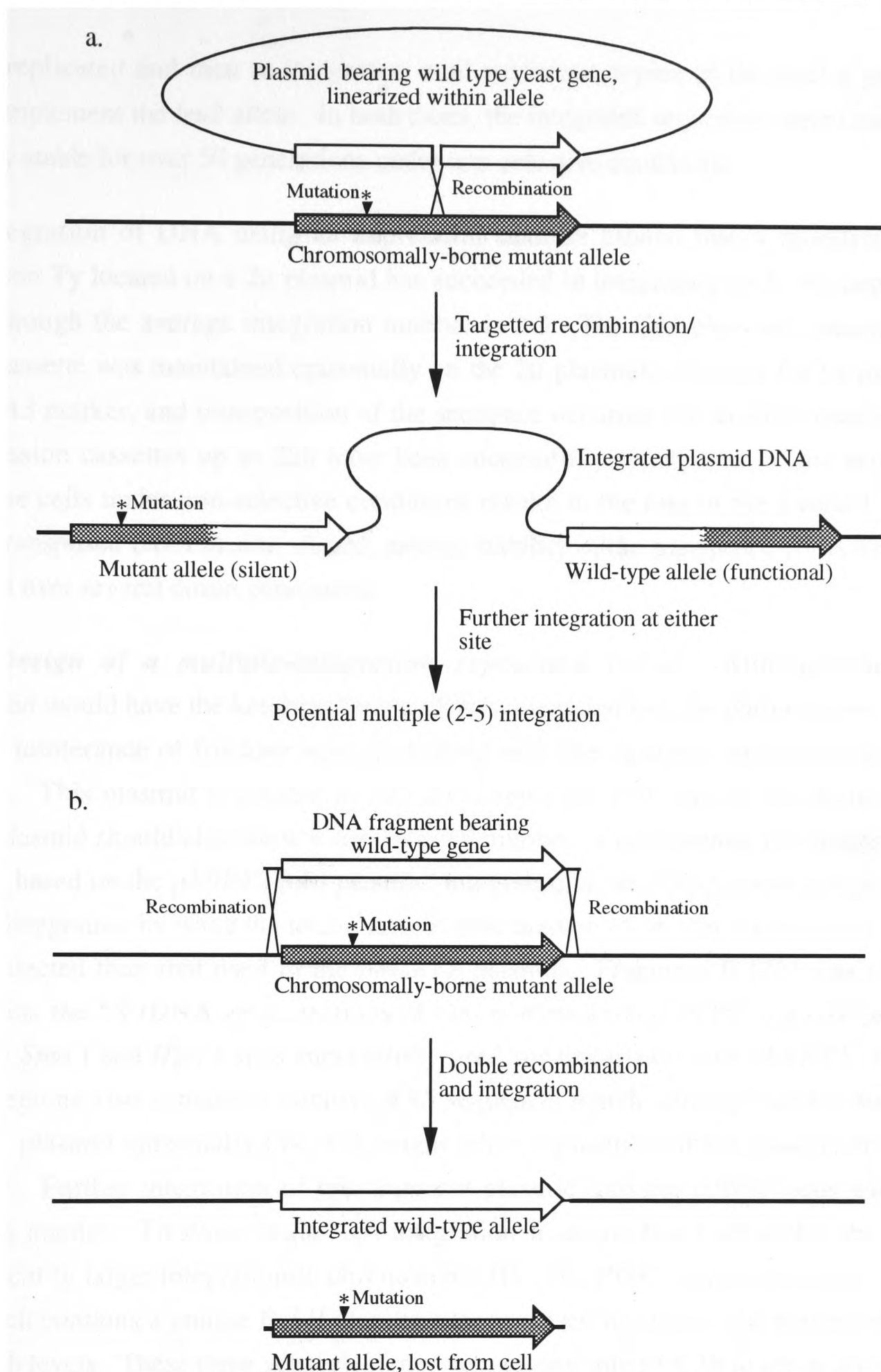


Figure 6.1 a. Integration of a linearized plasmid by homologous recombination into the yeast chromosome. A plasmid is first cut with a restriction enzyme within the sequence at which homologous recombination is targetted. This is often targetted at a mutant allele in the yeast chromosome (mutation marked as a *). The recombination event results in two copies of the gene, one wild-type, and one mutant, along with the plasmid DNA between them, being integrated into the chromosome. Further plasmid sequences can then integrate at either of the alleles, resulting in tandem multi-copy insertions. Inserted sequences can be lost, however, by a second recombination event within the chromosome, resulting in the plasmid sequence being looped-out, and subsequently lost from the cell. Following this event, the wild-type or mutant allele may remain within the chromosome.

b. Integration of a DNA fragment by a double recombination event, resulting in the replacement of the mutant allele with the wild-type gene. Such integration events lead to mitotically stable integrants, since only one copy of the DNA is present in the chromosome.

sequence is replicated and then re-integrates, until sufficient copies of the *leu2-d* gene are present to complement the *leu2* allele. In both cases, the integrated sequences were claimed to be mitotically stable for over 50 generations under non-selective conditions.

Multiple integration of DNA using an expression cassette cloned into a modified yeast retrotransposon Ty located on a 2 μ plasmid has succeeded in integrating up to ten copies per cell [65], although the average integration number is 2.3. The Ty1 element containing an expression cassette was maintained episomally on the 2 μ plasmid, selecting for its presence with the *URA3* marker, and transposition of the sequence occurred via an RNA intermediate [69]. Expression cassettes up to 2kb have been successfully transposed in this way [66]. Growth of the cells under non-selective conditions results in the loss of the plasmid, whilst leaving the transposed DNA *in situ*. Again, mitotic stability of the transposed DNA had been demonstrated over several dozen generations.

6.1.2.5 Design of a multiple-integration expression vector. Although the final selection strain would have the ketohexokinase cDNA integrated into the chromosome, initial trials to test intolerance of fructose were performed with the episomal expression plasmid *pMA91* [70]. This plasmid is present in 100-200 copies per cell, and so the design of an integration plasmid should also allow a similar copy number. Consequently the design of the plasmid was based on the *pMIRY2* [68] plasmid, integrating at the rDNA locus and selecting for multiple integration by using the *leu2-d* auxotrophic marker. A shorter region of the rDNA locus was selected than that used in the *pMIRY2* plasmid. Fragment B [71] was chosen, which contains the 5S rDNA gene, sections of two nontranscribed (NTS) regions (see Fig 6.2), and the *Sma* I and *Hpa* I sites successfully used for linearization of *pMIRY2*. One of these NTS regions also contains a putative *ARS* sequence, which although unable stably to maintain the plasmid episomally [54, 55], might allow replication of the plasmid from the chromosome. Further integration of this transient plasmid into the rDNA locus will then increase copy number. To allow single copy integration, a unique *Kpn* I site within the *leu2-d* gene can be cut to target integration to chromosome III. The *PGK* expression cassette from *pMA91*, which contains a unique *Bgl* II cloning site, was used to express the ketohexokinase cDNA at high levels. These three yeast elements were cloned into *pUC18* to allow replication in *E. coli*.

6.1.3 ³¹P-NMR of *in vivo* Yeast Cells.

In vivo NMR spectroscopy is now widely used to study cellular physiology and energetics in yeast, as well as other organisms [72]. This technique not only provides real-time information regarding the environment, localization, and levels of intracellular metabolites, but also does so non-invasively. The ability to study cellular events as they naturally occur provides *in vivo* information that cannot be obtained otherwise. Metabolites can be observed by NMR

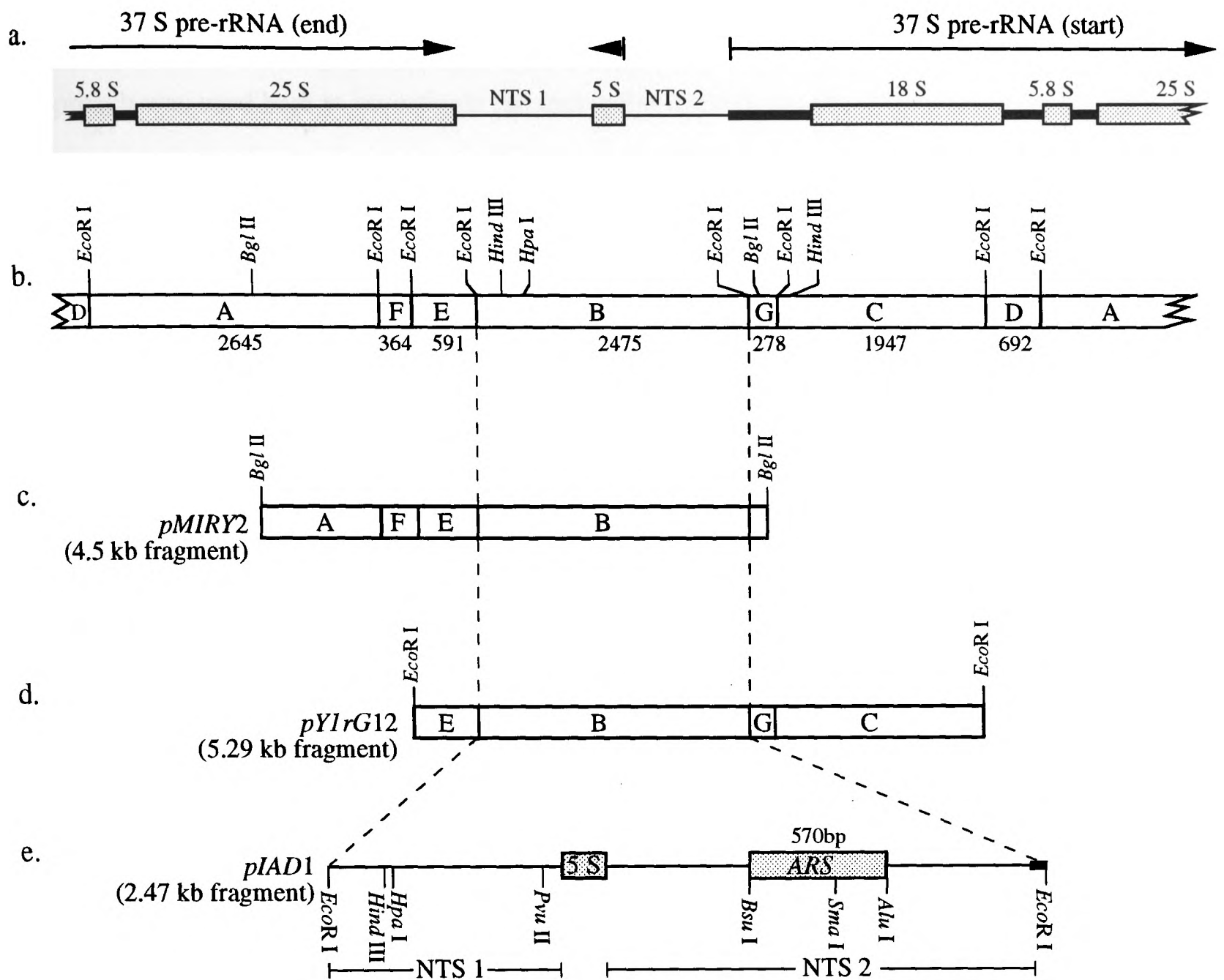


Figure 6.2 (a) A section of the tandemly-repeated rDNA locus from chromosome XII of *Saccharomyces cerevisiae* showing the rDNA genes, (b) a restriction map of the same region, showing the *EcoR* I fragments described in [71], (c) the *Bgl* II fragment used in the *pMIRY2* plasmid [68], (d) the *EcoR* I partial digest fragment in the *pYlrG12* plasmid [71] used as a source of the rDNA sequence for this work, and (e) a restriction map of the *EcoR* I-fragment B used in the construction of *pIAD1*. NTS, non-transcribed sequence; ARS, putative autonomously replicating sequence.

providing they are mobile and in sufficiently high concentrations to give a signal which is significantly higher than background noise. In yeast, phosphorous-containing metabolites such as inorganic phosphate (P_i), nucleotide tri-, di- and monophosphates (NTP, NDP and NMP), polyphosphates, sugar phosphates, nicotinamide adenine dinucleotide phosphate (NAD), and uridine diphosphoglucose (UDPG) can be measured by ^{31}P NMR [73]. Distinct signals corresponding to various ^{31}P -containing metabolites can often be resolved because nuclei in different chemical environments experience different fields and consequently produce signals at different shifts or frequencies. ^{31}P NMR can also be used to monitor intracellular pH, since the chemical shifts of several phosphorylated metabolites are sensitive to pH in the physiological range [74, 75]. *In vivo* ^{31}P -NMR spectroscopy provided a convenient non-invasive method to identify metabolic pools from which orthophosphate was drawn and

sequestered as fructose 1-phosphate in the study of the hepatic metabolism of fructose administered to healthy [76] and fructose intolerant [77] human subjects, and the same approach was used here to investigate the metabolic impairment of a fructose intolerant yeast.

6.2 *Materials and Methods*

6.2.1 *Yeast Plasmids.*

Two plasmids, *pMA91* [70] and *pYIrG12* [71], were used in the construction of vectors to express ketohexokinase. The former was used as a high copy number yeast expression vector, and was also used in the construction of the high copy of integration yeast expression vector (*pIAD1*), providing the *PGK* promoter/3' regions and the *leu2-d* auxotrophic marker. *pMA91* consists of the bacterial plasmid *pBR322*, allowing selection for ampicillin resistance in *E. coli*, along with three yeast elements, the 2 μ origin of replication, the *leu2-d* auxotrophic marker, and the yeast *PGK* promoter and 3' region, with a unique *Bgl* II cloning site for the expression of heterologous genes. The *pYIrG12* plasmid comprises a 5.29kb *EcoR* I partial digest fragment of the 9.1kb rDNA repeating unit in the bacterial *pBM9* plasmid, and was used as a source of the rDNA sequence for *pIAD1*.

The *pYIrG12* plasmid was also used to provide a region of rDNA sequence, unique to chromosome XII, to probe Southern blots of pulsed-field gel electrophoresis gels of yeast chromosomes. A third plasmid, *pMA1* [70], from which *pMA91* was constructed, was used in a similar fashion to provide a DNA fragment containing the *PGK* gene coding region, as a probe for chromosome III.

6.2.2 *Transformation of Yeast with Circular Plasmid DNA.*

Since methods for transformation of yeast are strain dependent, three methods were attempted to determine which was the most efficient.

6.2.2.1 *Transformation using lithium chloride.* A cascade of yeast cells is grown overnight in YEPD by incubating 35ml. from a plate, and transferring 10ml. to two subsequent tubes (in 50ml. Falcon tubes) containing 25ml. medium. A final dilution is then performed into 15ml. The cells are then harvested (3000 rpm, 5 minutes) and cells with a volume of ~0.5ml. are washed in 25ml. LiCl/TE

10mM. Tris.HCl, pH 8.0
1mM. EDTA
100mM. Lithium chloride

and resuspended in 0.5ml. the same [5]. This was incubated for one hour at 37°C with shaking and then placed on ice for at least 1 hour. 0.1ml. aliquots of this were then placed in an Eppendorf along with 25µl. sonicated salmon sperm DNA (25µg/ml.), 1µg (in 10-15ml.) plasmid DNA and 10µl. 100% ethanol. After incubating at 30°C for 30 minutes, 1ml. 40% w/v PEG-4000 in LiCl/TE was added, mixed well, and left at 30°C for a further 60 minutes. The cells were then heat-shocked at 42°C for 5 minutes, sedimented in a microfuge for 30 seconds, resuspended in 1ml. YEPD and incubated with shaking for 6 hours at 30°C. This step is only necessary when transforming with a *leu2-d*-based plasmid, since plasmid replication was necessary before the *leu2-d* is of sufficiently high copy number to complement the chromosomal *leu2* allele. The cells were then resuspended in 100µl. water before plating on selective SD media. The plates were then incubated for 3-5 days at 30°C.

6.2.2.2 Transformation of spheroplasts. Cells were grown overnight in YEPD to a concentration of 2×10^7 cells/ml., as determined using a hemocytometer. Each transformation requires 10ml. cell suspension, as described. The cells were washed with 10ml. 1M. sorbitol (spin for 6 minutes at 3,000 rpm), and resuspend in 1ml. of the same. Glusulase [NEN, DuPont] was then added to a final concentration of 3% v/v, and this was incubated for 1-2 hours at 30°C with gentle shaking. Formation of spheroplasts was confirmed by observing cell lysis under the light microscope following the addition of 10µl. 10% w/v SDS to a 50µl. aliquot of the cells. The spheroplasts were washed 3 times with 1M. sorbitol, and then in 10ml. STC.

1M. Sorbitol
10mM. Calcium chloride
10mM. Tris.HCl, pH 7.5

The spheroplasts were finally resuspended in 100µl. STC before the addition of 0.5µg plasmid DNA and 1µg carrier (sonicated salmon sperm) DNA. After incubating for 20 to 30 minutes at room temperature, ten volumes of PEG solution

20% w/v PEG-4000
10mM. Calcium chloride
10mM. Tris.HCl, pH 7.5

were added and the mixture incubated for a further 10 minutes. The cells were gently sedimented by low speed centrifugation and resuspended in 1ml. sorbitol before adding to 3ml. regeneration agar

1% w/v Glucose
2.5× Amino acid mix as required
1% v/v YEPD media
0.7% w/v Yeast nitrogen base w/o amino acids
2% w/v Ammonium sulphate
1M. Sorbitol
3% w/v Agar

and poured on to SD plates with amino acids as necessary. The plates were then incubated at 30°C and transformed cells were rescued from within the agar.

6.2.2.3 *Electro-transformation of yeast.* A 500ml. YEPD overnight yeast culture was grown with shaking at 30°C to give an A₆₀₀ of 1.3-1.5. This was centrifuged at 5000 rpm for 5 minutes at 4°C and the cells washed first in 500ml. and then 250 ml. water and finally in 20ml. 1M. sorbitol, each time centrifuging as before. The cells were finally resuspended in 0.5ml. 1M. sorbitol. 40μl. of this suspension was used for each transformation by adding ≤100ng plasmid DNA, transferring to a cold electroporation cuvette (0.2cm electrode gap), and pulsing at 1.5kV, 25μF, 200Ω. Immediately following the pulse, 1ml. cold YEPD/270mM. sorbitol was added, and the mixture transferred to a 15ml. polypropylene snap-cap tube. This was incubated for 1 hour (or 6 hours for a *leu2-d*-based plasmid) with shaking at 30°C, before plating on selective SD plates containing 1M. sorbitol.

6.2.3 *Transformation of Yeast with Linear DNA for Integration.*

To aid homologous recombination into the yeast chromosome, plasmid DNA was first cut at the desired site of recombination before transformation. This was achieved by restriction endonuclease digestion (see 4.2.7.2), incubated overnight to ensure complete digestion, prior to phenol/chloroform and chloroform extractions. This was then used to transform the yeast, using the lithium chloride method, and the transformed cells were incubated on selective medium for up to 2 weeks before transformants were observed. Since the integration plasmid cannot survive extra-chromosomally, no pre-incubation of the transformants in YEPD was performed when transforming with *leu2-d*-based plasmids.

6.2.4 *Miniprep of Yeast Plasmid DNA.*

Cells were grown in SD media, supplemented with amino acids as appropriate, to an A₆₀₀ of 0.6-0.7. The cells were then centrifuged, resuspended in 0.5ml. 1M. sorbitol, 20mM. EDTA, and transferred to an Eppendorf tube. 20μl. 2.5mg/ml. Zymolyase-60,000 [ICN Flow] was added and the mixture incubated at 37°C for 60 minutes, before centrifuging for 1 minute and resuspending in 0.5ml. 50mM. Tris.HCl, pH 7.4, 20mM. EDTA. 50μl. 10% w/v SDS was then added, mixed well, and incubated for 30 minutes at 65°C. To precipitate the SDS, membranes and proteins, 200μl. 5M. potassium acetate were added, and the mixture was left on ice for 60 minutes before centrifuging for 5 minutes. The supernatant was transferred to a fresh tube and 1 volume isopropanol added. After a 5 minute incubation at room temperature, this was centrifuged very briefly, the supernatant poured off, and the pellet allowed to air-dry. The pellet was dissolved in 300μl. TE, 10μl. 20μg/ml. RNAase A added, and incubated at 37°C for 30 minutes. 30μl. 3M. sodium acetate was added, and the mixture extracted with

phenol, phenol/chloroform and finally chloroform before adding 200 μ l. isopropanol. The DNA was then sedimented by centrifugation, and redissolved in 100 μ l. TE.

6.2.5 Pulsed-Field Gel Electrophoresis (PFGE) of Yeast Chromosomal DNA.

PFGE was performed essentially as described in Southern *et al* [78].

6.2.5.1 Preparation of chromosomal DNA in agarose plugs. A 10ml. saturated yeast culture was grown in selective SD medium at 30°C. Cells were sedimented by centrifugation at 1500 rpm for 10 minutes, washed in 10ml. 50mM. EDTA, pH 8.0, centrifuged again and resuspended in 400 μ l. plug prep buffer

1M. Sorbitol
20mM. EDTA, pH 8.0, added prior to use
14mM. β -Mercaptoethanol (freshly added)
1mg/ml. Zymolyase 20T [ICN Flow].

500 μ l. of 1% w/v LGT agarose [Sea Plaque, ICN Biomedicals] in 1M. sorbitol, 20mM. EDTA, pH 8.0, which has been melted and cooled to 37°C, was added and mixture was dispensed into Perspex plug molds (6mm $\times$ 1.7mm $\times$ 10mm) and allowed to set on ice for 10 min. Plugs were then pushed into 5ml. of plug prep buffer and after incubation at 37°C for 2 hours, the solution was replaced with 5ml. fresh yeast lysis solution

100mM. EDTA, pH 8.0
1% w/v Lithium dodecyl sulphate

and incubated at 50°C for 1 hour (plugs change appearance from opaque to translucent after a few minutes). This was replaced with 5ml. fresh solution and incubated overnight at 50°C. Sterile plugs can be stored at room temperature for several months. The lithium dodecyl sulphate precipitates at low temperatures.

6.2.5.2 Equilibration of pH prior to PFGE. For pulsed-field gel electrophoresis, plugs were soaked in 20ml. sterile TE, pH 7.4, in Falcon tubes at 4°C for several hours with intermittent mixing by inversion. The TE was then replaced with a fresh sample and plugs incubated for another 30 minutes when they were loaded on the PFGE gel. Appropriate concentrations of gel were prepared in $1/2$ TAE buffer which was also used as a running buffer. After soaking, gel plugs were loaded in to gel wells and covered with molten 0.5% w/v agarose. PFGE was run between 25-50V for 1% and 1.5% w/v gels, respectively, pulse time varied between 4 minutes and 50 seconds, respectively, and the running time between 50 and 27 hours. Gels were run with bacteriophage lambda DNA 48.5 oligomers (concatemers) [Bio-lab], and a standard yeast strain prepared using the same procedure as size markers. Gels were stained for 1-2 hours with ethidium bromide (0.5 μ g/ml., prepared from 10mg/ml. stock solution) and destained in distilled water.

6.2.5.3 Blotting and probing of chromosomal DNA. As PFGE gels contain large DNA molecules, it is essential to break up the DNA before transfer. Acid depurination (150mM. hydrochloric acid, 15 minutes), as opposed to UV irradiation, was used to achieve DNA breakage. Electrophoretic samples in an agarose gel were Southern blotted (see 5.2.2.1) and probed (see 5.2.2.3) with radioactively-labelled DNA probes (see 5.2.2.2).

6.2.6 Determination of Fructose Intolerance in Transformed Yeast.

6.2.6.1 Preliminary screens on Petri dishes. Transformants were screened for fructose intolerance by replica plating colonies on to SD leu- plates (containing all amino acids except leucine) and then SD leu- plates containing 0.5% w/v fructose (SD^{1/2}F). These were incubated overnight and colony growth compared. During sub-culturing of strains, this check was routinely performed to ensure the maintenance of the ketohexokinase gene.

6.2.6.2 Quantification of intolerance of fructose. To identify the fructose concentration which results in intolerance of fructose, cells were exposed to various fructose concentrations. An overnight culture of yeast in SD leu- was separated into six 50ml. Falcon tubes (30ml. each), and the growth monitored for several hours by removal of five 1ml. aliquots, and determination of their A₆₀₀. After 2.5 hours, fructose (from 13% w/v stock, filter sterilised) was added to give 0.5, 0.25, 0.1, 0.05 and 0.025% w/v final concentration, as well as a water control. Growth was monitored for up to 10 hours to observe the effects.

6.2.6.3 Recovery of cells from exposure to fructose. Cells exposed to fructose in the above experiment were sedimented after 3 hours exposure and washed in water. The cells were then resuspended in their original volume of SD leu-, without the fructose, and their growth again monitored. The cultures were then allowed to grow overnight and 20µl. aliquots were spread on SD leu- and SD^{1/2}F leu- plates. Cell densities on each plate were then visually compared.

6.2.6.4 Sensitivity of cells to sucrose. Since a fructose-intolerant^{strain} developed was to be used in the selection of mutant invertases capable of transferring the fructose moiety from sucrose to an acceptor substrate, the cells were tested for their tolerance of sucrose, in the absence of a cloned *SUC2* gene. SD leu- plates had 0.3ml. 20% w/v filter-sterilised Aristar sucrose spread over them, and were allowed to dry immediately prior to their use. Fructose intolerant cells were replica plated on to these, and their growth compared with cells plated on SD leu- plates. The spreading of sucrose helps prevent thermal- or acid-hydrolysis of the sucrose, that could occur if the sucrose were added prior to autoclaving of the media.

6.2.7 *Partial Purification of Ketohexokinase from Yeast.*

Cells were grown overnight in 500ml. SD leu- containing 0.2% w/v asparagine and 0.05% w/v potassium dihydrogen orthophosphate ('Enhanced synthetic dextrose', ESD) at 30°C with shaking. The cells were then harvested by centrifugation, washed with water, and their wet weight determined. The cells were then resuspended (1g wet weight/2ml.) in ice cold lysis buffer [79]

0.1 M. Sodium chloride
0.05M. Tris.HCl, pH 8.0
1% v/v Dimethyl sulphoxide
1mM. Phenylmethylsulphonyl fluoride
1mM. EDTA
1mM. EGTA
1mM. Benzamidine
1.5µg/ml. Leupeptin
3µg/ml. Pepstatin A

in a snap-cap polypropylene 15ml. Falcon tube. A volume, equal to that of the buffer added, of acid-washed glass beads were added, and the tube vortexed vigorously three times for one minute each, cooling on ice for 1 minute between each vortexing. The beads were allowed to settle and the lysed cells and the supernatant removed. The beads were washed twice in a small volume of lysis buffer, and this was pooled with the other sample. The lysate was then clarified by centrifugation at 100,000g_{av} for 1 hour at 4°C, and the supernatant collected.

A crude purification of the enzyme, to allow a simple characterization of the expressed ketohexokinase, was performed using similar steps to those in the purification of the rat enzyme. Proteins were selectively precipitated by addition of solid ammonium sulphate to 40% and 50% saturation, the enzyme being present in the 50% precipitate. This was redissolved in cell lysis buffer and heated for 1 minute at 65°C. The supernatant was then applied to a calibrated 53cm×1.7cm² S100 column, and eluted with 10mM. triethanolamine.HCl, adjusted to pH 7.2 with potassium hydroxide and 2.5ml. fractions collected. Eluted fractions were tested for ketohexokinase activity using the colorimetric assay (see 3.2.4.2), and the pooled sample, and other fractions assayed using the quantitative spectrophotometric method (see 3.2.4.1). Partially purified fractions were also analysed by SDS-PAGE (see 3.2.5).

6.2.8 *Enzymatic Identification of Fructose 1-Phosphate in Transformed Yeast.*

6.2.8.1 *Perchloric acid extraction of cell metabolites* [80, 81]. Cells were grown in selective ESD overnight before addition of fructose to 0.5% w/v final concentration, or water as a control, for three hours. The cells were then sedimented by centrifugation, washed with water and resuspended in a volume of water equal to their weight. This was

extruded though a syringe needle directly into liquid nitrogen in a pestle cooled with the same, and ground into a fine powder. A quarter volume of 3*M.* perchloric acid was added, and the mixture was again ground to a fine powder under liquid nitrogen. This was subjected to three freeze-thaw cycles, freezing in liquid nitrogen, and the precipitated proteins removed by centrifugation at 3000 rpm at 4°C for 10 minutes. EDTA to 5*mM.* was added to the supernatant, before adjusting the *pH* to 5.5-6.0 with 1*M.* potassium carbonate. The mixture was stirred for 20 minutes at 4°C before centrifuging as before, to remove the precipitated potassium perchlorate. The supernatant was passed though a 10*ml.* Chelex column to remove paramagnetic ions, before being lyophilized. Before assaying for fructose 1-phosphate, this was dissolved in 2*ml.* triethanolamine.KOH, *pH* 7.3, and left on ice, excess being stable for several months at -20°C. The Chelex column was regenerated by passage of 20*ml.* 1*M.* sodium chloride, followed by washing with 100*ml.* distilled water.

6.2.8.2 Enzymatic analysis of fructose 1-phosphate. Fructose 1-phosphate is assayed with a coupled reaction between aldolase B and glyceraldehyde 3-phosphate dehydrogenase (GAPDH), resulting in a stoichiometric oxidation of NADH [82]. Aldolase A and triosephosphate isomerase, and lactate dehydrogenase (LDH) are also added prior to aldolase B addition to remove fructose 1,6-bisphosphate and pyruvate, respectively, which would interfere with the assay. Since aldolase B is not available commercially, it first had to be purified from rat livers [82].

A rat was killed by cervical dislocation and exsanguinated by severing the carotid arteries, and the liver slowly perfused with 10*ml.* 150*mM.* potassium chloride, through the hepatic portal vein, to wash out remaining blood. The liver was homogenised in 4 volumes 150*mM.* potassium chloride at 4°C and centrifuged for 20 minutes at 30,000*g* at 2°C. Solid ammonium sulphate was added to 45% saturation, the *pH* adjusted to 7.8 with 2*M.* sodium hydroxide, and this was stirred for 60-90 minutes at 4°C. After centrifugation, more ammonium sulphate was added to the supernatant to 55% saturation, and after repeating the stirring and centrifugation, the pellet was dissolved in water (0.3*ml./g* tissue). The protein was dialysed overnight against water at 4°C, centrifuged, and EDTA added to 2.5*mM.* This was left to stand for 60 minutes at room temperature, and stored at -70°C until use.

To assay for fructose 1-phosphate, 3*ml.* reaction mixture

50 <i>mM.</i>	Tris.HCl, <i>pH</i> 7.4
70 <i>μM.</i>	NADH
20 U/ <i>ml.</i>	Lactate dehydrogenase
0.5 U/ <i>ml.</i>	Aldolase A
0.4 U/ <i>ml.</i>	GAPDH
0.5 U/ <i>ml.</i>	Triosephosphate isomerase
1.15 <i>ml.</i>	Fructose 1-phosphate sample
0.2 <i>ml.</i>	Aldolase B, added after 5 minutes

was incubated in a spectrophotometer at room temperature, and the diminution of A_{340} measured 10 minutes after the addition of aldolase B. Measuring against a reagent blank, the fructose 1-phosphate concentration was determined by $(A_{340} \text{ before AldB} - A_{340} \text{ after AldB}) \times 0.414 \text{ mM.}$, to produce a value from which the cellular concentration could be calculated.

6.2.9 Metabolic Studies Using ^{31}P NMR.

6.2.9.1 NMR of intact cells. Cells were grown aerobically at 30°C on Burkholder's defined media (see 2.2.2.6) with low phosphate (30mg per litre) and an amino acid mixture lacking leucine. The main cultures (500ml. in a 2l. Erlenmeyer flask) were inoculated with 1ml. samples of log-phase precultures and harvested 48-72 hours later, when the cells were in late log-phase, by low speed centrifugation. The cells were washed in 30ml. ice-cold water and then in ice-cold substrate-free MES buffer.

25mM. MES, pH 6.0
2mM. Magnesium sulphate
2mM. Potassium chloride
1.7mM. Sodium chloride

The cells were immobilized in agarose threads as described by Brindle and Krikler [83]. Yeast cells (0.5g) were mixed with 0.5ml. of 1.8% w/v LMP agarose [Gibco-BRL] dissolved in substrate-free MES buffer at 37°C. The agarose-yeast mixture was extruded through a chilled intravenous cannula, 0.5mm internal diameter [Portex, ref. 200/300/030] into ice-cold substrate-free MES buffer in a 20mm NMR tube. The agarose-yeast matrix was perfused with the same buffer except the 50mM. glucose was added, as described by Brindle and Krikler. The effect of fructose was monitored by its addition to the perfusion to final concentrations of 1, 5 or 25mM.

^{31}P -NMR spectra were recorded on a Bruker CXP300 spectrometer, operating at a phosphorus-31 frequency of 121.49 MHz and using a double-tuned $^{13}\text{C}/^{31}\text{P}$ probe head. All spectra were acquired with a spectral window of 8064 Hz. As fully relaxed spectra were not required, spectra were recorded using a 90° pulse angle and a 1.0 second recycle time. The spectra were acquired sequentially over the time course in files of 900 scans (equivalent to 15 minutes); each free induction decay contained 1K data points. Chemical shifts were measured relative to the signal from a capillary containing a 2% v/v aqueous solution of TEMDP, which is at 22.49 ppm relative to the signal from 85% v/v phosphoric acid. The TEMDP resonance was ^1H -decoupled in order to improve the lineshape. Spectral peaks were assigned by comparison with previously published spectra [81].

6.2.9.2 NMR of perchloric acid extracts from yeast cells. After the required NMR data had been collected, perchloric acid extracts of the yeast-agarose threads were prepared by the method described above (see 6.2.7.1). The lyophilized extracts were then

dissolved in 3ml. 50mM. TEA, pH 8.0, 5mM. EDTA. ^1H -decoupled ^{31}P -NMR spectra of the perchloric acid extracts were recorded in files of 1024 scans, each containing 8K data points. A 90° pulse angle and 4 second recycle time were used. The fructose 1-phosphate resonance was assigned by the addition of an internal standard of 2mM. fructose 1-phosphate [Sigma] dissolved in 50mM. TEA pH 8.0, 5mM. EDTA to the extracts, and also by observing the disappearance of the resonance on treatment of the extract with aldolase B (see 6.2.7.2), coupled to GPDH [Sigma] with NADH to drive the reaction towards fructose 1-phosphate degradation.

6.3 Results and Discussion

6.3.1 Heterologous Expression of Ketohexokinase on Episomally-Borne Plasmids in Yeast.

6.3.1.1 Cloning of the ketohexokinase coding region into pMA91. Rather than employing the cloned cDNA of ketohexokinase containing substantial regions of 3'- and 5'-untranslated sequence for the expression work, it was decided to amplify the region of the cDNA containing the coding region of the cDNA, using PCR primers incorporating the yeast 5'-untranslated consensus sequence, and *Bgl* II linkers at either end to facilitate ligation into the PGK expression cassette. The primers for PCR amplification were designed as 32-mers, containing 14 bases either upstream of the stop codon, or 14 bases downstream of the ATG start codon. The sequences of the primers are shown in Table 6.2.

Yeast 5' consensus sequence	5' (A/T) A(A/T) A(A/C) A <u>ATG</u> ^{3'}
<i>Bgl</i> II site	5' A↓GATCT ^{3'}
KHK 5' upstream sequence	5' TTG GGG AGC AGC CTC <u>ATG</u> GAA GAG AAG CAG ATC ^{3'}
PCR 5' primer	5' TTG GaG AtC taC aca <u>ATG</u> GAA GAG AGG CAG AT 3'
KHK 3' downstream sequence	5' TTT GAT GGC ATT GTG <u>TGA</u> GAG ATG AGC GGT GGG ^{3'}
PCR 3' primer	3' AA CTA CCG TAA CAC <u>ACT</u> CTC TAg aCG CCA CCC ^{5'}

Table 6.2 Design of PCR primers for the amplification of the ketohexokinase cDNA coding region. The yeast consensus sequence is shown with its sequence variance, and the cleavage site of the *Bgl* II site is shown by an arrow. The start and stop codons are shown underlined, and in the PCR sequences, the *Bgl* II sites are shown italicised, the consensus sequence in bold, and the mismatched base pairs in lower case. Note that the 3' primer is shown as the complementary sequence to the 3' downstream sequence.

The conditions used for PCR were as described in section 4.2.9.3, using 1mM. magnesium chloride, the 911 base pairs product was blunt-ended (4.2.7.5) and ligated into the Klenow-treated blunt-ended *Sph* I site of *pUC18*. A 693 base pairs *Mlu* I/*Sph* I fragment was excised from the centre of the cloned PCR product, and replaced with an equivalent fragment from *pUC* KHK. The resulting clone was named *pUC* KHK(PCR). In order to check for any residual PCR errors, the nucleotide sequence of the remaining PCR-generated flanking segments were confirmed using primers 1 and 4 (see Table 4.3) in the sequencing reactions.

The coding region for ketohexokinase was excised from *pUC* KHK(PCR) by *Bgl* II digestion and ligated into the *Bgl* II cloning site of the yeast expression vector *pMA91*. Clones with the KHK coding region in the correct orientation were termed *pMA91*-KHK, and those in the opposite orientation, *pMA91*-KHK(-).

6.3.1.2 Transformation of the yeast TDS II. No success was obtained for the transformation of the TDS II strain, with any of the *pMA91* plasmids of *pRS315* and 316, using either the spheroplast (6.2.2.2) or electro- transformation (6.2.2.3) methods, although transformation of the yeast was obtained at low levels with the lithium chloride method. Fifteen transformants were obtained per μ g with *pMA91*-KHK and 56 per μ g with *pMA91*-KHK(-). Control experiments using the *ARS/CEN* plasmids *pRS315* (*LEU2*) and *pRS316* (*URA3*) [84] gave c.1800 and 0 transformants/ μ g, respectively, indicating relatively efficient transformation. The discrepancies in the transformation efficiencies is probably due to the requirement of the *pMA91* based plasmids to be present at high copy number to allow the plasmid-borne *leu2-d* auxotrophic marker to complement the *leu2* mutation. The *pRS315* plasmid is a single-copy plasmid, and will therefore transform more efficiently. PCR were performed on yeast minipreps of all of the *pMA91*-KHK, and three of the *pMA91*-KHK(-) transformants using the same conditions and primers as 6.3.1.1. The presence of the ketohexokinase coding region was confirmed in all cases. Transformants bearing the plasmid *pMA91*-KHK were termed TDB I, and those bearing *pMA91*-KHK(-), TDB I(-).

6.3.1.3 Heterologous expression of ketohexokinase. The presence of the heterologously^{expressed} ketohexokinase in yeast was shown by making a crude extract of the transformed cells and assaying for activity. The protein was also visualized on SDS-PAGE. The colorimetric assay (3.2.4.2) revealed the presence of the enzyme in the TDB I strain, but not in either the parent TDS II strain or the TDB I(-) strain (see Fig. 6.3). Crude extracts from TDB I contained ketohexokinase with a specific activity of 0.62 μ mol/min/mg, which is almost nine times larger than that found in rat liver (0.07 μ mol/min/mg, see Table 3.5), as determined using the spectrophotometric assay (3.2.4.1). Control assays showed a minor ATPase activity in TDB I (0.069 μ mol/min/mg) which was absent from the parent TDS II strain. Curiously, in TDB I(-) this ATPase assumed a much higher activity (0.214 μ mol/min/mg).

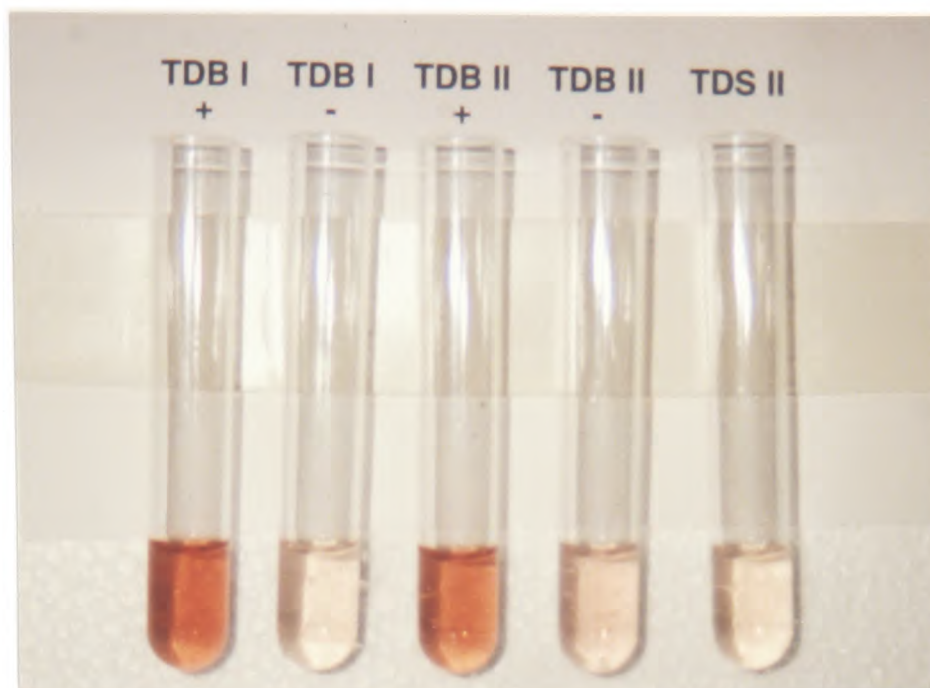


Figure 6.3 Photograph demonstrating the presence of active ketohehexokinase in extracts of yeast strains transformed with the ketohehexokinase cDNA correctly orientated in the *PGK* expression cassette. The '+' and '-' indicate this orientation ('+' being the correct orientation). The parent TDS II strain was not transformed, TDB I strains were transformed with the *pMA91*-based plasmids (6.3.1.2), and TDB II strains with the *pIAD1*-based plasmids (6.3.2.2). Extracts were assayed using the colorimetric assay (3.2.4.2), the red coloration indicating activity.

Due to co-migration of a large band of a yeast protein in the crude extracts, partial purification of positive and control extracts were necessary before a band corresponding to ketohehexokinase could be identified by SDS-PAGE. *Figure 6.4* identifies a 31/32 kDa band in the TDB I strain (and the TDB II strain, see 6.3.2.5), that corresponds to that for the rat ketohehexokinase standard. Application of the partially purified protein to a calibrated Sephacryl S100 gel filtration column revealed that, like the rat enzyme, the native protein has a molecular mass of c.130 kDa, consistent with its quaternary structure being a tetramer (see *Fig. 6.5*).

6.3.1.4 Determination of fructose intolerance. TDB I cells were demonstrated to be intolerant of the presence of fructose on SD^{1/2}F plates (see *Fig. 6.6a*), whereas neither the parent TDS II or TDB I(-) strains showed this intolerance, and grew well on this medium. That the ketohehexokinase gene was borne on the plasmid, and not integrated into the yeast genome, was confirmed by passage of the transformants through several generations under non-selective conditions (growing on YEPD medium), when both leucine independence and intolerance of fructose were lost in the TDB I strains. Intolerance of fructose in the TDB I strain was observed in liquid cultures at fructose concentrations as low as 0.025% w/v (see *Fig. 6.6b*), growth being arrested within ninety minutes. TDB I(-) and TDS II again were unaffected by the addition of fructose to their growth media. Interestingly, when the fructose treated TDB I cells were washed and resuspended in media lacking fructose, growth was shown to resume, and the cells retained their intolerance of fructose. The effect of fructose would appear not to kill the cells, but to impose a metabolic paralysis (See *Fig. 6.7*).

6.3.1.5 Enzymatic identification of fructose 1-phosphate in TDB I cells. Metabolite extracts by perchloric acid treatment of TDB I cells exposed to 0.5% w/v fructose in

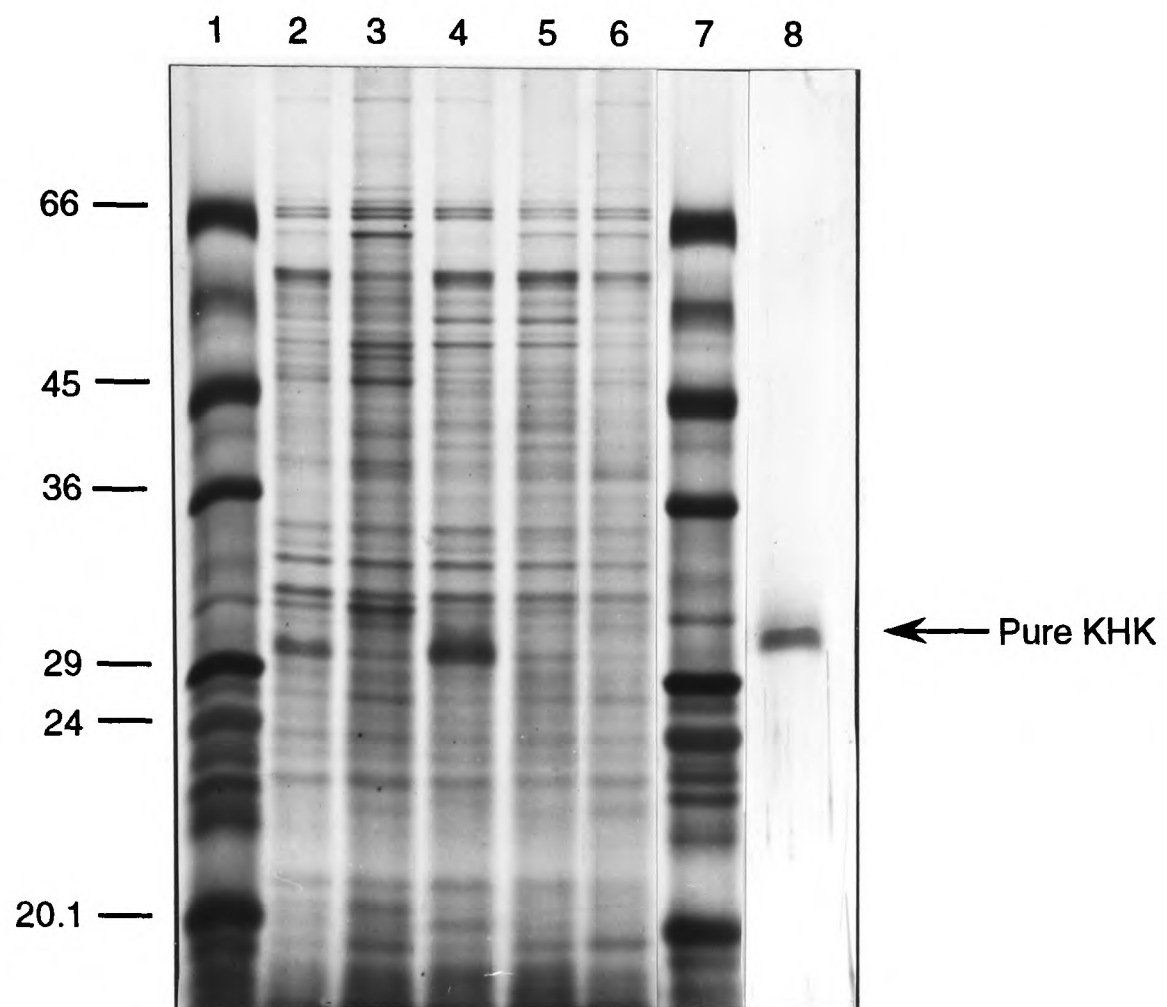


Figure 6.4 SDS-PAGE (12.5% w/v polyacrylamide) gel of partially purified yeast extracts. Lanes 1 and 7 were loaded with 3.3 μ g each of bovine albumin (66 kDa), egg albumin (45 kDa), GPDH (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (24 kDa), trypsin inhibitor (20.1 kDa) and α -lactalbumin (14.2 kDa). Lane 2 was loaded with a TDB II extract, Lane 3 contained a TDB II(-) extract, Lane 4 contained a TDB I extract, Lane 5 contained a TDB I(-) extract, Lane 6 contained a TDS II extract and Lane 8 contained approx. 60 μ g purified rat ketohexokinase. The arrow indicates the position of the ketohexokinase protein.

ESD medium for three hours showed the presence of fructose 1-phosphate at 30 μ moles/g wet weight cells. Fructose 1-phosphate was not detectable in TDB I(-) or TDS II cell extracts.

6.3.1.6 ^{31}P NMR studies of transformed TDB I strains. An external concentration of fructose as low as 1mM (c. 0.025% w/v) caused almost complete inhibition of growth within 90 minutes, and at 25mM (c. 0.5% w/v) growth ceased immediately after the addition of fructose (see *Fig. 6.6a*). ^{31}P NMR allowed simultaneous monitoring of the changes in the pools of a number of important phosphorylated metabolites. *Figure 6.8* shows the ^{31}P NMR spectra recorded from the fructose-intolerant strain during a control experiment in which glucose was the only carbon source. The spectra were similar to those of untransformed yeast cells [72, 83] and the only unusual feature was the relatively large size of the unassigned phosphodiester signal at -1.25 ppm. The cytoplasmic orthophosphate (P_i) and nucleoside triphosphate (NTP) signals increased slightly over the time-courses of 2-3 hours but this drift in intensity was not sufficient to complicate the interpretation of the fructose loading experiments described below. The polyphosphate signal at -22.6 ppm decreased during the control experiments, although it was also observed to increase on one occasion. The

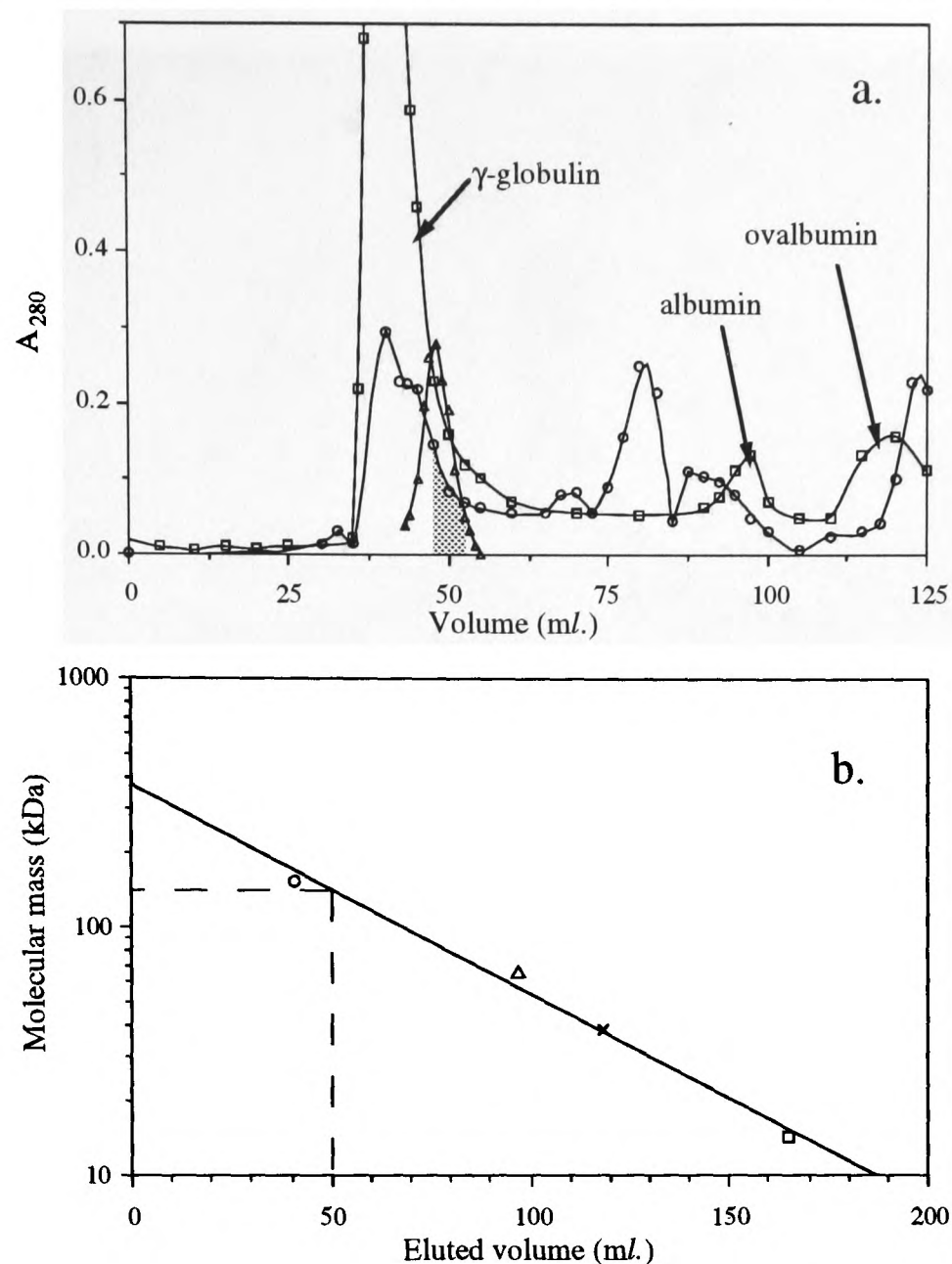


Figure 6.5 (a) Protein elution profiles from an S100 column, of protein standards, as indicated ($\square$) and a crude extract of TDB I ($\circ$). The elution of ketohehexokinase activity of the native rat enzyme (Δ) and of TDB I (shaded) are shown. (b) A semi-log plot of the protein standards γ -globulin (155 kDa, $\circ$), serum albumin (67 kDa, Δ), ovalbumin (38 kDa, $\times$) and cytochrome c (14 kDa, $\square$) eluted from a Sephacryl S100 column, indicating the elution volume of both rat liver and TDB I-expressed ketohehexokinase (dashed lines). The native molecular mass of both enzymes was determined to be 134 kDa.

physiological significance of changes in the intensity of this resonance are difficult to interpret because of the unknown proportion of NMR 'invisible' high molecular weight and precipitated polyphosphates. However, the change in intensity does indicate that the detectable polyphosphate pool is dynamic and thus capable in principle of contributing to the phosphorus requirements of the cells, even though TDBI lacks a functional vacuolar alkaline phosphatase. When a fructose-tolerant (TDB I(-)) strain was perfused with 50mM glucose, or 50mM glucose plus 25mM fructose (data not shown), no significant differences in spectra compared with those shown in *Figure 6.8* were observed.

When the fructose-intolerant strain (TDB I) was incubated with 1mM fructose, a signal appeared in the ^{31}P NMR spectrum at 4.8 ppm (*Fig. 6.9a-c*). This new signal was assigned to fructose 1-phosphate on the basis of spectrophotometric enzyme assays of cell extracts and the observation that addition of fructose 1-phosphate [Sigma] to the extract resulted in a resonance

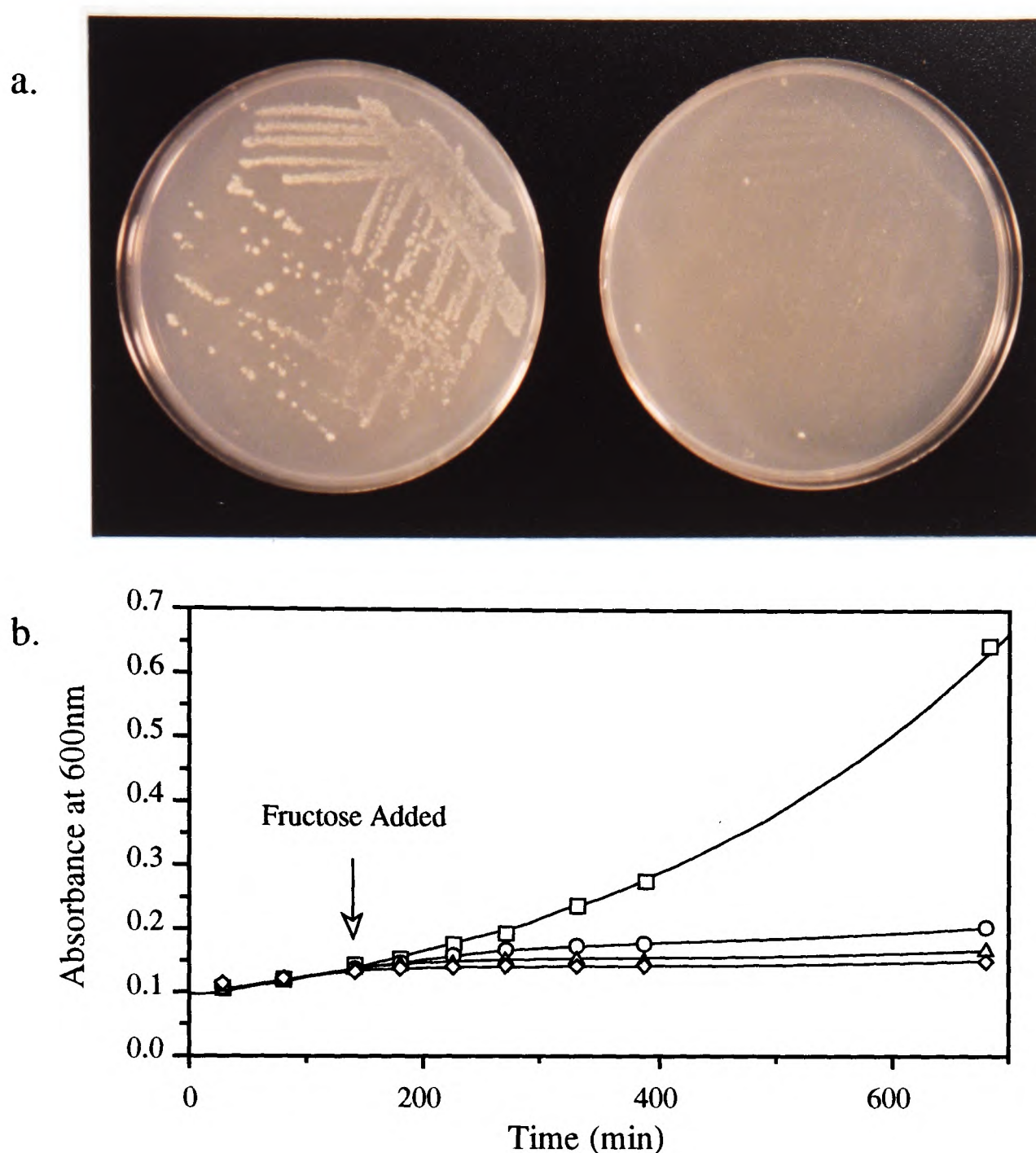


Figure 6.6 (a) Replica plates of TDB I cells on SD leu⁻ (left) and SD^{1/2}F leu⁻ (right) plates. Cells grew well on SD plates, but only three revertants were observed on the fructose-supplemented plates. (b) Growth profiles of the fructose-intolerant yeast strain TDB I upon addition of fructose to various concentrations. A exponential-phase culture of cells in SD leu⁻ medium was divided into equal aliquots, and the growth monitored by A₆₀₀ measurement. As indicated, fructose was added to the cultures to a final concentration of 0.5% w/v (◇), 0.1% w/v (Δ) or 0.025% w/v (○), or had water added as a control (□).

at the same chemical shift. Upon exposure to 5mM (*Fig. 6.9d*) or 25mM (*Fig. 6.10*) fructose it was observed that both the fructose 1-phosphate signal and the residual cytoplasmic P_i signal broadened and shifted upfield. These changes in chemical shift indicate that the cytoplasm became more acidic, most probably due to substrate limitation of the plasmalemma H⁺-ATPase, which is thought to play an important role in pH regulation in yeast cells [85], while the broadening of the signal indicates a heterogeneity in the response of the cell population to the fructose treatment. ³¹P NMR studies on perfused rat liver [86] with either 5mM or 10mM fructose also showed rapid falls in the concentrations of NTP, P_i and cytoplasmic pH, accompanied by a rapid rise in fructose 1-phosphate. In this case, the decrease in pH was attributed to the production of lactate from fructose.

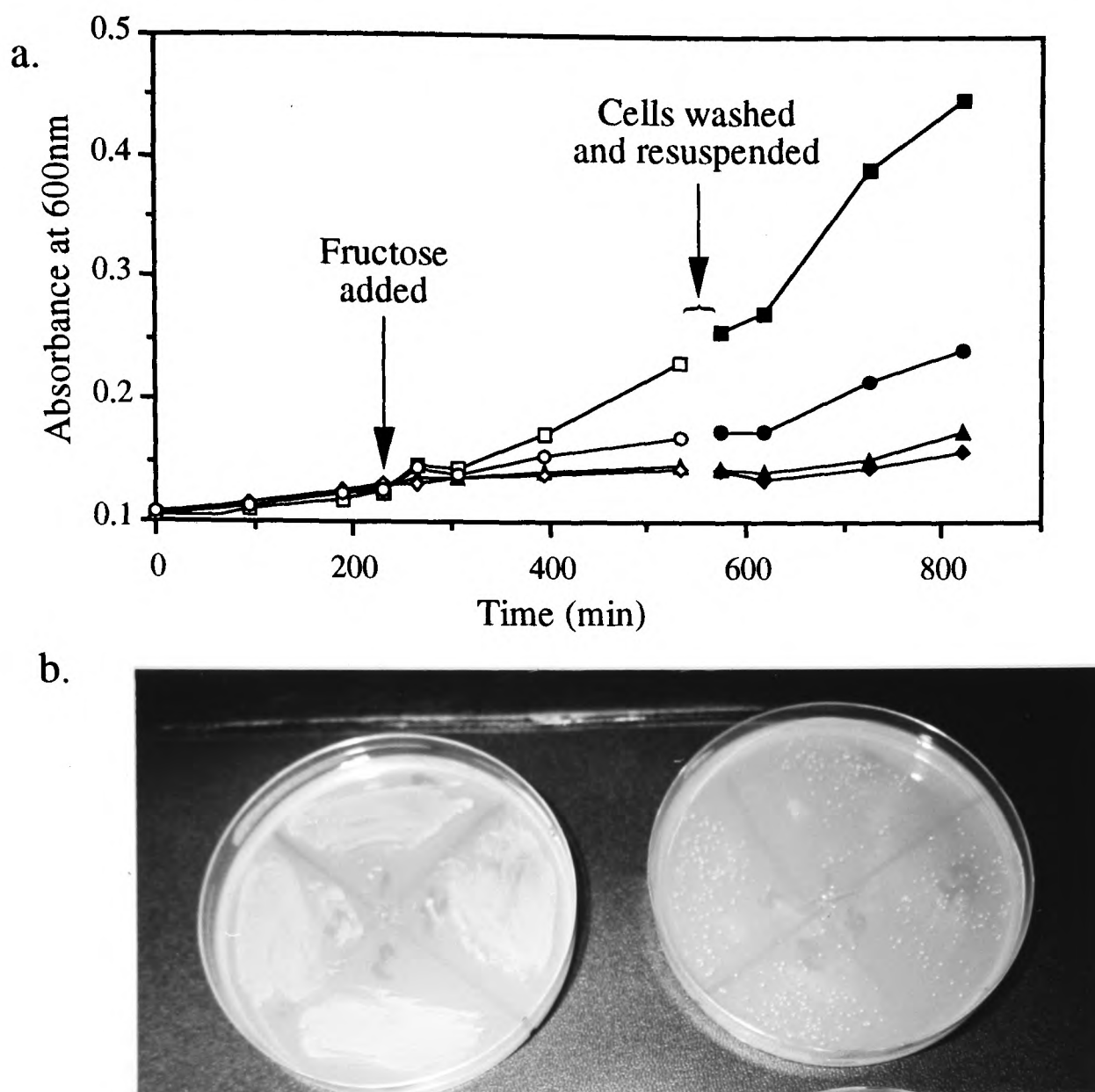


Figure 6.7 (a) Recovery of TDB I cell growth following addition of fructose. Fructose was added as described in Fig. 6.6b (open symbols), and five hours later, the cells were washed with water and resuspended in SD leu- (no fructose, filled symbols), and the growth monitored by measurement of absorbance at 600nm. (b) Cells from (a) were grown overnight, and 20 μ l. aliquots spread on SD leu- (left) or SD^{1/2}F leu- plates (right). 0.1% w/v fructose treatment; top and right quarters, and 0.5% w/v fructose treatment; bottom left quarters.

Two responses to different concentrations of fructose could be discerned in *Figures 6.9* and *6.10*. At high concentrations of fructose (5mM, *Fig. 6.9d* and 25mM, *Fig. 6.10*) the immediate cessation of growth on addition of fructose was attended by a severe disruption of the pools of phosphorylated metabolites. However, the arresting of cell division by a low concentration of fructose (1mM, *Fig. 6.9a-c*) was not accompanied by the same changes, although a depletion of the polyphosphate resonance was observed.

At the higher concentrations of fructose, the marked fall in the NTP and cytoplasmic P_i levels, to a point where the latter became undetectable, coupled with the loss of pH control, is likely to produce a condition in which growth cannot be sustained. In principle the TDB1 cells could draw on the polyphosphate pool during the accumulation of fructose 1-phosphate, but the existence of the polyphosphate pool did not prevent the depletion of the cytoplasmic P_i and NTP pools.

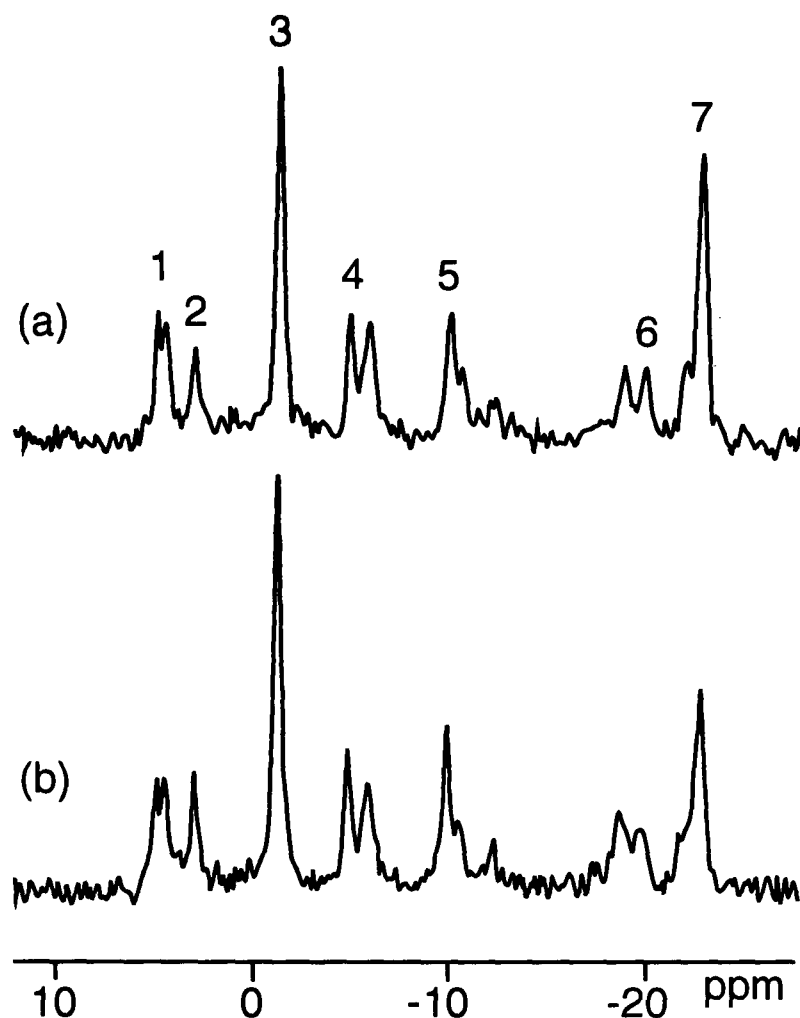


Figure 6.8. ^{31}P NMR spectra of the yeast transformant TDB I during a control experiment. The immobilized cells were bathed in a circulating medium containing 50mM glucose and the spectra were recorded (a) 30 minutes and (b) 4 hours after the start of the circulation. The numbered peaks correspond to: 1, several phosphomonoesters; 2, cytoplasmic phosphate; 3, an unidentified phosphodiester; 4, 5 and 6, the γ -, α -, and β - phosphates, respectively, of NTP; 7, polyphosphate.

The immediate effect of fructose on growth may however be more subtle, and could involve the cellular concentration of GTP, a metabolite involved in transmembrane and cytoplasmic signal transduction. It has been previously observed that ketohexokinase can utilize GTP to phosphorylate fructose [87], and that the symptoms of hereditary fructose intolerance in humans are due to the depletion of hepatic GTP. Although it is not possible to differentiate between ATP and GTP *in vivo* with NMR, the overall depletion of the NTP pool would suggest that GTP levels have fallen.

At lower concentrations, where fructose had little effect on the pool size of NTP and P_i (Figs. 6.9a-c), fructose was likely to be exerting its effect on TDB I by the continuous diversion of phosphate from energy metabolism into the production of fructose 1-phosphate at the expense of cell division. Under these conditions, it is possible that these pools were being maintained by mobilization of polyphosphates, and that the slow and restricted growth observed in Figure 6.6a may also have been drawing on this source of phosphate. Interestingly, addition of phosphate to the growth medium of TDB I did not relieve inhibition of growth by fructose at either low or high fructose concentrations (results not shown, but almost identical growth

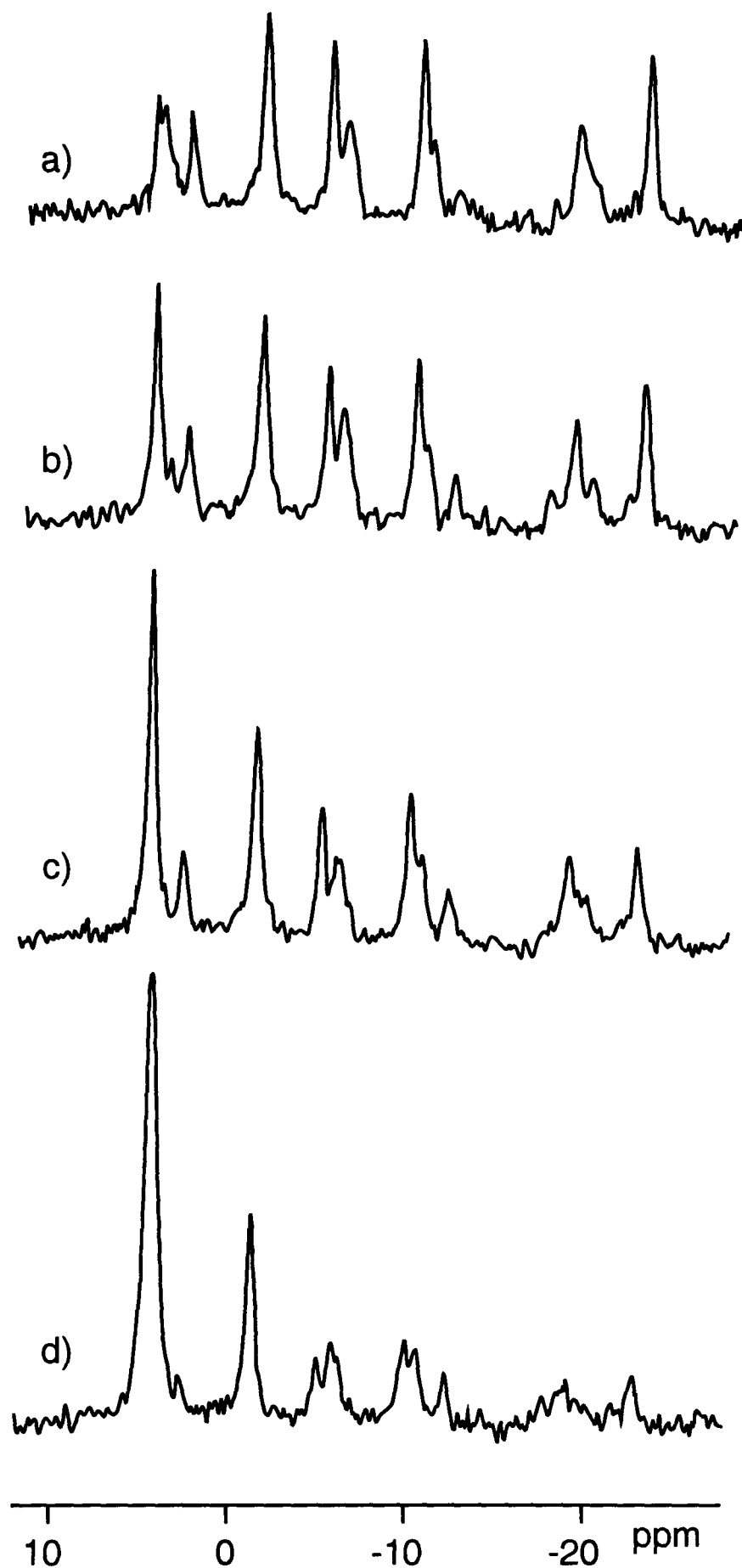


Figure 6.9 ^{31}P NMR spectra of the yeast transformant TDB I during a fructose loading experiment. After recording a spectrum (a) with the cells bathed in a circulating medium containing 50mM glucose, the medium was switched to a solution containing 50mM glucose/1mM fructose and spectra were obtained after (b) 1 hour and (c) 3 hours exposure to fructose. The fructose concentration was then increased to 5mM and a spectrum (d) was recorded 1 hour later. Increasing the fructose concentration caused the fructose 1-phosphate signal to shift upfield from 4.8 ppm (b, c) to 4.6 ppm (d).

profiles were observed), suggesting that the availability of phosphate is not the limiting factor in determining growth.

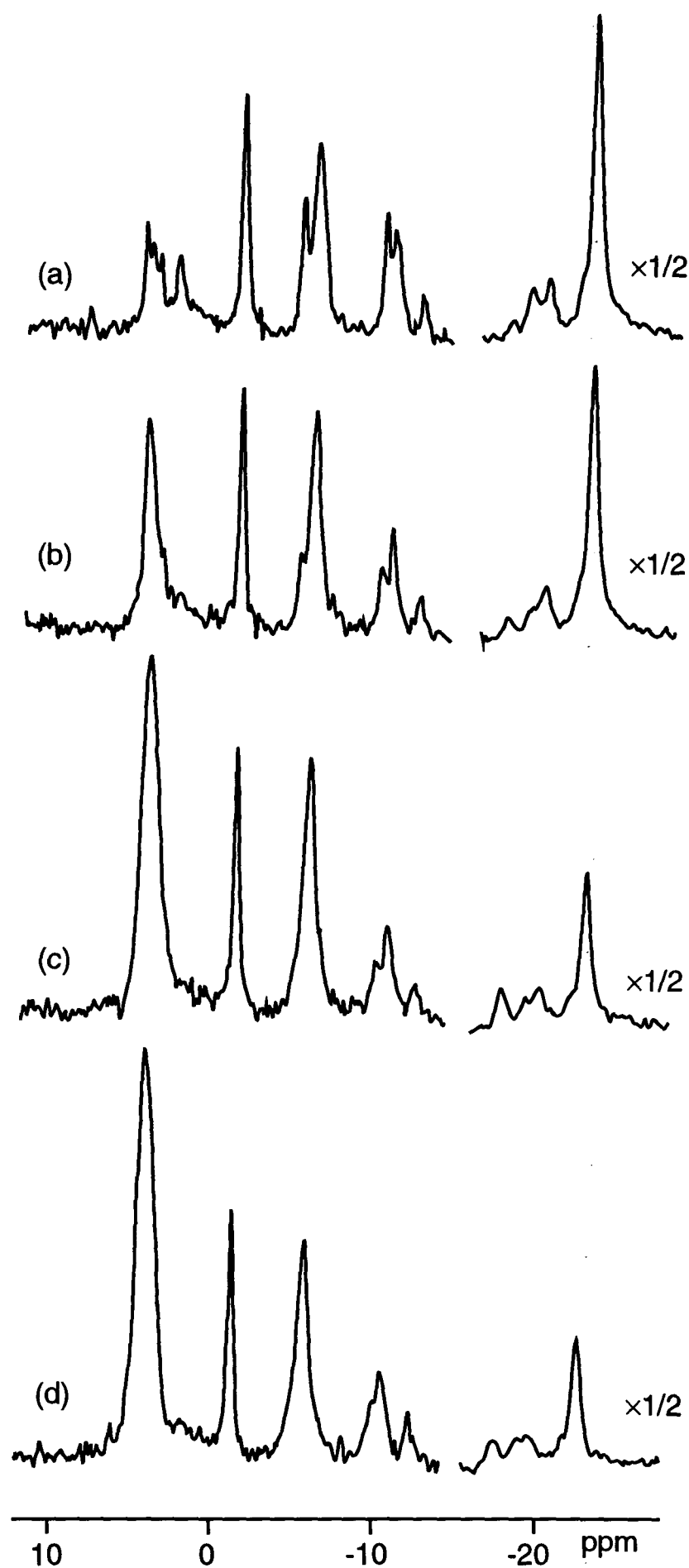


Figure 6.10 ^{31}P NMR spectra of the yeast transformant TDB I during a fructose loading experiment at high concentration. After recording a spectrum (a) with the cells bathed in a circulating medium containing 50mM glucose, the medium was switched to a solution containing 50mM glucose/25mM fructose and spectra were obtained (b) immediately after the switch and then after a further (c) 1 hour and (d) 2 hours exposure to fructose. The chemical shift of the broad fructose 1-phosphate signal is approximately 4.5 ppm in (b) and 4.2 ppm in (c) and (d).

6.3.2. Multiple Integration of the Ketohexokinase cDNA into the Yeast Chromosome.

6.3.2.1 Construction of the integration plasmid, *pIAD1*. The construction of the *pMIRY2*-like plasmid is shown in Fig. 6.11. The *pYlrG12* plasmid was digested with *EcoR* I and the fragments separated on a 0.7% w/v LMP agarose gel. The 2.47kb fragment (B) was removed from the gel and ligated into the *EcoR* I site of *pUC18*. This fragment is much smaller than that used in the *pMIRY2* plasmid, but had all the sequence that was required for integration into the yeast chromosomal rDNA locus. It contains the 5S region, along with the two non-transcribed regions (NTS1 and NTS2), the second of these having a putative ARS sequence. This ARS sequence was not foreseen to create a problem, since integration of the plasmid would require a *Sma* I digestion, which cuts the ARS sequence in the middle. Two orientations of insertion were found, which were distinguished by *Sma* I digestion. One of these (see Fig. 6.11) was subjected to *Sst* I digestion, and then Klenow treatment to generate blunt ends (4.2.7.5), and this was further digested with *BamH* I. This process removes the *Sma* I and *Kpn* I sites from the *pUC* multiple cloning site, so that the final plasmid has only one of each of these sites. Into these sites was directionally ligated the 3.1kb fragment generated by a *BamH* I/*Hpa* I double digest of *pMA91*, containing the *PGK* expression cassette and the *leu2-d* auxotrophic marker. This final construct was termed *pIAD1*, and it was into this that the coding region of the ketohexokinase cDNA was cloned, inserting it into the unique *Bgl* II site. Again, both orientations of KHK insert were obtained which termed *pIAD1*-KHK and *pIAD1*-KHK(-), in the same fashion as the *pMA91* inserts. As well as the unique *Bgl* II site in *pIAD1* for the cloning of DNA to be expressed, this plasmid contains two other unique sites for integration into the yeast chromosome. The *Sma* I site allowed targetted homologous multiple-copy integration into the rDNA locus on Chromosome XII, multiple integration being obtained by virtue of the need of multiple copies of the *leu2-d* auxotrophic marker to complement the chromosomal *leu2* mutation. The *Kpn* I site allowed targetted single-copy integration into the *leu2* locus on Chromosome III.

The size of *pIAD1* (8.26kb) was much smaller than the *pMIRY2* plasmid [68], and Lopes *et al.* [88] have suggested that the stability of integration decreases as the plasmid size increases. Even with the addition of the ketohexokinase cDNA, this plasmid (9.17kb) was still smaller than the *pMIRY2* plasmid, and well below the upper limit suggested for mitotic stability (10.2kb), suggesting the plasmid will be stably integrated. In addition, the presence of the *PGK* expression cassette with its unique *Bgl* II cloning site made *pIAD1* a more versatile plasmid than *pMIRY2*.

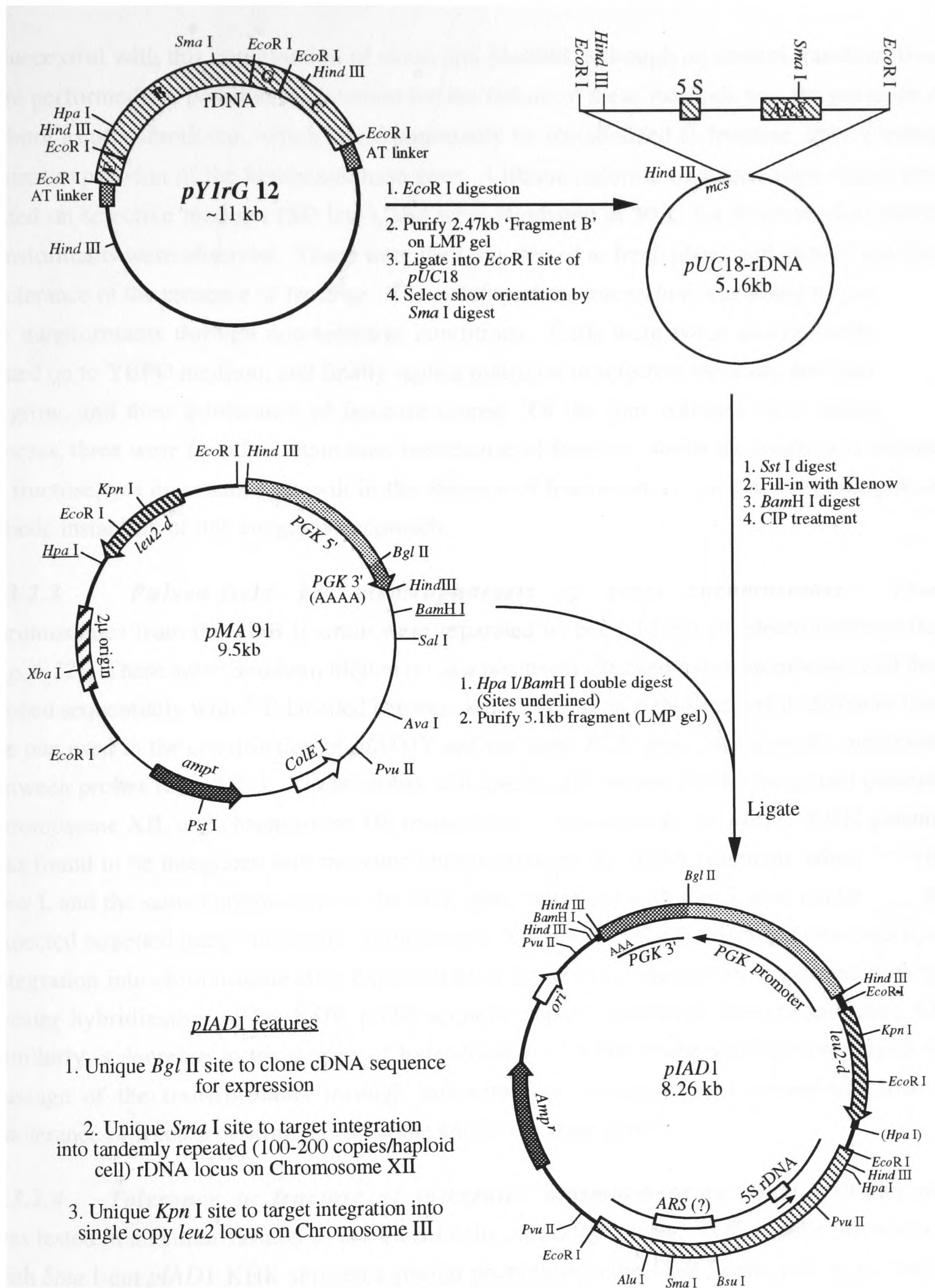


Figure 6.11 Construction of the multiple integration yeast expression vector *pIAD1*, as described in the text. Plasmids are not shown to scale.

6.3.2.2 Transformation and integration of *pIAD1*-KHK. The *pIAD1*-KHK and *pIAD1*-KHK(-) plasmids were digested to completion with either *Sma* I or *Kpn* I, and the purified digestion products were used to transform the TDS II strain of yeast using the lithium chloride method. Again, the electro- and spheroplast transformation methods proved

unsuccessful with this combination of strain and plasmid, although no control transformations were performed. A possible explanation for the failure of these methods was the presence of sorbitol as an osmoticum, which could potentially be metabolized to fructose, thus selecting against expression of the ketohexokinase gene. Lithium chloride transformation mixes were plated on selective medium (SD leu-), and were incubated at 30°C for fourteen days before transformants were observed. These were then transferred to fresh plates and used to test their intolerance of the presence of fructose. The stability of the integration was tested by passage of the transformants through non-selective conditions. Cells were twice successively replica plated on to YEPD medium, and finally replica plated on to selective medium, and their ability to grow, and their intolerance of fructose scored. Of the four colonies taken through this process, three were found to retain their intolerance of fructose, whilst the fourth was tolerant of fructose, yet maintained growth in the absence of leucine in the media, suggesting some genetic instability of this integrative approach.

6.3.2.3 Pulsed-field gel electrophoresis of yeast chromosomes. Yeast chromosomes from the TDB II strain were separated by pulsed-field gel electrophoresis (see Fig. 6.12). These were Southern blotted on to a positively-charged nylon membrane, and then probed sequentially with ³²P-labelled ketohexokinase DNA, an rDNA fragment (different from the one used in the construction of *pIAD1*), and the yeast *PGK* gene, stripping the membrane between probes (Fig. 6.12). These probes will specifically screen for the integrated plasmid, chromosome XII, and chromosome III, respectively. As expected, the *pIAD1*-KHK plasmid was found to be integrated into the same chromosome as the rDNA fragment, when cut with *Sma* I, and the same chromosome as the *PGK* gene, when cut with *Kpn* I, thus confirming the expected targetted integrations into chromosomes XII and III. The fainter band observed upon integration into chromosome III is expected for a single copy integration, compared with the greater hybridization of the KHK probe sequence upon integration into chromosome XII. Similarly, a decrease in the degree of hybridization of KHK probe was observed upon the passage of the transformants through non-selective conditions and subsequent loss of intolerance of fructose, owing to loss of the ketohexokinase gene.

6.3.2.4 Tolerance to fructose of integrated plasmid-bearing yeast. Intolerance was tested in a similar fashion to the TDB I cells on SD^{1/2}F plates. TDS II cells transformed with *Sma* I-cut *pIAD1*-KHK showed a similar phenotype to the TDB I cells, and were called TDB II cells. Cells transformed with *Kpn* I-cut *pIAD1*-KHK or *pIAD1*-KHK(-) cut with either enzyme did not show this intolerance. Growth of TDB II cells in liquid medium was monitored (see Fig. 6.13) in a similar fashion to TDB I cells, and although growth was stopped at high concentrations, at lower concentrations, the cells growth was observed at a reduced rate. In a similar fashion to the TDB I cells, TDB II cells exposed to fructose for three hours were found to continue growth when resuspended in fructose-free SD leu- medium (see Fig.

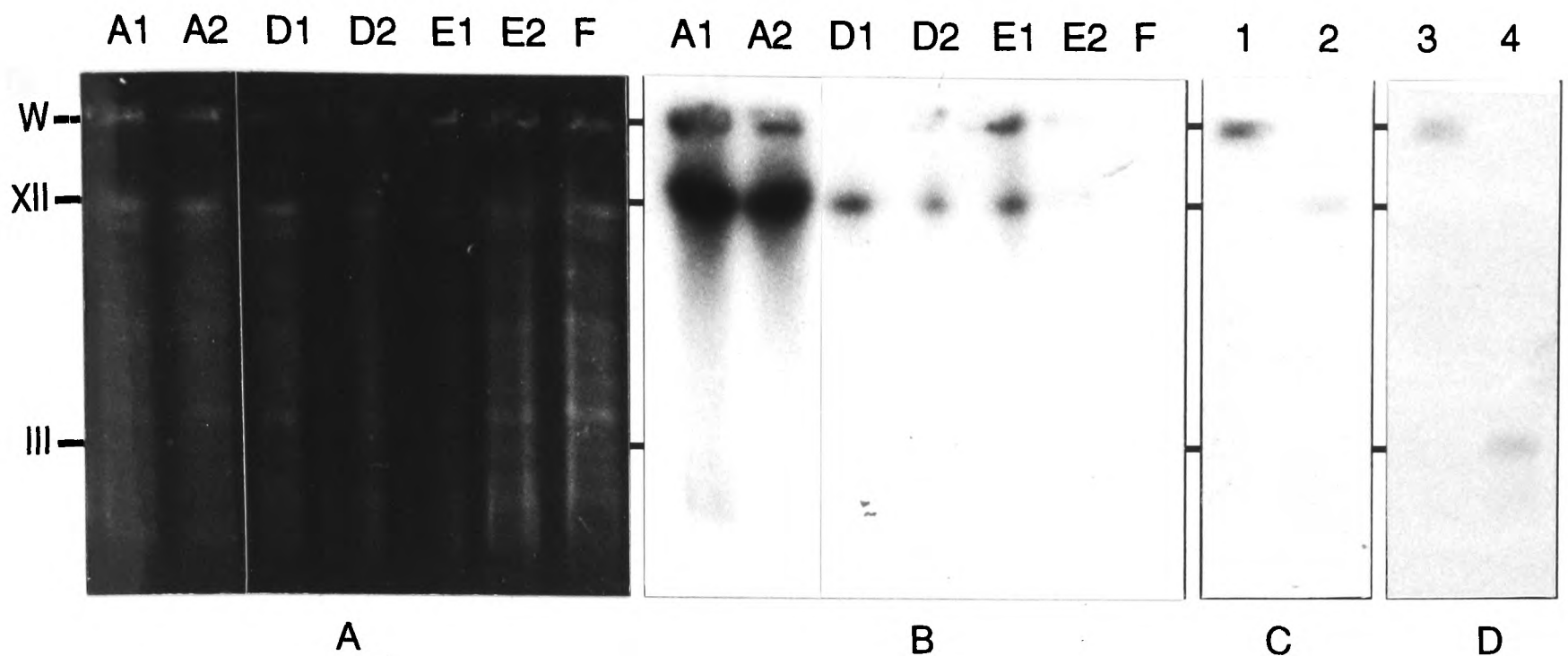


Figure 6.12 Pulsed-field gel electrophoresis of the chromosomes of several TDS II strains transformed with linearised *pAD1-KHK*. Pictures are ethidium bromide-stained chromosomes within the gel (A), Southern blot probed with labelled ketohexokinase cDNA (B), rDNA fragment 'C' (C), and the yeast *PGK* gene (D). Pictures A and B both show several strains either before (A1, D1 and E1), or after (A2, D2 and E2), passage of cells through non-selective conditions. Loss of the cDNA sequence from strain E is apparent from the intensity of the radioactive signal, and the tolerance of this strain to fructose. The plasmid was either targeted to the rDNA locus (strains A, D and E), or the *PGK* gene (strain F). Confirmation of chromosome assignments are shown in pictures C and D, where lanes 1 and 3 show the position of the loading well (W), lane 2, chromosome XII and lane 4, chromosome III. It was the transformed strain A1 that was used in the rest of this study as the TDB II strain.

6.14a). Plating of these cultures following overnight growth on SD^{1/2}F leu- plates led to the growth of a large number of fructose tolerant colonies (see Fig. 6.14b, compare with Fig. 6.7b), but although it was not possible to count the number of revertants because of the high cell density, it was apparent from the control SD plate that there were still a large number of fructose intolerant cells.

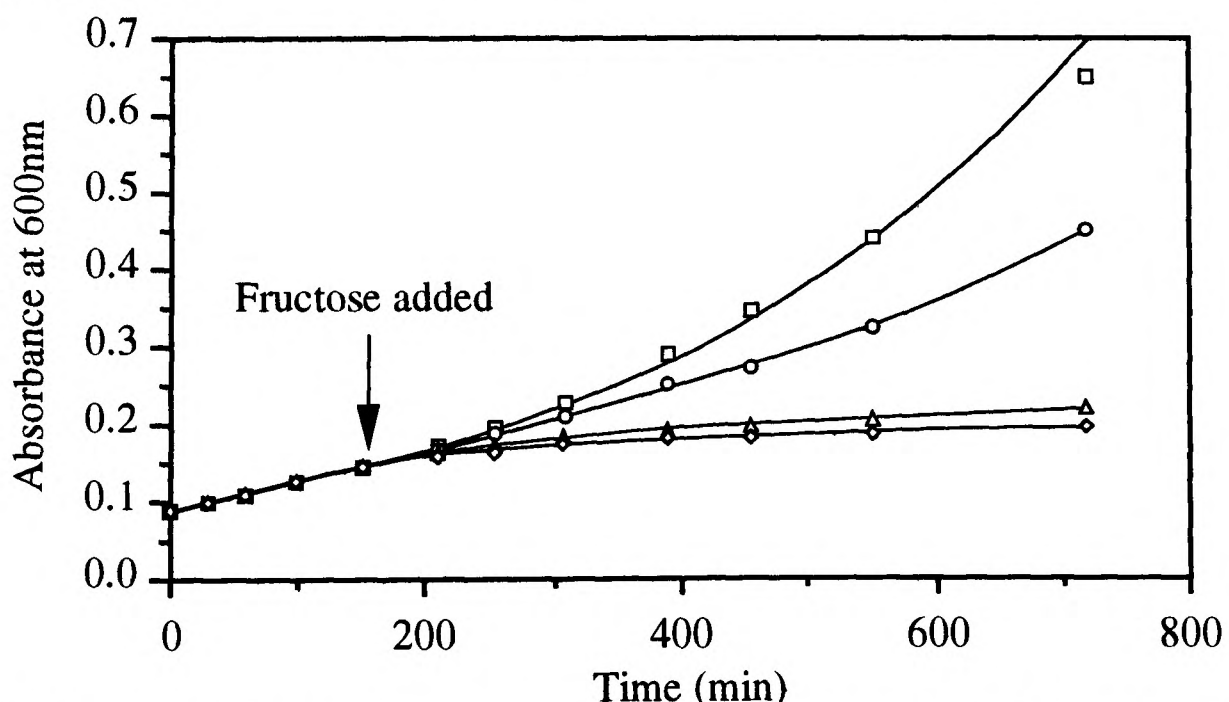


Figure 6.13 Growth profiles of the fructose-intolerant yeast strain TDB II upon addition of fructose to various concentrations, as described in Fig. 6.6b. As indicated, fructose was added to the cultures to a final concentration of 0.5% w/v (◇), 0.1% w/v (Δ) or 0.025% w/v (○), or water was added as a control (□).

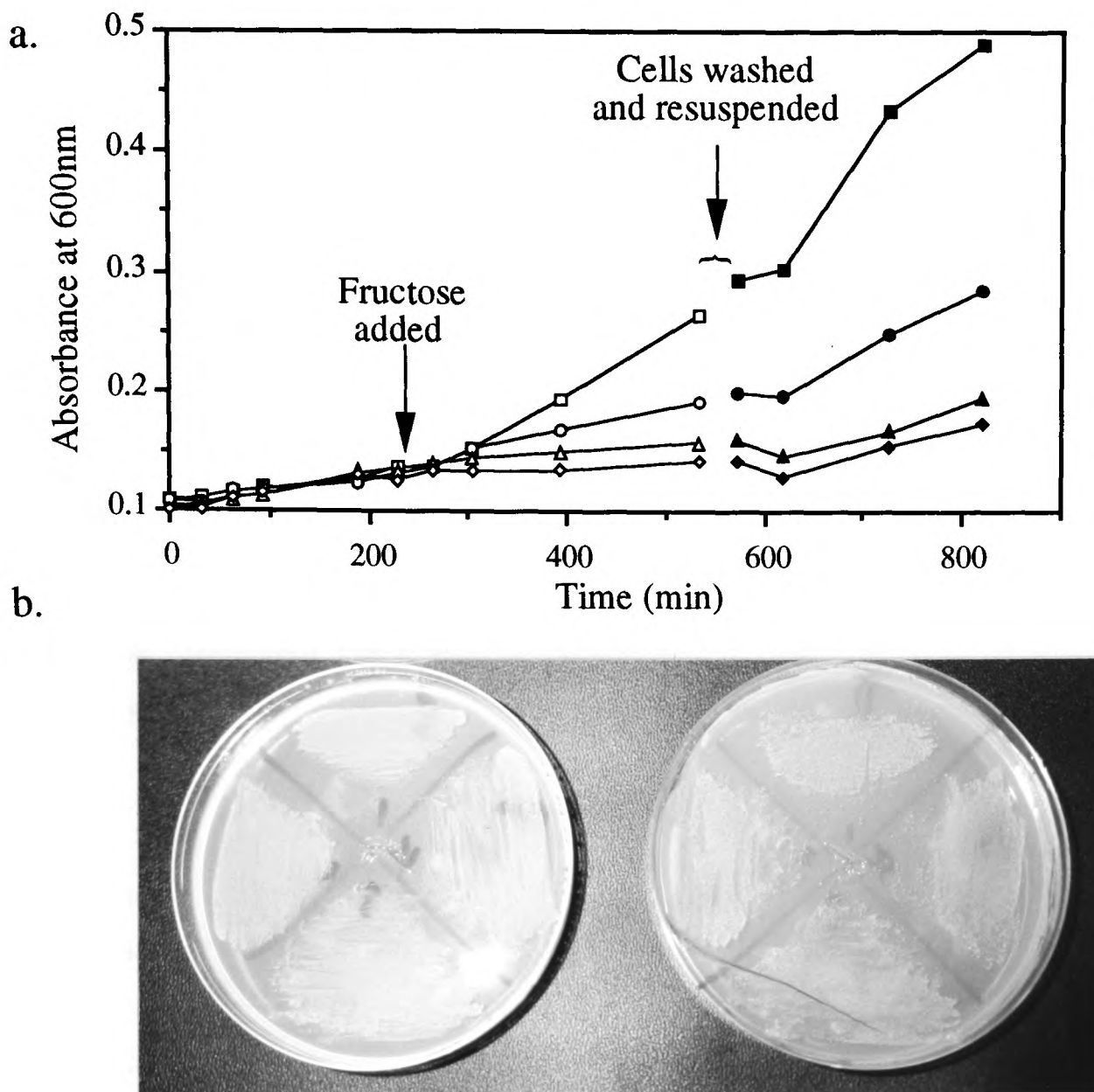


Figure 6.14 (a) Recovery of TDB II cell growth following addition and removal of fructose. Fructose was added as described in Fig. 6.13 (open symbols), and five hours later, the cells were washed with water and resuspended in SD leu- (no fructose, filled symbols), and the growth monitored by measurement of absorbance at 600nm. (b) Cells from (a) were grown overnight, and 20 μ l. aliquots spread on SD leu- (left) or SD $^{1/2}$ F leu- plates (right), fructose concentration as in Fig. 6.7b.

6.3.2.5 Heterologous expression of ketohexokinase from the integrated plasmid. In a similar fashion to the demonstration of ketohexokinase in the TDB I strain, the expression of the enzyme in TDB II was also determined (see Fig. 6.3), and the absence of activity in the control TDB II(-) was also shown. The protein was also partially purified from the yeast crude extract and analysed by SDS-PAGE (see Fig. 6.4), and again a protein was expressed in TDB II with comparable size to the rat enzyme. The activity of ketohexokinase in a TDB II extract was also determined using the spectrophotometric assay, and found to be 0.53 μ mol/min/mg, which is 85% of the activity expressed from the *pMA91-KHK* plasmid. This lower expression level of ketohexokinase in TDB II compared to TDB I may explain the greater tolerance of fructose at low concentrations.

6.3.3 Genetic Stability of the Integrated *pIAD1* Plasmid.

It is clear that when the cells are stressed with fructose under selective conditions, the genetic stability of this integrated plasmid is much lower than that of the 2μ -based plasmid *pMA91*, although the opposite is true when the transformants are transferred to non-selective media. This would suggest that the TDB II yeast can selectively lose the ketohexokinase coding region, and yet remain *LEU2*. The method of multiple integration at the rDNA locus is thought to be by an initial single integration event, followed by replication of the integrated region from the ARS region, and its re-integration at the same locus. Selection for multiple integration by using the *leu2-d* marker would favour cells where this process repeats itself until sufficient copies have integrated. If, however, the region were to remain dynamic with looping out and re-integration, then integrated DNA that is detrimental to the cell (*e.g.* ketohexokinase expression in the presence of fructose), might 'pop-out' and be lost.

6.3.4 Sucrose Tolerance of the Fructose Intolerant Yeast Strains.

Both the TDB I and TDB II strains showed no growth impairment when a glucose carbon source was supplemented with sucrose, confirming that the phenotype of these cells was still *mal*^o *suc*^o, and that the strains would be a suitable selection system for the selection of fructotransferase activities of genes introduced into these strains.

6.4 References

1. Tuite, M.F. (1992). Strategies for the Genetic Manipulation of *Saccharomyces cerevisiae*. *Crit. Rev. Biotech.*, **12**, 157-188.
2. Hinnen, A., Hicks, J.B. and Fink, G.R. (1978). Transformation of yeast. *Proc. Natl. Acad. Sci. USA*, **75**, 1929-1933.
3. Beggs, J.D. (1978). Transformation of yeast by a replicating hybrid plasmid. *Nature*, **275**, 104-109.
4. Harashima, S., Takagi, A. and Oshima, Y. (1984). Transformation of protoplasted yeast cells is directly associated with cell fusion. *Mol. Cell. Biol.*, **4**, 771-778.
5. Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163-168.
6. Schiestl, R.H. and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single-stranded nucleic acids as a carrier. *Curr. Genet.*, **16**, 339-346.
7. Hashimoto, H., Morijawa, H., Yamata, Y. and Kimura, A. (1985). A novel method for transformation of intact yeast cells by electroinjection of plasmid DNA. *Appl. Microbiol. Biotechnol.*, **21**, 336-338.

8. Karube, I., Tamiya, E. and Matsuoka, H. (1985). Transformation of *Saccharomyces cerevisiae* spheroplasts by high electric pulse. *FEBS Lett.*, **182**, 90-94.
9. Meilhoc, E., Masson, M.-J. and Teissié, J. (1990). High efficiency transformation of intact yeast cells by electric field pulses. *Bio/Technology*, **8**, 223-227.
10. Becker, D. M. and Guarente, L. (1991). High-Efficiency Transformation of Yeast by Electroporation. *Meths. Enzymol.*, **194**, 182-187.
11. Ratzkin, B. and Carbon, J. (1977). Functional expression of cloned yeast DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **74**, 487-491.
12. Rothstein, R. (1985). Cloning in Yeast, in *DNA cloning; A practical approach* (Glover, D.M, ed.), IRL Press, Oxford, vol 2, 45-66.
13. Chinery, S.A. and Hincliffe, E. (1989). A novel class of vector for yeast transformation. *Curr. Genet.*, **16**, 21-25.
14. Cox, M. (1983). The *FLP* protein of the yeast 2 μ plasmid: expression of a eukaryotic genetic recombination system in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **80**, 4223-4227.
15. Murray, J.A.H. (1987). Bending the rules: the 2 μ plasmid of yeast. *Mol. Microbiol.*, **1**, 1-4.
16. Erhart, E. and Hollenberg, C.P. (1983). The Presence of a Defective *LEU2* Gene on 2 μ DNA Recombinant Plasmids of *Saccharomyces cerevisiae* Is Responsible for Curing and High Copy Number. *J. Bacteriol.*, **156**, 625-635.
17. Lopes, T.S., Hakkaart, G.-J.A.J., Koerts, B.L., Raué, H.A. and Planta, R.J. (1991). Mechanism of high-copy-number integration of pMIRY-type vectors into the ribosomal DNA of *Saccharomyces cerevisiae*. *Gene*, **105**, 83-90.
18. Kaster, K.R., Burgett, S.G. and Ingolia, T.D. (1984). Hygromycin B resistance as dominant selectable marker in yeast. *Curr. Genet.*, **8**, 353-358.
19. Jimenez, A. and Davies, J. (1980). Expression of a transposable antibiotic resistance element in *Saccharomyces cerevisiae*. *Nature*, **287**, 869-871.
20. Webster, T.D. and Dickson, R.C. (1983). Direct selection of *Saccharomyces cerevisiae* resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin resistance gene of Tn 903. *Gene*, **26**, 245-252.
21. Fogel, S. and Welch, J.W. (1982). Tandem gene amplification mediates copper resistance in yeast. *Proc. Natl. Acad. Sci. USA*, **79**, 5342-5346.
22. Henderson, R.C.A., Cox, B.S. and Tubb, R. (1985). The transformation of brewing yeasts with a plasmid containing the gene for copper resistance. *Curr. Genet.*, **9**, 133-138.
23. Romanos, M.A., Scorer, C.A. and Clare, J.J (1992). Foreign Gene Expression in Yeast: A Review. *Yeast*, **8**, 423-488.
24. Beggs, J.D., Van den Berg, J., Van Ooyen, A. and Weissman, C. (1980). Abnormal expression of chromosomal rabbit β -globin gene in *Saccharomyces cerevisiae*. *Nature*, **283**, 835-840.

25. Henikoff, S. and Furlong, C. (1983). Sequence of a *Drosophila* DNA segment that functions in *Saccharomyces cerevisiae* and its regulation by a yeast promoter. *Nuc. Acids Res.*, **11**, 789-800.
26. Kiss, G.B., Pearlman, R.E., Cornish, K.V., Friesan, J.D. and Chan, V.L. (1982). The *Herpes simplex* virus thymidine kinase is not transcribed in *Saccharomyces cerevisiae*. *J. Bacteriol.*, **149**, 542-547.
27. Hitzeman, R.A., Hagie, F.F., Levine, H.L., Goeddel, D.W., Ammerer, G. and Hall, B. D. (1981). Expression of human gene for interferon in yeast. *Nature*, **293**, 717-723.
28. Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. *Ann. Rev. Biochem.*, **58**, 1051-1077.
29. Guarente, L. (1984). Yeast promoters: positive and negative elements. *Cell*, **36**, 799-800.
30. Serfling, E., Jasin, M. and Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. *Trends Genet.*, **1**, 224-230.
31. Benoist, C. and Chambon, P. (1981). *In vivo* sequence requirements of the SV40 early promoter region. *Nature*, **290**, 304-311.
32. Rosenberg, S., Coit, D. and Tekamp-Olson, P. (1990). Glyceraldehyde-3-phosphate dehydrogenase-derived expression cassettes for constitutive synthesis of heterologous proteins. *Meths. Enzymol.*, **185**, 341-351.
33. Tuite, M.F., Dobson, M.J., Roberts, N.A., King, R.M., Burke, D.C., Kingsman, S.M. and Kingsman, A.J. (1982). Regulated high efficiency expression of human interferon-alpha in *Saccharomyces cerevisiae*. *EMBO J.*, **1**, 603-608.
34. Johnston, S.A., Salmeron, J.M. and Dincher, S.S. (1987). Interaction of positive and negative regulatory proteins in the galactose-regulon of yeast. *Cell*, **50**, 143-146.
35. Price, V.L., Taylor, W.E., Clevenger, A., Worthington, M. and Young, E.T. (1990). Expression of heterologous proteins in *Saccharomyces cerevisiae* using *ADH2* promoter. *Meths. Enzymol.*, **185**, 308-318.
36. Hinnen, A., Meyhack, B. and Helm, J. (1989). Heterologous gene expression in yeast, in *Yeast Genetic Engineering* (Barr, P.J., Brake, A.J. and Valenzuela, P., eds.), Butterworths, 193-213.
37. Kramer, R.A., DeChiara, T.M., Schaber, M.D. and Hilliker, S. (1984). Regulated expression of a human interferon gene in yeast: control by phosphate concentration or temperature. *Proc. Natl. Acad. Sci. USA*, **81**, 367-370.
38. Etcheverry, T. (1990). Induced expression using yeast copper metallothionein promoter. *Meths. Enzymol.*, **185**, 319-329.
39. Brake, A.J., Merryweather, J.P., Coit, D.G., Heberlein, U.A., Masiarz, F.R., Mullenbach, G.T., Urdea, M.S., Valenzuela, P. and Barr, P.J. (1984). α -factor-directed synthesis and secretion of mature foreign protein in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*, **81**, 4642-4646.
40. Hitzeman, R.A., Leung, D.W., Perry, L.J., Kohr, W.J., Levine, H.L. and Goeddel, D.V. (1983). Secretion of human interferons by yeast. *Science*, **219**, 620-625.

41. Holland, J.P. and Holland, M.J. (1980). Structural comparison of two nontandemly repeated yeast glyceraldehyde-3-phosphate dehydrogenase genes. *J. Biol. Chem.*, **255**, 2596-2605.
42. Johnston, M. (1987). A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **51**, 458-476.
43. St. John, T.P. and Davis, R.W. (1981). The organisation and transcription of the galactose gene cluster of *Saccharomyces*. *J. Mol. Biol.*, **152**, 285-315.
44. Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic mRNA. *Nature*, **263**, 211-214.
45. Zaret, K.S. and Sherman, F. (1984). Mutationally altered 3' ends of yeast *CYC1* mRNA affect transcript stability and translational efficiency. *J. Mol. Biol.*, **177**, 107-136.
46. Henikoff, S. and Cohen, E.H. (1984). Sequences responsible for transcription termination on a gene segment in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **4**, 1515-1520.
47. Osborne, B.I. and Guarente, L. (1989). Mutational analysis of a yeast transcriptional terminator. *Proc. Natl. Acad. Sci. USA*, **86**, 4097-4101.
48. Irniger, S., Egli, C.M. and Braus, G.H. (1991). Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **11**, 3060-3069.
49. Broach, J.R. (1983). Construction of high copy number yeast vectors using 2 μ m circle sequences. *Meths. Enzymol.*, **101**, 307-325.
50. Parent, S.A., Fenimore, C.M. and Bostian, K.A. (1985). Vector systems for the expression, analysis and cloning of DNA sequences in *Saccharomyces cerevisiae*. *Yeast*, **1**, 83-138.
51. Armstrong, K.A., Som, T., Volkert, F.C, Rose, A. and Broach, J.R. (1989). Propagation and expression of genes in yeast using 2-micron circle vectors, in *Yeast Genetic Engineering* (Barr, P.J., Brake, A.J. and Valenzuela, eds.), Butterworths, 162-192.
52. Futcher, A.B. and Cox, B.S. (1983). Maintenance of the 2 μ m circle plasmid in populations of *Saccharomyces cerevisiae*. *J. Bacteriol.*, **154**, 612-622.
53. Tschumper, G. and Carbon, J. (1980). Sequence of a yeast DNA fragment containing a chromosomal replicator and the *TRP1* gene. *Gene*, **10**, 157-166.
54. Murray, A.W. and Szostak, J.W. (1983). Pedigree analysis of plasmid segregation in yeast. *Cell*, **34**, 911-970.
55. Spalding, A. and Tuite, M.F. (1987). Host plasmid interactions in *Saccharomyces cerevisiae*: effect of host ploidy on plasmid copy number and stability. *J. Gen. Microbiol.*, **135**, 1037-1045.
56. Rose, M.D., Novick, P., Thomas, J.H., Botstein, D. and Fink, G.R. (1987). A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. *Gene*, **60**, 237-243.
57. Clarke, L. and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, **257**, 504-509.

58. Szostak, J.W. and Blackburn, E.H. (1982). Cloning yeast telomeres on linear plasmid vectors. *Cell*, **29**, 245-255.
59. Murray, A.W. and Szostak, J.W. (1983). Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.
60. Burke, D.T., Carle, G.F. and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, **236**, 806-812.
61. Murray, A.W., Schultes, N.P. and Szostak, J.W. (1985). Chromosomes length controls mitotic chromosome segregation in yeast. *Cell*, **45**, 529-536.
62. Smith, D.R., Smyth, A.P. and Moir, D.T. (1990). Amplification of large artificial chromosomes. *Proc. Natl. Acad. Sci. USA*, **87**, 8242-8246.
63. Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R.J. (1983). Genetic applications of yeast transformations with linear and gapped plasmids. *Meths. Enzymol.*, **101**, 228-245.
64. Orr-Weaver, T.L. and Szostak, J.W. (1983). Multiple, tandem plasmid integration in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **30**, 747-749.
65. Boeke, J.D., Hua, X. and Fink, G.R. (1988). A General Method for the Chromosomal Amplification of Genes in Yeast. *Science*, **239**, 280-282.
66. Jacobs, E., Dewechin, M. and Boeke, J.D. (1988). Retrovirus-like vectors for *Saccharomyces cerevisiae*: integration of foreign genes controlled by efficient promoters into yeast chromosomal DNA. *Gene*, **67**, 259-269.
67. Sakai, A., Shimizu, Y. and Hishinuma, F. (1990). Integration of heterologous genes into the chromosome of *Saccharomyces cerevisiae* using a delta sequence of yeast retrotransposon Ty. *Appl. Microbiol. Biotechnol.*, **33**, 302-306.
68. Lopes, T.S, Klotwijk, J., Veenstra, A.E., Van der Aar, P.C., Van Heerikhuizen, H., Raué, H.A. and Planta, R.J. (1989). High-copy-number integration into the ribosomal DNA of *Saccharomyces cerevisiae*: a new vector for high-level expression. *Gene*, **79**, 199-206.
69. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985). Ty elements Transpose through an RNA Intermediate. *Cell*, **40**, 491-500.
70. Mellor, J., Dobson, M.J., Roberts, N.A., Tuite, M.F., Emtage, J.S., White, S., Lowe, P.A., Patel, T., Kingsman, A.J. and Kingsman, S.M. (1983). Efficient synthesis of enzymatically active calf chymotrypsin in *Saccharomyces cerevisiae*. *Gene*, **24**, 1-14.
71. Petes, T.D., Hereford, L.M. and Skryabin, K.G. (1978). Characterization of Two Types of Yeast Ribosomal DNA Genes. *J. Bacteriol.*, **134**, 295-305.
72. Campbell-Burk, S.L. and Shulman, R.G. (1987). High-resolution NMR studies of *Saccharomyces cerevisiae*. *Ann. Rev. Microbiol.*, **41**, 595-616.
73. Salhany, J.M., Yamane, T., Shulman, R.G. and Ogawa, S. (1975). High resolution ³¹P NMR studies of intact yeast cells. *Proc. Natl. Acad. Sci. USA*, **72**, 4966-4970.
74. Gillies, R.J., Alger, J.R., den Hollander, J.A. and Shulman, R.G. (1982). Intracellular pH measured by NMR: Methods and results, in *Intracellular pH: Its Measurement, Regulation, and Utilization in Cellular Functions* (Nuccitelli, R. and Deamer, D.W., eds.), Liss, New York, 79-104.

75. Gillies, R.J., Ugurbil, K., den Hollander, J.A. and Shulman, R.G. (1981). ^{31}P NMR studies of intracellular pH and phosphate metabolism during cell division cycle of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*, **78**, 2125-2129.
76. Oberhaensli, R.D., Galloway, G.J., Taylor D.J., Bore, P.J. and Radda, G.K. (1986). Assessment of human liver metabolism by phosphorous-31 magnetic resonance spectroscopy. *Br. J. Radiol.* **59**, 695-699.
77. Oberhaensli, R.D., Rajagopalalan, B, Taylor D.J., and Radda, G.K. (1987). Study of hereditary fructose intolerance by use of ^{31}P magnetic resonance spectroscopy. *Lancet* **2**, 931-934.
78. Southern, E.M., Anand, R., Brown, W.R.A. and Fletcher, D.S. (1987). A model for the separation of large DNA molecules by cross field gel electrophoresis. *Nuc. Acids Res.*, **15**, 5925-5943.
79. Jazwinski, S.M. (1990). Preparation of Extracts from Yeast. *Meths. Enzymol.* **182**, 154-174.
80. den Hollander, J.A., Ugurbil, K. and Shulman, R.G. (1986). ^{31}P and ^{13}C NMR studies of intermediates of aerobic and anaerobic glycolysis in *Saccharomyces cerevisiae*. *Biochemistry*, **25**, 212-219.
81. Brindle, K.M. (1988). ^{31}P NMR Magnetization-Transfer Measurements of Flux between Inorganic Phosphate and Adenosine 5'-Triphosphate in Yeast Cells Genetically Modified To Overproduce Phosphoglycerate Kinase. *Biochemistry*, **27**, 6187-6196.
82. Eggleston, L.V. (1983). D-Fructose 1-phosphate, in *Methods of Enzymatic Analysis* (Bergmeyer, H. U., Ed.), Verlag Chemie, Weinheim, vol. 2, 328-334.
83. Brindle, K. and Krikler, S. (1985). ^{31}P -NMR saturation transfer measurements of phosphate consumption in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta*, **847**, 285-292.
84. Sikorski, R.S. and Hieter, P. (1989). A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19-27.
85. Vallejo, C.G. and Serrano, R. (1989). Physiology of mutants with reduced expression of plasma membrane H^+ -ATPase. *Yeast*, **5**, 307-319.
86. Iles, R.A., Griffiths, J.R., Stevens, A.N., Gadian, D.G. and Porteous, R. (1980). Effects of fructose on the energy metabolism and acid-base status of the perfused starved-rat liver. *Biochem. J.*, **192**, 191-202.
87. Phillips, M.I. and Davies, D.R. (1985). The mechanism of guanosine triphosphate depletion in the liver after a fructose load. *Biochem. J.* **228**, 667-671.
88. Lopes, T.S., de Wijs, I.J., van Heerikhuizen, H. and Planta, R.J. (1992). Factors affecting the mitotic stability of pMIRY-type vectors integrating into the ribosomal DNA of *Saccharomyces cerevisiae*. *Yeast*, **8** (Special conference issue), S592.

Chapter 7

Mutagenesis and Screening of the SUC2 Gene

7.1 Introduction

The desired selection system has been successfully engineered to be intolerant of fructose, whilst being unaffected by the presence of sucrose in its growth medium. Expression of libraries of mutant invertase genes by this strain will therefore allow the simple isolation of those conferring a fructotransferase activity. Since the integrated ketohexokinase gene had shown some mitotic instability when the TDB II strain was grown in the presence of fructose, the fructose-intolerant phenotype of the cells isolated growing on sucrose as sole carbon source should be confirmed by their inability to grow in the presence of fructose.

7.1.1 Random Mutagenesis of DNA.

The generation of mutations at a specific site within a nucleotide sequence by *in vitro* mutagenesis has now become routine [reviewed in 1 and 2]. When used in combination with the structural information available for many proteins, site-directed mutagenesis provides a powerful technique for the manipulation of structures and sequences. However, with the exception of a few cases, this technique is often limited by the need to know which sites to investigate within the protein [3]. Often, the lack of tertiary structural detail makes the use of a site specific approach inappropriate, although techniques have been recently developed that allow putative surface residues of a protein to be targetted and subjected to mutagenesis, such as ‘charged-to-alanine’ scanning mutagenesis [4, 5, 6, 7] and ‘random-replacement’ mutagenesis [8, 9, 10]. Both of these techniques have succeeded in isolating functionally impaired mutants, often affecting intra-protein interactions, but both rely on the use of large numbers of custom synthesized oligonucleotides, making them relatively expensive protocols.

A third method has also been developed for the introduction of a single point mutation making use of a misincorporation of a nucleotide by reverse transcriptase [11, 12]. Four populations of DNA molecules are first generated by primer extension so that they terminate randomly, but always just before a known type of base, due to the limitation of this nucleotide in the reaction mixture. The DNA is then purified and resuspended in a separate misincorporation reaction, where the correct base is missing, and misincorporation occurs by a reverse transcriptase-

mediated reaction. This reaction is then chased by adding the missing nucleotide and Klenow fragment to generate a full length transcript. The site of mutation may be controlled to a small degree by the initial limiting concentration of nucleotide, and can be optimised to target a certain region of the gene to be mutated.

Random chemical mutagenesis may also be used to identify and investigate functionally interesting sites within a protein or nucleic acid sequence [3]. Over the past forty years, a large number of substances have been shown to react with nucleic acids, both *in vivo* and *in vitro*, and in so doing induce changes in the sequence of the DNA that are subsequently replicated *in vivo*. However, the majority of these compounds cause various types of damage, and consequently only five of the best understood treatments are currently routinely employed in generating randomly distributed point mutations *in vitro*. These are nitrous acid [13], hydroxylamine [14], sulphurous acid (in the form of a bisulphite salt) [15], hydrazine [16], and mild acid hydrolysis [17].

7.1.1.1 Nitrous acid mutagenesis. The reaction between nitrous acid and the amines of purines and pyrimidines initially leads to the formation of their corresponding diazonium derivatives (see Fig. 7.1) [13]. These are unstable in aqueous solution, and rapidly decompose into their ketolic/enolic hydrolysis products. Thus guanine is converted into xanthine, adenine into hypoxanthine (inosine), and cytosine into uracil. The rates of these deamination reactions vary (in the approximate ratios 1:2:6; guanine:adenine:cytosine) and increase rapidly with decreasing *pH*. On replication of nitrous acid-treated DNA, cytidine is incorporated base paired to hypoxanthine, thymidine base paired with xanthosine, and adenosine base paired to uridine. Nitrous acid therefore generates mutations of the transition (purine to purine and pyrimidine to pyrimidine) type.

7.1.1.2 Hydroxylamine mutagenesis. Under strongly acidic conditions (less than *pH* 2.5), hydroxylamine acts as a powerful hydroxyl donor, the overall result of which is to deaminate cytosine to uracil [14]. The naturally occurring derivatives of cytosine,

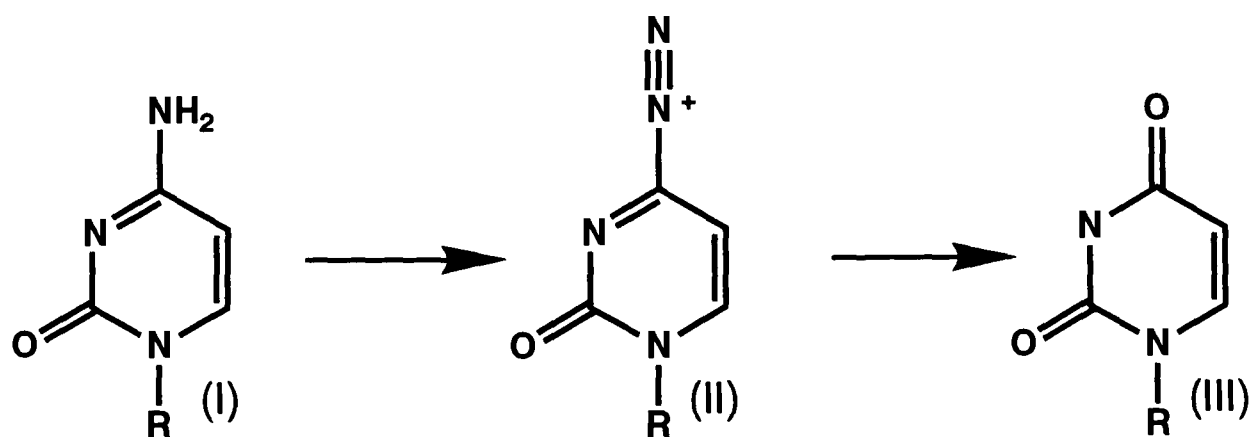


Figure 7.1 Mechanism of nitrous acid mutagenesis of cytosine (I), via a diazonium intermediate (II) to produce uracil (III).

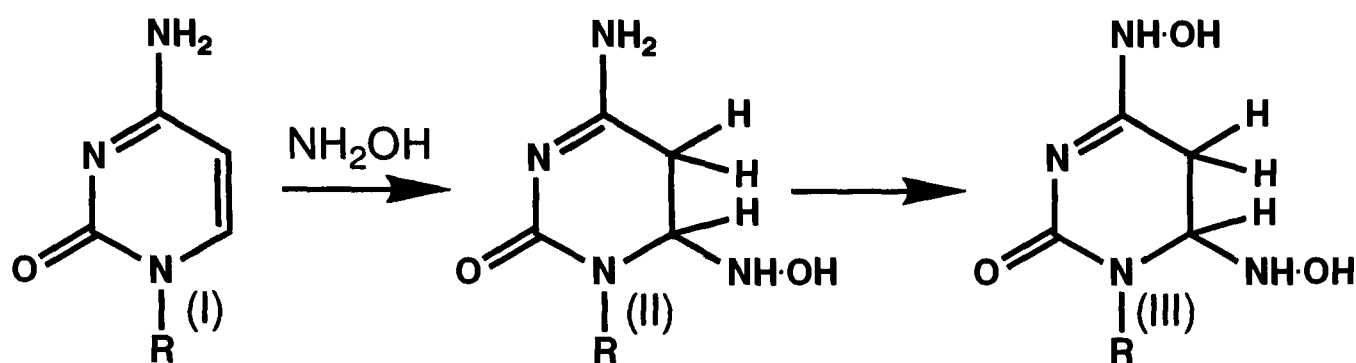


Figure 7.2 Reaction scheme of hydroxylamine treatment of the cytosine moiety of cytosine nucleosides. Cytosine (I) reacts with hydroxylamine to produce 6-hydroxylamino-5,6-dihydrocytosine residues (II) before the generation of hydroxyaminocytosine (III).

5-hydroxymethylcytosine and 5-[glycosyl]-hydroxymethylcytosine are also deaminated by hydroxylamine, although 5-methylcytosine is not. Under milder conditions ($\text{pH } 6$), an intermediate in the deamination of cytosine by hydroxylamine, hydroxyaminocytosine, is formed (see *Fig. 7.2*). This is stable and can form a non-Watson-Crick base-pair with adenine. Both mechanisms of hydroxylamine induced mutation give rise to C-G to T-A transition mutations.

7.1.1.3 Bisulphite mutagenesis. The reaction of bisulphite with cytosine (see *Fig. 7.3*) is approximately three orders of magnitude faster than the deamination reactions between bisulphite and purines, and is therefore only used for such mutations [15]. The treatment of DNA with bisulphite yields the same subset of transition point mutations as treatment with hydroxylamine.

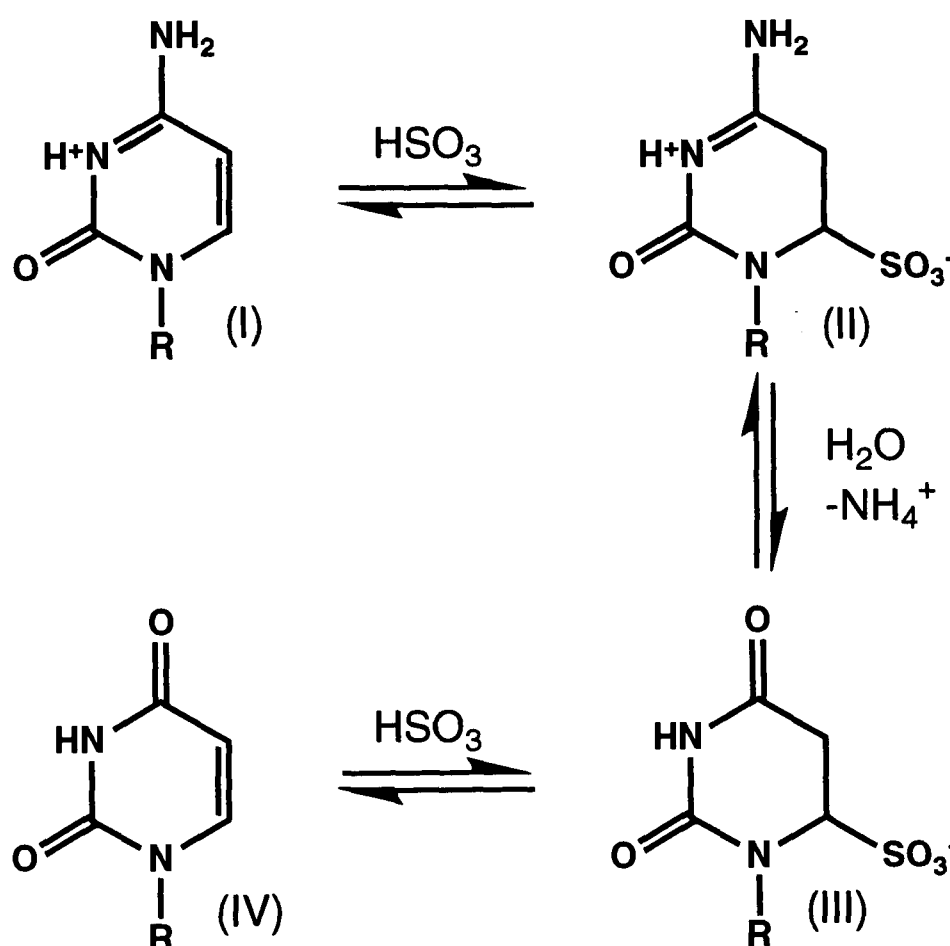


Figure 7.3 Specific deamination of cytosine to uracil by sodium bisulphite treatment. Cytosine (I) undergoes a bisulphite addition (cytosine desaturation, $\text{I} \rightarrow \text{II}$), followed by a deamination ($\text{II} \rightarrow \text{III}$) and an elimination of the bisulphite ($\text{III} \rightarrow \text{IV}$) to give uracil (IV).

7.1.1.4 Hydrazine mutagenesis. Anhydrous hydrazine and aqueous solutions of hydrazine hydrate react specifically with the pyrimidine bases of DNA, leading to cleavage of the heterocyclic ring system (see Fig. 7.4) [16]. Subsequent spontaneous hydrolysis results in depurination of the nucleic acid at the modified base. The actual product left in the polynucleotide chain following treatment of DNA with hydrazine is not deoxyribose, but its 1-hydrazone derivative. If necessary, deoxyribose can be regenerated by treatment of the nucleic acid with benzaldehyde, although this is not usually necessary for the simple formation of point mutations from depyrimidated DNA [3]. Cytosine and thymidine are attacked at differing rates, the latter being three times more susceptible to mutagenesis. The reaction is affected by ionic strength; the higher the cation concentration the faster the relative rate of attack of cytosine.

7.1.1.5 Acid depurination of DNA. The N7-glycosidic bond between the purine bases and the C1 of the deoxyribose ring is far more labile to mild acid hydrolysis than any of the other bonds in DNA [17]. By treatment with mild acid it is therefore possible to excise purines specifically from DNA. In practice, formic acid is generally used because of the ease with which all traces of this acid can subsequently be removed.

7.1.2 Generation of Single-Stranded DNA.

Those mutagens that have proven most useful for controlled mutagenesis react much more rapidly with single-stranded than double-stranded DNA [3]. Consequently, chemical mutagenesis of double stranded DNA is highly non-random, probably due to a differential 'breathing' of the duplex, exposing only certain regions to the mutagenic chemicals. This leads to a distribution of mutations being highly dependant on the nature of the target sequence. Attempts to target the mutations to the target sequence by excision of this fragment and recloning the mutated DNA have been hampered by the preferential mutation of the single-stranded restriction enzyme-generated overhangs, and consequent difficulty of cloning. These problems may be overcome by the initial generation of single-stranded DNA before exposure to the mutagen.

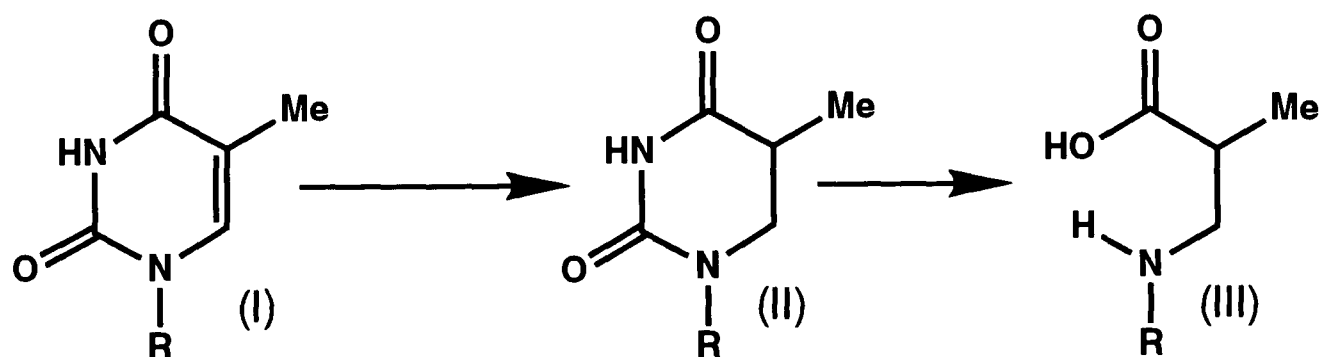


Figure 7.4 Hydrazine mutagenesis of thymidine (I), via 5,6-dihydrothymidine (II), to α -methyl- β -alanine (III).

7.1.2.1 Preparation of single-stranded DNA by filamentous phage. The cloning of the target sequence into a plasmid containing the filamentous phage F1 ori allows the relatively simple preparation of single-stranded DNA (see 5.1.1). The plasmid is used to transform a bacterial strain harbouring an F-factor, and consequent infection of the cells with a filamentous helper phage leads to single-stranded copies of the plasmid being packaged in the bacteriophage coat proteins. The helper phage lyse the bacteria, releasing the encapsulated single-stranded plasmid DNA, which can then be separated from the bacterial debris by centrifugation. The DNA is then purified by phenol/chloroform extraction, yielding the single-stranded DNA [3].

This may be used directly for chemical mutagenesis, but since all the plasmid genes are also potential targets for the mutagens, a large number of the mutations will not affect the target sequence. Such mutations may affect the various plasmid-borne genes, making retransformation of the DNA difficult. Consequently, it is desirable to base pair the plasmid sequence before exposing the DNA to the mutagen. This may be done in two ways. If the host plasmid is available with the F1 ori in both directions (*i.e.* F1 (+) and F1 (-) ori), single-strand copies of both strands of both the plasmid and plasmid with insert are made, and the complementary strands are mixed in equimolar quantities. The plasmid sequences are able to base-pair, leaving a single-stranded target sequence 'looped-out' of the plasmid sequence [18]. In a similar fashion, fragments of the inserted sequence are deleted by restriction enzyme digestion/religation. A heteroduplex is then generated in a similar fashion, and the region sequence deleted from one plasmid is then 'looped-out' [19]. This can then be subjected to 'deletion loop mutagenesis', in which the heteroduplex is exposed to the mutagen (usually sodium bisulphite to give cytosine to uracil mutations), and transformed into a bacterial strain lacking the enzyme uracil *N*-glycosylase (*ung*⁻). The mutated DNA is replicated replacing the mutated cytosine base with thymidine, and transformants are then selected that contain the mutated sequence.

If the plasmid is not available with both F1 origins, the single-stranded DNA can be mixed with excess of the host plasmid, heated to denature the duplex sequence, and the mixture allowed to cool. Some of the single-stranded plasmid will then anneal with the host plasmid to produce a similar deletion loop, which again can be subjected to mutagenesis.

7.1.2.2 Production of single-stranded DNA by asymmetric PCR. This modification of PCR technology utilizes an unequal, or asymmetric, concentration of the two amplification primers [20, 21]. During the initial 15 to 25 cycles, most of the product generated is double-stranded, as the PCR proceeds in a normal fashion. As the low-concentration primer becomes depleted, so further cycles of the PCR generate only the strand for which the primer is in excess. This single-stranded DNA accumulates linearly and is

complementary to the limiting primer. Since the production of the single-stranded DNA is linear once the limiting primer is depleted, the number of cycles is increased, usually to 40, such that the last 15 to 25 cycles produce this product. The ratio of excess of primer for asymmetric PCR is typically 50:1, the limiting primer being a fiftieth of the concentration usually used.

Since this reaction yields a significant amount of double stranded product (albeit 15 to 25-fold less than the single-stranded product), this must first be removed before the DNA can be subjected to chemical mutagenesis. This can be simply done by separation of the products by electrophoresis, the double-stranded product having a greater mobility than the single-stranded DNA. Under short-wave UV light (254nm), the ethidium bromide-stained single-stranded DNA's fluorescence is much reduced compared with double-stranded DNA, due to lower intercollation of the dye.

PCR induced errors [22] in this process do not represent a problem when the DNA is used for chemical mutagenesis. Following chemical mutagenesis, the complementary strand is synthesized using reverse transcriptase, this enzyme being able to introduce the appropriate base on the complementary strand when the point mutation is reached. The DNA may then be cloned prior to screening for the mutant activity.

7.1.3 Design of the Yeast Vector for *SUC2* Expression.

A similar set of criteria were used in the design of the second yeast vector for the expression of the mutated *SUC2* gene. In wild-type yeast, the *SUC2* gene is present only in a single copy in most cases, and so a single copy plasmid was thought to be adequate. The yeast *ARS/CEN* provided a convenient origin of replication, which also conveys very high mitotically stability to the plasmid (see 6.1.2.4). Standard selection markers (*i.e.* not those with deletions in their promoters) may be used with such an origin of replication, and the *TRP1* marker was selected to complement the *trp1* phenotype of TDB II. These two sequences are both present of the *pRS314* plasmid [23], which also has the bacterial sequences required for selection in *E. coli*. For the same reasons as mentioned in the design of *pIAD1*, a constitutive promoter was used, unlike the native *SUC2* promoter which is repressed by the presence of glucose. The *ADH1* expression cassette in the *pVT-L* plasmid [24] was selected for this purpose, which also has the advantage of a multiple cloning site between the 5' and 3' ends, containing the recognition sequences of seven enzymes.

7.2 Materials and Methods

7.2.1 Yeast Plasmids.

The *SUC2* yeast invertase gene was obtained from the *pRB58* [25] plasmid from the stocks of Dr. A. Greenland, Cranfield Institute of Biotechnology, University of Leicester. It was originally constructed by ligating fragments of greater than 10kb from a *Sau3A* I partial digest of yeast chromosomal DNA into *YEp24* and selecting for ability to condition invertase activity in a *suc^o* strain.

Two plasmids, *pVT-L* [24] and *pRS314* [23], were used in the construction of the expression vector *pIAD3*. The *pVT-L* plasmid is a yeast multi-copy plasmid with a 2 μ origin of replication and *leu2-d* auxotrophic marker selection system. It also contains the 5' and 3' flanking regions of the *ADH1* gene, linked by a multiple cloning site (MCS-7) containing the recognition sites for seven restriction endonucleases. The *pRS314* plasmid is a yeast / *E. coli* shuttle vector based on the *pBLUESCRIPT* plasmid with the yeast *TRP1* gene and *ARSH4/CEN6* origin of replication, which maintains the plasmid at single copy-number in yeast.

7.2.2 Direct Cloning of PCR Products by dT-Tailing the Plasmid.

The previous method for direct cloning of a PCR product involved first blunt-ending it and cloning into a blunt-ended plasmid (4.2.7.4 and .5). This is performed to remove the extra dA residue added to the PCR products 3' ends by *Taq* DNA polymerase. However, the PCR product may be ligated directly into a plasmid, if the plasmid is cut first to generate a blunt end, and then a complementary dT residue added to its 3' ends [modification of 26].

The plasmid was treated with a restriction enzyme which generates blunt ends (*e.g.* *Sma* I), and the reaction allowed to go to completion (4.2.7.2). The DNA was precipitated (4.2.7.1) and redissolved in

2 μ l.	10 $\times$ <i>Taq</i> reaction buffer (Promega)
1 μ l.	5mM. dTTP
0.5 μ l.	<i>Taq</i> DNA polymerase (5U/ μ l.)
16.5 μ l.	Water.

This was overlayed with paraffin oil, and incubated at 72°C for two hours. The reaction was cooled to room temperature, and ATP to 1mM. and 10U T4 polynucleotide kinase were added. This was mixed and incubated at 37°C for 30 minutes.

In a similar fashion, the amplified DNA was phosphorylated (4.2.7.4), and both insert and plasmid were purified using Promega 'Magic PCR Preps' DNA purification cartridges

according to manufacturers instructions, the DNA being eluted in 50 μ l. TE. Ligations were performed as described for 'sticky-ended' DNA (4.2.7.6).

7.2.3 *Direct Blotting of Bacterial DNA on to Charged Nylon Membranes.*

Cells to be screened were plated on to LB medium, containing ampicillin, on a 14cm plate and grown at 37°C until colonies were 1-2mm across. The plates were chilled at 4°C for two hours before placing the charged nylon membrane on to the colonies. The position of the filter relative to the plate was recorded for realignment by making holes through the filter into the agar with a fine needle and ink. The filters were left for 1 minute before lifting and placing on to filter paper soaked in 1% SDS, 0.5M. sodium hydroxide, 1.5M. sodium chloride, colony side up, for 1 minute. The filters were then neutralised with 0.5M. Tris.HCl, pH 7.4, 1.5M. sodium chloride for 1 minute, and then finally washed twice with 100mM. sodium phosphate, pH 7.2, 1.4M. sodium chloride for 2 minutes each. In all cases, the buffers were soaked into pads of filter paper, and the nylon membranes laid colony side up on to them.

The membranes were then dried, and probed with radioactively labelled DNA (5.2.2.2) as described before (5.2.2.3) to detect colonies possessing plasmids with the desired insert.

7.2.4 *Chemical In Vitro Mutagenesis of Single-Stranded DNA.*

7.2.4.1 *Amplification and purification of single-stranded DNA by asymmetric PCR.* Single-stranded DNA may be generated by using unequal, or asymmetric, concentrations of the two amplification primers. Generally, one is retained at the usual concentration (500pmol/ml.) and the other is present at a fiftieth of this. The PCR reaction mix was therefore made up in a similar fashion to the usual PCR amplification, except with this slight alteration:

10 μ l.	10 $\times$ amplification buffer (Promega)
1 μ l.	5mM. dNTP mixture
50pmol	Excess primer
1pmol	Limiting primer
2.5U	<i>Taq</i> DNA polymerase
1 μ l.	Template DNA
To 100 μ l.	Water

The target DNA was then amplified in the usual way, except 40 cycles of PCR were performed. By performing two reactions in parallel, but using opposite asymmetry of the primers in each reaction, single-stranded DNA can be amplified for both strands. The DNA fragments were then separated on a 1% LMP agarose gel (4.2.7.7). The single-stranded DNA runs slower than the double-stranded DNA product, and could easily be isolated from the double-stranded product by cutting its band out of the gel and purifying the DNA (4.2.7.8).

The DNA concentration is then estimated spectrophotometrically using 40µg/ml. single-stranded DNA gives an $A_{260}=1$.

7.2.4.2 Preparation of Sephadex G50 spun columns [27]. The barrel of a 1ml. syringe was removed and the end of the barrel plugged with a small piece of sterile glass wool. 1ml. of a sterile slurry of Sephadex G50 [Pharmacia LKB Biotechnology] in TE buffer was then added to the plugged barrel, and then placed in a 15ml. polystyrene Falcon tube. This was centrifuged for 3 minutes at 1600g at 4°C. The eluted buffer was removed from the tube, further slurry added and the centrifugation repeated. This was continued until the volume of the packed column was 0.9ml. The column was then washed five times with 100µl. TE buffer by centrifugation. All the eluted buffer was removed, and a sterile, capless Eppendorf tube placed at the bottom of the Falcon tube. A 100µl. DNA sample could then be loaded to the column, centrifuged as before, and the DNA eluted, in the void volume, in to the Eppendorf tube.

7.2.4.3 In vitro mutagenesis with formic acid. The method used was designed to damage one base per 1500 nucleotides in 10 minutes, and since it was intended to introduce multiple mutations in a target DNA 1900 nucleotides, a range of reaction times was used [3, 28]. 40µg single stranded DNA was mutated in 100µl. 12M. formic acid. (The DNA is diluted to 1µg/µl., and 40µl. of this was added to 60µl. 18M. formic acid). The reaction was allowed to proceed at room temperature, and 20µl. samples were taken after 0 hours (immediately after addition of acid), 1, 2, 4, and 8 hours, when the reaction was stopped by the addition of 20µl. 2.5M. sodium acetate, pH 4.2, 4µg tRNA and 40µl. water. The DNA was then precipitated by the addition of 1ml. absolute ethanol, before reprecipitation with ethanol two more times to remove all traces of mutagen.

7.2.4.4 In vitro mutagenesis with hydrazine. To induce a similar degree of damaged bases as the formic acid method, this method was repeated for hydrazine mutagenesis adding 60µl. 12M. hydrazine instead of formic acid. Otherwise, the reaction conditions were the same [3, 28].

7.2.4.5 In vitro mutagenesis with sodium bisulphite. 50µg single-stranded DNA was dissolved in 68µl. TE buffer and 320µl. 4M. sodium bisulphite reaction mixture (156mg sodium bisulphite and 64mg sodium sulphite dissolved in 430µl. water) and 12µl. 50mM. hydroquinone were added [29]. After 4, 9, and 23 hours at 37°C in the dark, 100µl. samples were taken and applied to a 1ml. Sephadex G50 spun column (see 7.2.4.2), and the void volume containing the DNA was collected and adjusted to 0.2M. Tris.HCl, pH 9.2, 50mM. sodium chloride, 2mM. EDTA. After a further 8-16 hours at 37°C, the samples were again applied to a G50 spun column, and the DNA precipitated by the addition of a one tenth volume 2.5M. sodium acetate, pH 4.5, and 2.5 volumes ethanol.

7.2.4.6 Amplification of mutated DNA for ligation. The complementary strand of the mutated DNA was synthesized using the primer that was limiting in the single-stranded DNA synthesis method. Second strand synthesis was catalysed by Moloney Murine Leukemia Virus reverse transcriptase (M-MLV RT) according to manufacturers instructions.

50mM.	Tris.HCl, pH 8.3
75mM.	Potassium chloride
10mM.	Dithiothreitol
3mM.	Magnesium chloride
0.5mM.	Each dNTP
10µg/ml.	Limiting primer from PCR step
20µg/ml.	Mutated single-stranded DNA
200U	M-MLV RT
To 20µl.	Water.

This was incubated at 37°C for 60 minutes before phenol/chloroform, chloroform extraction and ethanol precipitation. The DNA was then dissolved in 10µl TE buffer, and 1µl. used for amplification by PCR, before ligation into a plasmid.

7.3 Results and Discussion

7.3.1 Cloning and Expression of the SUC2 Gene for Invertase.

7.3.1.1 Construction of the yeast expression vector, *pIAD3*. A single copy constitutive expression vector was constructed for the synthesis of the *SUC2* gene as shown in Fig. 7.5. The plasmid was based on the *pRS314* yeast shuttle vector, which carries the *TRP1* selection marker, and the yeast *CEN6/ARS4* origin of replication. Most of the *E. coli* gene encoding the β-galactosidase was initially removed by complete digestion with *Pvu* II, generating blunt ends, which were treated with alkaline phosphatase to remove terminal phosphates (4.2.7.3). The *ADHI* constitutive expression cassette was removed from the *pVT-L* plasmid by a *Sph* I digestion, the termini were blunt-ended with the Klenow fragment (4.2.7.5), and purified on a LMP-agarose gel (4.2.7.8). This fragment was then ligated into the *Pvu* II-treated *pRS314* plasmid (4.2.7.6). This plasmid was termed *pIAD3*, and contained five unique sites in the expression cassette into which a gene may be cloned for expression; *Bam*H I, *Sst* I, *Xho* I, *Pst* I and *Pvu* II.

7.3.1.2 Cloning of the *SUC2* coding region. Due to the large size of the *pRB58* plasmid (c. 15.4kb), it was decided to remove the *SUC2* gene and flanking regions and ligate this smaller fragment into *pUC18*. The *pRB58* was therefore digested to completion with *Eco*R I, generating four fragments (approximately 5.6, 4.5, 3.3 and 2.2kb long), two of which were ligated into the *Eco*R I site of *pUC18*. Analysis of these ligations revealed that the *SUC2* gene was present on the 4.5kb fragment (see Fig. 7.6). These plasmids were termed

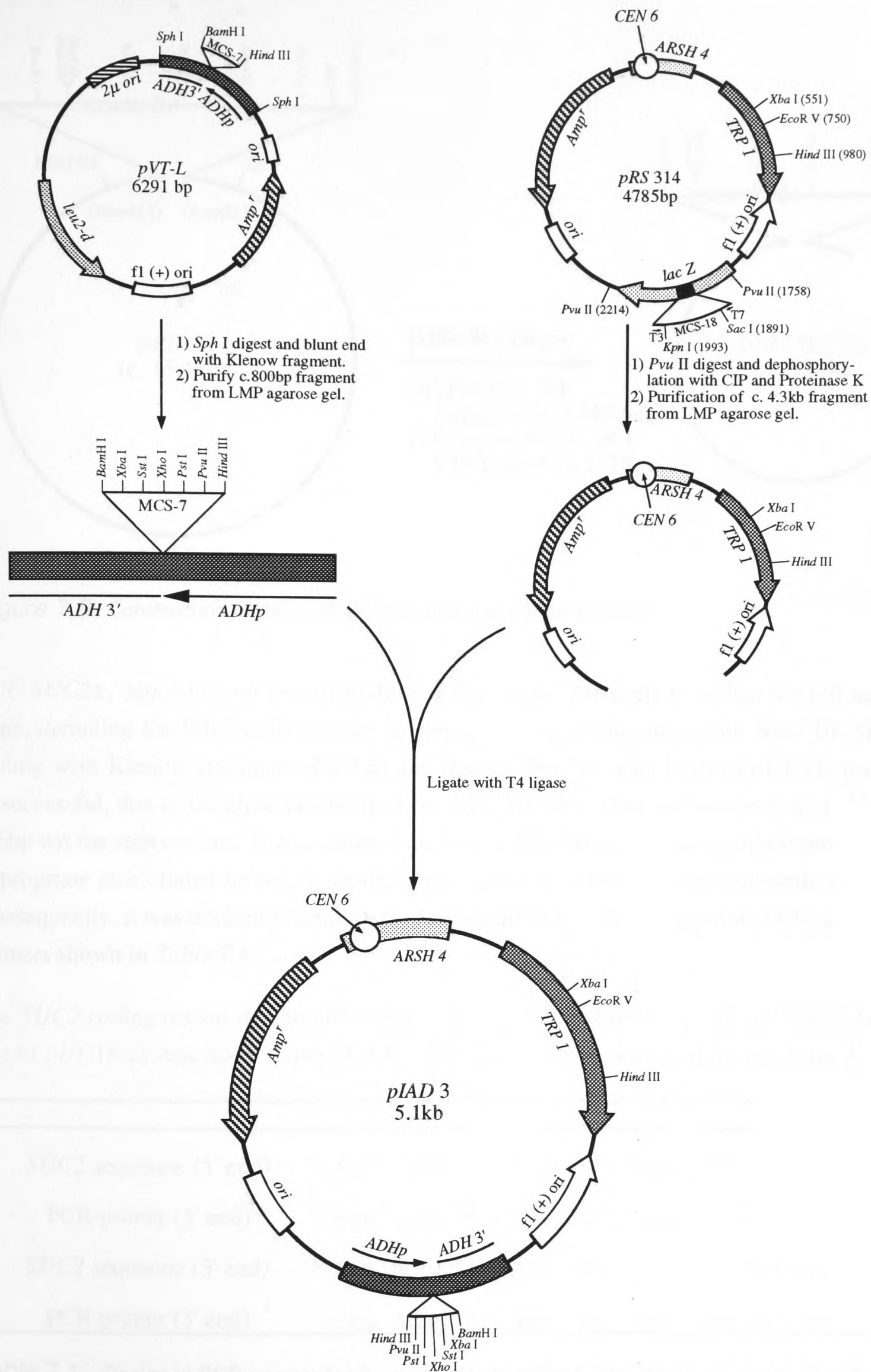


Figure 7.5 Strategy for the construction of the *pIAD3* plasmid. The *ADH1* expression cassette from the *pVT-L* plasmid was ligated into the *Pvu* II sites of the *pRS314* plasmid.

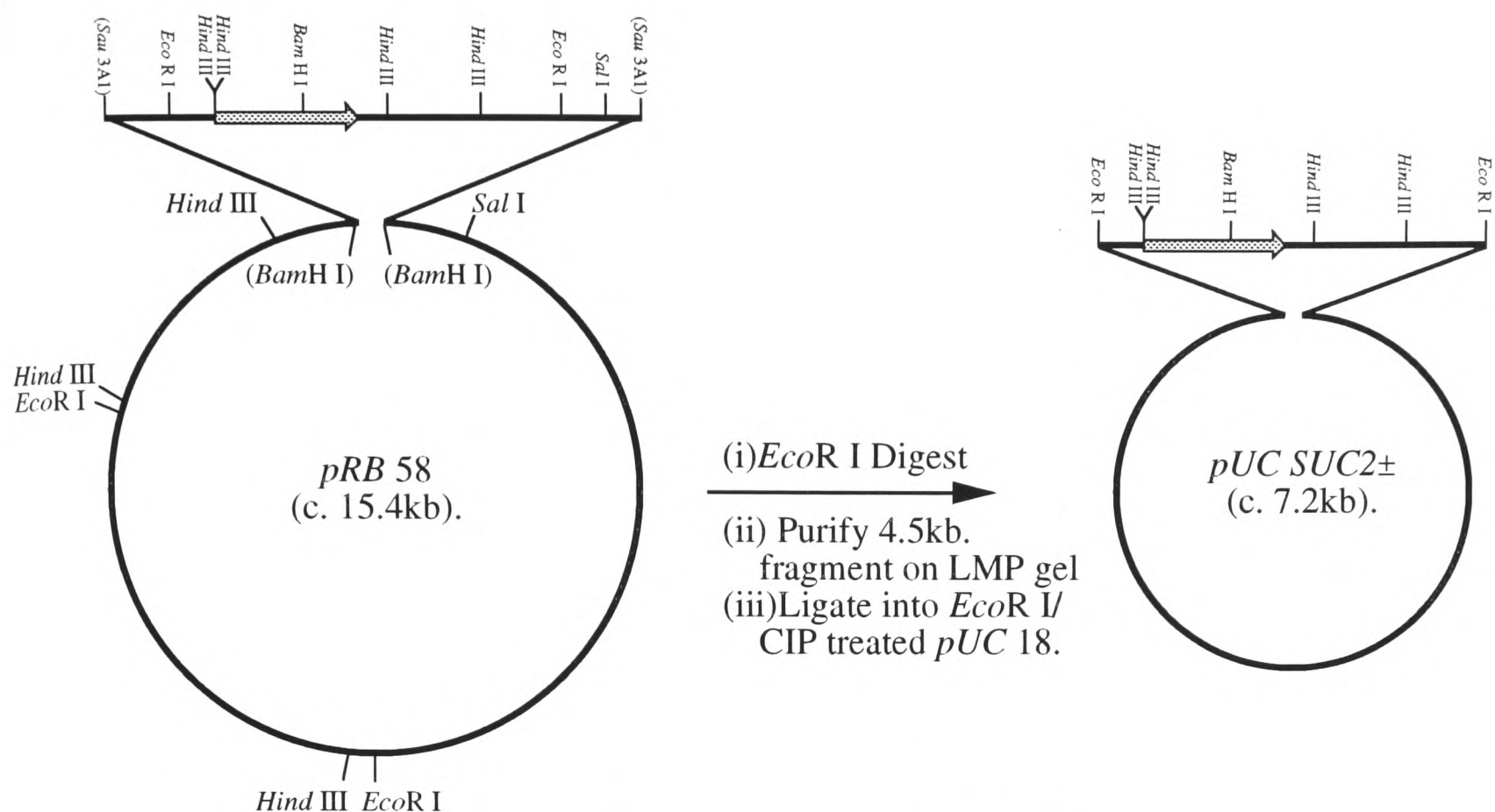


Figure 7.6 Construction of the *pUC SUC2±* from the *pRB58* plasmid.

pUC SUC2±, depending on the orientation of the insert. Attempts to isolate the full length gene, including the full length secretor sequence, by partial digestion with *Hind* III, blunt-ending with Klenow fragment (4.2.7.5) and ligation into the *Sma* I site of *pUC18*, proved unsuccessful, due to the close positions of the *Hind* III sites at the start of the gene (+13 and -26bp wrt the start codon). It was shown by a further *Hind* III digest that all digest products of appropriate size cloned to be the shorter gene, encoding a truncated, cytoplasmic product. Consequently, it was decided to amplify the full length *SUC2* coding region by PCR, using the primers shown in Table 7.1.

The *SUC2* coding region was amplified by PCR and cloned directly into the dT-tailed *Sma* I site of *pUC18* as described above (7.2.2). The ligated DNA was used to transform *E. coli*

<i>SUC2</i> sequence (5' end)	5'AGC TTT TCT TTT CAC TAA CGT ATA TG ^{3'}
PCR primer (5' end)	5' <i>ctg aga TCT</i> TTT CAC TAA CGT ATA TG ^{3'}
<i>SUC2</i> sequence (3' end)	5'AAA AGG TCC ATC CTA GTA GTG TAA GG ^{3'}
PCR primer (3' end)	5' <i>ctg AGa TCT</i> ATC CTA GTA GTG TAA GG ^{3'}

Table 7.1 Design of PCR primers for the amplification of the full length *SUC2* gene. The start codon of the signal peptide lies immediately 3' to the 5' primer, and the stop codon lies 42 base pairs upstream to the 3' primer. The introduced *Bgl* II sites are italicised, and the miss-matched bases in lower case.

(4.2.4), and the presence of the *SUC2* gene was shown by radioactively probing nylon membrane replicas of the transformed cells (7.2.3). All colonies found to be bearing plasmids containing the *SUC2* gene were rescued, and the orientation of the inserts determined by *Hind* III and *Bam*H I restriction enzyme analysis. Plasmids in which the *SUC2* gene was orientated in the opposite direction to the *lacZ* gene were termed *pUC SUC2-*, and those in the same direction, *pUC SUC2+*.

7.3.1.3 Expression of wild-type invertase. The unique restriction enzyme sites flanking the cloned *SUC2* gene within the *pUC18* plasmid allowed directional cloning of the gene into the *pIAD3* expression vector to generate a plasmid termed *pIAD3-SUC2+* (see Fig 7.7). A *Pst* I/*Sst* I double digest of the *pUC SUC2-* clones released a fragment encoding invertase that could be cloned into the same sites within *pIAD3* to allow correct transcription for expression. Since the PCR amplification of the *SUC2* gene could potentially introduce nucleotide misincorporation into the sequence, due to the infidelity of *Taq* DNA polymerase [22], ten of the *pUC SUC2-* clones were used to generate *SUC2*-containing *pIAD3* plasmids. These could then be used to transform the TDS II yeast to confirm their ability to express active invertase. A similar *Pst* I/*Sst* I-generated fragment from a *pUC SUC2+* plasmid was cloned into *pIAD3* to create a plasmid in which the *SUC2* gene was inserted an orientation in the expression cassette that does not allow correct transcription. This plasmid was termed *pIAD3 SUC2-* (see Fig 7.7), and was also used to transform the TDS II strain, as was the *pIAD3* plasmid with no inserted sequence.

Only TDS II cells (*i.e.* the parent strain into which the KHK gene was introduced to form the TDB strains) transformed with *pIAD3-SUC2+* were able to grow on synthetic medium supplemented with sucrose as sole carbon source (SS *trp*⁻). No growth was observed in cells transformed with either *pIAD3-SUC2-* or *pIAD3*, confirming that the ability to utilize sucrose as a carbon source was conferred by the episomally-borne *SUC2*. Of the ten different *pIAD3-SUC2* plasmids used to transform TDS II, all expressed an active invertase. Although PCR-induced nucleotide misincorporation cannot be ruled out in these invertase genes, their active products confirm the suitability of these clones to be subjected to *in vitro* chemical mutagenesis.

In a similar fashion, TDB II cells were also transformed with the *pIAD3-SUC2+*, *pIAD3-SUC2-* and *pIAD3* plasmids. Transformants were plated on SD *leu*⁻,*trp*⁻ plates supplemented with Aristar sucrose (6.2.6.4), and no growth was observed in cells transformed with *pIAD3-SUC2+*, although growth of cells transformed with either of the other two plasmids were unaffected. This confirmed the suitability of TDB II as a selection strain to screen for mutants of the invertase gene that possess a fructotransferase activity, since the fructose released by genes expressing a fructohydrolase activity was capable of suppressing

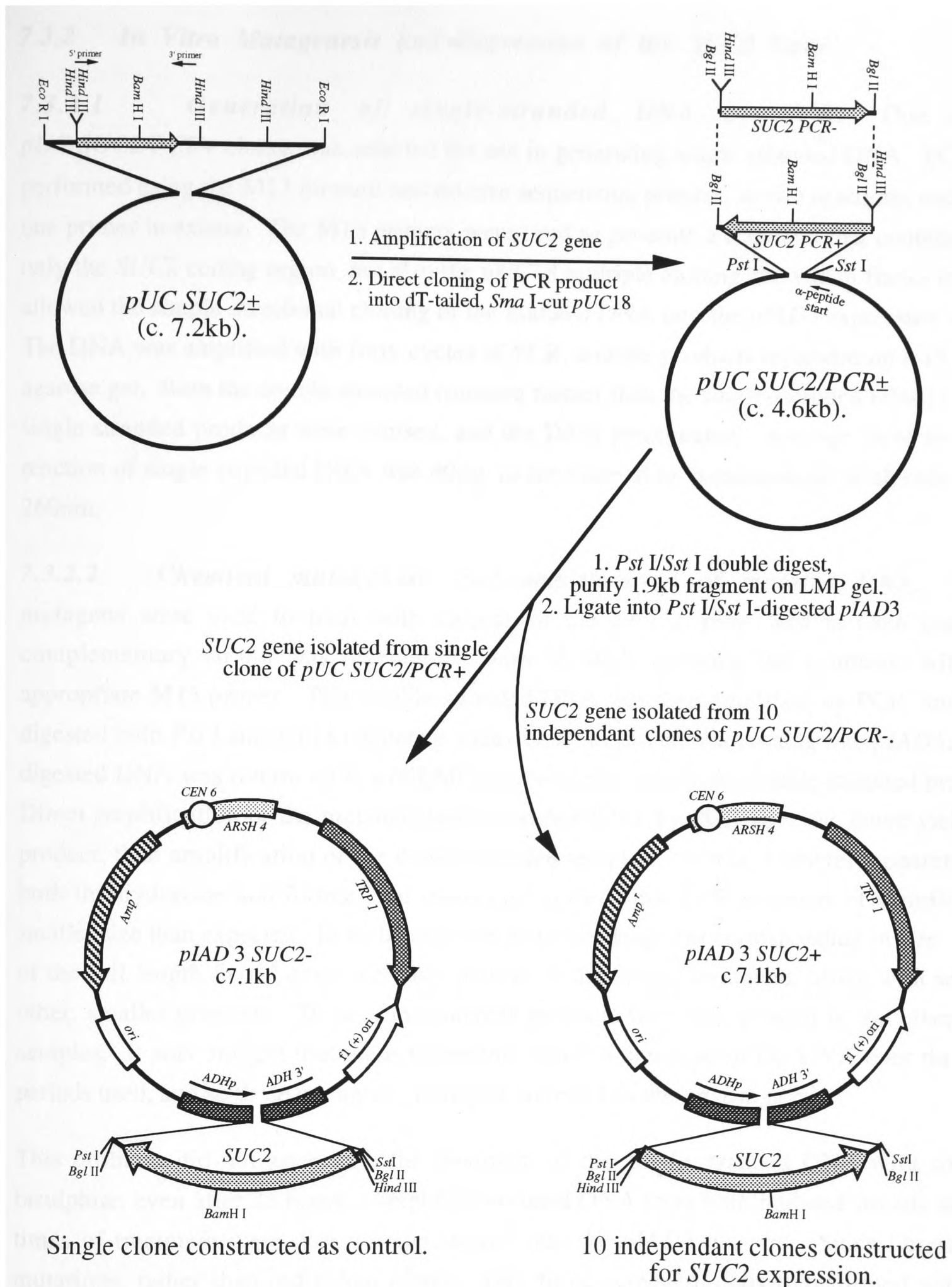


Figure 7.7 Directional cloning of the *SUC2* gene into the *pIAD3* expression cassette.

growth. Since a degree of mitotic instability of the ketohexokinase gene had been observed for cells exposed to fructose, revertants to fructose tolerance could easily be identified by screening such cells on SD^{1/2}F plates.

7.3.2 *In Vitro* Mutagenesis and Expression of the *SUC2* Gene.

7.3.2.1 *Generation of single-stranded DNA by PCR.* One of the *pUC SUC2/PCR+* clones was selected for use in generating single stranded DNA. PCR was performed using the M13 forward and reverse sequencing primers, in two reactions, each with one primer in excess. The M13 primers were used to generate a fragment that contained not only the *SUC2* coding region, but also the *pUC18* multiple cloning site which flanks it. This allowed the simple directional cloning of the mutated DNA into the *pIAD3* expression vector. The DNA was amplified with forty cycles of PCR, and the products separated on a 1% LMP agarose gel. Both the double-stranded (running further than the single-stranded DNA) and the single stranded products were excised, and the DNA precipitated. Average yield per PCR reaction of single stranded DNA was 40µg, as determined by measurement of absorbance at 260nm.

7.3.2.2 *Chemical mutagenesis and amplification of mutated DNA.* Three mutagens were used to treat both strands of the *SUC2* gene, and in each case the complementary strand was regenerated with M-MLV, priming the synthesis with the appropriate M13 primer. This double-stranded DNA was then amplified by PCR, and then digested with *Pst* I and *Sst* I to generate sticky ends for directional cloning into *pIAD3*. This digested DNA was run on a 1% w/v LMP agarose gel to purify the double stranded product. Direct amplification of the mutated single-stranded DNA by PCR gave far lower yields of product, than amplification of the double-stranded template. It was, however, apparent that both the hydrazine and formic acid treatment yielded final PCR products of significantly smaller size than expected. In both cases, the amplified fragment corresponding in size to that of the full length *SUC2* gene was only present in the 0 hour treatment, along with several other, smaller products. These same, smaller products were also present in the other time samples. It was thought that these treatments induced breakage of the DNA over the time periods used, and so these mutagenic strategies were not further investigated.

This problem did not arise with the treatment of the single-stranded DNA with sodium bisulphite, even after 23 hours. Amplified mutated DNA from both mutated strands and all times of treatment were directionally ligated into the *pIAD3* plasmid. Since libraries of mutations, rather than individual clones, were to be screened, cells transformed with the ligation mixture were not plated to isolate individual colonies (excepting for a small aliquot to determine efficiency of transformation), but were kept pooled and used for miniprep plasmid purifications directly. These libraries of mutated *SUC2* genes were then used to transform the TDB II cells.

7.3.2.3 *Expression of mutant *SUC2* genes and screening for fructotransferase activity.* Eight libraries of bisulphite-mutated *SUC2* genes (four time samples on both

strands) were used to transform the TDB II cells, and transformants spread on either SD leu⁻, trp⁻ or SS leu⁻, trp⁻ plates. Yeast colonies growing on SD plates were replica-plated on SS leu⁻, trp⁻ plates, and any colonies growing on these or the transformants plated directly on SS leu⁻, trp⁻ plates were replica-plated on SD^{1/2}F leu⁻, trp⁻ plates. All potential fructotransferase mutants identified by their ability to grow on sucrose as sole carbon source were, by this method, shown to have lost the ketohexokinase gene. No fructotransferase mutants were identified by this preliminary screen.

Although no transferase mutants were identified by this random chemical mutagenesis strategy, it is clear that this selection system offers an incredibly simple screen for such mutants. The probability of isolating a fructotransferase mutant which had lost all fructohydrolase activity in the small libraries used (it was estimated from the number of *E. coli* transformants of mutated *SUC2* genes that each library contained about 100 different clones) was very small, but a more extensive mutagenesis strategy could easily identify such a mutant with this selection system.

7.4 References

1. Leatherbarrow, R.J. and Fersht, A.R. (1986). Protein engineering. *Prot. Engin.*, **1**, 7-16.
2. McPherson, M.J. (1991). Directed Mutagenesis, A Practical Approach. IRL Press, Oxford.
3. Walton, C., Booth, R.K. and Stockley, P.G. (1991). Random chemical mutagenesis and the non-selective isolation of mutated DNA sequences *in vitro*, in *Directed Mutagenesis, A Practical Approach* (McPherson, M.J., Ed.), IRL Press, Oxford, 135-162.
4. Cunningham, B.C. and Wells, J.A. (1989). High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science*, **244**, 1081-1085.
5. Bennett, W.F., Paoni, N.F., Keyt, B.A., Botstein, D., Jones, A.J., Presta, L., Wurm, F.M. and Zoller, M.J. (1991). High resolution analysis of functional determinants on human tissue-type plasminogen activator. *J. Biol. Chem.*, 5191-5201.
6. Wertman, K.F., Drubin, D.G. and Botstein, D. (1992). Systematic Mutational Analysis of the Yeast *ACT1* Gene. *Genetics*, **132**, 337-350.
7. Gibbs, C.S. and Zoller, M.J. (1991). Identification of Functional Residues in Proteins by Charged-to Alanine Scanning Mutagenesis. *Methods*, **3**, 165-173.
8. Palzkill, T. and Botstein, D. (1991). Extracting Information from Protein Sequences Using Random Replacement Mutagenesis. *Methods*, **3**, 155-164.
9. Palzkill, T. and Botstein, D. (1992). Probing β -Lactamase Structure and Function Using Random Replacement Mutagenesis. *Proteins*, **14**, 29-44.
10. Palzkill, T. and Botstein, D. (1992). Identification of Amino Acid Substitutions That Alter the Substrate Specificity of TEM-1 β -Lactamase. *J. Bacteriol.*, **174**, 5237-5243.

11. Lehtovaara, P.M., Koivula, A.K., Bamford, J. and Knowles, J.K.C. (1988). A new method for random mutagenesis of complete genes; enzymatic generation of mutant libraries *in vitro*. *Prot. Eng.*, **2**, 63-68.
12. Holm, L., Koivula, A.K., Lehtovaara, P.M. Hemminki, A. and Knowles, J.K.C. (1990). Random mutagenesis used to probe the structure and function of *Bacillus stearothermophilus* alpha-amylase. *Prot. Eng.*, **3**, 181-191.
13. Schuster, H. (1960). The reaction of nitrous acid with deoxyribonucleic acid. *Biochem. Biophys. Res. Acta*, **2**, 320-323.
14. Lawley, P.D. (1967). Reaction of Hydroxylamine at High Concentration with Deoxycytidine or with Polycytidylic Acid: Evidence that Substitution of Amino Groups in Cytosine Residues by Hydroxylamino is a Primary Reaction, and the Possible Relevance to Hydroxylamine Mutagenesis. *J. Mol. Biol.*, **24**, 75-81.
15. Shapiro, R., Braveman, B., Louis, J.B. and Servis, R.E. (1973). Nucleic Acid Reactivity and Conformation. II. Reaction of Cytosine and Uracil with Sodium Bisulphite. *J. Biol. Chem.*, **248**, 4060-4064.
16. Brown, D.M., McNaught, A.D. and Schell, P. (1966). The Chemical Basis of Hydrazine Mutagenesis. *Biochem. Biophys. Res. Commun.*, **24**, 967-971.
17. Shapiro, H.S. and Chargaff, E. (1957). Studies on the nucleotide arrangement in deoxyribonucleic acids. II. Differential analysis of pyrimidine nucleotide distribution as a method of characterization. *Biochim. Biophys. Acta*, **26**, 608-623.
18. Kalderon, D., Oostra, B.A., Ely, B.K. and Smith, A.E. (1982). Deletion loop mutagenesis: a novel method for the construction of point mutations using deletion mutants. *Nuc. Acids Res.*, **17**, 5161-5171.
19. Peden, K.W.C. and Nathans, D. (1982). Local mutagenesis within deletion loops of DNA heteroduplexes. *Proc. Natl. Acad. Sci. USA*, **79**, 7214-7217.
20. Gyllenstein, U.B. and Erlich, H.A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. *Proc. Natl. Acad. Sci. USA*, **85**, 7652-7656.
21. McCabe, P.C. (1990). Production of single-stranded DNA by asymmetric PCR, in *PCR Protocols: A Guide to Methods and Applications* (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., Eds.), Academic Press, San Diego, 76-83.
22. Tindall, K.R. and Kunkel, T.A. (1988). Fidelity of DNA Synthesis by the *Thermus aquaticus* DNA Polymerase. *Biochemistry*, **27**, 6008-6013.
23. Sikorski, R.S. and Hieter, P. (1989). A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19-27.
24. Vernet, T. Dignard, D. and Thomas, D.Y. (1987). A family of yeast expression vectors containing the phage f1 integenic region. *Gene*, **52**, 225-233.
25. Carlson, M. and Botstein, D. (1982). Two differentially Regulated mRNAs with Different 5' Ends Encode Secreted and Intracellular Forms of Yeast Invertase. *Cell*, **28**, 145-154.
26. Holton, T.A. and Graham, M.W. (1991). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nuc. Acids Res.*, **19**, 1156.

27. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual*, 2nd Edn, Cold Spring Harbor Laboratory Press.
28. Myers, R.M., Lerman, L.S. and Maniatis, T. (1985). A General Method for Saturation Mutagenesis of Cloned DNA Fragments. *Science*, **229**, 242-247.
29. Weiher, H. and Schaller, H. (1982). Segment-specific mutagenesis: Extensive mutagenesis of a *lac* promoter/operator element. *Proc. Natl. Acad. Sci. USA*, **79**, 1408-1412.

General Conclusions: The Future Use for the Fructose-Intolerant Yeast

8.1 SUC2 Screening Strategy

Although no fructotransferase mutants of the *SUC2* gene were isolated in the preliminary screen attempted, the suitability of the selection system is clear. The small libraries of mutations screened, unsurprisingly, did not reveal any interesting mutants, and it is likely that a more extensive screen would be required to succeed in these aims.

To isolate fructotransferase mutants, much larger libraries would have to be constructed, and to achieve this, the chemical mutation methods would have to be optimised. The formic acid and hydrazine methods were the least promising of those attempted, but even prolonged bisulphite mutagenesis did not degrade the DNA so that it could not be amplified by PCR. Whether treatment with hydroxylamine or nitrous acid could successfully expand the range of mutations, remains to be investigated. The limiting factor in the initial construction of the library would be the transformation of the DNA into the *E. coli* host. Again, this could be optimised by electro-transformation, which can yield transformation efficiencies of 10^{12} transformants/ μ g DNA. These libraries could then be used to transform the TDB II strain, and the desired mutants selected.

Following isolation of mutants showing fructotransferase activity that have been generated by random mutagenesis, these would be characterized to determine whether they show any mutational 'hot-spots' that confer fructotransferase activity. If a region is identified, this might then be targetted for by, first, random mutagenesis of this region and then, finally, site-directed mutagenesis. In order to identify such a region, an analytical system based on recombination of the mutant plasmid with a set of nested deletions of the wild-type gene would be employed (see 8.2). Since the secondary or tertiary structure of invertase are not known, it is unclear whether amino acid residues involved in determining the substrate and catalytic specificity have a close proximity in the primary structure, or require tertiary folding to bring them together. Consequently, although it is not possible to predict whether mutations of interest will lie in a small region within the gene, or are scattered over the total gene, the characterization of mutations in this way may shed some light on this.

Primer 1	5' TTC ACT AAC GTA TAT GAT GC ³ '
Primer 2	5' GTA CCC CAT ACG GTG TCA TT ³ '
Primer 3	5' CAC TCC TGA AAG TGA AGA GC ³ '
Primer 4	5' GGA CTT CAA GTC ATC AGA GG ³ '
Primer 5	5' GGA TCC TTC AAT GGT ACT CA ³ '
Primer 6	5' TAG CTT GAT ATT CAG TGT TC ³ '
Primer 7	5' CTG TTA ACA CCA CAC AAA CC ³ '
Primer 8	5' GGT CGT TCT CAG ACT TGA AT ³ '
Primer 9	5' AGC AAC GAG TGA CGA ATG TA ³ '
Primer 10	5' CTG AGA TCT ATC CTA GTA GTG TAA GG ³ '

Table 8.1 Sequence of the proposed ten primers required for the characterization of the *SUC2* mutations. Primer 10 is the same as used to amplify the *SUC2* region (see Table 7.1). Primer 1 is also in our possession.

8.2 A Strategy for the Characterization of *SUC2* Mutants

A convenient technique for the rapid determination of the site of mutation within the *SUC2* gene involves the reversion to wild-type activity of the mutant enzyme following transformation and recombination of short sequences of the wild-type *SUC2* gene, in a similar fashion to that described by Kunes *et al.* [1]. By synthesizing eight new oligonucleotides (see Table. 8.1), and using two primers already synthesized to amplify *SUC2*, it would be possible to amplify fifteen fragments of varying length that cover the *SUC2* gene (see Fig. 8.1), using the wild-type sequence as a template.

The fructotransferase-bearing plasmid would be rescued from the TDB II yeast and grown up in *E. coli*. Both the TDB II and TDS II strains would then be co-transformed with each of the fifteen PCR fragments and the rescued mutant *SUC2* gene on *pIAD3*. Colonies that subsequently grow under tryptophan selective conditions will necessarily have been transformation competent, and are liable to have been transformed with both the plasmid and the PCR fragment. Recombination could then occur between the *SUC2* gene on *pIAD3* and each of the PCR fragments. Since recombination by replacement (see Fig 6.1*b*) would result in the plasmid-borne region of the *SUC2* gene being replaced by the PCR fragment, a wild-type fructohydrolase activity would be obtained if the mutation occurred within this region. That the recombination has occured at the plasmid borne *SUC2* gene, rather than the silent, chromosomal *suc*^o allele, could easily be confirmed by plasmid rescue and transformation into the fructose intolerant yeast, followed by selection on sucrose or glucose. Only the PCR-generated fragments that replace the plasmid-borne mutated DNA will confer fructohydrolase activity after recombination, and the site of the mutation could be located to

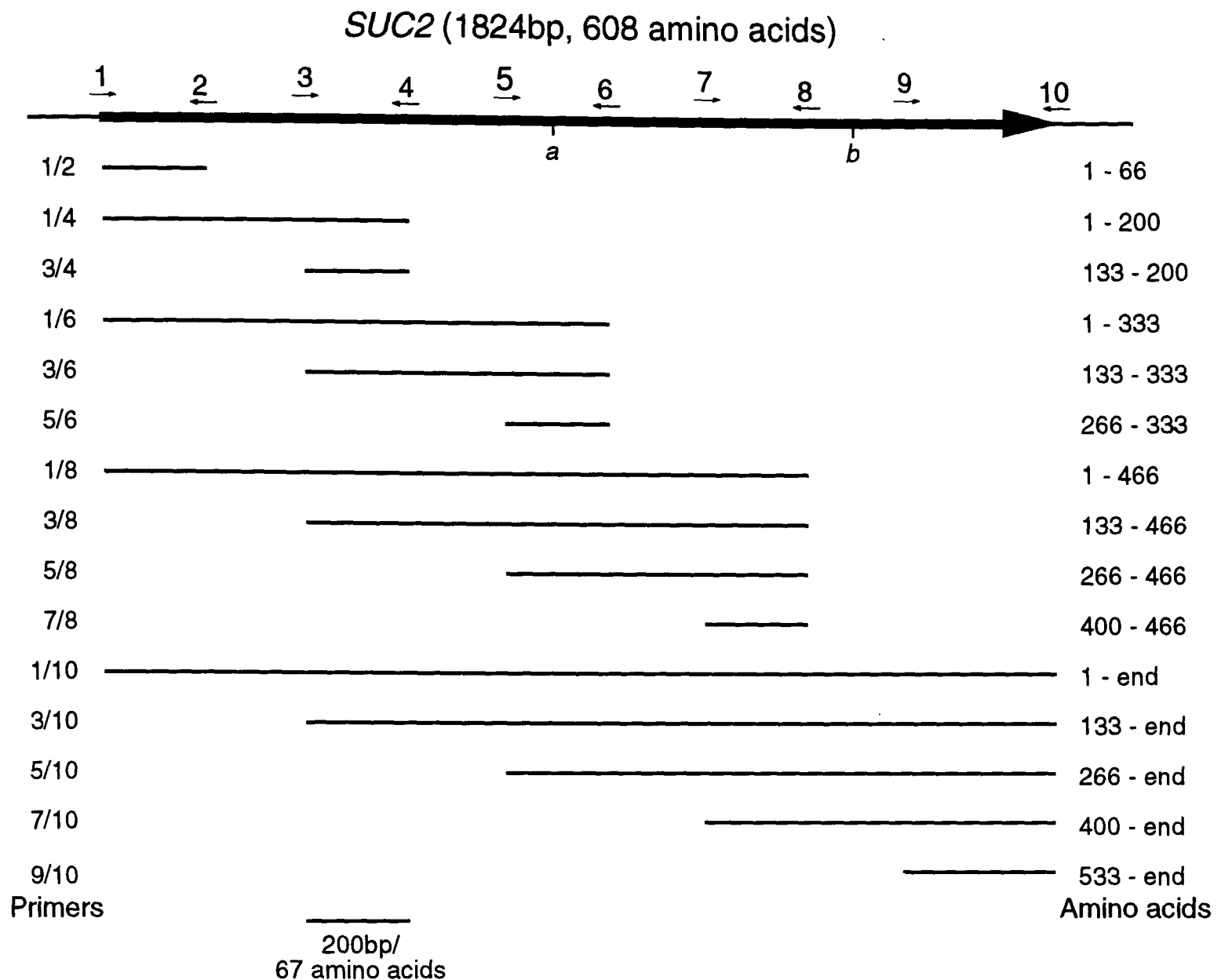


Figure 8.1 PCR-generated *SUC2* gene fragments proposed to localize mutations within the gene. The broad arrow indicates the *SUC2* gene, with the position of the ten primers shown in Table 8.1 indicated above it, arrowed to indicate 5' to 3'. Below the gene are shown the PCR-generated fragments, the primer pairs required for their synthesis, and the region of the invertase protein the fragments encode are indicated in the left- and right-hand columns, respectively. Two hypothetical mutation sites are indicated (a and b) to show the rationale for identifying their location. A single mutation at site 'a' would be complemented by nine different PCR-generated fragments, 1/6, 3/6, 5/6, 1/8, 3/8, 5/8, 1/10, 3/10 and 5/10, localizing the mutation to within amino acids 266-333. Similarly, a mutation at 'b' would be complemented by fragments 1/10, 3/10, 5/10 nor 7/10, but neither 7/8 or 9/10, localizing the mutation to amino acids 466-533. A double mutation at both sites, together conferring the mutant phenotype would be complemented by fragments 1/10, 3/10, 5/10 or 5/6 and 7/10 together, but the combinations of 5/6 and 9/10 or 5/6 and 7/8 would not complement the mutation, thereby localizing both sites as before. In all cases, the mutations are localized to within 200bp, allowing a simple sequencing reaction to identify the important mutations.

within 200 base pairs (see legend, Fig 8.1). Then, by using the appropriate pair of primers, the mutated DNA would be sequenced on both strands to identify the precise location of the mutation causing the phenotype.

Such a protocol requires screening of the mutated *SUC2* gene in both the TDB II and TDS II strains only when the acceptor substrate for the fructose moiety is sucrose. When other acceptor substrates are used, an ability of the PCR fragment-transformed TDS II/*pIAD3-SUC2*

strain to grow on sucrose in the absence of the acceptor substrate would be evidence of a fructohydrolase activity. However, those invertase mutants that synthesize fructans would be able to sustain growth in a transformed TDS II strain when grown on sucrose. Consequently, the recombination must also be screened in the TDB II strain, in which expression of a fructohydrolase activity would impair growth on sucrose. Whether or not such a recombination event has conferred a null-activity would be shown by the phenotype of the TDS II strain.

8.2.1 *Region-Specific Mutagenesis of 'Hot-Spots'.*

A mutational approach similar to that described in section 7.2.4 might be taken with the shorter sequences of the gene, generating a single-stranded copy of this region by PCR, the primers for which would be flanked by convenient restriction enzyme sites to allow simple insertion into the gene. Targetting a smaller region of the gene would allow the generation of multiple point mutations that are less likely to be inactive.

A smaller region of DNA would also be more suitable for mutation by reverse transcriptase mediated misincorporation method [2, 3] described in section 7.1.1. It has been suggested that this method obtains good random molecular populations ranging to c.250 nucleotides from the 3' end of the primer, making it an ideal candidate method for mutation of a short sequence. Again, this may narrow down the residues defining the enzyme specificity, and so these residues may be targetted by either site-directed or 'random-mutagenesis' [4, 5].

A 'random-mutagenesis' method has recently been developed to introduce a completely random sequence into a gene at a specified site, this region usually being eight base-pairs long. If a library of sufficient size is generated, every possible mutation would be covered (the 8 base pairs mutated will cover three codons, the third of which would only have two nucleotides changed, so the total library size would be somewhat smaller than 20^3). Expression of this library could be screened for wild-type, mutant and null activity, and a selection of each subjected to sequence analysis to determine the amino acid 'information content' [6] of this region. This would then greatly increase our understanding of the active site of invertase, as well as isolating fructotransferase mutants.

In addition to providing information about the active site of invertase, the generation of a variety of fructotransferase mutants should allow a variety of acceptor substrates to be used, thus increasing the use of invertase in the preparation of fructofuranoside compounds. Wild-type invertase is currently used in the preparation of such compounds, but the yields are very low due to the fructohydrolase activity, so the isolation of mutants would be of great use in these fields.

8.3 Potential Improvements for Integration of Ketohexokinase

Surprisingly, the TDB II selection strain was shown to be mitotically less stable than TDB I, with respect to the retention of the ketohexokinase cDNA upon stressing the cells with fructose. Although reversion to fructose-tolerance could be tested during the selection of fructotransferase mutants by replica plating potential mutants on fructose supplemented plates, ideally a mitotically stable multiple integration of the ketohexokinase would be highly desirable.

As discussed in section 6.1.2.4, multiple integration has also been achieved using the yeast Ty element, either by directing integration into the δ -sequence that is randomly distributed throughout the yeast chromosome [7], or inserting the expression cassette into an active Ty element and allowing integration via an RNA intermediate [8]. Experiments had shown that single copy integration of *plAD1-KHK* into the *leu2* locus of Chromosome III had not resulted in complete intolerance to the presence of fructose in the growth medium, thus demonstrating that the presence of multiple copies of the gene are necessary. However, both the strategies achieving multiple integration using the Ty element have achieved, at most, 10 copies per cell, and on average, a copy number of 2-5.

Integration at the rDNA locus has been suggested to yield a copy number of 100 per cell, and a similar level could be attempted with the Ty methods, were auxotrophic markers with deficient promoters (*e.g. leu2-d*) employed for selection. However, integration via an RNA intermediate using a modified retrotransposon Ty element would not appear to be a very promising candidate for two reasons. Firstly, integration by this method is entirely random. Such integration to copy numbers of 100 per cell would almost certainly result in interruption of open reading frames which may render the strain unviable. Secondly, this method has only been reported for the integration of insert DNA up to 2kb length, whereas, if it were to include the *PGK* expression cassette, the ketohexokinase cDNA and the *leu2-d* auxotrophic marker, the insert size required would be over 4kb.

Integration into the δ -sequence may therefore offer a more promising strategy, since this would target the integration to 30-50 sites, distributed over the yeast genome. Since the δ -sequences do not appear to confer a function, strain viability would not be impaired. Indeed, the authors of the original communication have suggested that high copy number integration may be achieved if a selection marker with a deleted promoter (*e.g. leu2-d*) were to be used [7].

Whether any of these strategies would enable a mitotically stable fructose-intolerant yeast to be engineered by multiple integration into the chromosome is unclear, since no evidence is available on the stability of such multiple integrants. It would, however, be useful to improve upon the stability of the integrated ketohexokinase gene displayed by the TDB II strain, and these strategies do possess the potential to provide this. It may, however, be sufficient to use

the TDB I strain, with the 2 μ -based *pMA91-KHK* episomal plasmid to act as a selection strain, with the *SUC2* gene borne on the *ARS/CEN* episomal plasmid *pIAD3*. If kept under selective conditions, both plasmids may be retained within the cell, without competitive loss. One would envisage employing TDB I in this way, performing transformation with *pIAD3* in the presence of *pMA91-KHK*, in order to supplement any loss of the latter from the yeast during transformation.

8.4 *Potential for Isolation of Other Fructotransferase Genes*

This fructotransferase selection system is, of course, not restricted to the selection of mutants of the *Saccharomyces cerevisiae* *SUC2* gene that confer this activity, but can be used for the isolation of fructotransferase genes from a variety of organisms. Initial attempts to express the *B. subtilis* *sacB* gene [9] in *S. cerevisiae* initially met with no success, although more extensive investigations of this possibility were hindered by time constraints.

8.4.1 *Expression and Mutation of the Aspergillus niger SUC1 Gene.*

The invertase of *Aspergillus niger* is already commercially exploited to produce kestose and fructo-oligosaccharides from sucrose [10, 11]. This enzyme has a fructotransferase: fructohydrolase activity ratio of 14.2. Although it has not been conclusively demonstrated to condition this enzyme, the *SUC1* gene from *A. niger* encodes an invertase, and has recently been kindly given to our laboratory by Dr. Christian Barreau of the Laboratoire de Genetique, Universite de Bordeaux II, prior to publication of its sequence. This gene was cloned from a genomic library, and its heterologous expression *S. cerevisiae* has not yet been attempted. The presence of short in-frame introns can not be excluded (C. Barreau, personal communication). Whether this is the gene responsible for the fructotransferase used commercially has not been established, but the fructose-intolerant yeast might provide an ideal system to determine this.

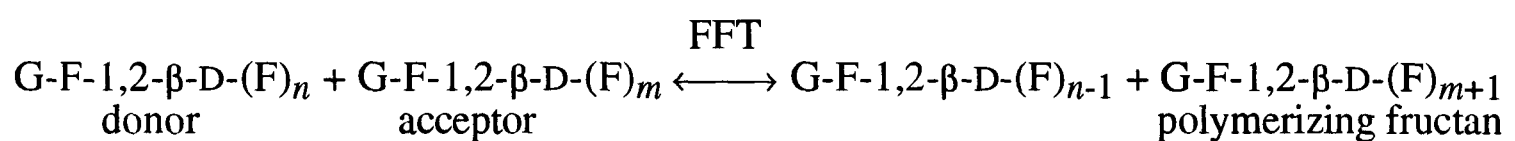
The sequence of the *A. niger* *SUC1* gene reveals a putative signal sequence, but it is unclear whether it might be functional in *S. cerevisiae* [12, 13, 14, 15, 16]. Expression of the *SUC1* gene in both the TDS II and a fructose-intolerant strain would reveal the nature of the activity conditioned by this gene. A sucrose hydrolase activity (either β -fructofuranosidase or α -glucosidase) would yield sucrose-viable transformants only in the TDS II strain, whilst a fructotransferase activity would allow both transformed strains to grow on sucrose. A hydrolase could be mutated to a transferase, or *vice versa*, in order to reveal domains or residues associated with one or the other activity.

8.4.2 *Screening of an Aspergillus niger Library for a Fructotransferase Gene.*

An alternative approach to the isolation of the fructotransferase gene of *A. niger* might involve the screening of an *A. niger* cDNA expression library in the fructose-intolerant strain. Such a library might be constructed from *A. niger* induced to produce the fructotransferase gene by its growth on sucrose. Using a strategy similar to that to produce λ uniZAP II libraries, the cDNA library might be directionally cloned into the *pIAD3* yeast expression vector, and this library used to transform the fructose-intolerant strain. Plating of the transformants on selective media containing sucrose as sole carbon source would then act as a rapid screen for the fructotransferase-encoding gene.

8.4.3 *Isolation of the Fructan Synthesizing Proteins of Compositae Plants.*

Certain plants are also known to produce fructans as carbohydrate storage compounds, which perform a similar function to starch (see 1.1.1). In the *Compositae* family of plants, such as chicory and Jerusalem artichoke, fructan is synthesized by two proteins, sucrose:sucrose fructosyl transferase (SST, EC 2.4.1.99) and fructan:fructan fructosyl transferase (FFT, EC 2.4.1.100) [17, 18]. These proteins catalyse kestose production, and fructan chain elongation, respectively:



[G = glucosyl; F = fructosyl; $m, n = 1$ or >1].

The fructose-intolerant selection strain again offers an ideal system to isolate the genes encoding these proteins from a cDNA library. Young chicory roots or Jerusalem artichoke tubers, actively sequestering fructose as fructans, would be ideal sources of mRNA with which to make such a cDNA library. The plant root library might be made with the cDNAs ligated directionally in the *ADH1* expression cassette of *pIAD3*. By screening this library in the selection strain, on sucrose as a sole carbon source, fructotransferase-encoding cDNA might be isolated. As it stands, the selection strain would be restricted to selection of sucrose:sucrose fructosyl transferase cDNA and would not allow selection of the cDNA for fructan:fructan fructosyl transferase, proposed to be responsible for the formation of fructans of DP > 6 in higher plants. A selection system could be devised for FFT cDNA by constitutively expressing a cDNA for SST in the fructose-intolerant yeast and then selecting clones transformed with a cDNA library for growth on kestose (see Fig. 1.2). The rationale behind this scheme is that

this selection strain could grow on sucrose as a carbohydrate source because fructose would not be released, and that, under these growth conditions, only fructosyl transfer from kestose by FFT could furnish sucrose without the release of fructose. (Hydrolysis of kestose by either invertase or inulinase would result in fructose poisoning.) This system, of course, requires that the SST constitutively expressed in the selection strain has negligible ability to catalyse fructosyl transfer from kestose.

8.5 References

1. Kunes, S., Ma, H., Overbye, K., Fox, M.S. and Botstein, D. (1987). Fine Structure Recombinational Analysis of Cloned Genes Using Yeast Transformation. *Genetics*, **115**, 73-81.
2. Lehtovaara, P.M., Koivula, A.K., Bamford, J. and Knowles, J.K.C. (1988). A new method for random mutagenesis of complete genes; enzymatic generation of mutant libraries *in vitro*. *Prot. Eng.*, **2**, 63-68.
3. Holm, L., Koivula, A.K., Lehtovaara, P.M. Hemminki, A. and Knowles, J.K.C. (1990). Random mutagenesis used to probe the structure and function of *Bacillus stearothermophilus* alpha-amylase. *Prot. Eng.*, **3**, 181-191.
4. Palzkill, T. and Botstein, D. (1992). Probing β -Lactamase Structure and Function Using Random Replacement Mutagenesis. *Proteins*, **14**, 29-44.
5. Palzkill, T. and Botstein, D. (1992). Identification of Amino Acid Substitutions That Alter the Substrate Specificity of TEM-1 β -Lactamase. *J. Bacteriol.*, **174**, 5237-5243.
6. Reidhaar-Olsen, J.F. and Sauer, R.T. (1988). Combinational Cassette Mutagenesis as a Probe of the Information Content of Protein Sequences. *Science*, **241**, 53-57.
7. Sakai, A., Shimizu, Y. and Hishinuma, F. (1990). Integration of heterologous genes into the chromosome of *Saccharomyces cerevisiae* using a delta sequence of yeast retrotransposon Ty. *Appl. Microbiol. Biotechnol.*, **33**, 302-306.
8. Boeke, J.D., Hua, X. and Fink, G.R. (1988). A General Method for the Chromosomal Amplification of Genes in Yeast. *Science*, **239**, 280-282.
9. Steinmetz, M., Le Coq, D., Aymerich, S., Gonzy-Tréboul, G. and Gay, P. (1985). The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. *Mol. Gen Genet.*, **200**, 220-228.
10. Hidaka, H., Hirayama, M. and Sumi, N. (1988). A Fructooligosaccharide-producing Enzyme from *Aspergillus niger* ATCC 20611. *Agric. Biol. Chem.*, **52**, 1181-1187.
11. Hirayama, M., Sumi, N. and Hidaka, H. (1989). Purification and Properties of a Fructooligosaccharide-producing β -Fructofuranosidase from *Aspergillus niger* ATCC 20611. *Agric. Biol. Chem.*, **53**, 667-673.
12. Perlman, D. and Halvorson, H.O. (1983). A Putative Signal Peptidase Recognition Site and Sequence in Eukaryotic and Prokaryotic Signal Peptides. *J. Mol. Biol.*, **167**, 391-409.

13. Innis, M.A., Holland, M.J., McCabe, P.C., Cole, G.E., Wittman, V.P., Tal, R., Watt, K.W.K., Gelfand, D.H., Holland, J.P. and Meade, J.H. (1985). Expression, Glycosylation, and Secretion of an *Aspergillus* Glucoamylase by *Saccharomyces cerevisiae*. *Science*, **228**, 21-26.
14. Kaiser, C.A. and Botstein, D. (1986). Secretion-Defective Mutations in the Signal Sequence for *Saccharomyces cerevisiae* Invertase. *Mol. Cell. Biol.*, **6**, 2382-2391.
15. Kaiser, C.A., Preuss, D., Grisafi, P. and Botstein, D. (1987). Many Random Sequences Functionally Replace the Secretion Signal Sequence of Yeast Invertase. *Science*, **235**, 312-317.
16. Ngsee, J.K., Hansen, W., Walter, P. and Smith, M. (1989). Cassette Mutagenic Analysis of the Yeast Invertase Signal Peptide: Effects on Protein Translocation. *Mol. Cell. Biol.*, **9**, 3400-3410.
17. Edelman, J. and Jefford, T.G. (1968). The mechanism of fructosan metabolism in higher plants as exemplified in *Helianthus tuberosus*. *New Phytol.*, **67**, 517-531.
18. John, P. (1991). Fructan quality and fructan synthesis. *Biochem. Soc. Trans.*, **19**, 569-572.

Appendix

Publications and Presentation of Data

1. Abstract of poster presented at the 16th International Conference of Yeast Genetics and Molecular Biology, Vienna, Austria, August 1992. (Doyle, T.C., Donaldson, I.A., Matas, N. and Cox, B.S. (1992). Genetic engineering of a fructose-intolerant yeast strain to screen for fructotransferase mutants of invertase. *Yeast*, **8**, S573.)
2. Donaldson, I.A., Doyle, T.C. and Matas, N. (1993). Expression of rat liver ketohexokinase in yeast results in fructose intolerance. *Biochem. J.*, **291**, 179-186
3. Doyle, T.C., Spickett, C.M., Donaldson, I.A. and Ratcliffe, R.G. (1993). Metabolic Studies of a Fructose-Intolerant Yeast by *in vivo* ³¹P-NMR Spectroscopy. *Yeast*, In press.

Abstract to poster, presented at 16th International Conference on Yeast Genetics.
(*Yeast*, 8, S573.)

GENETIC ENGINEERING OF A FRUCTOSE-INTOLERANT YEAST STRAIN TO SCREEN FOR FRUCTOTRANSFERASE MUTANTS OF INVERTASE.

T. C. Doyle, I. A. Donaldson, N Matas¹, B. S. Cox²

¹Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K.

²Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RA, U.K.

A selection system for yeast cells expressing mutated invertase (*SUC2*) with altered fructotransferase activity has been developed based on the survival of a fructose-intolerant strain in the presence of suitable acceptor substrates and sucrose. Cells of such a strain which express wild-type hydrolase activity will die owing to the release of free fructose from sucrose; cells expressing inactive mutations of *SUC2* will not grow, because they cannot cleave sucrose which will be the sole source of carbon provided; only cells displaying fructotransferase activity will form colonies, because the fructosyl moiety is not freed, but the glucose released is available as a carbon source.

A strain of *Saccharomyces cerevisiae* has been engineered to be intolerant of the presence of fructose in its growth media. This was achieved by inducing a condition in the cells similar to human hereditary fructose intolerance, where the absence of the enzyme aldolase B (EC 4.1.2.13) results in the accumulation of the product of the enzyme ketohexokinase (EC 2.7.1.3), fructose 1-phosphate, which leads to the depletion of cytoplasmic phosphate and ATP in tissues possessing this enzyme (liver, kidney, small intestine and pancreas). Thus, cells in which ketohexokinase phosphorylates fructose, but which lack aldolase 'B', are intolerant of the hexose. Yeast possesses neither ketohexokinase nor aldolase 'B' and so the expression of the former enzyme in yeast results in fructose intolerance.

To achieve this condition in the yeast cells a strain was bred specifically for this purpose. Four criteria were used in the design of this yeast strain; (i) the enzymes involved in the primary metabolism of fructose in yeast (hexokinases 1 and 2) should be absent, so the strain should be *hxx1 hxx2*, although the strain should still ferment glucose using the substrate specific glucokinase (*GLK*), (ii) the strain should not be able to remobilize the phosphate sequestered as fructose 1-phosphate, so the non-specific phosphatases should be absent, and for this purpose the strain is *pho3 pho5 pho8*, (iii) since the strain is to be used to screen mutant invertase genes for fructotransferase activity, wild type enzymes conditioning sucrose hydrolysis should be absent, and hence the strain is *mal^o suc^o*, (iv) amino acid auxotrophic markers should be present to allow the selection of cells containing plasmids bearing the ketohexokinase and mutated invertase genes (*leu2* and *trp1* respectively). Several strains were bred containing all these mutations by classical mating, and are described as TDS-II strains.

Ketohexokinase was purified to homogeneity from rat liver and partial protein sequence data obtained were exploited to amplify a section of the coding region by PCR. This was used to probe a rat liver cDNA library. By this route the cDNA was cloned, sequenced and introduced into the TDS strain of yeast on an expression vector pMA91. As predicted, the transformed yeasts showed extreme sensitivity to the presence of fructose in their growth media. growth of a log phase culture was completely arrested within 90 minutes by the addition of 0.1% fructose. To produce a fructose-intolerant selection strain a vector was constructed that allowed integration of the expression cassette containing the ketohexokinase gene into the yeast genome by targetted homologous recombination with the untranscribed region of the rRNA genes on chromosome XII.

Expression of rat liver ketohexokinase in yeast results in fructose intolerance

Iain A. DONALDSON,* Timothy C. DOYLE and Nada MATAS†

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Rat liver ketohexokinase (ATP:D-fructose 1-phosphotransferase; EC 2.7.1.3) was purified to homogeneity and the molecular mass of the protein was found by mass spectrometry to be 32800 Da. The enzyme was cleaved and the amino acid sequences of seven peptides, comprising 24% of the total sequence, were determined. This sequence information was used to design oligonucleotide primers for a PCR using rat liver single-stranded cDNA as a template. The 224 bp PCR product was used as a probe to screen a rat liver cDNA library. A cDNA sequence of 1342 bp was obtained from three positive clones. This contained the entire coding region for ketohexokinase, and all seven peptides were identified in the predicted amino acid sequence. When ketohexokinase was expressed in *Saccharomyces cerevisiae* using the yeast expression vector pMA91, the cells

became intolerant of the presence of fructose in their growth media. The growth of an exponential-phase culture was completely arrested within 90 min by the addition of fructose to a final concentration as low as 0.1% (w/v). This response is associated with an accumulation of fructose 1-phosphate. The cDNA for ketohexokinase encodes a protein composed of 299 amino acids with a combined molecular mass of 32728 Da. This is in close agreement with the value for the isolated protein determined by mass spectrometry. The primary structure does not show any significant homology with those of other eukaryotic hexokinases, but it contains a highly conserved region that is present in three prokaryotic phosphotransferases that have furanose substrates.

INTRODUCTION

Ketohexokinase (ATP:D-fructose 1-phosphotransferase; EC 2.7.1.3) is the enzyme responsible for the primary metabolism of dietary fructose in mammals [1,2]. The enzyme is most abundant in the liver, but is also present in the kidney, small intestine and pancreas [3]. All of these tissues possess a specialized pathway employing ketohexokinase, aldolase B (EC 4.1.2.13) and triokinase (EC 2.7.1.28), which together convert fructose into intermediates for glycolysis and gluconeogenesis. Two inherited abnormalities of this pathway have been recognized in humans. Hereditary fructose intolerance is a potentially lethal condition caused by a deficiency of aldolase B [4]. As a result of the activity of ketohexokinase, ingestion of fructose by individuals with this disorder is followed by accumulation of fructose 1-phosphate and depletion of ATP in tissues possessing this pathway (liver, kidneys, small intestine and pancreas). Essential fructosuria, in contrast, is a benign defect caused by the absence of ketohexokinase, which results in an inability to synthesize fructose 1-phosphate [5]. Under these circumstances, uptake of dietary fructose leads to a transient increase in the concentration of blood fructose. This gradually decreases, and approx. 80–90% of the ingested fructose is eventually metabolized, the rest being excreted in urine, indicating that, under these circumstances, fructose is being utilized by other pathways [2].

Mammalian hexokinases (EC 2.7.1.1) are also able to phosphorylate fructose [6]. In rats there are four isoforms of this enzyme, belonging to a family of kinases that display antigenic cross-reactivity and regions of amino acid homology [7]. Brain contains hexokinase type I only, skeletal muscle contains type II,

and adipose tissue and heart contain both of these. Kidney contains I and III, whereas the intestine and liver contain isoenzymes I, II and III. Isoenzymes I, II and III have affinities for glucose roughly ten times those for fructose, phosphorylate a range of other hexoses *in vitro*, and are strongly inhibited by glucose 6-phosphate [8]. Hexokinase type IV (glucokinase; EC 2.7.1.1), the predominant isoenzyme in the liver, is not inhibited by its product, and is regulated by insulin and glucagon [6,8]. It shows a similar substrate specificity to the other hexokinases, but has a much lower affinity for its substrates, and is not responsible for the primary metabolism of fructose *in vivo*. Nonetheless, this enzyme has recently been shown to have substantial regions of amino acid homology with isoenzymes I–III, and so was confirmed to be a member of the hexokinase family. The family also includes the yeast isoenzymes PI and PII [7].

The properties of ketohexokinase show substantial differences from those of isoenzymes belonging to the hexokinase family. It phosphorylates its substrate at the C-1 position rather than the C-6 position [9,10], has an absolute requirement for monovalent cations (physiologically, K⁺) [11], will only phosphorylate furanose substrates and, unlike glucokinase, is not induced hormonally or by its substrate. In common with hexokinases, it does show marked preference for Mg²⁺-ATP as a nucleotide, and is inhibited by ADP.

Liver ketohexokinase has been isolated from a number of sources [10,12–14]. Extensive studies of the enzyme's kinetic properties and substrate specificity have been performed. However, a primary structure of ketohexokinase has not been reported.

Saccharomyces cerevisiae has been used to express a number of

Abbreviations used: TEA, triethanolamine; PTH-amino acid, phenylthiohydantoin-amino acid; TFA, trifluoroacetic acid.

* To whom correspondence should be addressed.

† Permanent address: Department of Molecular Biology, Faculty of Natural Sciences and Mathematics, University of Zagreb, Rooseveltov trg 6, P.O. Box 933, YU-41001 Zagreb, Croatia.

The nucleotide sequence data reported in this paper have been submitted to the Genbank and EMBL Data Banks under accession numbers M86235 and X63658 respectively.

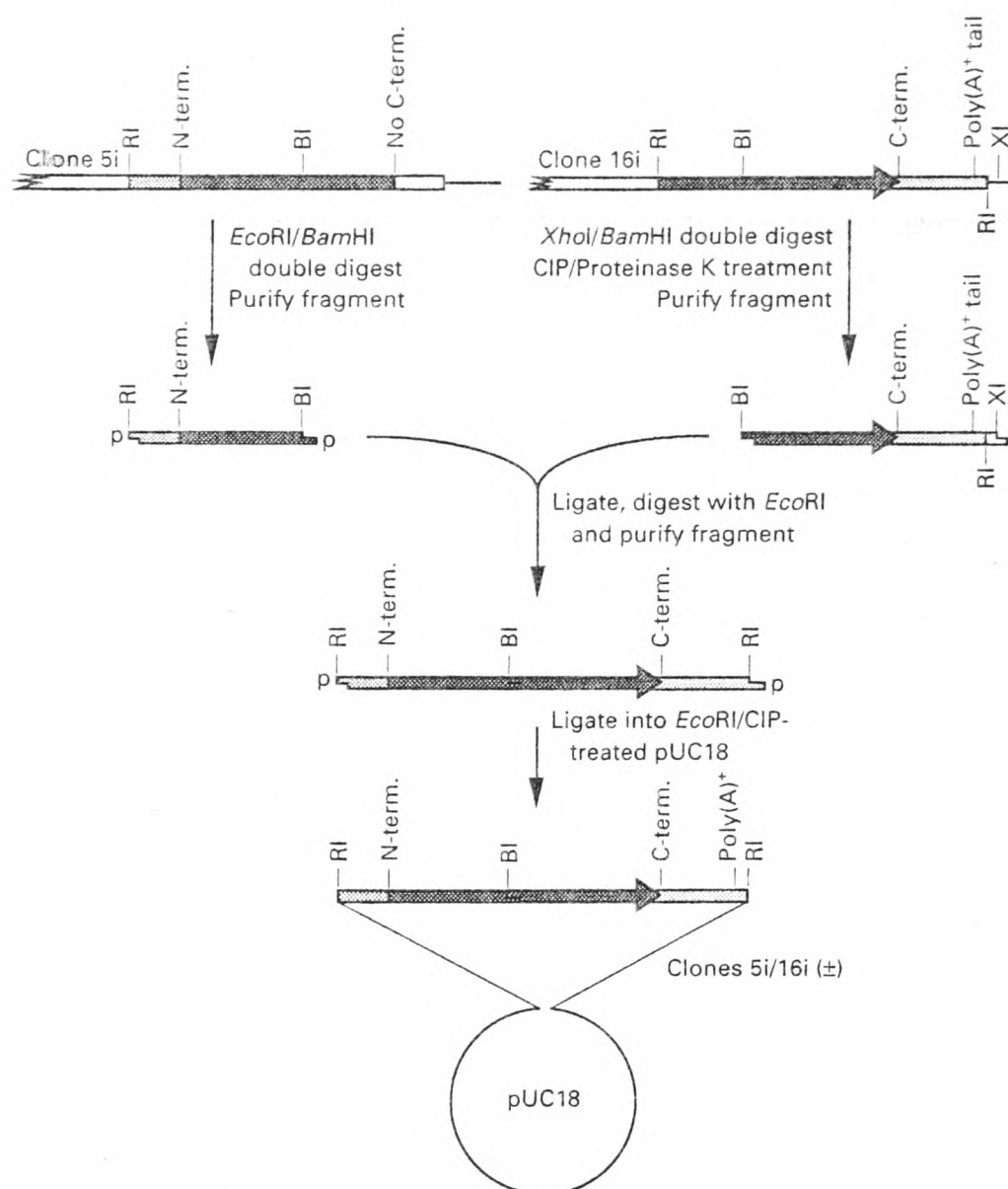


Figure 1 Strategy for the construction of hybrid clone 5i/16i in pUC18

The co-ligated cDNA was removed from the ketohexokinase cDNA at the *EcoRI* sites (RI), and the 5i and 16i clones were spliced at the *Bam*HI site (BI). The clone 16i was originally cut with *Xho*I (XI) and *Bam*HI, before treatment with calf intestinal phosphatase (CIP). The heavily shaded boxes represent the ketohexokinase coding region, the lightly shaded boxes represent 5' or 3' untranslated regions, the open boxes represent unassociated, co-ligated cDNA, and the thin lines represent plasmid DNA. C-term and N-term indicate C- and N-termini respectively.

active mammalian proteins [15]. Since yeast does not contain aldolase B, it was of interest to us to investigate whether or not the expression of active ketohexokinase in this organism would condition fructose intolerance. Such a strain would be of great interest to us for selection of fructotransferase mutants of invertase.

We report the cloning and complete nucleotide sequence for the cDNA of rat liver ketohexokinase and the effects that the expression of this enzyme in yeast has upon the response of the yeast to the provision of fructose in the growth medium.

EXPERIMENTAL

Purification of rat liver ketohexokinase

Eight male Sprague-Dawley rats (250–300 g) were killed by cervical dislocation and exsanguinated by severing the carotid arteries. Their livers were used immediately for enzyme purification.

The procedure used to purify ketohexokinase was a modification of previously published methods [10,12–14], using improved column materials and conditions. Gel filtration was performed on a 108 cm × 4 cm² Sephacryl S100 (Pharmacia LKB Biotechnology) column eluted with triethanolamine (TEA) buffer

(10 mM TEA/HCl, pH 7.3, 1 mM dithiothreitol) at a flow rate of 3 ml/min. Anion-exchange chromatography was performed on a 19 cm × 5 cm² Fractogel TSK DEAE-650(S) (Merck) column eluted with a 0–200 mM KCl gradient in TEA buffer at 2 ml/min. Chromatofocusing was performed on a 15 cm × 0.75 cm² PBE94 column (Pharmacia LKB Biotechnology) eluted at a flow rate of 0.33 ml/min with Polybuffer 74/HCl, pH 4.5 (Pharmacia LKB Biotechnology).

Enzyme activity in fractions of column eluent was assayed as described below, and the activity of pooled samples was measured by the method of Adelman et al. [12]. For the colorimetric assay, a 5 µl aliquot of sample was added to 1 ml of reaction mixture (75 mM potassium phosphate, pH 7.2, 300 mM KCl, 20 mM MgCl₂, 15 mM ATP, 5 mM phosphoenolpyruvate, 0.2 mM TPP, 8 µM FAD, 10 mM phenol, 8 mM 4-aminopyrene and 5 mM fructose) containing 2 units of peroxidase, 2 units of pyruvate oxidase (Boehringer Mannheim) and 3 units of pyruvate kinase (Sigma). The reaction was incubated at 37 °C for 60 min prior to quenching with 100 µl of 110 mM EDTA. The absorbance at 510 nm was then read spectrophotometrically. Protein concentrations were determined by the method of Bensadoun and Weinstein [16]. SDS/PAGE was performed according to Hames [17].

The molecular mass of the purified protein was accurately determined by matrix-assisted laser desorption on a Finnigan MAT LASERMAT 'time-of-flight' mass spectrometer [18]. A 0.2 μ l or 1.0 μ l sample of the purified protein (approx. 800 μ g/ml) was mixed with 0.2 μ l of sinapinic acid matrix on a gold-plated stainless steel target and allowed to dry under an infra-red lamp. This was then put in the path of the ionizing laser of the mass spectrometer and the relative molecular masses of ions generated during irradiation were measured.

Preparation and sequencing of peptides of ketohexokinase

A sample (approx. 30 μ g) of purified ketohexokinase in 45 μ l of 100 mM ammonium carbonate/acetic acid buffer, pH 8.1, was diluted with 50 μ l of 100 mM ammonium bicarbonate, pH 8.3, containing 5% (v/v) acetonitrile. To this, 2 μ g each of endoproteinase Glu-C and trypsin (Boehringer, Sequencing Grade) were added and the reaction was incubated for 60 min at 37 °C. A control reaction containing proteinase only was also performed. After digestion, 100 μ l of 8 M guanidine hydrochloride/0.5 M Tris/0.35 mM EDTA was added to each reaction. The mixtures were then reduced by the addition of 2 μ l of a 100 mg/ml solution of dithiothreitol and incubated at 37 °C for 60 min. In order to alkylate the cysteine residues, a 2 μ l volume of 4-vinylpyridine was then added to each reaction and the mixtures were incubated at 37 °C for 30 min. Both samples were then chromatographed by reverse-phase h.p.l.c. on a Brownlee Aquapore RP300 (C8) 220 mm $\times$ 2 mm column, loading with 0.1% (v/v) trifluoroacetic acid (TFA) in water and then applying a 5–50% (v/v) gradient of 0.09% (v/v) TFA in 80% (v/v) acetonitrile over 45 min, followed by a 50–90% (v/v) gradient over 15 min. The sample and control were then compared to eliminate peptides derived from the proteinases themselves, and the six major peaks were sequenced on an Applied Biosystems 470A protein sequencer with on-line phenylthiohydantoin (PTH) analysis.

Cloning of the ketohexokinase probe

Unless specifically mentioned, all procedures were performed according to Sambrook et al. [19].

Preparation of PCR primers

Because the relative positions of peptides KHK 45 and 49 (see Figure 3) were not known, four oligonucleotide primers were synthesized to account for both possible orientations (primer 1, ATACTRTGDTTRGASGGWCTACA; primer 2, TATGAY-ACHAAYCTSCCWGATGT; primer 3, AARGTYGASAAA-CCBATAACWCT; primer 4, TTYCARCTSTTTGGVTATG-GWGA, using the letter codes recommended by the IUPAC/IUB Biochemical Nomenclature Commission). The primers were prepared on an Applied Biosystems DNA synthesizer and purified on oligonucleotide purification cartridges (Applied Biosystems 400771). The concentrations of the oligonucleotides were determined spectrophotometrically.

Amplification and cloning of the ketohexokinase probe from a single-stranded cDNA template

A rat liver was dissected as described above. Isolation of RNA from 5 g of rat liver tissue was performed by directly scaling up the acid guanidinium thiocyanate/phenol/chloroform (AGPC) extraction method of Chomczynski and Sacchi [20]. Single-stranded cDNA was prepared from 20 μ g of crude RNA by employing the method for purified mRNA described in section 8.60–8.61 of Sambrook et al. [19] using Moloney murine

leukaemia virus reverse transcriptase (Gibco BRL) and oligo(dT)_{18–20} (Pharmacia LKB Biotechnology) as a primer. PCR was performed directly on 20 μ l of this cDNA reaction mixture as described in section 7.26–7.29 of Sambrook et al. [19], using oligonucleotide primers 1 and 4 or 2 and 3. Single primer control reactions were also performed for all oligonucleotides, and three MgCl₂ concentrations (1.5 mM, 3.0 mM and 5.0 mM) were used for all reactions. Thirty cycles of PCR were carried out on a Perkin-Elmer Cetus DNA thermal cycler. Cycles were: 94 °C, 1.5 min; 50 °C, 2 min; 72 °C, 2 min (10 min for the last cycle). Cloning of the PCR products into pUC18 [21] was performed by the method of Bonthron [22], followed by transformation into *Escherichia coli* strain DH5 α . The nucleotide sequence of both strands was determined by the dideoxy method [23] using the Pharmacia T7 Sequencing Kit (27-1682-01) employing deoxyadenosine 5'-[α -³⁵S]thiotriphosphate (Amersham).

Labelling of the probe

The ketohexokinase fragment was excised from pUC18 and used for the synthesis of a mixture of randomly primed probes, labelled with [³²P]dCTP (3000 Ci/mmol) (Amersham), by the method described in section 10.14–10.17 of Sambrook et al. [19]. The mixture of labelled probes was purified from the reaction mixture using Nensorb 20 columns (NEN Research Products).

Screening of a cDNA library in λ ZAP

The rat liver cDNA library was kindly provided by Professor Brian Burchell, University of Dundee. It was prepared from Sprague-Dawley rat liver mRNA and cloned into λ ZAP [24] with a phage titre of 10⁹ plaque-forming units/ml. The R408 f1 helper phage and *E. coli* strains XL1-Blue and BB4 were gifts from Dr. Tamar Enoch (Microbiology Unit, University of Oxford). Confluent plates with BB4 cells as hosts were blotted in duplicate with Hybond N+ nylon filters (Amersham), and treated according to sections 2.112–2.113 of [19], substituting 100 mM sodium phosphate, pH 7.2, for the 0.5 M Tris/HCl used in the neutralizing solution. Hybridization was performed according to sections 2.114–2.117 of [19], substituting 100 mM sodium phosphate (pH 7.2)/1% (w/v) SDS/1 M NaCl for the SSC, and blocking with 200 μ g/ml salmon sperm DNA and 5 $\times$ Denhart's solution (section 9.49 [19]). The filters were then washed five times at high stringency with 10 mM sodium phosphate, pH 7.2, and 1% (w/v) SDS at 65 °C and allowed to dry before autoradiography. Clones hybridizing the probe were isolated by successive replating and screening. Single clones were finally suspended in 1 ml of 100 mM NaCl, 8 mM MgSO₄, 0.01% (w/v) gelatin and 50 mM Tris/HCl, pH 7.5 (section A.7 [19]).

The sizes of the inserts in the clones were determined by amplifying this region by PCR using the M13 forward and reverse primers flanking the cloning site of λ ZAP [25]. Thirty cycles of PCR were performed as described above. The DNA was separated on a 1% agarose gel, Southern-blotted (section 9.31–9.46 in [19]) and the membrane probed at high stringency with the ³²P-labelled DNA probe. An autoradiograph revealed the size of the clonal inserts. Clones were rescued into the phagemid pBluescript SK using R408 helper phage, which was used to transform *E. coli* XL1-Blue [24].

Rescued clones were mapped using the restriction enzymes *Apa*I, *Bam*HI, *Eco*RI, *Hind*III, *Kpn*I, *Not*I, *Pst*I, *Sma*I, *Sst*I, *Xba*I and *Xho*I (Gibco BRL), unique cleavage sites for which exist within the multiple cloning site of the pBluescript SK. Approx. 300 bp of nucleotide sequence of each of the clones was

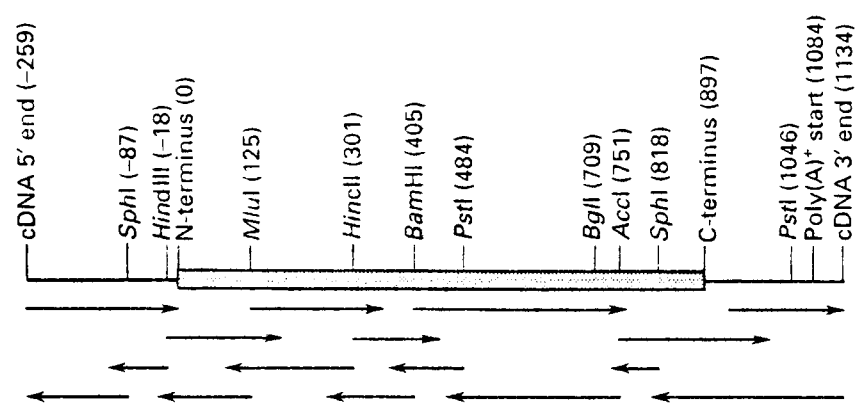


Figure 2 Sequencing strategy for the complete cDNA of ketohexokinase

The arrows represent the regions of the gene sequenced by each deleted clone. The restriction enzyme sites of importance within the cDNA are marked, along with their position (in bp) from the first nucleotide in the start codon.

Unidirectional deletion of clone 16i by exonuclease treatment was performed by a modification of the method described in section 13.34–13.41 of [19], substituting mung-bean nuclease for S1 nuclease. The clone was cut with *Xho*I and *Kpn*I and digested with *Exo*III according to the published method. To each of the aliquots taken, 10 × concentrated mung-bean buffer [22] and mung-bean nuclease to 1 unit/μl were added. Mixtures were then incubated at room temperature for 20 min and quenched with 6 × stop solution (800 mM Tris/HCl, pH 8.0, 20 mM EDTA, 80 mM MgCl₂). The samples were then analysed and treated as in the published method [19] and the nucleotide sequences were determined as described above.

Clones 16i and 5i were cut and re-ligated according to Figure 1 to produce a clone (16i/5i) containing the coding region and the 3' and 5' untranslated regions. Clones were isolated which contained the hybrid ketohexokinase gene inserted into the plasmid in both directions. The co-ligated, unassociated DNA was removed from clone 5ii by cleaving with *Eco*RI and ligating the fragment containing ketohexokinase cDNA into pUC18 to give clone 5ii/a. Endonuclease digestion and re-ligation of each orientation of the two clonal inserts 16i/5i and 5ii/a yielded a set of deleted clones, enabling the nucleotide sequence on both strands to be determined using the M13 forward and reverse primers (Figure 2).

Expression of ketohexokinase in *Saccharomyces cerevisiae*

The coding region of the ketohexokinase cDNA was amplified from clone 5ii/a by PCR using oligonucleotide primers with the following sequences: upstream primer, 5' TTGGAGATCTACACAATGGAAGAGAAGCAGAT 3'; downstream primer 5' CCCACCGCAGATCTCTCACACAATGCCATCAA 3'. Both primers incorporate non-complementary *Bgl*II sites and the upstream primer also contains the yeast consensus sequence [26] immediately 5' to the start codon. The start and stop codons in the sequences are underlined. The conditions used for PCR were as described above, using 1 mM MgCl₂, and the 911 bp product was blunt-end-ligated into the Klenow-treated *Sph*I site of pUC18. A 693 bp *Mlu*I/*Sph*I fragment was excised from the centre of the cloned PCR product and replaced with the equivalent fragment from clone 5ii/a. The resulting clone was named pUC KHK-G₇. In order to check for any residual PCR errors, the nucleotide sequences of the remaining PCR-generated flanking segments were confirmed using primers 1 and 4 in the sequencing reactions.

The coding region for ketohexokinase was excised from pUC KHK-G₇ with *Bgl*II and ligated into the *Bgl*II cloning site

of the yeast expression vector pMA91 [27]. Clones in the correct orientation for expression were termed pMA91-KHK and those in the opposite orientation were termed pMA91-KHK(–). These plasmids were used to transform *Saccharomyces cerevisiae* strain TDSII (*leu2-3*, *leu2-112*, *trp1*, *suc*^o, *mal*^o, *hxx1*, *hxx2*, *GLK*, *pho8-2*, *pho3Δ1*, *pho5Δ1*, MATα) using a modification of the method of Ito et al. [28]. The TDS II cells were prepared essentially as published, using lithium acetate, but the transformation mixture contained 100 μl of cell suspension, 10 μl of carrier DNA (sonicated salmon sperm DNA; 25 mg/ml), 1 μg of plasmid DNA and 10 μl of absolute ethanol. This mixture was incubated for 30 min at 30 °C before the addition of 1 ml of 40 % (w/v) poly(ethylene glycol) 4000, 100 mM CH₃CO₂Li, 10 mM Tris/HCl, pH 8.0, and 1 mM EDTA, and the incubation was continued at 30 °C for 1 h. The cells were then pelleted and washed with water before being resuspended in 1 ml of YEPD [1 % (w/v) yeast extract, 1 % (w/v) bactopectone, 2 % (w/v) glucose] and incubating at 30 °C for 6 h without shaking. This allows the cells to obtain a high copy number of the plasmid. The cells were then plated on selective media as published [28].

The presence of the ketohexokinase gene in the transformants was confirmed by the preparation of yeast plasmid DNA minipreps and by the amplification of the coding region by PCR using the same primers and conditions originally used to generate this fragment.

Intolerance of fructose was demonstrated by replica plating of transformants on to SD Leu[–] plates [0.17 % (w/v) yeast nitrogen base without amino acids, 0.5 % (w/v) ammonium sulphate, 1 % (w/v) glucose, 20 μg/ml each of adenine sulphate, uracil, L-tryptophan, L-histidine-HCl, and 2 % (w/v) agar], either with or without 0.5 % (w/v) fructose. The plates were then incubated for at least 48 h at 30 °C. Colonies possessing the plasmid pMA91-KHK showing intolerance were termed TDB I, and those transformed with pMA91-KHK(–) and showing leucine-independence were termed TDB I(–). The effect of various fructose concentrations on exponential-phase cells growing in SD Leu[–] liquid culture was also investigated. Early exponential-phase cultures of TDB I and TDB I(–) were divided into 30 ml aliquots in 50 ml tubes, and growth (at 30 °C with shaking) was followed by taking 1 ml samples and measuring *A*₆₀₀. After 2 h, fructose was added to various concentrations (0.025–0.5 %, w/v) and the *A*₆₀₀ measurements were continued for several hours. Metabolites from TDB I cells, treated or not with fructose, were extracted with perchlorate by the method of den Hollander et al. [29]. The presence of fructose 1-phosphate in these extracts was detected by the method of Eggelston [30].

Cell extracts from yeast strains TDB I, TDB I(–) and TDS II were prepared by vortexing a cell suspension with glass beads [31]. The supernatants were partially purified by ammonium sulphate fractionation and heat treatment as described above, and all fractions were tested for the presence of ketohexokinase by the colorimetric methods described above and assayed by the method of Adelman et al. [12].

RESULTS

Purification and characterization of ketohexokinase

Rat liver ketohexokinase was purified to homogeneity, with a specific activity of 11.6 μmol/min per mg of protein. A 166-fold purification, though low when compared with previous reports, resulted from the use of a purer starting homogenate than those methods employing bladed blenders which break open organelles. The mild homogenization technique that we employed allows the sedimentation of intact organelles, producing a cytoplasmic fraction that is minimally contaminated with proteins released

GCAGCCAGACGCCAGGCTGGATGAAATGAAGAAGGCAATGGCAGGCAGTGGTGCCAGCCTCTGAGGTCT																							69	
CGAGCTGCTACCTCCCACCGCTCATCTCTCACACAATGCCATCAAACCCCTGCAAGCCACACTTCTTGCCAGCCACCTGGCACCCGAATCTCAGG																							164	
GCCTCCTGCATGCTGTTTCCCTTGGAGAGGCTGAAGAGACAGAGTCTAGCAGGAATCCCCTCCGCTTGCAGGTAGGAAGCTTGGGGAGCAGCCTC																							259	
KHK 32																								
Met	Glu	Glu	Lys	Gln	Ile	Leu	Cys	Val	Gly	Leu	Val	Val	Leu	Asp	Ile	Ile	Asn	Val	Val	Asp	Lys	Tyr	Pro	24
ATG	GAA	GAG	AAG	CAG	ATC	CTG	TGC	GTG	GGG	CTG	GTG	GTG	CTG	GAC	ATC	ATC	AAT	GTG	GTG	GAC	AAA	TAC	CCA	331
KHK 46																								
Glu	Glu	Asp	Thr	Asp	Arg	Arg	Cys	Leu	Ser	Gln	Arg	Trp	Gln	Arg	Gly	Gly	Asn	Ala	Ser	Asn	Ser	Cys	Thr	48
GAG	GAA	GAC	ACG	GAT	CGC	AGG	TGC	CTA	TCC	CAG	AGA	TGG	CAG	CGT	GGA	GGC	AAC	GCG	TCC	AAC	TCC	TGC	ACT	403
Val	Leu	Ser	Leu	Leu	Gly	Ala	Arg	Cys	Ala	Phe	Met	Gly	Ser	Leu	Ala	His	Gly	His	Val	Ala	Asp	Phe	Leu	72
GTG	CTT	TCC	TTG	CTC	GGA	GCC	CGC	TGT	GCC	TTC	ATG	GGC	TCG	CTG	GCC	CAT	GGC	CAT	GTT	GCC	GAC	TTC	CTG	475
Val	Ala	Asp	Phe	Arg	Arg	Arg	Gly	Val	Asp	Val	Ser	Gln	Val	Ala	Trp	Gln	Ser	Gln	Gly	Asp	Thr	Pro	Cys	96
GTG	GCC	GAC	TTC	AGG	CGG	AGG	GGT	GTG	GAT	GTG	TCT	CAA	GTG	GCC	TGG	CAG	AGC	CAG	GGA	GAT	ACC	CCT	TGC	547
KHK 45																								
Ser	Cys	Cys	Ile	Val	Asn	Asn	Ser	Asn	Gly	Ser	Arg	Thr	Ile	Ile	Leu	Tyr	Asp	Thr	Asn	Leu	Pro	Asp	Val	120
TCC	TGC	TGC	ATC	GTC	AAC	AAC	TCC	AAT	GGC	TCC	CGT	ACC	ATT	ATT	CTC	TAC	GAC	ACG	AAC	CTG	CCA	GAT	GTG	619
Ser	Ala	Lys	Asp	Phe	Glu	Lys	Val	Asp	Leu	Thr	Arg	Phe	Lys	Trp	Ile	His	Ile	Glu	Gly	Arg	Asn	Ala	Ser	144
TCT	GCT	AAG	GAC	TTT	GAG	AAG	GTC	GAT	CTG	ACC	CGG	TTC	AAG	TGG	ATC	CAC	ATT	GAG	GGC	CGG	AAT	GCA	TCG	691
KHK 22																								
Glu	Gln	Val	Lys	Met	Leu	Gln	Arg	Ile	Glu	Gln	Tyr	Asn	Ala	Thr	Gln	Pro	Leu	Gln	Gln	Lys	Val	Arg	Val	168
GAA	CAG	GTA	AAG	ATG	CTA	CAG	CGG	ATA	GAA	CAG	TAC	AAT	GCC	ACG	CAG	CCT	CTG	CAG	CAG	AAG	GTC	CGG	GTG	763
KHK 49																								
Ser	Val	Glu	Ile	Glu	Lys	Pro	Arg	Glu	Glu	Leu	Phe	Gln	Leu	Phe	Gly	Tyr	Gly	Glu	Val	Val	Phe	Val	Ser	192
TCC	GTG	GAG	ATA	GAG	AAG	CCC	CGA	GAG	GAA	CTC	TTC	CAG	CTG	TTC	GGC	TAT	GGA	GAG	GTG	GTG	TTT	GTC	AGC	835
Lys	Asp	Val	Ala	Lys	His	Leu	Gly	Phe	Arg	Ser	Ala	Gly	Glu	Ala	Leu	Lys	Gly	Leu	Tyr	Ser	Arg	Val	Lys	216
AAA	GAT	GTG	GCC	AAG	CAC	CTG	GGG	TTC	CGG	TCA	GCA	GGG	GAG	GCC	CTG	AAG	GCC	TTG	TAC	AGT	CGT	GTG	AAG	907
KHK 43b																								
Lys	Gly	Ala	Thr	Leu	Ile	Cys	Ala	Trp	Ala	Glu	Glu	Gly	Ala	Asp	Ala	Leu	Gly	Pro	Asp	Gly	Gln	Leu	Leu	240
AAA	GGG	GCT	ACG	CTC	ATC	TGT	GCC	TGG	GCT	GAG	GAG	GGA	GCC	GAT	GCC	CTG	GGC	CCC	GAC	GGC	CAG	CTG	CTC	979
KHK 43a																								
His	Ser	Asp	Ala	Phe	Pro	Pro	Pro	Arg	Val	Val	Asp	Thr	Leu	Gly	Ala	Gly	Asp	Thr	Phe	Asn	Ala	Ser	Val	264
CAC	TCA	GAT	GCC	TTC	CCA	CCA	CCC	CGA	GTA	GTA	GAC	ACT	CTC	GGG	GCT	GGA	GAC	ACC	TTC	AAT	GCC	TCT	GTC	1051
Ile	Phe	Ser	Leu	Ser	Lys	Gly	Asn	Ser	Met	Gln	Glu	Ala	Leu	Arg	Phe	Gly	Cys	Gln	Val	Ala	Gly	Lys	Lys	288
ATC	TTC	AGC	CTC	TCC	AAG	GGA	AAC	AGC	ATG	CAG	GAG	GCC	CTG	AGA	TTC	GGG	TGC	CAG	GTG	GCT	GGC	AAG	AAG	1123
Cys Gly Leu Gln Gly Phe Asp Gly Ile Val ***																							298	
TGT GGC TTG CAG GGG TTT GAT GGC ATT GTG TGA GAGATGAGCGGTGGGAGGTAGCAGCTCGACACCTCAGAGGCTGGCACCCT																							1207	
GCCTGCCATTGCCTTCTTTCATTTTCATCCAGCCTGGCGTCTGGCTGCCAGTTCCCTGGGCCAGTGTAGGCTGTGGAACGGGTCTTCTGTCTCTT																							1302	
CTCTGCAGACACCTGGAGCAAAATAAATCTTCCCTGAGCCAA																							1393	

Figure 3 DNA and amino acid sequence of rat liver ketohexokinase

Bases are numbered from the 5' end of the cDNA sequence in clone 5i. The in-phase termination codon is designated by ***, the in-frame upstream stop codon is in bold underlined letters, and the putative polyadenylation signal is underlined. The region of the structural cDNA that was used to probe the cDNA library is in italics, and the site of the point mutation (A to G) in clone 5i is arrowed. Amino acids are numbered starting with the initiator Met. Amino acids identified by peptide sequencing are indicated by overlines and the peptide name is shown above each line.

from organelles. This is evinced by the considerably higher specific activity of our starting homogenate (0.07 $\mu\text{mol}/\text{min}$ per mg of protein) when compared with previous reports [10,12-14]. A 166-fold purification suggests that ketohexokinase constitutes approx. 0.6% of the cytoplasmic protein of rat livers.

The purified protein displayed a relative mobility on SDS/PAGE consistent with a molecular mass of 31 kDa, but a more accurate value of 32.8 kDa was obtained by mass spectrometry. The mass spectrum also revealed a K^+ peak, which, since the buffer does not contain potassium, we assume was bound to the protein. This is consistent with the requirement for potassium by the enzyme [11].

Amino acid sequencing and design of PCR primers

The N-terminus of ketohexokinase was not accessible to Edman

Combined treatment with trypsin (EC 3.4.21.4) and endo-proteinase Glu-C (EC 3.4.21.19) eventually yielded peptides of a suitable size for amino acid sequencing. Fractions from the six largest peptide peaks resolved by reverse-phase h.p.l.c. were sequenced. One of these, KHK 43, was found to contain two peptides, yielding a pair of PTH-amino acid products for seven of the 21 cycles of Edman degradation. This ambiguity was later easily resolved from the cDNA, which allowed the distinction of peptides KHK 43a and KHK 43b.

Two peptides, KHK 45 and KHK 49, were found to have regions of sufficiently low degeneracy to allow the design and synthesis of small pools of oligonucleotides complementary to their amino acid sequences. Rather than employ these degenerate pools to screen a cDNA library directly, it was decided to use them as primers for the amplification of the region of cDNA that they flank. This strategy allows the generation of a PCR product

to screen the cDNA library at very high stringency from the outset.

Synthesis and characterization of the PCR product

Total rat liver RNA was used as template for an oligo(dT)-primed synthesis of single-stranded cDNA. This was then employed as a template for PCR. Primers 2 and 3 (designed against peptides KHK 45 and 49) yielded a 224 bp PCR product. This was found to contain a section complementary to the amino acid sequence of a further peptide, KHK 22. The remaining sections of KHK 45 and 49 that were not represented in the oligonucleotide primers, but appeared in the PCR product, were consistent with the amino acid sequences obtained (Figure 3). We believed that these characteristics adequately confirmed the identity of the PCR product as being a fragment of the coding region of ketohexokinase cDNA, and proceeded to use it as a probe to screen the cDNA library.

Screening of the cDNA library

Twelve independent isolates, with inserts ranging in size from 1 kb to 2.5 kb, were obtained by plaque hybridization of approx. 10^6 clones. The three largest (2.3 kb, 2.4 kb and 2.5 kb) were rescued from the phage to yield the clones in the pBluescript SK(−) phagemid. Restriction mapping of the clones and preliminary sequencing of *ExoIII*-deleted constructs of clone 16i revealed that all three contained substantial amounts of co-ligated, unassociated cDNA attached by distinctive *EcoRI* linkers (Figure 1).

Characterization and sequencing of clones

Both clones 16i and 5ii contained poly(A)⁺ tails at their 3'-termini, which identified the cDNA as being a copy of fully or partially processed mRNA. The consensus polyadenylation signal AATAAA was identified 20 nucleotides upstream of each poly(A)⁺ tract [32]. It was initially unclear whether clone 5ii contained the start codon and any associated 5' untranslated region. Clone 5i contained the longest section of 5' untranslated region, but did not contain the C-terminus of the coding region (Figure 1).

The hybrid clone 5i/16i (Figure 1) contained a section of cDNA encompassing the coding region and both the 3' and 5' untranslated regions, free of co-ligated cDNA. Sequencing revealed a point mutation in the parent clone 5i (AAT to GAT), resulting in a change of Asn-18 to Asp-18. Two lines of evidence supported the identification of this as a mutation. Firstly, the sequences of clone 5ii/a and the *ExoIII*-deleted constructs of 16i encoded Asn-18; secondly, the amino acid sequence of peptide KHK 32 revealed an Asn at this position. The probability of this Asn to Asp change being a polymorphism is very slight, since the cDNA library and protein sequence were both obtained from the same strain of rat.

The complete nucleotide sequence and the corresponding predicted amino acid sequence of ketohexokinase are shown in Figure 3. The coding region has been verified using data obtained from both strands of at least two independent clones. The translation initiation codon was assigned using several criteria. Firstly, its position relative to the 3' termination codon would generate a protein with a molecular mass (32728 Da) which is in close agreement with that determined by mass spectrometry (32800 Da). Secondly, all seven peptide sequences from the purified ketohexokinase lie in the open reading frame created by this start codon. Thirdly, it is the only in-frame ATG codon between the region encoding peptide KHK 32 and the in-frame

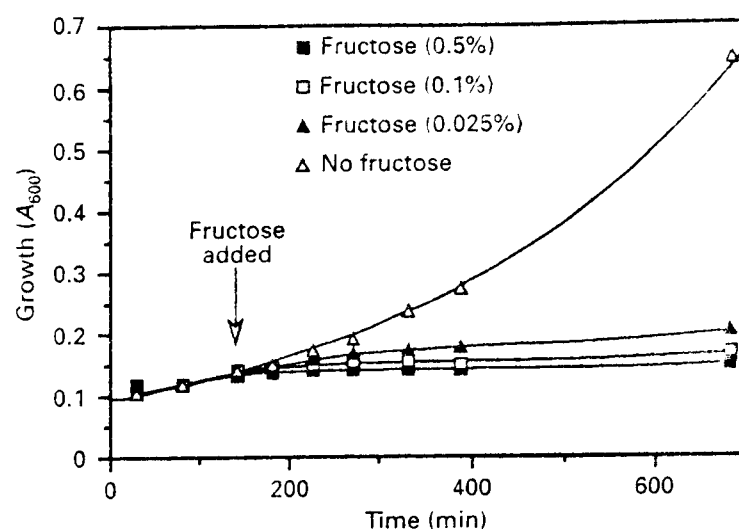


Figure 4 Growth profiles of the fructose-intolerant yeast strain TDB I upon addition of fructose at various concentrations

An exponential-phase culture of cells in SD Leu[−] medium was divided into equal aliquots, and the growth was monitored by measurement of A_{600} . After 2 h, fructose was added to the cultures to a final concentration of 0.5% (w/v) (■), 0.1% (w/v) (□) or 0.025% (w/v) (▲), or water was added as a control (△).

stop codon 198 nucleotides upstream of this initiation site. Fourthly, the overwhelming predominance of cytidine at the −1 and −4 positions of the initiation site sequences of rat cDNAs [33] is conserved within the ketohexokinase cDNA when it is assigned this start codon.

Expression of ketohexokinase in *Saccharomyces cerevisiae*

Fifteen transformants of the yeast TDS II bearing the plasmid pMA91-KHK (termed TDB I), and three bearing plasmid pMA91-KHK(−) [termed TDB I(−)], were inoculated on to SD Leu[−] medium. The presence of the ketohexokinase gene in transformants was confirmed by amplification of its coding region from plasmid DNA minipreps of these strains by PCR. Confirmation that the leucine-independence was borne epistemally, and was not the result of reversion, was given by the loss of this phenotype after passage of the cells through non-selective YEPD medium.

Fructose-intolerance of TDB I strains was shown by their inability to grow when replica-plated on to SD Leu[−] medium containing 0.5% (w/v) fructose. TDB I(−) strains did not show intolerance when tested in this manner. The effects of various concentrations of fructose on the growth of a TDB I strain in liquid culture are shown in Figure 4. A similar culture of TDB I(−) was unaffected by the addition of fructose (results not shown). Metabolite extracts of TDB I cells exposed to 0.5% (w/v) fructose for 3 h showed the presence of fructose 1-phosphate (30 μ mol of fructose 1-phosphate/g wet weight of cells). Fructose 1-phosphate was not present in TDB I(−) cell extracts.

Crude extracts from TDB I contained ketohexokinase with a specific activity of 0.62 μ mol/min per mg of protein, which is almost nine times greater than that found in rat liver (0.07 μ mol/min per mg of protein). Ketohexokinase activity was absent from extracts of both the TDS II and TDB I(−) strains. Control assays showed a minor ATPase activity in TDB I (0.069 μ mol/min per mg of protein) which was absent from the parent TDS II strain. Curiously, in TDB I(−) this ATPase assumed a much higher activity (0.214 μ mol/min per mg of protein).

Due to the co-migration of a large band of yeast protein in the crude extracts, partial purification of positive and control extracts

Table 1 Comparison of proteins using the MULTALIGN program

The percentage identity between the proteins is indicated above the diagonal, and the similarity score is shown below the diagonal. The similarity scores were determined by randomizing the sequence 100 times, and comparing each randomized sequence. The means and the S.D.s of the distribution of comparisons were calculated, and the similarity is expressed in terms of the number of units the S.D. is away from the mean of the random distribution. Abbreviations: KHK, ketohexokinase; pfk2, *E. coli* minor phosphofructokinase; FK, *V. alginolyticus* fructokinase; RK, *E. coli* ribokinase; GK and HK, rat liver glucokinase and hexokinase respectively.

	KHK	pfk2	FK	RK	GK	HK
KHK		21.4	18.24	19.66	20.27	21.21
pfk2	4.48		20.41	22.67	20.78	21.17
FK	3.24	6.17		23.21	16.99	19.61
RK	5.82	9.00	8.15		22.08	18.45
GK	0.03	-0.68	-0.33	-0.49		52.11
HK	1.24	-1.94	1.26	-0.50	59.23	

was necessary before a 31 kDa band corresponding to that for the ketohexokinase standard could be identified by SDS/PAGE.

Comparison of ketohexokinase sequence with other protein sequences

The National Biomedical Research Foundation (NBRF) Protein Identification Resource (PIR) V.29 databases were searched with the ketohexokinase protein sequence, which proved to be unique. Ketohexokinase does not contain sequence complementary to either the ATP- or hexose-binding sites that are common to members of the hexokinase family of proteins [7]. Pairwise comparison of amino acid sequences using the Alignment of Multiple Sequence (AMPS) program MULTALIGN [34] showed that no single protein had a prominently similar sequence identity with ketohexokinase, the greatest percentage identity being ~ 20 %. Of the protein sequences showing this degree of identity, three furanose sugar kinases were selected for detailed com-

parison, because of their similar size, function and substrate specificity. These proteins were the *E. coli* minor phosphofructokinase (EC 2.7.1.11) [35], *E. coli* ribokinase (EC 2.7.1.15) [36] and *Vibrio alginolyticus* fructokinase (EC 2.7.1.4) [37].

Multiple alignment of the ketohexokinase sequence with those of these three proteins was performed using the MULTALIGN program. This program uses the Needleman and Wunsch algorithm, modified to allow quick comparison of more than two sequences [34]. Rat hexokinase and glucokinase were also compared using this method, and the results are summarized in Table 1.

The multiple alignment of the four proteins revealed a highly conserved motif, TX(A/G)AGDX₃L, where the amino acids X are nearly always conservative amino acid replacements (Figure 5). This consensus was used to search the PIR databases and was also found in a fifth protein, *E. coli* DNA topoisomerase I [38].

DISCUSSION

The calculated molecular mass of rat liver ketohexokinase reported here (32 728 Da) differs somewhat from the value obtained from the protein (28 kDa) by Sánchez et al. using density gradient centrifugation [13], but we are confident that our value is correct because it is in close agreement with that obtained by mass spectrometry. Although the protein has six potential glycosylation sites (Asn-Xaa-Ser/Thr), the mass spectrum showed no evidence of glycosylation.

The fructose-intolerance displayed by yeast expressing ketohexokinase is associated with a large, and essentially irreversible, accumulation of fructose 1-phosphate. The level of fructose 1-phosphate in TDB I cells treated with fructose was approximately three times the level of total sugar phosphates in yeasts performing aerobic glycolysis [39]. This accumulation of fructose 1-phosphate is similar to that observed in mammalian liver in subjects suffering from hereditary fructose intolerance [2]. In both cases, this is due to the absence of aldolase B activity when ketohexokinase remains active. Administration of fructose to rat

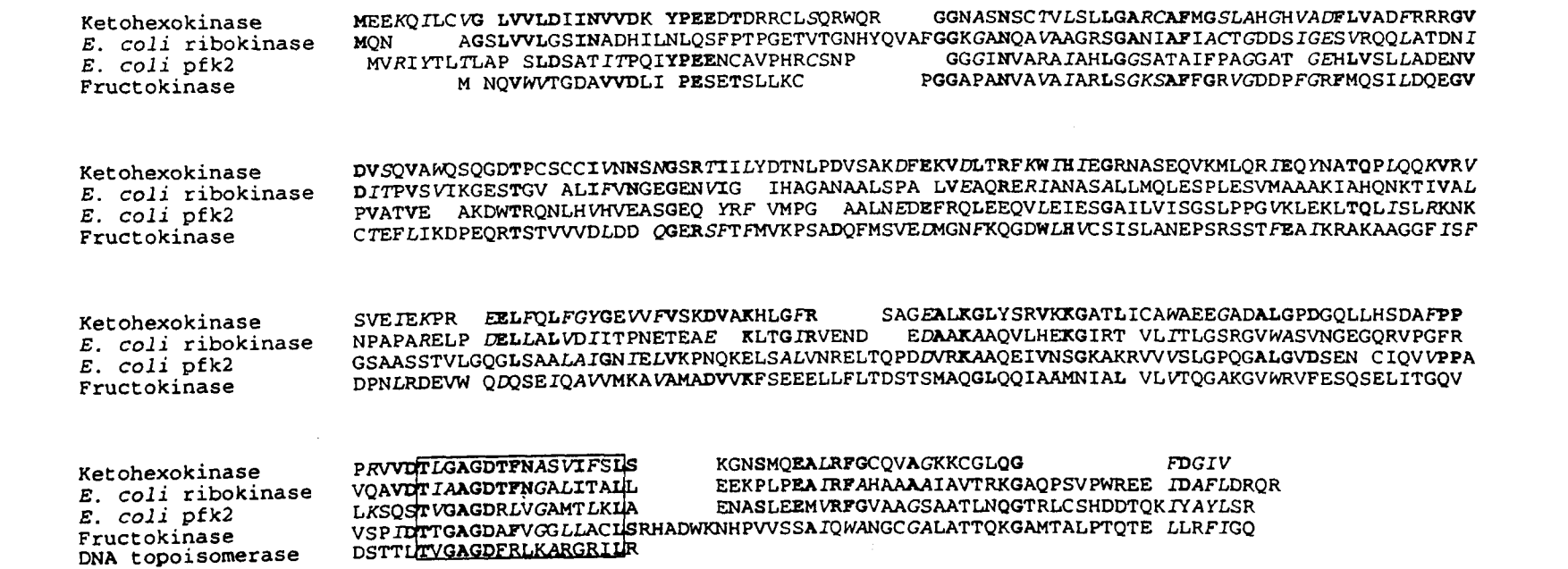


Figure 5 Comparison of amino acid sequences of ketohexokinase and other furanose kinases

The sequences were compared using the MULTALIGN program. The amino acid similarities of ribokinase, phosphofructokinase 2 (pfk2) and fructokinase are compared with those of ketohexokinase. Those found to be identical are in bold, and those with conservative replacements are italicized. The boxed sections contain a region of high sequence identity, the consensus of which was used to search the PIR databases. This led to the identification of DNA topoisomerase I, the relevant sequence of which is indicated below these sections (amino acid residues 414–435). The sequences

liver results in a transient depletion of ATP and cytoplasmic orthophosphate in the tissue owing to the sequestration of phosphate in fructose 1-phosphate. The absence of aldolase B in patients suffering from hereditary fructose intolerance will prevent recovery from such depletions. This causes a severe compromise of energy metabolism, although an exact locus at which metabolic paralysis is imposed has not been identified. It is clear that the fructose-intolerant yeast (TDB I) suffers equally severe metabolic impairment on exposure to fructose, and could serve as a metabolic model to study this type of disorder.

It is our intention to exploit the fructose-intolerance of yeast expressing ketohexokinase to select for mutations of invertase that retain fructotransferase activity [40], but will not hydrolyse sucrose. The genotype of the TDS II strain renders it incapable of utilizing sucrose as a carbon source. A fructose-intolerant strain derived from this will only survive on sucrose as its sole carbon source if it is expressing a mutated invertase that sequesters fructose as a fructofuranoside and only releases glucose. Under these circumstances, cells expressing wild-type hydrolase activity will die due to the release of fructose, and cells expressing inactive mutations of invertase will not grow due to the lack of a utilizable carbon source.

Interestingly, the amino acid sequence of ketohexokinase reveals that it is much more closely related to the prokaryotic furanose kinases examined than it is to the eukaryotic hexokinase family. Table 1 predictably shows a high percentage identity when a direct comparison is made between glucokinase and hexokinase. The percentage identities obtained with all other pairwise comparisons are ~20%. However, the similarity scores clearly show that, although glucokinase and hexokinase are very closely related to each other, they show no similarity with the other proteins. The similarity scores obtained for all possible comparisons between the other proteins show that they are related to one another, although, unsurprisingly, the mammalian ketohexokinase appears to be the most evolutionarily divergent protein within the group. The most striking similarity between the proteins is the conserved region depicted in Figure 5. Although these proteins are all ATP-dependent kinases, the possibility that this region encodes a common binding site for ATP would seem to be discounted by its conservation in the structural gene for DNA topoisomerase I, which does not employ ATP or NAD⁺ as a substrate. The proposed mechanism for *E. coli* topoisomerase I [41] involves the transient breakage of a DNA backbone bond and simultaneous formation of a phosphotyrosine bond between the protein and the cut 5' phosphoryl group. The cleavage specificity suggests that the enzyme is interacting with both sides of the point it cleaves, and therefore bridges the breakpoint, binding non-covalently with the 3' deoxyribose sugar. The conserved motif TX(A/G)AGDX₉L may be involved in the non-covalent binding of either the 5' deoxyribose phosphate group or the 3' sugar. The conservation of this region may therefore reflect a common function in the binding of furanose sugars.

Ketohexokinase has recently been reported in the halophilic archaeobacterium (*Haloarcula vallismortis*) [42]. This is the first report of an ATP-dependent D-fructose 1-phosphotransferase in prokaryotes. It would be interesting if future studies could show whether this protein belongs within the same group of furanose kinases revealed in this paper, as this would lend support to our belief that this represents an ancient family of proteins.

We are indebted to Dr. David Bonthron of the University of Edinburgh for his technical advice and for critically reading this manuscript. We also thank Dr. David Harvey for assistance with the mass spectrometry, Mr. Tony Willis for performing the amino acid sequence analysis, and Dr. Geoff Barton for his assistance with the sequence comparisons. This work was supported by the University of Oxford. T.C.D. received a Science and Engineering Research Council studentship. N.M. was supported by scholarships from the University of Oxford and the British Council.

REFERENCES

- Cori, G. T., Ochoa, S., Stein, M. W. and Cori, C. F. (1951) *Biochim. Biophys. Acta* **7**, 304–317
- Van den Berghe, G. (1986) *Prog. Biochem. Pharmacol.* **21**, 1–32
- Malaisse, W. J., Malaisse-Lagae, F., Davies, D. R. and Van Schaftingen, E. (1989) *FEBS Lett.* **255**, 175–178
- Cross, N. C. P., Tolan, D. R. and Cox, T. M. (1988) *Cell* **52**, 881–885
- Schapira, F., Schapira, G. and Dreyfus, J. C. (1961/1962) *Enzym. Biol. Clin.* **1**, 170–175
- Cárdenas, M. L., Rabajille, E. and Niemeyer, H. (1984) *Biochem. J.* **222**, 363–370
- Griffin, L. D., Gelb, B. D., Wheeler, D. A., Davison, D., Adams, V. and McCabe, E. R. B. (1991) *Genomics* **11**, 1014–1024
- Walker, D. G. (1968) *Essays Biochem.* **2**, 33–67
- Rauschel, F. M. and Cleland, W. W. (1973) *J. Biol. Chem.* **248**, 8174–8177
- Rauschel, F. M. and Cleland, W. W. (1977) *Biochemistry* **16**, 2169–2175
- Sánchez, J. J., González, N. S. and Pontis, H. G. (1971) *Biochim. Biophys. Acta* **227**, 79–85
- Adelman, R. C., Ballard, F. J. and Weinhouse, S. (1967) *J. Biol. Chem.* **242**, 3360–3365
- Sánchez, J. J., González, N. S. and Pontis, H. G. (1971) *Biochim. Biophys. Acta* **227**, 67–78
- Bais, R., James, H. M., Rofe, A. M. and Conyers, R. A. J. (1985) *Biochem. J.* **230**, 53–60
- Romanos, M. A., Scorer, C. A. and Clare, J. J. (1992) *Yeast* **8**, 423–488
- Bensadoun, A. and Weinstein, D. (1976) *Anal. Biochem.* **70**, 241–250
- Hames, B. D. (1981) in *Gel Electrophoresis of Proteins* (Hames, B. D. and Rickwood, D., eds.), pp. 1–86, IRL Press, Oxford
- Beavis, R. C. and Chait, B. T. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6873–6877
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103–119
- Bonthron, D. T. (1990) *Gene* **91**, 101–105
- Sanger, F. (1981) *Science* **214**, 1205–1210
- Short, J. M., Fernandez, J. M., Sorge, J. A. and Huse, W. D. (1988) *Nucleic Acids Res.* **16**, 7583–7600
- Friedman, K. D., Rosen, N. L., Newman, P. J. and Montgomery, R. R. (1990) in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J., eds.), pp. 253–257, Academic Press, San Diego
- Hamilton, R., Watanabe, C. K. and de Boe, H. A. (1987) *Nucleic Acids Res.* **15**, 3581–3593
- Mellor, J., Dobson, M. J., Roberts, N. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J. and Kingsman, S. M. (1983) *Gene* **24**, 1–14
- Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) *J. Bacteriol.* **153**, 163–168
- den Hollander, J. A., Ugurbil, K. and Shulman, R. G. (1986) *Biochemistry* **25**, 212–219
- Eggleson, L. V. (1983) in *Methods of Enzymatic Analysis* (Bergmeyer, H. U., ed.), vol. 2, pp. 328–334, Verlag Chemie, Weinheim
- Jazwinski, S. M. (1990) *Methods Enzymol.* **182**, 154–174
- Proudfoot, N. and Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214
- Kozak, M. (1984) *Nucleic Acids Res.* **14**, 857–872
- Barton, G. J. (1990) *Methods Enzymol.* **183**, 403–428
- Daldal, F. (1984) *Gene* **28**, 337–342
- Hope, J. N., Bell, A. W., Hermodson, M. A. and Groarke, J. M. (1986) *J. Biol. Chem.* **261**, 7663–7668
- Blatch, G. L., Schoile, R. R. and Woods, D. R. (1990) *Gene* **95**, 17–23
- Tse-Dihn, Y.-C. and Wang, J. C. (1986) *J. Biol. Chem.* **261**, 321–331
- Brindle, K. M. (1988) *Biochemistry* **27**, 6187–6196
- Edelman, J. (1956) *Adv. Enzymol.* **17**, 189–232
- Wang, J. C. (1985) *Annu. Rev. Biochem.* **54**, 665–697
- Altekar, W. and Rangaswamy, V. (1991) *FEMS Microbiol. Lett.* **83**, 241–246

Metabolic Studies of a Fructose-Intolerant Yeast by *in vivo* ^{31}P -NMR Spectroscopy.

Timothy C. Doyle¹, Corinne M. Spickett^{2*}, Iain A. Donaldson^{1†} and R. George Ratcliffe².

¹ Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU U.K.

² Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB U.K.

Key Words: ketohexokinase; fructose-intolerance; sugar phosphates; ^{31}P -NMR; *Saccharomyces cerevisiae*.

In press, 'Yeast'

Submitted: 15th February, 1993, accepted without revision: 29th March, 1993

† Corresponding author: I. A. Donaldson, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU U. K., Fax 0865 275259, Tel. 0865 275221

* Current address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW U.K.

Abstract

The metabolic effects of the administration of fructose to a yeast expressing the cDNA for rat liver ketohexokinase has been investigated by ^{31}P NMR. Cessation of growth suffered by the yeast on exposure to 5 and 25mM fructose was accompanied by a large accumulation of fructose 1-phosphate at the expense of cytoplasmic orthophosphate and NTP. Shifts in resonances were consistent with a drop in cytoplasmic pH. Arresting growth with 1mM fructose, however, did not result in these changes, although a large accumulation of fructose 1-phosphate occurred which may have been supported by the mobilization of polyphosphates.

1. Introduction

A strain of yeast, TDB I, has recently been described (Doyle *et al.*, 1992; Donaldson *et al.*, 1993) which has been rendered intolerant of the presence of fructose in its growth media by constitutive expression of the cDNA for rat liver ketohexokinase (ATP D-fructose 1-phosphotransferase; EC 2.7.1.3) borne on the expression vector *pMA* 91 (Mellor *et al.*, 1983). This enzyme normally catalyzes the primary step in a pathway that converts the dietary fructose in mammals into intermediates of glycolysis and gluconeogenesis (reviewed in Van de Berghe, 1986). In this pathway, fructose is phosphorylated at the C-1 position and the product undergoes direct cleavage by aldolase B (EC 4.1.2.13) to dihydroxyacetone phosphate and glyceraldehyde. The latter intermediate is then phosphorylated by triokinase (EC 2.7.1.28).

It is our intention to exploit fructose intolerance in yeast to select for mutations of invertase that retain fructotransferase activity, but will not hydrolyse sucrose. The phenotype of TDB I renders it unable to utilize sucrose as a carbon source (*mal*^o *suc*^o) unless an episomally-borne fructotransferase mutant of *SUC2* is expressed. In the presence of a suitable acceptor substrate, such mutants will release free glucose as a utilizable carbon source, retaining the fructose moiety as a fructofuranoside. Wild-type mutants will not grow due to a release of free fructose, and inactive mutants will not be able to grow due to the lack of a carbon source. A greater understanding of the metabolic effects of fructose on TDB I is therefore of interest to us.

The growth of a log-phase culture of the yeast TDB I, which expresses ketohexokinase but contains no aldolase B, was completely arrested within 90 minutes by the addition of fructose to a final concentration as low as 5mM. The response was found to be associated with a large and essentially irreversible accumulation of fructose 1-phosphate (Donaldson *et al.*, 1993). This accumulation is similar to that observed in human liver suffering hereditary fructose-intolerance (Gitzelmann *et al.*, 1989). In both cases the build up of fructose 1-phosphate is due to the absence of aldolase B activity, when ketohexokinase remains active. This causes a severe compromise of energy metabolism in the presence of fructose, although the exact locus at which metabolic paralysis is imposed has not yet been

identified. *In vivo* ^{31}P -NMR spectroscopy provided a convenient non-invasive method to identify metabolic pools from which orthophosphate was drawn and sequestered as fructose 1-phosphate in the study of the hepatic metabolism of fructose administered to healthy (Oberhaensli *et al.*, 1986) and fructose intolerant (Oberhaensli *et al.*, 1987) human subjects, and the same approach was used here to investigate the metabolic impairment of TDB I.

2. Experimental

2.1. Cell preparation

The *Saccharomyces cerevisiae* strain TDB I (*leu2-3*, *leu2-112*, *trp1*, *suc*^o, *mal*^o, *hxx1*, *hxx2*, *GLK*, *pho8-2*, *pho3* Δ 1, *pho5* Δ 1, MAT α , bearing the plasmid *pMA* 91-KHK) was prepared as previously described (Donaldson *et al.*, 1993). Cells were grown aerobically at 30°C on defined media (Burkeholder, 1943) with low phosphate (30mg/L, KH_2PO_4) and an amino acid mixture lacking leucine. The main cultures (500ml in a 2L Erlenmeyer flask) were inoculated with 1ml samples of log-phase (overnight) precultures and harvested in mid log phase, by low speed centrifugation. The cells were washed in 30ml ice-cold water and then in ice-cold substrate-free MES buffer (25mM MES pH 6.0, 2mM MgSO_4 , 2mM KCl, 1.7mM NaCl) before immobilization in agarose threads as described by Brindle and Krikler (1985). The agarose-yeast matrix was perfused with the same buffer except that 50mM glucose was added and the effect of fructose was monitored by its addition to the perfusion buffer to final concentrations of 1, 5 or 25mM.

2.2. ^{31}P -NMR spectroscopy of immobilized cells

^{31}P -NMR spectra were recorded on a Bruker CXP300 spectrometer, operating at a frequency of 121.49 MHz with a double-tuned $^{13}\text{C}/^{31}\text{P}$ probehead. All spectra were acquired with a spectral window of 8064 Hz. As fully relaxed spectra were not required, spectra were recorded using a 90° pulse angle and a 1.0 s recycle time. The spectra were acquired sequentially over the time course in files of 900 scans (equivalent to 15 minutes) and each free

induction decay contained 1K data points. Chemical shifts were measured relative to the signal from a capillary containing a 2% (v/v) aqueous solution of the tetramethyl ester of methylene diphosphonate (TEM DP). The TEM DP signal occurs at 22.49 ppm relative to the signal from 85% H₃PO₄ and it was ¹H-decoupled in order to improve the lineshape. Spectral peaks were assigned by comparison with previously published spectra (Brindle and Krikler, 1985; Campbell-Burke and Shulman, 1987).

2.3. ³¹P-NMR of perchloric acid extracts of yeast-agarose threads

Perchloric acid extracts of the yeast-agarose threads were prepared by the method of Brindle (1988), and the lyophilized extracts were dissolved in 3ml 50mM TEA, pH 8.0, 5mM EDTA. These extracts were either used directly for the collection of spectra, in which the assignment of the fructose 1-phosphate peak was confirmed by the addition an authentic sample to 2mM (Sigma), or for enzymatic analysis. The enzymatic identification of fructose 1-phosphate involved the observation of the disappearance of the fructose 1-phosphate resonance on treatment of the extract with c.0.4mg/ml aldolase B (prepared from rat liver according to Eggleston, 1983) coupled to 1.25 U/ml GPDH (Boehringer) and 225μM NADH to drive the reaction towards fructose 1-phosphate degradation. ¹H-decoupled ³¹P-NMR spectra of the perchloric acid extracts were recorded in files of 1024 scans, each containing 8K data points. A 90° pulse angle and 4 second recycle time were used.

2.4. Characterization of heterologously expressed ketohexokinase.

Protein extractions of the transformed TDB I and TDB I(-) yeast strains, and the untransformed TDS II parent strain were performed and partially purified according to Donaldson *et al.* (1993). These were then analysed by SDS-PAGE (Hames, 1981), with a standard of purified rat ketohexokinase. The partially purified TDB I extract was then applied to a 1.7cm² × 53cm Sephacryl S100 (Pharmacia LKB Biotechnology) column, and eluted with 10mM triethanolamine/KOH, pH 7.3, at a flow rate of 0.5ml/min. The purified rat enzyme was also analysed in a similar fashion. In both cases, fractions containing ketohexokinase were determined using the colorimetric assay of Donaldson *et al.* (1993). The

column was calibrated under similar conditions with a mixture of γ -globulin, serum albumin, ovalbumin and cytochrome c.

3. Results and Discussion

As previously reported for growth on other media (Donaldson *et al.*, 1993), fructose had a marked effect on the growth profile of the yeast transformant TDB I in liquid culture. An external concentration as low as 1mM caused almost complete inhibition of growth within 90 minutes, and at 25mM growth ceased immediately after the addition of fructose (fig. 1). ^{31}P NMR allowed simultaneous monitoring of the changes in the pools of a number of important phosphorylated metabolites.

Figure 2 shows the ^{31}P NMR spectra recorded from the fructose-intolerant strain during a control experiment in which glucose was the only carbon source. The spectra were similar to those of untransformed yeast cells (Brindle and Krikler, 1985; Campbell-Burke and Shulman, 1987) and the only unusual feature was the relatively large size of the unassigned phosphodiester signal at -1.25 ppm. The cytoplasmic orthophosphate (Pi) and nucleoside triphosphate (NTP) signals increased slightly over the time-courses of 2-3 hours but this drift in intensity was not sufficient to complicate the interpretation of the fructose loading experiments described below. The polyphosphate signal at -22.6 ppm decreased during the control experiments, although it was also observed to increase on one occasion. The physiological significance of changes in the intensity of this resonance are difficult to interpret because of the unknown proportion of NMR "invisible" high molecular weight and precipitated polyphosphates. However, the change in intensity does indicate that the detectable polyphosphate pool is dynamic and thus capable in principle of contributing to the phosphorus requirements of the cells, even though TDBI lacks a functional vacuolar alkaline phosphatase (*pho8⁻*). When a fructose-tolerant (TDB I(-)) strain was perfused with 50mM glucose, or 50mM glucose plus 25mM fructose (data not shown), no significant differences in spectra compared with those shown in Figure 2 were observed.

When the fructose-intolerant strain (TDB I) was incubated with 1mM fructose, a signal appeared in the ^{31}P NMR spectrum at 4.8 ppm (fig. 3). This new signal was assigned to fructose 1-phosphate on the basis of spectrophotometric enzyme assays of cell extracts and the observation that addition of fructose 1-phosphate (Sigma) to the extract resulted in a resonance at the same chemical shift. Upon exposure to 5mM (fig. 3d) or 25mM (fig. 4) fructose it was observed that both the fructose 1-phosphate signal and the residual cytoplasmic Pi signal broadened and shifted upfield. These changes in chemical shift indicate that the cytoplasm became more acidic, most probably due to substrate limitation of the plasmalemma H^+ -ATPase, which is thought to play an important role in pH regulation in yeast cells (Vallejo and Serrano, 1989), while the broadening of the signal indicates a heterogeneity in the response of the cell population to the fructose treatment. ^{31}P NMR studies on perfused rat liver (Iles *et al.*, 1980) with either 5mM or 10mM fructose also showed rapid falls in the concentrations of NTP, Pi and cytoplasmic pH, accompanied by a rapid rise in fructose 1-phosphate. In this case, the decrease in pH was attributed to the production of lactate from fructose.

Two responses to different concentrations of fructose could be discerned in Figures 3 and 4. At high concentrations of fructose (5mM, fig. 3d and 25mM, fig. 4) the immediate cessation of growth on addition of fructose was attended by a severe disruption of the pools of phosphorylated metabolites. However, the arresting of cell division by a low concentration of fructose (1mM, fig. 3a-c) was not accompanied by the same changes, although a depletion of the polyphosphate resonance was observed.

At the higher concentrations of fructose, the marked fall in the NTP and cytoplasmic Pi levels, to a point where the latter became undetectable, coupled with the loss of pH control, is likely to produce a condition in which growth cannot be sustained. In principle the TDB1 cells could draw on the polyphosphate pool during the accumulation of fructose 1-phosphate, but the existence of the polyphosphate pool did not prevent the depletion of the cytoplasmic Pi and NTP pools. Analogous observations have been made in higher plant tissues using mannose and galactose to sequester phosphate [Loughman *et al.*, 1989a; 1989b)]. The

accumulation of the slowly metabolized mannose 6-phosphate and galactose 1-phosphate in these tissues depleted the cytoplasmic Pi and NTP pools even in the presence of a large vacuolar Pi pool, indicating that Pi transport across the tonoplast was slow in relation to the demand for phosphate that had been imposed on the cytoplasm.

The immediate effect of fructose on growth may however be more subtle, and could involve the cellular concentration of GTP, a metabolite involved in transmembrane and cytoplasmic signal transduction. It has been previously observed that ketohexokinase can utilize GTP to phosphorylate fructose (Phillips and Davies, 1985), and that the symptoms of hereditary fructose intolerance in humans are due to the depletion of hepatic GTP. Although it is not possible to differentiate between ATP and GTP *in vivo* with NMR, the overall depletion of the NTP pool would suggest that GTP levels have fallen.

At lower concentrations, where fructose had little effect on the pool size of NTP and Pi (figs. 3a-c), fructose was likely to be exerting its effect on TDB I by the continuous diversion of phosphate from energy metabolism into the production of fructose 1-phosphate at the expense of cell division. Under these conditions, it is possible that these pools were being maintained by mobilization of polyphosphates, and that the slow and restricted growth observed in Figure 1 may also have been drawing on this source of phosphate. Interestingly, addition of phosphate to the growth medium of TDB I did not relieve inhibition of growth by fructose at either low or high fructose concentrations (results not shown), suggesting that the availability of phosphate is not the limiting factor in determining growth.

The *in vivo* NMR data support the view that active ketohexokinase has been expressed in the yeast strain TDB1. The recombinant enzyme has been shown to have a number of characteristics identical to rat liver ketohexokinase. It produces fructose 1-phosphate from ATP and fructose, and it confers fructose intolerance as a result of the cells' inability to metabolize fructose 1-phosphate, because they lack aldolase B. In the light of this similarity, it was unsurprising that the heterologously expressed protein was also found to have similar physical characteristics to the native rat enzyme by both SDS-PAGE and size exclusion chromatography (fig. 5). The yeast ketohexokinase was shown to be present in the TDB I

strain only, and to have a subunit size of approximately 31 kDa (fig. 5a), the same value obtained for the rat enzyme (actual size is 32728 Da). Gel filtration of the yeast extracts and purified rat enzyme (fig. 5b) show that both have a native molecular mass of approx. 134 kDa, consistent with a tetrameric quaternary structure.

We are therefore confident that the heterologously expressed ketohexokinase is capable of undergoing similar folding and subunit association as the rat enzyme, as well as exhibiting a similar activity and conferring fructose intolerance in the absence of the aldolase B gene. These results confirm the suitability of such a system as a selection strain for screening of mutants of the *SUC2* gene for invertase that express a fructotransferase activity, and therefore do not release free fructose into the medium. However, since it is intended that libraries of *SUC2* mutants are to be introduced into the selection strain on a plasmid, the next objective must be to integrate multiple copies of the ketohexokinase coding region into the yeast chromosome, thus avoiding any potential plasmid competition.

4. *References*

- Brindle, K. M. (1988). ^{31}P NMR magnetization-transfer measurements of fluxes between inorganic phosphate and adenosine 5'-triphosphate in yeast genetically modified to overproduce phosphoglycerate kinase. *Biochem.* **27**, 6187-6196.
- Brindle, K. and Krikler, S. (1985). ^{31}P -NMR saturation transfer measurements of phosphate consumption in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **847**, 285-292.
- Burkholder, P. R. (1943). Vitamin deficiencies in yeasts. *American J. Botany* **30**, 206-211.
- Campbell-Burke, S.L. and Shulman, R.G. (1987). High-resolution NMR studies of *Saccharomyces cerevisiae*. *Ann. Rev. Microbiol.* **41**, 595-616.
- Donaldson, I.A., Doyle, T.C. and Matas, N. (1993). Expression of rat liver ketohexokinase in yeast results in fructose intolerance. *Biochem. J.* in press.
- Doyle, T.C., Donaldson, I. A., Matas, N. and Cox, B. S. (1992). Genetic engineering of a fructose-intolerant yeast strain to screen for fructotransferase mutants of invertase. *Yeast* **8** S573.

- Eggleston, L.V. (1983). D-Fructose 1-phosphate. In Bergmeyer, H. U., (Ed), *Methods of Enzymatic Analysis*, vol. 2. Verlag Chemie, Weinheim, pp. 328-334.
- Gitzelmann, R., Steinmann, B., Van den Berghe, G. (1989). Disorders of fructose metabolism. In Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (Eds), *The metabolic basis of inherited disease*, vol. 1, ch. 11, 6th edn., McGraw-Hill, New York.
- Hames, B.D. (1981). An introduction to polyacrylamide gel electrophoresis. In Hames, B.D. and Rickwood, D. (Eds), *Gel Electrophoresis of Proteins*, ch. 1. IRL Press, Oxford.
- Iles, R.A., Griffiths, J.R., Stevens, A.N., Gadian, D.G. and Porteous, R. (1980). Effects of fructose on the energy metabolism and acid-base status of the perfused starved-rat liver. *Biochem. J.* **192**, 191-202.
- Loughman, B.C., Ratcliffe, R.G. and Schwabe, J.W.R. (1989a). Galactose metabolism in *Zea mays* root tissues observed by ^{31}P -NMR spectroscopy. *Plant Sci.* **59**, 11-23.
- Loughman, B.C., Ratcliffe, R.G. and Southon, T.E. (1989b). Observation on the cytoplasmic and vacuolar orthophosphate pools in leaf tissue using *in vivo* ^{31}P -NMR spectroscopy. *FEBS Lett.* **242**, 279-284.
- Mellor, J., Dobson, M. J., Roberts, N. A., Tuite, M. F., Emtage, J. S., White, S., Lowe, P. A., Patel, T., Kingsman, A. J. and Kingsman, S. M. (1983). Efficient synthesis of enzymatically active calf chymotrypsin in *Saccharomyces cerevisiae*. *Gene* **24**, 1-14.
- Oberhaensli, R.D., Galloway, G.J., Taylor D.J., Bore, P.J. and Radda, G.K. (1986). Assessment of human liver metabolism by phosphorous-31 magnetic resonance spectroscopy. *Br. J. Radiol.* **59**, 695-699.
- Oberhaensli, R.D., Rajagolpalan, B, Taylor D.J., and Radda, G.K. (1987). Study of hereditary fructose intolerance by use of ^{31}P magnetic resonance spectroscopy. *Lancet* **2**, 931-934.
- Phillips, M.I. and Davies, D.R. (1985). The mechanism of guanosine triphosphate depletion in the liver after a fructose load. *Biochem. J.* **228**, 667-671.

Vallejo, C.G. and Serrano, R. (1989). Physiology of Mutants with reduced expression of plasma membrane H⁺-ATPase. *Yeast* **5**, 307-319.

Van de Berghe, G. (1986). Fructose: metabolism and short-term effects on carbohydrate and purine metabolic pathways. *Prog. Biochem. Pharmacol.* **21**, 1-32.

Figure Legends

Figure 1. Growth profile of the fructose-intolerant yeast strain TDB I upon the addition of fructose. Cells were grown as described to early log phase, with shaking before the addition of fructose to 25mM (○), 5mM (Δ) or 1mM (□) after 2 hours. A control culture (◇) had no addition of fructose. Growth was monitored before and after addition of fructose by A₆₀₀ measurement. Supplementing the growth medium with phosphate (1.5g/L, KH₂PO₄, data not shown) did not effect growth profiles.

Figure 2. ³¹P NMR spectra of the yeast transformant TDB I during a control experiment. The immobilized cells were bathed in a circulating medium containing 50mM glucose and the spectra were recorded (a) 30 minutes and (b) 4 hours after the start of the circulation. The numbered peaks correspond to: 1, several phosphomonoesters; 2, cytoplasmic phosphate; 3, an unidentified phosphodiester; 4, 5 and 6, the γ-, α-, and β- phosphates, respectively, of NTP; 7, polyphosphate.

Figure 3. ³¹P NMR spectra of the yeast transformant TDB I during a fructose loading experiment. After recording a spectrum (a) with the cells bathed in a circulating medium containing 50mM glucose, the medium was switched to a solution containing 50mM glucose/1mM fructose and spectra were obtained after (b) 1 hour and (c) 3 hours exposure to fructose. The fructose concentration was then increased to 5mM and a spectrum (d) was recorded 1 hour later. Increasing the fructose concentration caused the fructose 1-phosphate signal to shift upfield from 4.8 ppm (b, c) to 4.6 ppm (d).

Figure 4. ³¹P NMR spectra of the yeast transformant TDB I during a fructose loading experiment at high concentration. After recording a spectrum (a) with the cells bathed in a circulating medium containing 50mM glucose, the medium was switched to a solution

containing 50mM glucose/25mM fructose and spectra were obtained (b) immediately after the switch and then after a further (c) 1 hour and (d) 2 hours exposure to fructose. The chemical shift of the broad fructose 1-phosphate signal is approximately 4.5 ppm in (b) and 4.2 ppm in (c) and (d).

Figure 5. (a) SDS-PAGE (12.5% w/v polyacrylamide) gel of partially purified yeast extracts, as described in the text. Lane 1 contains 3.3 μ g each of bovine albumin (66 kDa), egg albumin (45 kDa), GPDH (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (24 kDa), trypsin inhibitor (20.1 kDa) and α -lactalbumin (14.2 kDa). Lane 2 contains TDB I extract, Lane 3 contains TDB I(-) extract, Lane 4 contains TDS II extract and Lanes 5 contains approx. 60 μ g purified rat ketohexokinase. The arrow indicates the position of the ketohexokinase protein. (b) A semi-log plot of the protein standards γ -globulin (155 kDa, $\circ$), serum albumin (67 kDa, Δ), ovalbumin (38 kDa, $\times$) and cytochrome c (14 kDa, $\square$) eluted from a Sephacryl S100 column, indicating the elution volume of both rat liver and TDB I-expressed ketohexokinase. The native molecular mass of both enzymes was determined to be 134 kDa.

Figure 1.

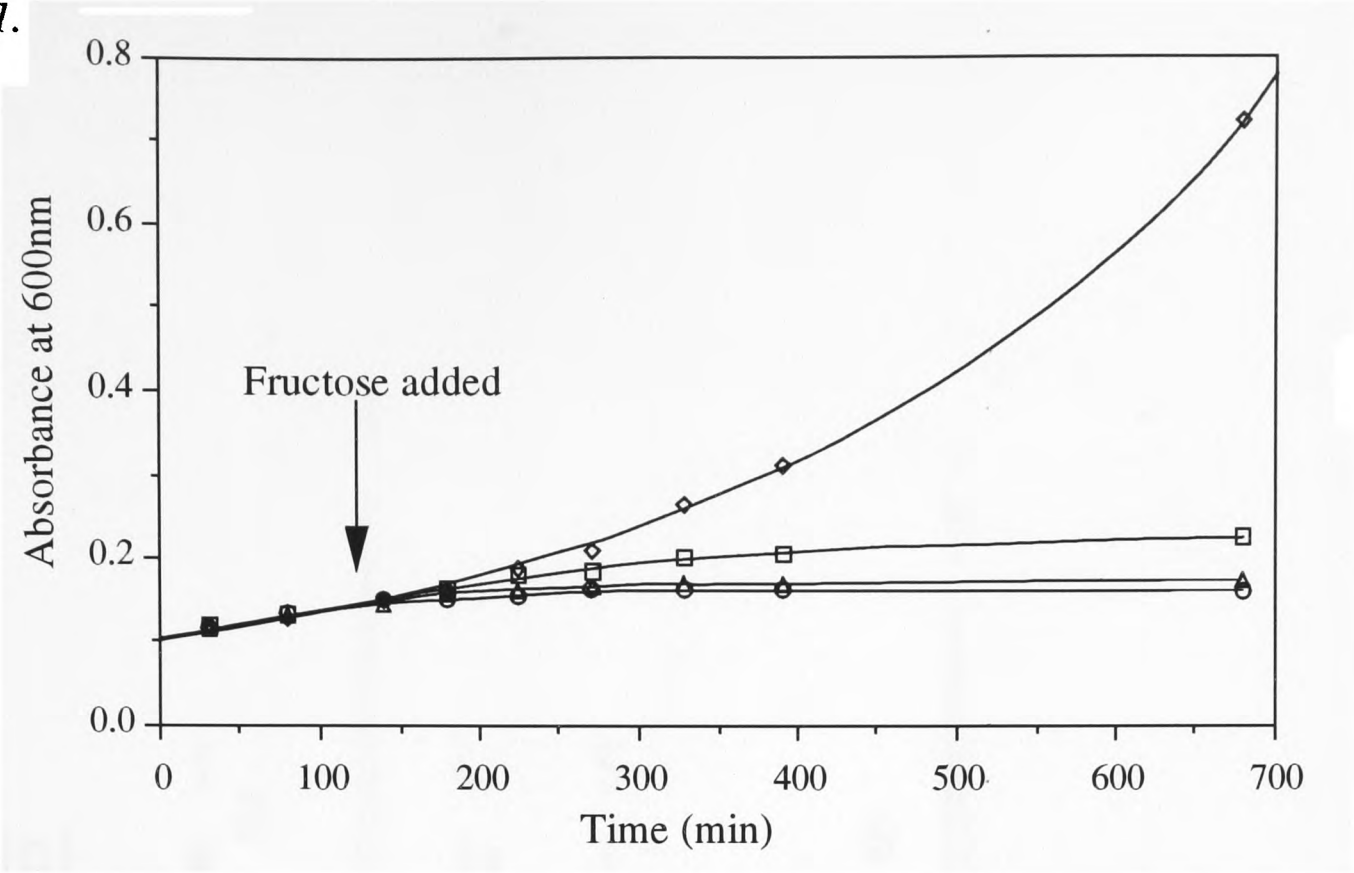


Figure 2.

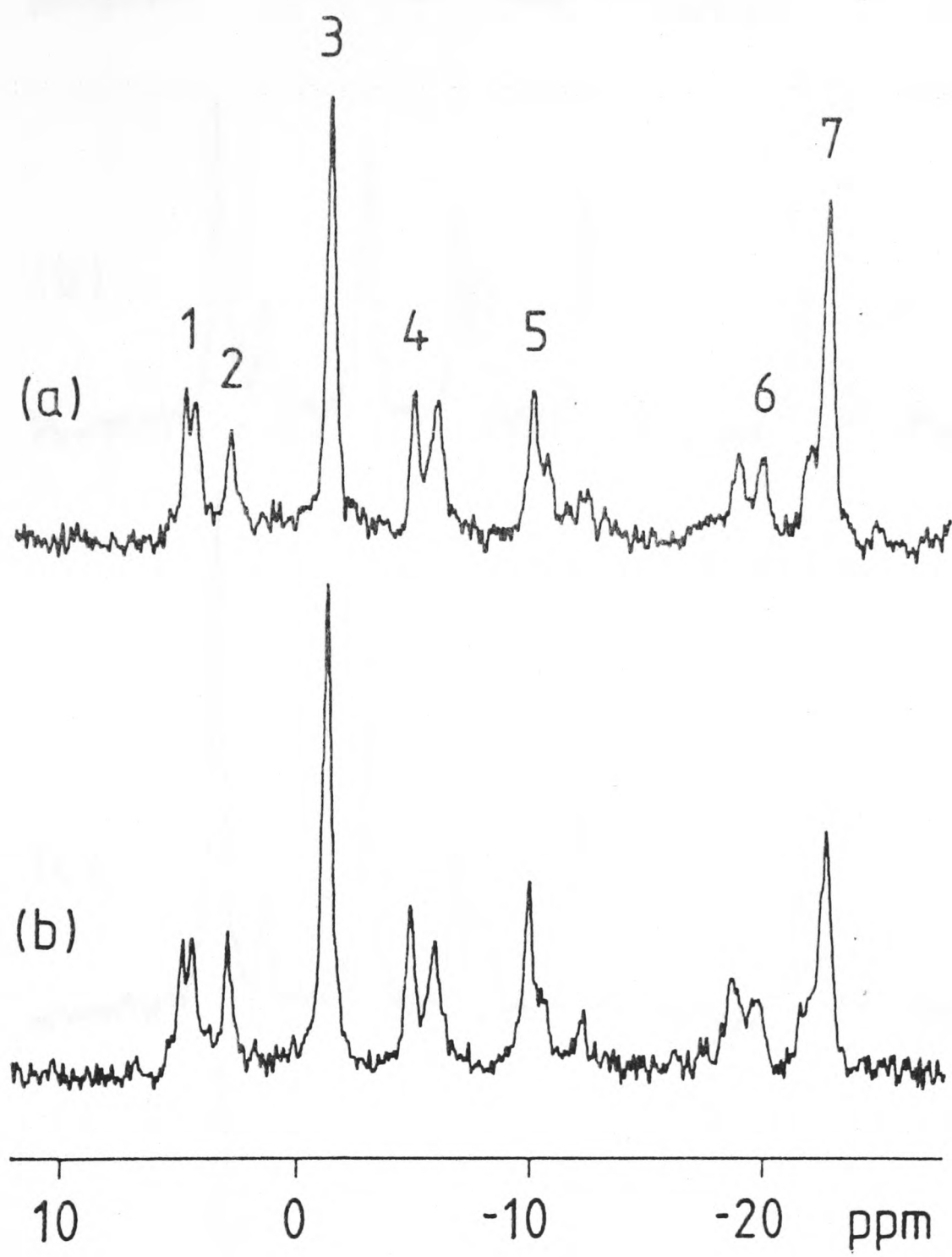


Figure 3.

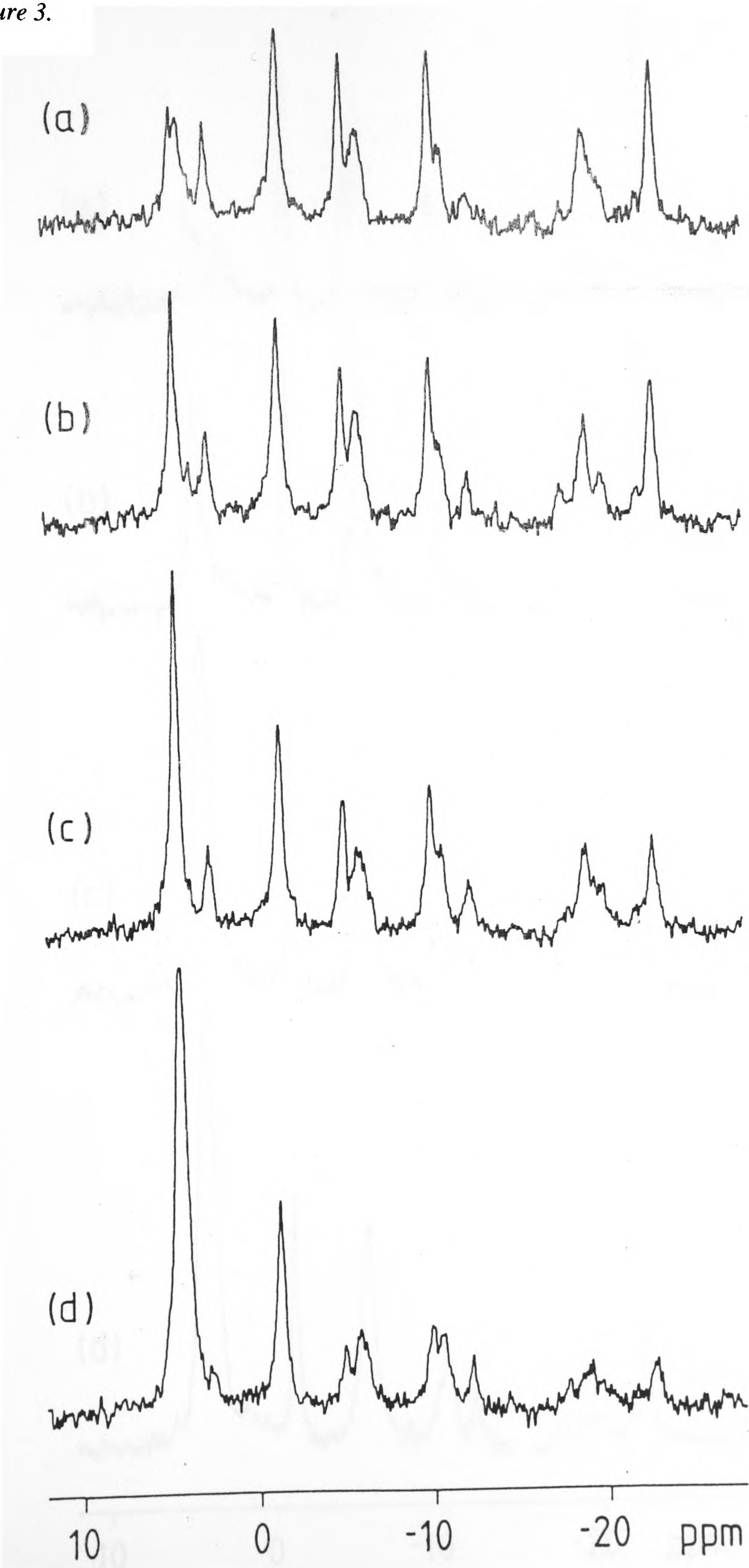


Figure 4.

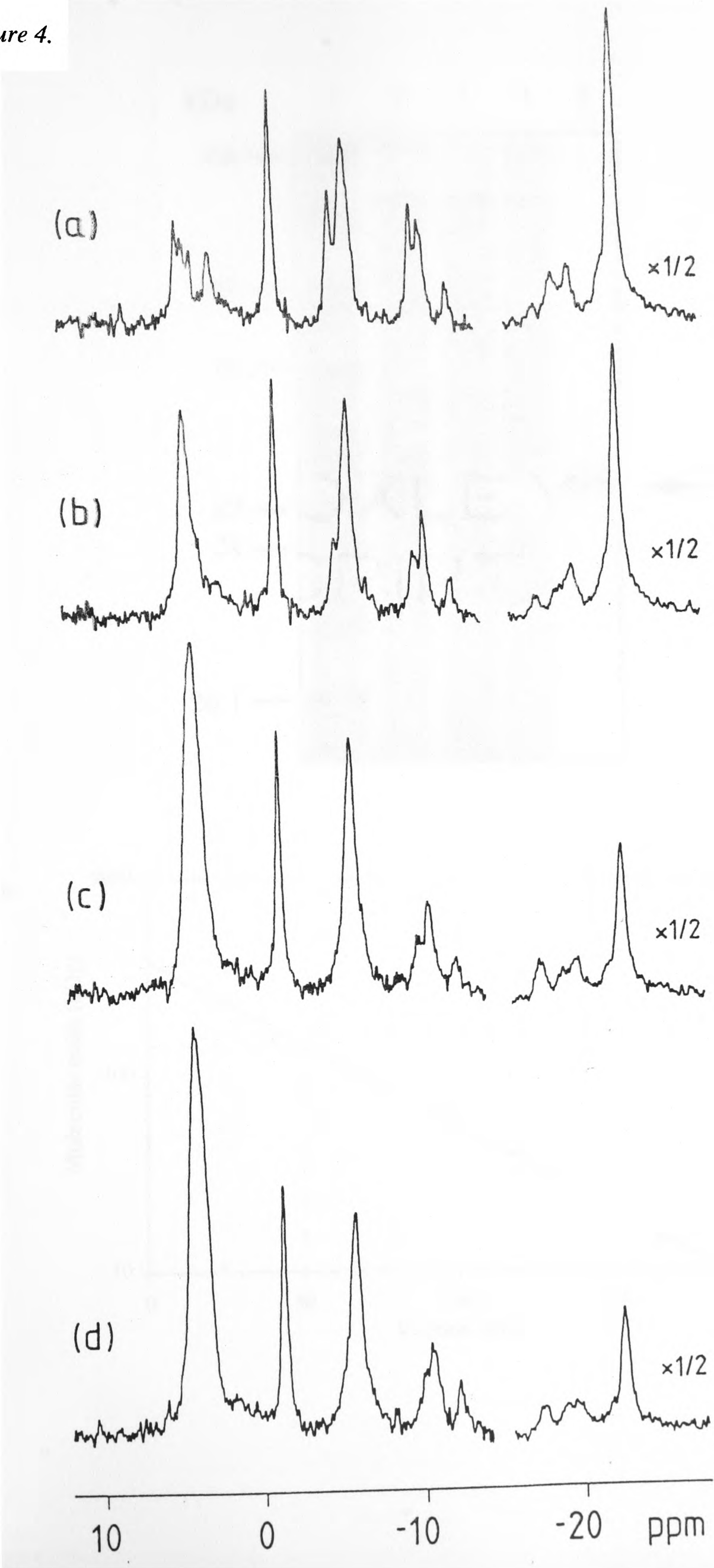
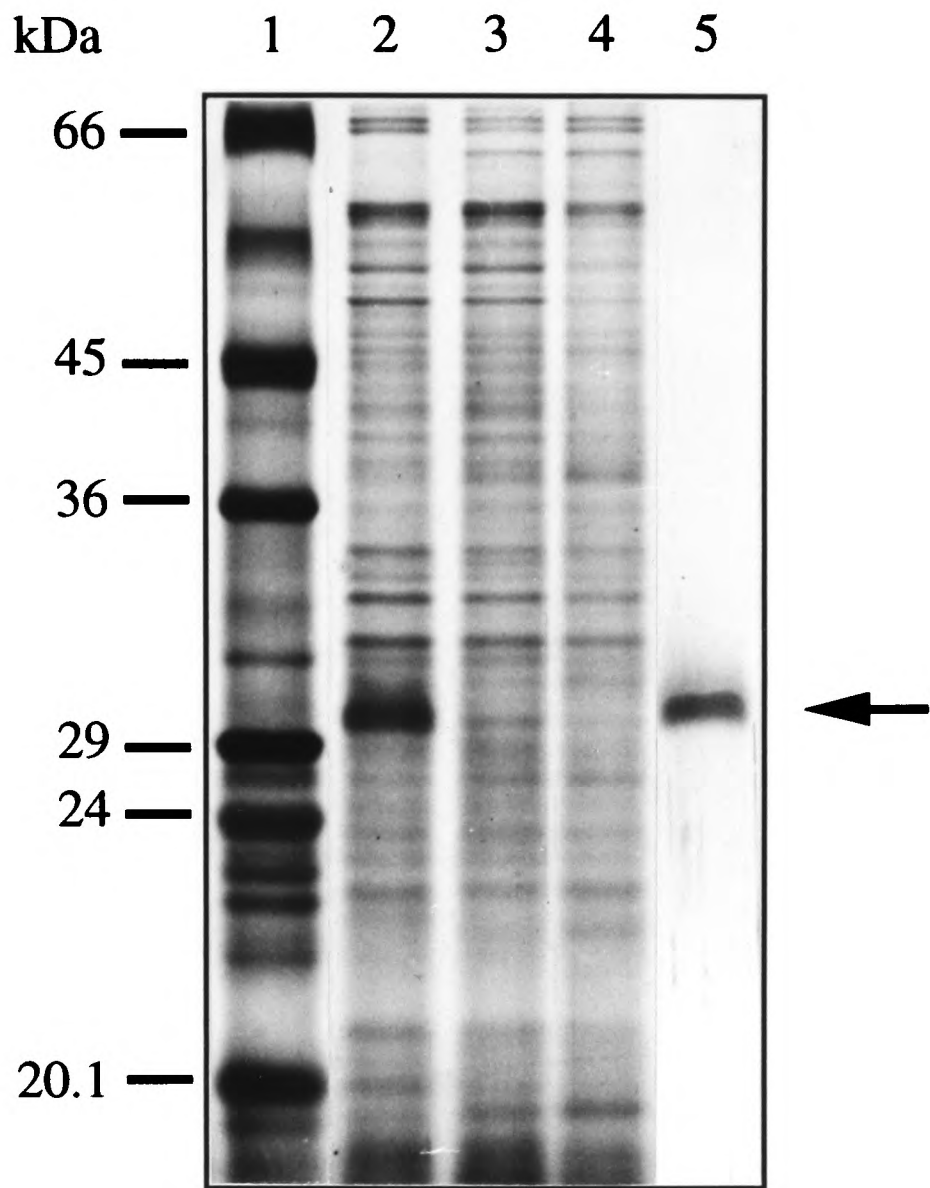


Figure 5.

a.



b.

