Subcloning, purification and characterization of the
GLY1 gene product of Saccharomyces cerevisiae

by

Gholam Hossien Kiani Amineh

A Thesis submitted in conformity with the requirements for the Degree of Master
of Science
Graduate Department of Molecular and Medical Genetics
University of Toronto

© Copyright 1997 by G. H. Kiani
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-28782-3
Subcloning, purification and characterization of the \textit{GLY1} gene product of \textit{Saccharomyces cerevisiae}

Gholam Hossien Kiani Amineh, Master of Science, 1997

University of Toronto, Department of Microbiology

\textbf{ABSTRACT}

In \textit{Saccharomyces cerevisiae} the inactivation of both the cytosolic and mitochondrial isozymes of serine hydroxymethyltransferase, as well as a third gene, designated \textit{GLY1}, are necessary to yield glycine auxotrophs.

The functional characterization of the \textit{GLY1} gene was approached by assaying for threonine aldolase activity catalyzing the conversion of threonine to glycine and acetaldehyde using the \textit{GLY1} enzyme purified on a nickel column. No such activity was obtained with a coupled assay to alcohol dehydrogenase using either the purified HIS-tagged \textit{GLY1} enzyme or the nickel column-purified HIS-tagged SHMT2. In the pursuit of showing that the expression of the \textit{GLY1} gene might be up-regulated by threonine and down-regulated by glycine, a yeast shuttle vector in which the promoter and the 5' coding region of the \textit{GLY1} gene were fused to a reporter gene (\textit{Lac Z}) was used to transform the yeast strain YM 22 to leucine prototrophy. The transformed yeast cells were grown in the presence and absence of L-threonine in minimal medium. A $\beta$-galactosidase assay was done on the crude extracts. A more than 1.5 times increase in $\beta$-galactosidase ($\beta$-gal) activity was observed for the yeast cells grown in the presence of threonine than in its absence, an increase that is not conclusive for
regulation by threonine. Similarly, the results of the above assay with or without the glycine supplement indicate that \textit{GLY1} gene is not down-regulated by glycine.

We next examined the hypothesis that \textit{GLY1} may be part of the "glyoxylate anaplerotic pathway" coding for an enzyme with a glyoxylate aminotransferase activity. It had been previously reported that a \textit{ser1} yeast strain (which has a mutation in the phosphoserine aminotransferase gene, \textit{SER1}, an enzyme required for serine synthesis from glucose) could grow without serine on a non-fermentable carbon source (such as ethanol or acetate) due to the induction of the glyoxylate pathway. The glyoxylate can be transaminated to glycine which can be used to make serine using the glycine cleavage system and mitochondrial SHMT enzymes.

To characterize the possible aminotransferase activity of the \textit{GLY1} protein, a yeast shuttle vector containing the entire \textit{GLY1} gene sequence was used to transform the \textit{ser1} yeast strain to uracil prototrophy. The \textit{ser1} yeast transformants grew on glucose but the over-production of \textit{GLY1} on a multicopy plasmid did not stimulate the growth on acetate, suggesting the enzyme is not part of the glyoxylate anaplerotic pathway.

To confirm the growth study results, a different genetic approach, gene disruption, was employed based on the reasoning that if the \textit{GLY1} gene were involved in the glyoxylate pathway, its disruption would prevent the growth of the \textit{ser1} yeast on acetate in the absence of serine. \textit{GLY1} gene disruption
experiments were done on the ser1 yeast utilizing direct gene replacement and mating strategies.

The GLY1 gene replacement test was done using a gly1::URA3 disruption plasmid which was previously used to inactivate GLY1 (where the GLY1 promoter, part of the 5' non-coding region and the first three amino acids of the gene are replaced by a 1.1 kb URA3 fragment). The ser1 cells were transformed with the above plasmid and the transformants were selected for uracil prototrophy. A number of uracil prototrophs were obtained; these were shown by Southern blot analysis to be wild type at the GLY1 locus.

The GLY1 gene disruption was attempted using a mating strategy in which the ser1 yeast strain YM14: α was mated with the yeast YM11: a gly1::URA3. The resultant diploids sporulated and the spores were plated onto rich YPD plates to reach their haploid state. The haploids were then screened for uracil prototrophy and serine auxotrophy. No such haploid was obtained using the mating technique; however, during one of these mating experiments, a number of cells were encountered which were serine auxotrophic and uracil prototrophic. These cells were tested for the GLY1 gene disruption using Southern blot analysis. The results of Southern analysis indicated that these cells were in fact ser1/ser1 &gly1::URA3/GLY1 diploids.

A number of tetrad dissection studies were performed on the above diploids with results indicating that the disruption of the GLY1 gene in the ser1 background of the yeast strain YM 14 is lethal.
Acknowledgments

I would like to thank Dr. Andy Bognar for giving me the opportunity to work in his laboratory as well as guiding me and advising me throughout the course of this research. I would like to express my appreciation to the members of my committee, Dr. J. Campbell and Dr. C. Lingwood for their moral and scientific support and assistance upon preparing this thesis.

I would like to thank Dr. J.B. McNeil and Dr. Ron Pearlman for providing me with the necessary cell strains and plasmids to make this work possible.

Finally, I would like to take this opportunity to express my utmost appreciation to my parents and my brother for their moral and financial support.
Dedication

I would like to dedicate this thesis to my parents, Esmat and Ali Kiani, to my brother Hamid Kiani and to my wife, Nathalie Kiani.
## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xvii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xviii</td>
</tr>
</tbody>
</table>

## INTRODUCTION:

Overview of folate mediated one-carbon metabolism in eukaryotes          3
Subcellular distribution of folates                                      4
Sources of one-carbon units                                              7
Folate-binding proteins                                                  8
One-carbon metabolism is compartmentalized in eukaryotes                10
The cytoplasmic pathways                                                 11
The mitochondrial pathways                                               14
A general overview of SHMT                                               17
A general overview of the glycine cleavage system  19
Glycine synthesis in eukaryotes  20
General amino acid control in yeast  21
Transmembrane proteins  28
Use of synthetic peptides in raising antibody to a protein  38
Gene fusion for purpose of expression  42
Goals of this project  49

**MATERIALS AND METHODS:**  52

Materials  52
Media  52
Buffers  54
Stock solutions and reagents  55
Methods  64
Routine procedures  64
Isolation of plasmid DNA  64
Restriction digests  65
Agarose gel electrophoresis  65
Ligations  66
Preparation of fresh competent *E. coli* cells using calcium chloride  66
Transformation of plasmid DNA into competent cells  67
SDS-PAGE  68
PCR protocol for the amplification of the GLY1 gene  71
<table>
<thead>
<tr>
<th>PCR programs</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein assay</td>
<td>72</td>
</tr>
<tr>
<td>Protein purification using IMAC</td>
<td>72</td>
</tr>
<tr>
<td>P1cm transduction</td>
<td>74</td>
</tr>
<tr>
<td>Threonine aldolase assays</td>
<td>77</td>
</tr>
<tr>
<td>Assay of β-galactosidase in yeast</td>
<td>78</td>
</tr>
<tr>
<td>Preparation of ssDNA for yeast transformation</td>
<td>80</td>
</tr>
<tr>
<td>Yeast transformation</td>
<td>81</td>
</tr>
<tr>
<td>Isolation of yeast genomic DNA</td>
<td>81</td>
</tr>
<tr>
<td>Gene disruption</td>
<td>82</td>
</tr>
<tr>
<td>Southern blot analysis</td>
<td>82</td>
</tr>
<tr>
<td>Yeast mating</td>
<td>83</td>
</tr>
<tr>
<td>Tetrad analysis</td>
<td>84</td>
</tr>
</tbody>
</table>

**RESULTS:**

The coding region of the GLY1 gene was successfully amplified

The GLY1 amplicons with 5' Ndel site and 3' BamHI site were subcloned into pET 15b expression vector

The GLY1 protein was successfully expressed in BL21 cells

Most of the GLY1 protein expressed in BL21 cells is produced in an insoluble state in the form of inclusion bodies
Construction of $\text{glyA}^-$ BL21 by P1 cm transduction of GS 459

E.coli cell strain as donor and BL 21 as recipient

The glycine auxotrophy of $\text{glyA}^-$ BL 21 could be weakly overcome provided these cells are transformed with the PG construct and induced by IPTG

The GLY1 protein was expressed in $\text{glyA}^-$ BL 21 and purified using immobilized ion affinity chromatography on the HIS-BIND resin

Anti-GLY1 antibody could not be raised against the purified GLY1 protein

Yeast cytoplasmic SHMT was also expressed in $\text{glyA}^-$ BL 21 and purified on nickel column utilizing IMAC

Threonine aldolase assays of GLY1 protein

Glyoxylate anaplerotic pathway is functional in yeast and can provide the $\text{ser}1$ mutant yeast strain with serine when grown on a non-fermentable carbon source without serine supplements

Overexpression of the GLY1 gene and relief of serine Auxotrophy in a $\text{ser}1$ yeast strain
Genetic probing of the possible regulation of the GLY1 gene expression by threonine and glycine in a yeast setting 100

Disruption studies of the GLY1 gene 101

Tetrad analysis of ser1/ser1;gly1::URA3/GLY1 YM 14 diploids 104

The GLY1 gene lacks the Gcn4p consensus sequence 106

Hydrophobicity analysis of the GLY1 protein 109

DISCUSSION AND CONCLUSIONS 111

FUTURE DIRECTIONS 117

Raising antibody against the GLY1 protein 117

A more sensitive threonine aldolase assay 118

The lethality hypothesis 118

Transmembrane protein identity confirmation of the GLY1 protein 120

APPENDICES 121

REFERENCES 133
Abbreviations

Ab - antibody
ac - activity
ADE3 - the gene encoding cytoplasmic trifunctional C$_1$-THF-synthase enzyme in S. cerevisiae
AGAT - alanine-glyoxylate aminotransferase
AICAR - aminoimidazole carboxamide ribonucleotide
AIDS - acquired immunodeficiency syndrome
Amp - ampicillin
APS - ammonium persulphate
ATP, ADP - adenosine triphosphate, adenosine diphosphate
β-gal - β-galactosidase
bp - base pair(s)
BRL - Bethesda Research Laboratories
BSA - bovine serum albumin
C$_1$ - one-carbon unit
CHCl$_3$ - chloroform
chl$^s$ - chloramphenicol sensitive
CHO - Chinese hamster ovary
$^\circ$C - degree Celsius
DHF - dihydrofolate

DHFR - dihydrofolate reductase enzyme

DNA - deoxyribonucleic acid

dNTPs - mixture containing all four deoxynucleotides (dATP, dGTP, dCTP, dTTP)

dUMP - deoxyuridine monophosphate

dUTP - deoxythymidine triphosphate

E - eluent

EDTA - ethylene diaminetetraacetic acid

EMS - ethyl methane sulfonate

ER - endoplasmic reticulum

FAD - flavin adenine dinucleotide

FOR' - the gene encoding cytoplasmic SHMT in *N. crassa*

FPGS - the gene encoding the folylpoly-γ-glutamate synthetase enzyme in yeast

FT - flow-through

GAR - glycinamide ribonucleotide

GCN1-4 - the positive transcription regulatory factor for amino acid control in *S. cerevisiae* (GCN = General Control Nonderepressive)

GCD1-13 - the negative effectors in general amino acid control (GCD = General Control Derepressive)

GCV1 - the gene encoding the T-protein of the glycine cleavage complex in budding yeast
**GLY A** - the gene encoding SHMT in *E. coli*

**GLY1** - the gene in *S. cerevisiae* that appears to control the major pathway for the generation of glycine

**H₄PteGluₙ** - tetrahydropteroyl polyglutamate

**HIV** - human immunodeficiency virus

**Hz** - hertz

**IMAC** - immobilize metal ion chromatography

**IPTG** - isopropyl β-D-galactopyranoside

**Kan** - kanamycin

**kbp** - kilobase pair(s)

**kDa** - kilodaltons

**Kₘ** - Michaelis constant

**M** - minimal bacterial medium

**MC** - magnesium chloride-calcium chloride solution

**MCS** - multiple cloning site

**μCi** - microcurie

**MIS7** - the gene encoding mitochondrial trifunctional Cl-THF-synthase enzyme in *S. cerevisiae*

**NAD⁺ & NADH** - nicotinamide adenine dinucleotide and its reduced form

**NADP⁺ & NADPH** - nicotinamide adenine dinucleotide phosphate and its reduced form

**NaOAc** - sodium acetate
OD - optical density

ONPG - o-nitrophenyl-β-D-galactoside

PCR - polymerase chain reaction

PEG - polyethylene glycol

pET - plasmid for expression of T7 RNA polymerase

PG - pET 15b-GLY1 construct

PLP - pyridoxal-5'-phosphate

PMSF - phenylmethylsulfonylfluoride

PS - pET 15b-SHM2 construct

rxn - reaction(s)

SA - synthetic minimal yeast medium where the sole carbon source is acetate

SAG - synthetic minimal yeast medium where the carbon sources are acetate and sodium glyoxylate

SAM - S-adenosylmethionine

SC - synthetic complete

SDS - sodium dodecyl sulphate

SDS-PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

SER1 - the gene encoding phosphoserine aminotransferase in yeast

SHM1 - the gene encoding the mitochondrial SHMT from baker's yeast

SHM2 - the gene encoding the cytoplasmic SHMT from baker's yeast

SHMTc - cytoplasmic serine hydroxymethyltransferase

SHMTm - mitochondrial serine hydroxymethyltransferase
sp - specific
SSC - sodium chloride-sodium citrate solution
ssDNA - single-stranded DNA
TCA - trichloroacetic acid
TBE - Tris-boric acid-EDTA
TE - Tris-HCl-EDTA buffer
TEMED - N,N,N',N' tetramethylethylene diamine
TFIID - transcription factor II-D
THF - tetrahydrofolate
Thr. Ald. - threonine aldolase
TM - transmembrane
Tn - transposon
Tris - tris (hydroxymethyl) aminomethane
uORFs - upstream open reading frames
W1-3 - wash stages
WT - wild-type
5-CH$_3$-THF - 5-methyltetrahydrofolate
5,10-CH$_2$-THF - 5,10-methylenetetrahydrofolate
5,10-CH$^+$-THF - 5,10-methyltetrahydrofolate
10-CHO-THF - 10-formyltetrahydrofolate
5-CHO-THF - 5-formyltetrahydrofolate
<table>
<thead>
<tr>
<th>List of Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reactions catalyzed by serine hydroxymethyltransferase</td>
<td>18</td>
</tr>
<tr>
<td>2. Gene fusion systems used to facilitate protein purification</td>
<td>45</td>
</tr>
<tr>
<td>3. Chemical and enzymatic agents used to cleave fusion proteins site-specifically</td>
<td>48</td>
</tr>
<tr>
<td>4. Bacterial strains used in this study</td>
<td>60</td>
</tr>
<tr>
<td>5. Yeast strains used in this study</td>
<td>61</td>
</tr>
<tr>
<td>6. Plasmids used in this study</td>
<td>62</td>
</tr>
<tr>
<td>7. Single-stranded oligonucleotides used in this study</td>
<td>63</td>
</tr>
<tr>
<td>8. Growth of glycine auxotrophic mutants on different media</td>
<td>92</td>
</tr>
<tr>
<td>9. Glycine complementation of glyA&lt;sup&gt;−&lt;/sup&gt; BL 21 by utilization of different constructs</td>
<td>93</td>
</tr>
<tr>
<td>10. Direct threonine aldolase assay</td>
<td>97</td>
</tr>
<tr>
<td>11. Growth of the haploid progeny after tetrad dissection of ser1/ser1;gly1::URA3/GLY1 diploids on different selective media</td>
<td>105</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Molecular structure of tetrahydrofolic acid</td>
<td>4a</td>
</tr>
<tr>
<td>2.</td>
<td>Organization of the enzymes of one-carbon metabolism</td>
<td>4b</td>
</tr>
<tr>
<td>3.</td>
<td>Two major pathways for serine/glycine biosynthesis in yeast</td>
<td>21a</td>
</tr>
<tr>
<td>4.</td>
<td>Protein sorting pathways</td>
<td>31a</td>
</tr>
<tr>
<td>5.</td>
<td>Classification of integral membrane proteins</td>
<td>36a</td>
</tr>
<tr>
<td>6.</td>
<td>Examples of gene fusions used for expression of recombinant protein (x)</td>
<td>43a</td>
</tr>
<tr>
<td>7.</td>
<td>Basic concepts for using gene fusions to purify the gene product (x) by the use of an affinity column</td>
<td>46a</td>
</tr>
<tr>
<td>8.</td>
<td>Schematic diagram of pET 15b and its cloning and expression regions</td>
<td>59a</td>
</tr>
<tr>
<td>9.</td>
<td>Schematic diagram of pJM1308</td>
<td>59b</td>
</tr>
<tr>
<td>10.</td>
<td>Schematic diagram of Yep24-GLY1</td>
<td>59c</td>
</tr>
<tr>
<td>11.</td>
<td>Schematic diagram of GLY1 knock-out plasmid used in gene disruption studies</td>
<td>59d</td>
</tr>
<tr>
<td>12.</td>
<td>Nucleotide sequence of the <em>Saccharomyces cerevisiae</em> GLY1 gene and flanking regions</td>
<td>86a</td>
</tr>
<tr>
<td>13.</td>
<td>PCR amplification of the GLY1 gene</td>
<td>86b</td>
</tr>
<tr>
<td>14.</td>
<td>Subcloning of the GLY1 coding region into pET 15b</td>
<td>87a</td>
</tr>
<tr>
<td>15.</td>
<td><em>Hind</em> III digestion of a putative PG construct and pET 15b</td>
<td>87b</td>
</tr>
</tbody>
</table>
16. Induction of the GLY1 gene analyzed by SDS-PAGE

17. The major portion of the GLY1 protein expressed in BL 21 cells is in the form of inclusion bodies

18. Generalized transduction of BL 21 E.coli strain with the P1cm lysate of GS 459 E.coli strain containing the glyA::Tn5 element

19. GLY1 protein expression and purification from the glyA^- BL 21 background analyzed by 13% SDS-PAGE

20. 13% SDS-PAGE analysis of the overexpressed SHMTc in glyA^- BL 21 cell strain transformed with pET 15b-SHM2

21. 13% SDS-PAGE analysis of purified yeast cytoplasmic SHMT

22. Threonine aldolase assay of purified GLY1 and SHMTc proteins

23. Growth of ser1 YM14 on SD (Synthetic Defined) medium

24. Growth of ser1 YM14 on SA (Acetate) and SAG (Acetate+Sodium Glyoxylate) media

25. Growth of wild type YM14 and Yep24-GLY1 transformed YM14 on SD (Synthetic Defined) with histidine, leucine and uracil supplements

26. Growth of wild type YM14 and Yep24-GLY1 transformed YM14 on SA+histidine+leucine+uracil

27. β-Galactosidase assay in yeast

28. Southern blot analysis of the uracil prototrophic yeast haploids of YM09 and YM14 obtained through the gene replacement strategy using pJM1223 GLY1 knock-out plasmid
29. Southern blot analysis of the uracil prototrophic YM14 haploids obtained from the gene replacement experiment with the pJM1223 plasmid

30. Southern blot analysis of the uracil prototrophic YM14 cells obtained by mating strategy

31. Tetrad dissection of four asci

32. Vector NTI search of the GLY1 and SHM2 genes for the Gcn4p consensus sequences
Introduction

Folate-dependent one-carbon metabolism plays an important role in a number of essential cellular processes including the biogenesis of purines, thymidylate, methionine, glycine and formylmethionyl-tRNA (Blakely and Benkovic, 1984). Polyglutamate derivatives of the coenzyme tetrathyrofolate (THF) function as carriers of single carbon units in both degradative and biosynthetic metabolic pathways (Mackenzie, 1984). The majority of intracellular one-carbon substituted folates are believed to be produced through the cleavage of serine and glycine to \( \text{N}^5,\text{N}^{10}\text{-methylenetetrahydrofolate (5,10-CH}_2\text{-THF)} \). This single-carbon substituted tetrahydrofolate can then be used directly for the \textit{de novo} synthesis of thymidylate or it may be oxidized or reduced for use in other cellular processes. For example, the reduced form, 5-methyltetrahydrofolate, is used in methionine synthesis, while the oxidized form, 10-formyltetrahydrofolate, is used in purine genesis.

The involvement of folate metabolism in major cellular processes has made it a key target in the development of chemotherapeutic drugs for use in the treatment of cancer and a number of microbial diseases. Two of the best established target enzymes for anti-cancer chemotherapy, dihydrofolate reductase and thymidylate synthase, are folate-dependent enzymes. Methotrexate, an anticancer drug (Jackson and Grindey, 1984), and trimethoprim, an antibacterial agent (Groom et al., 1991), are folate analogues which inhibit the enzyme dihydrofolate reductase (DHFR). DHFR, a key enzyme
in thymidylate synthesis, catalyzes the conversion of folate and dihydrofolate to THF. The biosynthesis of thymidylate from dUMP involves the oxidation of 5,10-CH₂-THF to dihydrofolate. DHFR prevents the intracellular pool of THF derivatives from becoming depleted by catalyzing the conversion of dihydrofolate to THF. Inhibition of DHFR by methotrexate, combined with the activity of thymidylate synthase, leads to the depletion of cellular THF derivatives and subsequently results in both decreased thymidylate and purine synthesis. However, this may not account entirely for the cytotoxicity of methotrexate. It has been proposed that increases in cellular levels of dUTP, which have been shown to accompany methotrexate treatment, may lead to significant misincorporation of dUMP into DNA. Repeated cycles of dUMP incorporation and removal by excision repair may lead to fragmentation of the DNA and ultimately to cell death (Goulian et al., 1980). Recently, it has been postulated that antifolate drugs such as methotrexate are potent inducers of programmed cell death or apoptosis (Mountz et al., 1994). The DNA damage induced by the administration of methotrexate leads to the activation of two powerful tumor suppressors, p53 and retinoblastoma (Rb) proteins. Activation of these suppressors leads to the arrest of the cell cycle at the G1 phase which helps limit the formation and propagation of damaged genomes (Almasan et al., 1995). Following the cell cycle arrest, if the DNA damage is irreparable, the cell undergoes plasma membrane blebbing, extensive chromatin condensation, nuclear fragmentation and formation of apoptotic bodies (da Silva et al., 1996).
Other antifolates, such as pyrimethamine, are being used to control infections by eukaryotic pathogens including *Toxoplasma gondii* and the fungus *Pneumocystis carinii*, both important secondary pathogens in patients with AIDS (Kovacs et al., 1989 & Kovacs et al., 1990). Antifolates have also recently been used to treat psoriasis and rheumatoid arthritis.

In light of the success of methotrexate as a therapeutic agent, other folate-dependent enzymes are being investigated as possible targets for antifolate drug development. However, a complete understanding of the roles of the enzymes involved in one-carbon metabolism is necessary to design more efficacious drugs.

(A) Overview of folate mediated one-carbon metabolism in eukaryotes

Folate coenzymes are present in virtually every known organism and cell type. One-carbon transfers mediated by folate coenzymes play essential roles in several major cellular processes including nucleic acid biosynthesis, mitochondrial and chloroplast protein biogenesis, amino acid metabolism, methyl group biosynthesis and vitamin metabolism. The diversity of reactions that involve these one-carbon units comes from the ability of the coenzyme 5,6,7,8 tetrahydrofolate (H₄folate) to carry activated one-carbon units at several oxidation states and the ability of the cell to interconvert these forms readily.
The major cellular forms of folate coenzymes contain poly-γ-glutamate tails attached to the ρ-aminobenzoic acid moiety (Fig. 1). In most organisms a single enzyme, folylpolyglutamate synthetase (FPGS), is responsible for the intracellular conversion of folates to their poly-γ-glutamyl derivatives (Fig. 2, rxns 2 and 2') (McGuire and Coward, 1984). The length of the tail differs from one cell type to another, and even within different organelles of a given cell. The predominant polyglutamate tail lengths in most eukaryotic cells are Glu_{5} and Glu_{6} derivatives, but, Glu_{7} is the predominant form in yeast (Cossins, 1984). The physiological function of the polyglutamates has not been well established, but evidence has indicated that in vivo polyglutamation of folates results in retention of the activated folates within the cells, increased substrate efficiency and substrate channeling (MacKenzie and Baugh, 1980 & 1983; McGuire and Coward, 1984; Shane, 1989). Tetrahydrofolate (THF) is the carrier for the active one-carbon units which bind to the N^5 and/or N^{10}. The active one-carbon unit can exist in several oxidation states: methyl (CH_{3}), methylene (CH_{2}-), methenyl (CH^{+}-), formyl (CHO-) and formimino (NHCH-) substituents. The substituted THF molecules serve as donors of one-carbon units for the biosynthetic processes mentioned above.

(B) Subcellular Distribution Of Folates

Intracellular folates are localized primarily in the cytosol (50-80%) and mitochondria (20-50%) of eukaryotic cells. For example, the levels of folates in
Fig. 1
Tetrahydrofolic Acid
Fig. 2 Organization of the Enzymes of One-Carbon Metabolism

Reaction 1 is catalyzed by dihydrofolate reductase (EC 1.5.1.3). Reactions 2 and 2' are catalyzed by the cytoplasmic and mitochondrial FPGS respectively. Reactions 3 and 3' are catalyzed by the cytoplasmic and mitochondrial SHMT isozymes (EC 2.1.2.1)respectively. Reactions 4-6 and 4'-6' are catalyzed by the cytoplasmic and mitochondrial C1-THF complex composed of 5,10-methylene-THF dehydrogenase(catalyzing reactions 4 and 4', EC 1.5.1.5), 5,10-methenyl-THFcyclohydrodrolase (catalyzing reactions 5 and 5', EC 3.5.4.9) and 10-formyl-THF synthetase (catalyzing reactions 6 and 6' (EC 6.3.4.3'). Reaction 7 is catalyzed by GAR transformylase(EC 2.1.2.2) and AICAR (EC 2.1.2.3)transformylase. .Reaction 8 is catalyzed by thymidylate synthase. Reactions 9 and10 are catalyzed by 5,10-methylene -THF reductase (EC 1.5.1.20) and by methionine synthase (EC 2.1.1.14) respectively. Reactions 13, 12 and 14 are strictly mitochondrial and are catalyzed by the glycine cleavage system (EC 2.1.2.10), by sarcosine dehydrogenase(EC 1.5.99.1 )and by methionyl tRNA formyltransferase (EC 2.1.2.9) respectively. Reactions 7, 8 and 10 are strictly cytoplasmic. Reaction 11 is catalyzed by the GLY1protein.
cytosolic, mitochondrial, nuclear, lysosomal and microsomal fractions of rat liver were 102, 78, 30, 54, and 22 ng/mg protein respectively. Furthermore, the levels of particular folate derivatives differ in the cytosol and mitochondria. In rat liver cytosol the distribution is as follows: 45% CH$_3$-THF, 19% 10-CHO-THF, 9%5-CHO-THF and 27% unsubstituted THF, whereas in mitochondria CH$_3$-THF represents only 7% of the total folate pool with the remainder distributed among 5-CHO-THF (11%), 10-CHO-THF (33%) and THF (48%) (Horne et al., 1989).

5-CH$_3$-THF monoglutamate (the predominant form of folate in mammalian plasma) (Cossins, 1984; Horne et al., 1989) can be taken up by most cells (Horne et al., 1978). Horne and coworkers have shown that in liver cells uptake of this compound is a complex process composed of a saturable system with an apparent Km of 0.89μM and a second system that was not saturable at concentrations up to 20μM. Using isolated hepatocytes and purified basolateral membrane vesicles, Horne was able to show that the uptake of 5-CH$_3$-THF is dependent on the presence of a transmembrane hydrogen ion gradient suggesting that this substituted folate is incorporated into the liver cells along with protons (Horne, 1990).

Uptake of 5-CH$_3$-THF monoglutamate is not the only way to transport folates into cells. For example, neoplastic cells generally use one major folate transport pathway which exhibits high affinity for 5-substituted folates, whereas transport in normal cells may involve one or more systems which express different relative affinities for folic acid and 5-substituted folates (Ratnam and
Freisheim, 1990). It seems that folate uptake may occur in a cell specific manner (Goldman, 1971; Henderson, 1986; Henderson, 1990; Ratnam and Freisheim, 1990). However, 5-CH$_3$-THF can not be interconverted directly to other forms of substituted folates (MacKenzie, 1984). 5-CH$_3$-THF contributes to the cellular folate pool via a reaction catalyzed by homocysteine methyltransferase in which the methyl group of 5-CH$_3$-THF is transferred to homocysteine to generate methionine (Fig. 2, rxn. 10) and THF (Blakely, 1969; Mattews, 1984). The studies of Kamen and coworkers (1990) suggest that the uptake of folates (monoglutamates) occurs via a novel form of endocytosis termed receptor-coupled transmembrane transport (RCT). This transport process takes place in four steps:

1. high affinity binding of 5-CH$_3$-THF to the receptor which is an integral membrane protein held in the membrane by a glycosyl-phosphatidylinositol linkage;

2. translocation of the ligand-receptor complexes into internal, membrane bound compartments called caveolae, derived from uncoated pits;

3. dissociation of the folate in response to an acidic environment and movement across the membrane into the cytoplasm through an anion carrier; and

4. covalent addition of multiple glutamic acid residues to the 5-methyltetrahydrofolate through the action of FPGS gene products.

Following folic acid uptake by eukaryotic cells, oxidized folates have to be reduced to dihydrofolate (DHF) and then tetrahydrofolate (THF) by dihydrofolate
reductase (Fig. 2, rxn. 1) (DHFR, tetrahydrofolate-NADP oxidoreductase [EC 1.5.1.3] (Hitchings and Burchall, 1965; Blakely, 1984) which has been found in the cytoplasm of all eukaryotic cells tested (Zelikson and Luzzati, 1977; Brown et al., 1965).

(C) Sources Of One-Carbon Units

There are three major pathways by which one-carbon (C₁) groups can enter THF-dependent metabolism. The primary source is from the β-carbon of serine (MacKenzie, 1984) to yield glycine and 5,10-methylene-THF via serine hydroxymethyltransferase (SHMT) [EC 2.1.2.1] reaction (Fig. 2, rxns 3 and 3'). When this enzyme is inactivated in E.coli, the strain becomes a glycine auxotroph. Glycine also provides C₁ units to THF by the glycine cleavage complex (GCV), comprising four enzymes, which breaks down glycine into NH₃, CO₂ and 5,10-methylene-THF (Fig. 2, rxn 13). Formate can also provide C₁ units as 10-formyl-THF via formyl-THF synthetase activity of C₁-THF synthase. There are also minor sources of one-carbon units one of which includes histidine breakdown (MacKenzie, 1984) in mammalian cells yielding N-formiminoglutamate, from which a formimino group is then transferred to THF to form 5,10-CH⁺-THF in reactions catalyzed by glutamate-formiminotransferase [EC 2.1.2.5] and formimino-tetrahydrofolate cyclodeaminase [EC 4.3.1.4] (Shane and Stokstad, 1984). One-carbon units are also available from catabolism of choline (MacKenzie, 1984). Dimethylglycine, an intermediate of choline catabolism,
contributes one methyl group to the one-carbon pool through the action of dimethylglycine dehydrogenase [EC 1.5.99.2]. The other product of the above reaction, sarcosine, is metabolized by the activity of sarcosine dehydrogenase [EC 1.5.99.1] (Fig. 2, rxn. 12) to generate 5,10-CH$_2$-THF as well as glycine (Wittwer and Wagner, 1980, 1981a & 1981b). This glycine, however, could be used as a donor of C$_1$ units or be converted by the mitochondrial SHMT to serine and transported out of mitochondria, resulting in the transfer of C$_1$ units between the two compartments (Cybulski and Fisher, 1981).

**(D) Folate-Binding Proteins**

Folate-binding proteins are located in both the cytosol and the mitochondria of eukaryotic cells (Wagner, 1982; Matherly et al., 1990). They can be either membrane associated proteins which may function in transportation of folates (Horne, 1993) or soluble cellular proteins, some of which were identified as enzymes involved in folate-mediated one-carbon metabolism. Such enzymes include 10-CHO-THF dehydrogenase (Min et al., 1988), glycine N-methyltransferase (Cook and Wagner, 1984), cytosolic and mitochondrial SHMT (Stover and Schirch, 1991), mitochondrial dimethylglycine dehydrogenase and sarcosine dehydrogenase (Wittwer and Wagner, 1981b).

Evidence suggests that most of the intracellular folates exist in a protein-bound state. Within the cell, the polyglutamated tetrahydrofolates (McGuire and Coward, 1984) have high affinity for most folate-requiring enzymes, exhibiting
Km values in the range of 0.1-1 μm (MacKenzie and Baugh, 1980; Matthews, 1984; Strong et al., 1989). The concentration of THF-polyglutamate is similar to the concentration of the enzymes which utilize this compound (Schirch and Strong, 1989). These two observations suggests that in vivo most of the intracellular THF is protein bound and that the concentration of the free THF is fairly low. Strong and Schirch (1989) calculated the concentration of SHMT as well as that of 10-CHO-THF dehydrogenase and proposed that because of the high levels of these two enzymes and their low dissociation constants for folates, they could alone potentially bind approximately two-thirds of the total cellular folate. Moreover, 5-CHO-THF, which represents approximately 10% of the total folate pool in most organisms (Horne et al., 1989) is probably all bound to SHMT in cells (Stover and Schirch, 1991).

These associations are potentially important for the regulation of one-carbon metabolism. For example, 5-CHO-THF tightly binds to SHMT, an enzyme which serves as the major entry point to the one-carbon pool. This binding may play an important role in folate metabolism (Strong and Schirch, 1989).
One-Carbon Metabolism Is Compartmentalized In Eukaryotes

There is considerable evidence (Appling, 1991) that folate coenzymes are compartmentalized in eukaryotic cells, primarily in the cytoplasm and mitochondria. Drugs which have profound effects on the cytoplasmic folate metabolism, such as nitrous oxide (Horne, Patterson and Cook, 1989) and trimetrexate (Trent, Seither and Goldman, 1991) have no effects on the mitochondrial folate pools. Separate isozymes of the trifunctional enzyme methylenetetrahydrofolate dehydrogenase (Fig. 2, rxns. 4 and 4'), methenyl-THF cyclohydrase (Fig. 2, rxns. 5 and 5') and 10-formyl-THF synthetase (Fig. 2, rxns. 6 and 6'), also called C1-THF synthase, have been found in the cytoplasm (coded by ADE3 gene) and mitochondria (coded by MllS1 gene) of yeast (Shannon and Rabinowitz, 1988; Staben and Rabinowitz, 1986) (Fig. 2, rxns. 4, 4', 5, 5', 6' and 6'). Cytoplasmic and mitochondrial isozymes of SHMT are found in both yeast and human cells (McNeil et al., 1994; Garrow et al., 1993). Both enzymes are encoded by separate nuclear genes. The two compartments appear to have parallel enzyme systems for synthesis and interconversion of folate -bound one-carbon units.
(F) The Cytoplasmic Pathways

(F1) Amino Acid Metabolism

(F1a) Serine Metabolism

Serine can be produced by the reverse function of cytosolic SHMT from glycine and 5,10-CH$_3$-THF (Schirch, 1984). It has been demonstrated that the in vitro conversion of formate and glycine to serine can be effected by the activities of 5,10-methylene-THF dehydrogenase, 5,10-methenyl-THF cyclohydrolase, 10-formyl-THF synthetase and SHMT (Pasternack et al., 1992). Formate (a product of mitochondrial folate-dependent metabolism) (McNeil et al., 1996) can be acted on by the cytoplasmic C$_1$-THF synthase complex and the cytoplasmic SHMT (its reverse reaction) to yield serine. Glycine can also be acted upon by the glycine cleavage system (GCV) and the mitochondrial SHMT to make serine. The conversion of glycine into serine in mitochondria requires the GCV and $SHM1$ gene products since disruption of either gene in a $shm2$ background leads to the inability of glycine to overcome serine auxotrophy in a $ser1$ strain (McNeil et al., 1996). $ser1$ strains have a mutation in phosphoserine aminotransferase, an enzyme required for serine synthesis from the glycolytic intermediates (Jones & Fink, 1982).
**(F1b) Methionine Metabolism**

Methionine biogenesis is effected through the action of folate-dependent one-carbon metabolism since a methyl group from 5-CH$_3$-THF is transferred to homocysteine by homocysteine methyltransferase (methionine synthase) (Fig. 2, rxn. 10) to yield methionine which can be used in protein synthesis (Blakely, 1969). Methionine can also be converted to S-adenosylmethionine (SAM) which is a methyl donor for a variety of acceptors (Blumenstein and Williams, 1960; Mudd et al., 1980).

**(F2) Pyrimidine Synthesis**

Thymidylate synthase catalyzes the formation of thymidylate (dTMP) from deoxyuridylate (dUMP) by the transfer of a methylene group from 5,10-CH$_2$-THF onto the 5 position of the pyrimidine ring of dUMP, followed by a hydride transfer (Blakely, 1969; Pogolotti et al., 1986; Finer-Moore et al., 1990) (Fig. 2, rxn. 8). The byproduct, dihydrofolate, must be regenerated to tetrahydrofolate by the action of DHFR (Blakely, 1984).

**(F3) Purine Synthesis**

It was initially believed that the C$_2$ and C$_8$ atoms of purines were derived from 5,10-CH$^+$-THF and 10-CHO-THF (Wyngaarden & Kelley, 1978). However, more recent work has demonstrated that 10-CHO-THF is the sole donor (Rowe,
1984, review). For purine synthesis, 10-CHO-THF is used for the biosynthesis of phosphoribosylformylglycinamide (FGAR) and phosphoribosylformamidoimidazole (FAIGAR) catalyzed by the glycinamide ribonucleotide (GAR) [EC 2.1.2.2] formyltransferase and aminoimidazole carboxamide ribonucleotide (AICAR) [EC.2.1.2.3] formyltransferase (Fig. 2, rxn. 7) (Gots, 1977; Patterson et al., 1981; Benkovic and Young, 1987) respectively. Associated with another two enzymatic activities required for purine synthesis, the GAR transformylase is a part of a trifunctional enzyme reported in eukaryotes such as yeast (White et al., 1985), avian and mammalian liver (Shild et al., 1990; Daubner et al., 1985; Aimi et al., 1990). This protein may associate with other proteins to form a multienzyme complex (Henikoff, 1987). This trifunctional protein was copurified with the NADP-dependent trifunctional methylene-THF dehydrogenase, methenyl-THF cyclohydrolase, formyl-THF synthetase and SHMT from chicken liver (Caperelli et al., 1980).
Mitochondrial Pathways

Formylation of Methionyl-tRNA\textsuperscript{fmet}

Supplying one-carbon units for synthesis of amino acids, as mentioned previously, is not the only contribution to protein synthesis made by folate-mediated C\textsubscript{1} metabolism. Mitochondria in eukaryotes use a prokaryotic type of protein synthesis where a modified genetic code is utilized. This is in contrast to protein synthesis in a cytosolic environment (Wallace, 1982). The similarity of protein synthesis in eukaryotic mitochondria and prokaryotes arises from the fact that both systems require formylmethionyl-tRNA\textsuperscript{fmet} as the initiator of protein synthesis. The formylmethionyl-tRNA\textsuperscript{fmet} is synthesized by similar steps in prokaryotes and eukaryotic mitochondria: the tRNA\textsuperscript{fmet} is charged with methionine by methionyl-tRNA synthetase, then methionyl-tRNA formyltransferase [EC.2.3.2.9] (Fig.2, rxn. 14) transfers a formyl group from 10-formyl-THF to yield the final product (Staben and Rabinowitz, 1994). The methionyl-tRNA formyl transferase is presumed to occur in all prokaryotic and eukaryotic organelles, but is not present in the cytoplasm of eukaryotes (Caskey \textit{et al.}, 1967). Mitochondrial metabolism has the sole responsibility of supplying formyl groups for the initiation of protein synthesis in that organelle since 10-formyl-THF can not come directly from the cytoplasm.
(G2) Synthesis of Amino Acids

Glycine synthesis appears to be mitochondrial in mammalian cells. A Chinese hamster ovary (CHO) cell mutant lacking mitochondrial SHMT requires glycine for growth (Chasin et al., 1974). CHO cells lacking FPGS (AUXB1 mutants) activity require glycine, adenine and thymidine due to the loss of both cytoplasmic and mitochondrial folate pools, because the coenzymes are not polyglutamated and not retained in the cell (McBurney and Whitmore, 1974; Taylor and Hanna, 1977). Expression of the E.coli FPGS in these mutants complements the adenine and thymidine requirements but the glycine requirement is only complemented if a mitochondrial leader sequence is added to the E.coli FPGS gene (Garrow et al., 1992). In Neurospora crassa, mutations at the FOR locus, encoding the cytosolic isozyme of SHMT, cause a nutritional requirement for formate (McClung et al., 1992), pointing to the fact that formate can supply one-carbon units in mutants lacking cytoplasmic SHMT activity upon conversion to 10-CHO-THF by the mitochondrial C1-THF synthase.

Given the role of SHMT in the production of cellular glycine, it would be expected that eukaryotic organisms possessing mutations in genes for both mitochondrial and cytoplasmic enzymes would be auxotrophic for glycine. However, this is not the case in S.cerevisiae. McNeil et al., (1994) cloned and sequenced genes encoding both the cytoplasmic (SHM2) and mitochondrial (SHM1) SHMT isozymes from this yeast. They demonstrated that inactivation of SHM1 or SHM2 by gene disruption does not result in either glycine or formate
auxotrophy. Similar results were obtained by inactivating both $SHM1$ and $SHM2$ in the same strain. The $shm1$ $shm2$ double disrupted mutants were treated with ethyl methane sulfonate (EMS) to yield glycine auxotrophs. Two groups of glycine auxotrophs were identified; those that grew when supplemented with formate alone and those that required glycine when supplemented with exogenous formate. The latter complementation group was designated $gly1$. $SHM1$, $SHM2$ and $GLY1$ were able to restore glycine prototrophy in the $gly1$ mutants. Disruption of all three genes was found to be necessary to yield glycine-requiring mutants of $S. cerevisiae$.

The $S. cerevisiae$ $GLY1$ is a gene of 2705 bp on chromosome 5 (juxtaposed to the $SUF19$ and $MAK10$ genes). The coding region of this gene extending from nucleotide 1438 to 2599 (designated ORF 35 in the EMBL citation) codes for a protein of 387 amino acids with a molecular weight of 42,797 daltons which functions in glycine biogenesis of the yeast cells.

Disruption of the $GLY1$ gene alone causes a reduction in doubling time, whereas disruption of both $SHM$ genes does not affect the growth rate to the same extent as that of the $GLY1$ gene alone (McNeil et al., 1996). The $SHM2$ disruption in an $ADE3$ background increases the doubling time to 3.5 hrs in comparison to the wild type doubling time of 2 hrs whereas, $SHM1$ disruption in the same background does not affect the doubling time. $SHM1$ & $SHM2$ double disruptions in the above background cause the doubling time to 5 hrs in comparison to that of wild type. The $GLY1$ pathway (Fig. 2, rxn. 11) therefore
appears to be the major source of glycine in *S. cerevisiae* (McNeil et al., 1994). A comparison of the relative enzyme activities of the cytoplasmic (SHMT<sub>c</sub>) and mitochondrial (SHMT<sub>m</sub>) SHMT isozymes demonstrated that SHMT activity is primarily cytoplasmic in yeast. Mutants in *SHM2* possess only 2-15% of the parental SHMT activity, whereas mutants in *SHM1* possess SHMT activities similar to that of parental wild-type strain (McNeil et al., 1994).

The GLY1 gene product does not show significant similarity to either SHMT isozymes of *S. cerevisiae*, or to any other known protein in the protein database (McNeil et al., 1994).

(H) A General Overview of SHMT

Serine hydroxymethyltransferase (SHMT) has been shown to catalyze a gamut of reactions and to have a wide range of substrate specificities. The primary physiological role of this enzyme appears to be the reversible interconversion of serine and glycine with THF serving as the one-carbon carrier molecule. In addition to the physiological reaction, all forms of the enzyme catalyze the aldol cleavage of a variety of 3-hydroxyamino acids (Table 1, rxns. 2-4). In these reactions, no THF carrier is required. SHMT is also involved in transamination, decarboxylation and racemization reactions with several amino acids (Table 1, rxns 5, 6 and 7 respectively) (Shostak et al., 1988; Palekar et al., 1973). More recently, SHMT has also been shown to catalyze the hydrolysis of 5,10-CH<sup>+</sup>-THF to 5-CHO-THF (Table 1, rxn 8).
### Reactions Catalyzed by Serine Hydroxymethyltransferase (SHMT)

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. L-Serine + H₄PteGluₙ</td>
<td>Glycine + 5,10-CH₂-H₄PteGluₙ</td>
</tr>
<tr>
<td>II. Allothreonine</td>
<td>Glycine + Acetaldehyde</td>
</tr>
<tr>
<td>III. L-Threonine</td>
<td>Glycine + Acetaldehyde</td>
</tr>
<tr>
<td>IV. erythro-B-Phenylserine</td>
<td>Glycine + Benzaldehyde</td>
</tr>
<tr>
<td>V. D-Alanine + Pyridoxal-P</td>
<td>Pyruvate + Pyridoxamine -P</td>
</tr>
<tr>
<td>VI. Aminomalonate</td>
<td>Glycine + CO₂</td>
</tr>
<tr>
<td>VIII. 5,10-CH⁺ -THF</td>
<td>5-CHO-THF</td>
</tr>
</tbody>
</table>

### TABLE 1

Reactions catalyzed by serine hydroxymethyltransferase (SHMT)
Reactions I-IV: Aldolytic cleavages; Reaction V: transamination; Reaction VI: decarboxylation; Reaction VII: racemization; Reaction VIII: hydrolysis of 5, 10 CH⁺ -THF to 5-CHO-THF
(I) A General Overview of the Glycine Cleavage System

The glycine cleavage system (GCV), also called glycine decarboxylase, functioning in prokaryotes and in the mitochondria of eukaryotic cells, effects the oxidation reaction of its substrate, glycine, into CO₂, NH₃, NADH and 5,10-CH₂-THF (Clandinin et al., 1975). This system is comprised of four subunits: P (containing pyridoxal phosphate), H (a carrier protein containing lipoamide), T (a transferase responsible for producing 5,10-CH₂-THF) and L (a lipoamide dehydrogenase to complete the cycle by regenerating oxidized lipoamide on the H-protein). These subunits form a loose complex with an estimated ratio of L/P/H/T=1:2:27:9 (Oliver et al., 1990).

Each reaction catalyzed by the four protein components of GCV is fully reversible. The initial event in the glycine cleavage reaction is the formation of a Schiff base between the carbonyl group of P subunit (PLP, pyridoxal phosphate) and glycine resulting in the release of CO₂. The lipoamide containing H-protein being flexible then reacts with glycine bound to the P-protein to form a methylamine-H-protein intermediate. During the above reaction, the α-carboxyl group of glycine is lost as CO₂. The T-protein catalyzes the transfer of the methylamine group to 5,6,7,8-THF₄PteGlu₅, resulting in the formation of 5,10-CH₂-THF and the release of the amino group of glycine as NH₃. The reduced lipoamide from the H-protein is reoxidized back to lipoamide by the FAD coenzyme bound to the dihydrolipoamide dehydrogenase (L-protein).
At this point, it is warranted for a number of related topics to be discussed since the complete understanding of these subjects will facilitate comprehending different concepts introduced and steps taken during the course of this work. These include different metabolic pathways leading to glycine genesis, regulatory elements in the synthesis of amino acids in yeast, transmembrane proteins and the use of synthetic peptides made according to the most conserved hydrophilic sequences of a target protein in raising antibody.

(J) Glycine Synthesis in Eukaryotes

One-carbon units can be supplied to eukaryotic cells in the form of glycine through a variety of metabolic reactions.

Glycine is available to the cell from the catabolism of choline (MacKenzie, 1984). Dimethylglycine, an intermediate of choline catabolism, contributes one methyl group to the one-carbon pool through the action of dimethylglycine dehydrogenase. The other product of the reaction, sarcosine, is metabolized by the activity of sarcosine dehydrogenase to generate 5,10-CH₂-THF and glycine (Wittwer and Wagner, 1980). However, glycine synthesis via the above reactions seems to be of minor significance and has not been characterized in yeasts as yet.

Serine coming from the glycolytic intermediates can be converted into glycine by the action of SHMT. This pathway proves to be significant in glycine synthesis in eukaryotes.
Glycine can also come from a transamination reaction in which glyoxylate is being transaminated by the alanine-glyoxylate aminotransferase to generate glycine. This pathway is called the glyoxylate anaplerotic shunt (Fig. 3). Enzymes of the glyoxylate shunt are derepressed when cells are grown on acetate rather than glucose (Jones and Fink, 1982). In this pathway, isocitrate lyase catalyzes the conversion of isocitrate to glyoxylate. In yeasts, glyoxylate is acted upon by an alanine-glyoxylate aminotransferase to yield glycine.

Threonine cleavage can also give rise to glycine. In this reaction threonine is acted upon by threonine aldolase [EC 4.1.2.5] to yield glycine and acetaldehyde.

(K) General Amino Acid Control in Yeast

The regulation of genes encoding amino acid biosynthetic enzymes in S. cerevisiae involves a complex interplay among numerous control mechanisms. Most of these genes are subject to a cross-pathway regulatory system known as general amino acid control that increases their expression under conditions of amino acid starvation. This regulatory mechanism is effected through the use of transcriptional and translational control elements.

In S. cerevisiae, the expression of at least 35 genes encoding enzymes in 12 different amino acid biosynthetic pathways is coregulated. Transcription of these genes in response to starvation for any of several amino acids is stimulated from 2-10 fold (Mosch et al., 1991). Starvation for any one of 11
Fig. 3 Two major pathways for serine/glycine biosynthesis in the yeast *S. cerevisiae*

On fermentable carbon sources serine is mainly synthesized via the glycolytic pathway from 3-phosphoglycerate. On non-fermentable carbon sources the gluconeogenic pathway, also called the glyoxylate anaplerotic pathway, starting from glyoxylate is utilized.
amino acids including histidine, arginine, lysine, isoleucine, valine, leucine, serine, phenylalanine, tryptophan, methionine, and proline will cause enzyme derepression mediated by the amino acid general control system (Delforge et al., 1975). Whether starvation for any other amino acid elicits such a response has not been tested.

(K1) Trans-acting Regulatory Factors In General Amino Acid Control

(K1a) A hierarchy of factors regulates the transcription of amino acid biosynthetic genes in response to amino acid availability

(K1a1) Identification of Positive and Negative Regulators in General Amino Acid Control

Mutations have been isolated in numerous unlinked genes that alter the expression of enzymes subject to general amino acid control. On the basis of their phenotypes, these mutations fall into two categories. A recessive mutation in any one of nine GCN genes (GCN1-GCN9) impairs enzyme derepression under conditions of amino acid starvation and results in a Gcn' phenotype (general control nonderepressible). Because gcn mutations are recessive, their nonderepressible phenotype suggests that the GCN gene products are positive effectors of gene expression (Hinnebusch, 1988).

A recessive mutation in any of 12 GCD genes (GCD1, GCD2 [the same as GCD12], GCD3-GCD11 & GCD13) results in elevated expression of enzymes subject to general control under non-starvation conditions. This Gcd' phenotype
(general control derepressed) suggests that the GCD gene products are negative regulators of gene expression. Like GCN factors, the products of \textit{GCD1, GCD2, GCD7, GCD10, GCD11, and GCD13} genes regulate amino acid biosynthetic gene expression at the transcriptional level (Hinnebusch, 1988).

**K1a2-** Gcn4p is the proximal positive regulator in general amino acid control

It has been shown that the product of the \textit{GCN4} gene has a positive regulatory function in amino acid enzyme derepression and Gcn1p, Gcn2p, and Gcn3p act indirectly as positive regulators by antagonism or repression of one or more of the GCD factors (Hinnebusch and Fink, 1983a; Harashima and Hinnebusch, 1986; Myers et al., 1986).

**(K1b) Regulation of \textit{GCN4} Expression at the Translational Level**

**(K1b1)** Upstream open reading frames regulate the translation of \textit{GCN4} mRNA in response to amino acid availability.

The positive regulators Gcn2p and Gcn3p and the negative factors Gcd1p, Gcd2p, Gcd10p, Gcd11p, and Gcd13p have been shown to regulate \textit{GCN4} expression at the translational level through a mechanism involving four short open reading frames (uORFs) in the leader of \textit{GCN4} mRNA. Most yeast mRNAs do not contain even a single uORF, and insertion of a uORF in the
mRNA leader of a yeast gene invariably reduces translation of the downstream coding sequence. These observations strongly suggest that the scanning model for translation initiation operates in yeast, as in other eukaryotes (Abastado et al., 1991). According to this model, 40S ribosomal subunits bind to the 5' end of the mRNA and scan in the 3' direction until the first AUG codon is encountered, whereupon translation usually begins. Thus, when a uORF is present in the mRNA, translation preferentially occurs at that site and precludes initiation at the downstream start codons, apparently because reinitiation is inefficient in eukaryotes (Kazak, 1989).

A deletion of all four uORFs (Hinnebusch 1984, Thireos et al., 1984), or point mutations in their four ATG start codons (Mueller and Hinnebusch, 1986), results in high constitutive GCN4 expression independent of the GCN and GCD factors that normally regulate GCN4 expression. These mutations do not alter the 5' end or abundance of GCN4 mRNA, implying that the uORFs inhibit GCN4 expression at the translational level by restricting the flow of scanning ribosomes from the 5' end of the transcript to the GCN4 start codon. Removing the uORFs has little or no effect on GCN4 expression in all gcn and gcd mutants tested thus far (Hinnebusch, 1985; Mueller and Hinnebusch, 1986). This latter observation led to the conclusion that the inhibitory effect of the uORFs on GCN4 translation requires Gcd1p, Gcd2p, Gcd10p, Gcd11p and Gcd13p and that the positive effectors Gcn1p, Gcn2p and Gcn3p are required to overcome the translational barrier to GCN4 translation imposed by the uORFs. These conclusions are
supported by the fact that insertion of the GCN4 uORFs into the leader of a GAL1-lacZ transcript caused expression of β-galactosidase to be regulated by Gcn2p, Gcn3p and Gcd1p in the same fashion observed for that of a GCN4-lacZ fusion (Mueller et al., 1987). Unlike gcn2 and gcn3 mutations, mutations in GCN6 and GCN7 lower the steady state of GCN4 mRNA (Greenberg et al., 1986), suggesting that Gcn6p and Gcn7p stimulate GCN4 expression at the level of transcription rather than translation.

When present alone in the GCN4 mRNA leader, uORF1, 3 or 4 represses GCN4 translation; however, whereas uORF 3 or 4 reduces GCN4 translation to only a few percent of that which occurs in the absence of all four uORFs, solitary uORF1 lowers GCN4 expression by only about 50%. In addition, removal of uORF1 in the presence of uORF 3 or 4 leads to a reduction, not an increase, in GCN4 expression (Tzamarias et al., 1986). Thus, in the wild type gene, uORF1 acts as a positive control element. Because uORF1 stimulates GCN4 expression only in the presence of uORF 3 or 4, it appears that the positive function of uORF1 enables ribosomes to overcome the translational barrier imposed by uORF3 and 4 and to initiate at the GCN4 start codon when cells are starved for amino acids. A GCN4 allele containing only uORFs 2 and 4 shows nearly wild type regulation, indicating that uORFs 2, 3 and 4 are redundant negative elements (Mueller and Hinnebusch, 1986).

(K2) Cis-acting Regulatory Elements at Structural Genes Subject to General Amino Acid Control
(K2a) Gcn4p-binding sites and basal control elements at amino acid biosynthetic genes.

Steady state amounts of mRNAs encoded by the genes (such as those genes involved in the biosynthesis of histidine [HIS1, HIS2, HIS3, and HIS4] and tryptophan [TRP1, TRP2, TRP3, TRP4, and TRP5]) subject to the general control increase under starvation conditions by an amount sufficient to account for the observed derepression of the corresponding enzyme activities. In the case of TRP3 and TRP4, the increase in the mRNA levels was shown to occur without any change in the mRNA half-life, implying that transcription initiation is the most important regulated parameter at these genes (Furter et al., 1988). This conclusion was demonstrated more directly for TRP5 mRNA by measuring the instantaneous rates of mRNA synthesis under normal and starvation conditions by pulse-labeling cells with radioactive RNA precursors (Zalkin and Yanofsky, 1982). Mutational analysis of several other genes like HIS3 and HIS4 led to similar conclusions. It was established thereby that a short nucleotide sequence located upstream of the start site of transcription at these genes mediates their transcriptional activation by Gcn4p element in response to amino acid starvation. This sequence, approximately 12 base pairs in length and containing the highly conserved hexanucleotide core 5'-TGACTC-3', is the binding site for Gcn4. The TGACTC sequence was initially recognized as a repeated motif at HIS3, HIS4, and TRP5 promoters (Jones and Fink, 1982;
Struhl, 1982a; Hinnebusch and Fink, 1983a). In the case of the *HIS3* gene it was shown that its 5'- noncoding region contains multiple TGACTC sequences. Small deletions of the copy located closest to the mRNA start site abolishes *HIS3* derepression in response to amino acid starvation. Thus this sequence is absolutely required for regulation of *HIS3* expression by the general control system even in the presence of TGACTC related sequences present further upstream (Struhl and Hill, 1987). In addition, certain mutations in nucleotides flanking the core sequence (TGACTC) reduces *HIS3* derepression almost as severely as core element substitutions.

On the basis of these results and a naturally occurring TGACTC sequences, the following consensus sequence was proposed for an efficient positive regulatory site for the general amino acid control: 5' - RRTGACTCATTTT-3', where R designates a purine nucleotide. More recently, Sellers *et al.* (1990) showed that in the *HIS3* gene, Gcn4p will bind efficiently to the sequence 5' - ATGACGTCAT- 3', containing a G nucleotide inserted in the middle of the optimal recognition site, whereas no binding occurred to the sequence 5' - ATGAGGTCAT-3', in which G is inserted at the adjacent position. On the basis of these results, it was concluded that Gcn4p binds to the 5-bp half sites which is the sequence ATGAC.
(K3) Role of the TATA Box and General Transcription Factor TFIIID in Transcriptional Activation by Gcn4p

In addition to the Gcn4p-binding site, transcriptional activation by Gcn4p requires an additional promoter element located close to the start site of transcription. For instance, deletions made in the region between -70 and -23 at the HIS4 gene lower basal expression considerably and impair its derepression under starvation conditions. All of these deletions remove the sequence 5' - TATATA- 3' at -60, which is closely related to the TATA element that functions in promoter activity and initiation site selection in mammalian cells. TATA sequences are the recognition sites for the general transcription factor TFIIID. The binding of this factor to the promoter is thought to be an early step in forming an RNA polymerase II initiation complex. Yeast contains a functional homologue of TFIIID that binds with high affinity to consensus TATA sequences (Hahn et al., 1989).

(L) Transmembrane proteins

(L1) Peripheral versus Integral Membrane Proteins

On the basis of relatively simple experiments in which membranes were subjected to various salts and/or pH conditions, it was suggested that they possessed essentially only two major classes of proteins (Findlay, 1987). Those which were removed from the bilayer under different salts and/or pH conditions were regarded as peripheral or extrinsic because solubilization did not require
disruption of bilayer integrity. The presumption is that these proteins interact either with protein, lipid or carbohydrate exposed at the membrane surface and that no significant part of their covalent structure penetrates the intramembranous hydrophobic domain. Also since there may be a variety of modes of association with the membrane surface, this definition appears to hold true. A good test is the brief exposure of membranes to ice-cold NaOH at pH 11.0 when, even though the membrane may vesiculate, essentially all the peripheral proteins are removed and the integrated species remain bilayer-bound (Findlay, 1987). Thus, even though the interaction between proteins may have a major hydrophobic element, denaturation will destroy this, thereby liberating the polypeptide.

The majority of peripheral proteins appear to have a cytosolic orientation. Those that are seen on the outer surface of the membrane are usually subunits of larger complexes: for example, β₂ microglobulin.

When removal of the protein from the lipid constituents of the membrane cannot be achieved by salt or high pH treatment but requires proteolysis or disintegration of the bilayer with detergents or chaotropic agents, such polypeptides are usually referred to as integral or intrinsic membrane proteins. The presumption is that some part of their covalent structure penetrates into the hydrophobic milieu of the bilayer, providing a near permanent anchor in the membrane (Wells and Findlay, 1980).
(L2) Targeting Signals of Transmembrane Proteins

All cells need to be able to sort proteins among a number of subcompartments. In *E.coli*, there are at least 5 well-defined compartments: the cytoplasm, the inner and outer membrane, the periplasm, and the extracellular medium. A membrane protein must thus somehow "know" that it should not remain in the cytoplasm and further whether it is supposed to go in the inner or the outer membrane. The highly complex subcellular structure of eukaryotic cells makes protein sorting a very complicated matter, in which proteins must not only be routed to a correct organelle but further sorted to the correct intraorganellar compartment.

In all cases known, the targeting information is encoded within the nascent polypeptide, often as an N-terminal extention that is removed by appropriately located proteases once the correct compartment has been reached (von Heijne, 1990a). The targeting signal in the nascent proteins is recognized by receptors in the cytoplasm or on the surface of the organelles, and the protein is translocated across one or more membranes and finally delivered to its site of action.

(L3) Sorting Between Organelles

Major recipients of cytoplasmically synthesized proteins are mitochondria, chloroplasts, the nucleus, peroxisomes, and the organelles in the secretory
pathway. Targeting signals specific for each of these organelles have been defined by, for example, gene fusion studies, and the basic designs have been elucidated by a combination of experimental and statistical techniques (Fig. 4).

Mitochondrial targeting peptides are in most cases found as N-terminal extensions that are cleaved by mitochondrial matrix peptidase. They are rich in positively charged amino acids (Arg in particular) and hydroxylated (Ser & Thr) amino acids, but lack negatively charged residues (Asp & Glu). An apparently very important property is their ability to form amphiphilic α-helices with one highly charged and one hydrophobic face (von Heijne, 1986b; Gavel and von Heijne, 1990a).

Chloroplast transit peptides from higher plants also have N-terminal extensions which are removed in the chloroplast by a stromal processing peptidase. They are characterized by an extremely high content of hydroxylated amino acids (~30% Ser+Thr) and contain few if any negatively charged residues (von Heijne et al., 1989 ; Gavel and von Heijne, 1990b). It is not clear if they are designed to have any particular conformational preferences (von Heijne and Nishikawa, 1991). Transit peptides from the green algae, *Chlamydomonas reinhardtii*, in contrast, are strikingly similar to the mitochondrial targeting peptides and probably form amphiphilic α-helices (Franzen et al., 1990).

Nuclear localization signals are not removed from the mature protein and are generally found in internal positions (Silver, 1991). Clusters of positively charged residues presumably exposed on the surface of the folded molecule
Fig. 4 Protein Sorting Pathways
signal nuclear import, and the presence of multiple copies of an import signal in a protein chain often lead to enhanced levels of import.

Peroxisomal targeting signals are less well understood, but a C-terminal tripeptide: Ser-Lys-Leu (SKL) has been shown to promote peroxisomal import in a number of cases (Gould et al., 1989, 1990; Miyazawa et al., 1989); However, many peroxisomal proteins lacking this signal are known.

Secretory signal peptides, finally, have a tripartite design, with an N-terminal positively charged region, a central hydrophobic region, and a C-terminal region that specifies the cleavage site (von Heijne, 1985, 1990b; Gierasch, 1989). This basic structure is similar throughout nature, from bacteria to humans.

(L4) Sorting in the Secretory Pathway

The secretory pathway comprises a number of subcompartments that are traversed in sequel from the ER through the Golgi stacks and the trans-Golgi network to the plasma membrane or to the lysosome (Breitfeld et al., 1989; Pugsley, 1989). Each subcompartment has its own specific complement of resident proteins that are thought to be actively retained in or recycled to that compartment in response to the retention signal present in their amino acid sequence. Thus, lumenal ER proteins have a C-terminal tetrapeptide retention signal (Lys-Asp-Glu-Leu or KDEL) (Pelham, 1990). In yeast and some plants, however, the consensus motif is HDEL. Resident ER membrane proteins seem
to have a retention signal located in their cytoplasmically exposed parts: one or two lysines in their cytoplasmic tails have been implicated in retention, but this is still a somewhat controversial point (Gabathuler and Kvist, 1990; Jackson et al., 1990).

A final aspect of membrane protein sorting is that of endocytosis. Many plasma membrane receptors continually cycle back and forth between the plasma membrane and the trans-Golgi network via endocytosis through coated pits. A critical tyrosine placed in the cytoplasmic tail near the membrane has been implicated as being part of the endocytotic signal (Jing et al., 1990; Ktistakis et al., 1990; Bansal and Gierasch, 1991; Eberle et al., 1991).

(L5) Topological Signals of Transmembrane Proteins

Once the protein has reached its target membrane, it needs to insert into that membrane in its correct orientation. Once inserted, changes in the orientation of the whole protein or of individual transmembrane segments would seem to be impossible on energetic grounds and have never been observed experimentally. It is during the insertion process that the number of transmembrane segments in the final structure as well as their orientation are decided. For a multispanning (polytopic) protein with most of its chain embedded in the membrane, the membrane insertion event is thus the most important step on the folding pathway; what remains after this step has been
completed is only the final packing of transmembrane segments against each other.

Most membrane proteins use one of the machineries normally used to translocate proteins across the membrane for their insertion. From this point of view, a membrane protein can be regarded as a partially translocated protein - a molecule that, in addition to the normal targeting signal(s), contains another signal that can not be translocated across the membrane but rather gets stuck and provides a permanent transmembrane anchoring.

In the process of protein export, a central role is played by the signal sequence: an N-terminal segment that somehow initiates export whereupon it may or may not be cleaved from the mature protein. Three structurally dissimilar regions have been recognized in all transmembrane signal sequences: a positively charged **N-terminal region**, (N region) a central **hydrophobic region** (H region), and a more polar **C-terminal region** (C region) that seems to define the cleavage site. The hydrophobic residues Phe, Ile, Leu, Met, Val and Trp are enriched in the H region of both prokaryotic and eukaryotic sequences (forming α-helices), and drop sharply at the H/C boundary. Conversely, the charged and polar amino acids Asp, Glu, Arg, Lys, His, Gly, Pro, Gin, Asn, Ser, Thr and Tyr are absent (most of the time) in the H region, but dominate the C region (especially tyrosine). Ala, which is very abundant in the prokaryotic signal sequences, does not vary appreciably in incidence across the H/C boundary. The overall hydrophobicity seems to be the one governing principle in this
region, and indeed, substitutions of more hydrophobic for less hydrophobic residues seem to make a big difference in the H regions that are close to the minimal length (8-residue long H region): one Ser, Gly, Thr or Pro residue obviously can be tolerated in a minimum H region but not two. It has been claimed that the distribution of amino acids in the H region is non random (Inouye and Hallegoua, 1980; Perlman and Halvorson, 1983); however, von Heijne (1985) has shown that there is no convincing pattern for this non randomness theory. Heijne claims that the H region varies in length and that there are no regular variations in amino acid composition with length. The length variation observed in the H region (~8 to ~20 residues) may indicate that this region spans the 25-30 Å thick non-polar interior of the membrane as a structure composed partly of α-helix and partly of extended chain depending on its length. Eight residues in a fully extended structure will have a length of 27 Å, close to the length of a helix 20 -residue long. The N region is always positively charged and is one residue long in eukaryotes and two residues long in prokaryotes. The C region, composed of polar residues 5 amino acids long in eukaryotes and 6 amino acids long in prokaryotes, is clearly involved in defining the cleavage site (von Heijne, 1981a).

Transmembrane signal sequences can be predicted by a variety of programs (i.e., TMPRED: TransMembrane PREDiction) which are based on hydrophobicity analyses that detect stretches of α-helices in proteins.
One useful classification based on the kinds of signals present in a protein is shown in figure 5 (von Heijne, 1988). Class I proteins have a cleavable signal peptide (SP) and a second hydrophobic stop-transfer sequence (ST); the orientation of the mature protein is $N_{\text{out}}-C_{\text{in}}$. Class II proteins have an uncleavable signal peptide (uSP); i.e., a signal peptide lacking a C-terminal cleavage domain, and hence get anchored to the membrane in the $N_{\text{in}}-C_{\text{out}}$ orientation. Class III proteins also lack a cleavable signal peptide and have a hydrophobic transmembrane segment close to N terminus (a reverse signal peptide, rSP) (Dalbey, 1990), but this segment is now oriented $N_{\text{out}}-C_{\text{in}}$; i.e., opposite to the class II proteins. Class IV proteins have multiple transmembrane segments, and examples are known with all possible combinations of $N_{\text{in}}, N_{\text{out}}, C_{\text{in}},$ and $C_{\text{out}}$ orientations.

(L6) Membrane Protein Topogenesis in the ER

Von Heijne and Gavel (1988) have shown that the positive inside rule holds also for membrane proteins that insert into the ER membrane and traverse the secretory pathway, although the bias in the distribution of Arg and Lys residues is less extreme than in bacterial proteins. A more detailed statistical study has suggested that the charge bias across the first transmembrane segment is an even better predictor of the topology than the total charge bias over the entire molecule (Hartmann et al., 1989). This would be consistent with the fact that protein translocation across the ER membrane is obligatorily
Fig. 5  Classification of integral membrane proteins. SP, signal peptide; ST stop transfer; uSP, uncleaved signal peptide; rSP, reverse signal peptide.
cotranslational (as opposed to the situation in \textit{E.coli}), which makes it more likely that insertion proceeds according to the sequential model discussed above, with the topology being determined by the first transmembrane segment(s).

There are as yet no good experimental data on this point, since the topological role of charged residues has only been tested on single-spanning proteins so far. For these proteins, however, it has been shown that positively charged residues play a similar role as in \textit{E.coli} (von Heijne and Manoil, 1990). The role of negatively charged amino acids has not been systematically studied, but they do seem to have some effects, at least in particular contexts (Haeuptle \textit{et al.}, 1989; Parks and Lamb, 1991).

\textbf{(L7) Membrane Protein Topogenesis in Chloroplasts and Mitochondria}

Thylakoid membrane proteins follow the positive inside rule (Gavel \textit{et al}, 1991), with the more highly charged loops facing the stromal compartment (i.e., not being translocated). Likewise, mitochondrial inner membrane proteins have a strong charge bias, with the matrix-facing parts having a high content of positively charged residues (Gavel and von Heijne, 1992). Considering the evolutionary relationships between chloroplasts, mitochondria, and bacterial cells, it seems highly likely that the mechanism of organellar membrane protein insertion is similar to that in \textit{E.coli}. As far as insertion into the outer membranes of mitochondria and chloroplasts and into the inner envelope chloroplasts is
concerned, there are very few data available, either from known sequences or from biochemical studies.

**M) Use of Synthetic Peptides in Raising Antibody to a Protein**

Antibodies (polyclonal or monoclonal) raised against a protein of interest can be used for a variety of purposes among which is the detection of site of localization of that protein. To raise antibody against a protein of interest, one can simply purify the protein, mix it with adjuvants and introduce it to a laboratory animal (rabbit or mouse). Following a number of booster shots, the animal is bled and the antibody can be purified. However, one might experience a problem where the protein of interest is not immunogenic enough to raise antibody. This was the case with the GLY1 protein. There is a way whereby one can circumvent this hurdle by using a synthetic peptide corresponding to an epitope of the putative protein, predicted on the basis of its primary structure (amino acid sequence).

In order for the synthetic peptides to be made, one must be able to predict the correct antigenic determinants of the respective proteins. Prediction of antigenicity is thus used to indicate which peptide (usually 10-15 amino acid long) within the protein should be synthesized in order to obtain anti-peptide antibodies that will recognize, i.e. cross react with, the intact protein. Such predictions rely on the existence of so-called continuous or sequential epitopes (antigenic determinants) which correspond to linear peptide fragments of the
that are antigenically related to the parent protein. Sequential epitopes consist of a number of consecutive residues (4-10 residues) in the protein sequence and may be part of a larger so-called discontinuous epitope composed of residues distant in the sequence that are brought together by folding of the polypeptide chain (Atassi and Smith, 1978). These discontinuous epitopes can only be defined or predicted when data on the three-dimensional structure of protein are available and accessible. Without that knowledge, prediction of discontinuous epitopes is only possible with some success when structural data of a related protein with defined homologies in the amino acid sequences and similar functions are known. In the majority of cases, sequential epitopes have to be predicted from the amino acid sequence without additional knowledge of three-dimensional structures.

Sequential epitopes consist of small regions of amino acids accessible for antibody interaction on the surface of a protein. Minimal length has been described as four residues: most frequently it is assumed that those epitopes are defined by about 6-10 neighboring amino acids (Tang et al., 1988). Of course, not only is the sequence of the amino acids specifically recognized by any antibody, but the three-dimensional structure of this region with its particular distribution of electrochemical properties (Barlow et al., 1986), which is based on the defined atomic composition of this area of protein, is also recognized.

Accumulated knowledge concerning the antigenic structure of a small number of well-studied proteins has suggested the possibility of predicting the
location of continuous, sequentially defined epitopes from certain features of the protein primary structure (Van Regenmortel, 1988, 1989). Many studies have shown that the location of sequential epitopes tends to correlate with features of segments of an amino acid chain such as their hydrophilicity (Hopp and Woods, 1981; Kyte and Doolittle, 1982), atomic mobility (Westhof et al., 1984) and flexibility (Karplus and Schulz, 1985), surface accessibility (Janin et al., 1978; Emini et al., 1985) and secondary structures such as locations of turns in the protein (Schulze et al., 1985; Dyson et al., 1985). Therefore, one possible way to predict continuous epitopes is to analyse the amino acid sequence and to predict antigenic sites based on computer programs linking theoretical and empirical data of chemical properties the amino acids with structural conformation parameters. It is important that individual parameters such as hydrophilicity, flexibility and surface accessibility are averaged over segments of 6 to 7 residues and are combined in a weighted manner with the probabilities for secondary structure for successful prediction of sequential antigenic sites (Wolf et al., 1988). These predictions of epitopes using an algorithm for the antigenic index of an amino acid sequence were successfully applied to a high number of viral and various other proteins (Jameson and Wolf, 1988; Modro and Wolf, 1990).

Hydrophilicity analysis of protein sequences has been used to predict antigenic determinants or epitopes in proteins. As one might intuitively expect, most antigenic determinants in proteins (though not all) are found in regions
exposed on the surface of the protein molecule. Hopp and Woods (1981) modified a set of hydrophilicity values for amino acids ultimately derived from their octanol-water partition coefficient, used them to calculate the hydrophilicity of protein sequences over a six-residue window, and plotted the results against the residue number. They found a good correlation between peaks of hydrophilicity and the known antigenic sequences of several proteins, although not all highly hydrophilic segments corresponded to known antigenic regions, and not all antigenic determinants corresponded to the highest hydrophilic peaks.

Eisenberg et al. (1984) calculated hydrophobic moments for a series of helices in 26 different proteins of known structure and compared those values to the average hydrophobicity. The moment and the average hydrophobicity were calculated for the entire helix sequence, so each different helix sequence gave a single value for moment and hydrophobicity. Buried or membrane-spanning helices tended to have large hydrophobicity and low hydrophobic moment, as one might expect.
Gene Fusion for Purpose of Expression

Recombinant DNA technology has allowed in vitro fusions of genes or gene fragments in a simple and predictable manner. Gene fusions were used in the first described systems for heterologous bacterial expression of small peptides, such as somatostatin and insulin (Goeddel et al., 1979). The β-galactosidase system used in these experiments has been extensively utilized in gene fusions, not only for expression but also as a probe for translational and transcriptional activities. In addition, gene fusions have been used to follow cellular localization of gene products (Marston, 1986) and to construct bifunctional proteins having the activity of each of the two gene products (Uhlen et al., 1983).

There are several reasons to use gene fusions for expression of recombinant proteins in heterologous hosts. First, foreign proteins are often rapidly degraded by host proteases, and this may sometimes be avoided by a gene fusion strategy. Second, general and efficient purification schemes may be obtained which allow rapid recovery of gene product. Third, the proteins can be localized to different compartments of the host cell (e.g., periplasm, cell wall or culture medium) through specific peptide fused to the protein (Marston, 1986).
(N1) Gene Fusion Strategies

In figure 6, some examples of gene fusion strategies are outlined, with emphasis on soluble gene fusion products. The most simple fusion involves splicing the recombinant gene X directly after a suitable signal sequence (Fig.6, I). One potential advantage of this strategy is that if the signal peptide is correctly processed during transport, it is possible to produce recombinant protein with a native N terminus. This was achieved in the production of human growth hormone in *E. coli*, using the alkaline phosphatase signal sequence (Gray et al., 1985).

Another conceptionally simple fusion strategy is to fuse a gene product to itself (Fig.6, II). Proinsulin was found to be more stable in *E. coli* using such an approach (Shen, 1984). Similarly, the yield of human insulin growth factor -1 expressed intracellularly in *E. coli* was increased by a factor of 200, accomplished by an increase of the half-life from approximately 1-2 min for monomeric IGF-I to more than 60 min for the fusion protein (Shen, 1984). In both of these cases, the effect was probably caused by changes in the folding of the gene products, resulting in the formation of inclusion bodies. Thus, the self-polymerization strategy may be an alternative to other types of fusions to obtain insoluble aggregates.

Two of the most common fusion strategies are C-terminal (Fig.6, III) or N-terminal (Fig.6, IV) fusions, often used where the fusion partner A encodes an affinity handle to facilitate the purification. An advantage with C-terminal fusion,
Fig. 6  Examples of gene fusions used for expression of recombinant protein (X). Promoters are symbolized by arrows and transcription terminator sequences by rectangles (T). The boxes represent signal sequence (S), different affinity "handles" (A and B), and a domain for insertion into membranes or cell wall (I).
where the recombinant product X is positioned at the C-terminal side of the fusion partner A, is that the promoter and the translation initiation signals are integrated in the 5' end of the gene and are thus not changed by different fusions at the 3' end. The expression level is therefore relatively predictable and several different promoters can rapidly be tested for each gene product. N-terminal fusion has the disadvantage that a product-specific translational and transcriptional start must be engineered in the 5' end of the gene X. In addition, when chemical methods are used to release X, a cleavage rest is usually obtained in the C terminus, thus giving a nonnative protein. Advantages with N-terminal fusions include the ease with which direct N-terminal sequencing of the gene fusion product can be performed and the possibility of designing various biological expression assays, such as HIV frameshift studies (Jacks et al., 1988)

(N2) Gene Fusion Systems

The choice of gene fusion system depends on the properties and final use of the gene product to be expressed. Table 2, outlines the gene fusion systems used routinely to facilitate the purification of desired recombinant proteins.

(N3) pET Vectors and Their Use in Gene Fusion and Purification

pET vectors (Plasmid for Expression by T7 RNA polymerase) are plasmid vectors developed for cloning and expressing target DNAs under the control of a
# Table 2

**Gene Fusion Systems Used to Facilitate Protein Purification**

<table>
<thead>
<tr>
<th>Gene Product</th>
<th>Origin</th>
<th>Molecular Weight (X10³)</th>
<th>Secretion (Sec.)</th>
<th>Ligand</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-Galactosidase</td>
<td><em>E. coli</em></td>
<td>116</td>
<td>-</td>
<td>TPEG, APTG</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Protein A</td>
<td><em>S. aureus</em></td>
<td>31</td>
<td>+</td>
<td>IgG</td>
<td>4, 5, 6</td>
</tr>
<tr>
<td>CAT</td>
<td><em>E. coli</em></td>
<td>24</td>
<td>+</td>
<td>Chloramphenicol</td>
<td>7</td>
</tr>
<tr>
<td>Poly(Arg)</td>
<td>Synthetic</td>
<td>1-3</td>
<td>-</td>
<td>Ion-exchange</td>
<td>8</td>
</tr>
<tr>
<td>Streptavidin</td>
<td>Synthetic</td>
<td>13</td>
<td>+</td>
<td>Biotin</td>
<td>9</td>
</tr>
<tr>
<td>Poly(Glu)</td>
<td>Synthetic</td>
<td>7</td>
<td>+</td>
<td>IgG</td>
<td>11, 12</td>
</tr>
<tr>
<td>PhoS</td>
<td><em>E. coli</em></td>
<td>36</td>
<td>+</td>
<td>Hydroxylapatite</td>
<td>13</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Synthetic</td>
<td>&lt;1</td>
<td>+</td>
<td>Thiol</td>
<td>14</td>
</tr>
<tr>
<td>Protein G</td>
<td>Streptococci</td>
<td>28</td>
<td>+</td>
<td>Albumin</td>
<td>15</td>
</tr>
<tr>
<td>MBP</td>
<td><em>E. coli</em></td>
<td>40</td>
<td>+</td>
<td>Starch</td>
<td>16</td>
</tr>
<tr>
<td>GST</td>
<td><em>E. coli</em></td>
<td>26</td>
<td>-</td>
<td>Glutathione</td>
<td>17</td>
</tr>
<tr>
<td>Flag peptide</td>
<td>Synthetic</td>
<td>2-5</td>
<td>+</td>
<td>Specific IgG</td>
<td>18</td>
</tr>
<tr>
<td>Poly(His)</td>
<td>Synthetic</td>
<td>1-7</td>
<td>+</td>
<td>Zn²⁺, Cu²⁺, Ni²⁺</td>
<td>19</td>
</tr>
</tbody>
</table>

**Refs**

1. Ullman, 1984
2. Germino et al., 1984
3. Scholtissek et al., 1988
4. Neilson et al., 1990
5. Neilson et al., 1985
6. Neilson et al., 1985
7. Bennett et al., 1984
8. Sassenfeld et al., 1984
9. Meade et al., 1986
10. Dalboge et al., 1987
11. Moks et al., 1987
12. Hammarberg et al., 1990
13. Anba et al., 1987
14. Carter et al., 1987
15. Nygren et al., 1988
16. diGuan et al., 1988
17. Smith et al., 1988
18. Hopp et al., 1988
19. Smith et al., 1988

A few references for each gene fusion are listed. The molecular weight of the most common fusion part is indicated, as well as if the secretion (Sec.) of the fusion proteins has been demonstrated (+). CAT, Chloramphenicol acetyltransferase; Z, IgG-binding fragment based on staphylococcal protein A; PhoS, phosphate-binding protein; MBP maltose-binding protein; GST, glutathionine S-transferase; TPEG, (p-aminophenyl-β-D-thiogalactosidase); APTG, (p-aminophenyl-β-D-thiogalactoside).
T7 promoter (Rosenberg et al., 1987). They contain a T7 promoter inserted into the BamHI site of the multicopy plasmid pBR322. In the absence of T7 RNA polymerase, transcription of target DNAs by *E. coli* RNA polymerase is low enough that very toxic genes can be cloned in these vectors. However, some expression can be detected, so it is possible that an occasional gene may be too toxic to be cloned in them.

There are two types of pET vectors, 1- transcription and 2- translation vectors (such as pET 15b, Fig. 5). Transcription vectors are designed for expression of target genes that already carry their own prokaryotic ribosome binding site and ATG start codon. Translation vectors contain the highly efficient ribosome binding site from the phage T7 major capsid protein.

In general, translation vectors are used for the expression of target genes derived from eukaryotic sources, whereas transcription vectors are used for prokaryotic genes which usually carry compatible ribosome binding sites.

(N4) Affinity Purification of Fusion Proteins

(N4a) Concepts

The basic concepts for using gene fusion for affinity purification of recombinant proteins are shown in figure 7. A cell lysate or a culture medium containing the fusion protein, consisting of the desired protein X fused to an affinity handle A, is passed through a column containing a ligand L that specifically interacts with the affinity handle. The fusion protein AX thus binds to
Fig. 7 Basic Concepts for Using Gene Fusions to Purify a Gene Product (X) by the Use of An affinity Column.

A crude extract expressing the fusion protein (AX) is passed through the column (a) containing a specific ligand (L) for the affinity handle (A). The fusion protein (AX) will bind to the column, and a pure and immobilized fusion protein is obtained (b). Elution is performed by an appropriate method (c), and the fusion protein is site-specifically cleaved (d). The cleavage mixture is again passed through the column
the ligand while all other proteins can be washed out of the column using an appropriate buffer. After elution, a chemical or enzymatic method (Table 3) is used to cleave the purified fusion protein at the junction of the two moieties. The cleavage mixture is again passed through the column to allow the affinity handle to bind, while the desired product X is collected in the flow-through fraction.

(N4b) IMAC

Immobilized metal ion affinity chromatography (IMAC, Porath et al., 1975) is an established technique for the purification of proteins on a laboratory scale. Histidine, cysteine, tyrosine, tryptophan and lysine residues of most proteins (Sulkowski et al., 1981, Conlon & Murphy, 1976, Yoshimoto & Laskowski, 1981; Porath, 1979, Scully & Kakkar, 1981) can form complexes with metal ions (such as Cu$^{2+}$, Zn$^{2+}$, Ni$^{2+}$, Fe$^{2+}$, Fe$^{3+}$ and Co$^{2+}$) provided that those residues are exposed to the metal ions (Sulkowski, 1985). Some special features of IMAC of proteins can be summarized as follows:

1. Exposure of certain amino acid residues on the surface of the proteins is required for the adsorption of proteins.

2. The steric arrangement of the protein chain plays an important role, which means that molecules with similar properties with respect to charge, molecular size, amino acid composition, but with differences in their secondary structure and tertiary structure can be separated.
TABLE 3

CHEMICAL AND ENZYMATIC AGENTS USED TO CLEAVE FUSION PROTEINS SITE-SPECIFICALLY

<table>
<thead>
<tr>
<th>Cleavage method</th>
<th>Recognition sequence</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanogen bromide</td>
<td>-Met- ▼</td>
<td>1,2,3</td>
</tr>
<tr>
<td>Formic acid</td>
<td>-Asp- ▼ Pro-</td>
<td>4,5,6</td>
</tr>
<tr>
<td>Hydroxylamine</td>
<td>-Asn- ▼ Gly-</td>
<td>7,8</td>
</tr>
<tr>
<td>Enzymatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagenase</td>
<td>-Pro-Val- ▼ Gly-Pro-</td>
<td>9,10</td>
</tr>
<tr>
<td>Enterokinase</td>
<td>-Asp-Asp-Asp-Lys- ▼</td>
<td>11</td>
</tr>
<tr>
<td>Factor Xa</td>
<td>-Ile-Glu-Gly-Arg- ▼</td>
<td>12,13</td>
</tr>
<tr>
<td>Thrombin</td>
<td>-Gly-Pro-Arg- ▼</td>
<td>14,15</td>
</tr>
<tr>
<td>Trypsin</td>
<td>-Arg- ▼</td>
<td>16</td>
</tr>
<tr>
<td>Clostripain</td>
<td>-Arg- ▼</td>
<td>16</td>
</tr>
<tr>
<td>Ala-subtilisin</td>
<td>-Gly-Ala-His-Arg- ▼</td>
<td>17</td>
</tr>
</tbody>
</table>

Refs:
1. Itakura et al., 1977
2. Goeddel et al., 1979
3. Hammarberg et al., 1989
4. Szoka et al., 1986
5. Neilson et al., 1985
6. Allan et al., 1985
7. Moks et al., 1987
8. Moks et al., 1987
9. Germino et al., 1984
10. Scholitssek et al., 1988
11. Hopp et al., 1988
12. Nagai et al., 1984
13. Smith et al., 1988
14. Smith et al., 1988
15. Doi et al., 1986
16. Varadaragan et al., 1985
17. Carter et al., 1987

A few references for each cleavage method are listed. The recognition sequences used in the references are shown and the cleavage sites are indicated with arrowheads.
3. Simple ionic adsorption and other complicating factors can be suppressed or modified by buffers of high ionic strength. For instance, high salt concentrations (0.1-1.0 M NaCl) have been used to suppress the protein-protein interactions and interactions between sample and the gel during IMAC.

4. Binding is influenced by pH. Low pH causes elution of adsorbed substances. Exceptions to this are known (Andersson & Porath, 1986).

5. Several elution techniques are available (pH gradient, competitive ligands, organic solvents and chelating agents). After washing away unbound material, bound substances are recovered by changing conditions to favour desorption. A gradient or step-wise reduction in pH to 3 or 4 is often suitable. Alternatively, competitive elution with a gradient of increasing concentration of, for example, ammonium chloride, glycine, histamine, histidine or imidazole, may be used. The columns should be saturated with imidazole prior to adsorption of the sample protein and imidazole should be included in the starting buffer if imidazole is to be used for elution.

(O) Goals Of This Project

The application of recombinant DNA technology and molecular genetic approaches to the study of one-carbon metabolism in yeast has led to new insights into the function and regulation of this system and will hopefully determine the inter-relatedness of the pathways in the cytosol and the mitochondria. These findings may prove significant for cancer therapy because
the enzymes studied have not yet been exploited as chemotherapy targets, possibly because they occur in mitochondria or both compartments. These studies may lead into the design of novel therapies using these enzymes as targets, provided that these studies have supplied us with a comprehensive view of one-carbon metabolism and a full knowledge on the enzymes involved in it.

As was previously mentioned, in the mammalian setting, glycine synthesis is effected solely through the action of the mitochondrial SHMT. However, in yeast the scenario is different in the sense that the products of three genes, \textit{SHM1, SHM2} and \textit{GLY1}, contribute to the glycine biogenesis. Among the three, the \textit{GLY1} pathway proves to be the major glycine provider for the yeast cells.

\(O1\) \textbf{The aims of this work were as follows:}

\begin{itemize}
  \item The coding region of the \textit{GLY1} gene was to be subcloned into a bacterial expression vector (pET 15b).
  \item The expressed \textit{GLY1} gene product was to be purified on a nickel column to be used for two purposes:
    \begin{itemize}
      \item Biochemical analysis in order to characterize the function of the \textit{GLY1} gene product
      \item Raising antibody against it to be used in indirect immunofluorescence microscopy experiments to locate the site of the \textit{GLY1} protein;
    \end{itemize}
  \item The \textit{GLY1} gene was to be disrupted by gene replacement or mating strategies in a \textit{ser1} yeast strain and the resultant \textit{gly1} mutants were be used in growth
study experiments in the presence or absence of fermentable and non-fermentable carbon sources to provide us with a genetic approach on the characterization of the GLY1 gene product.
Materials and Methods

Materials:

A-Media

1- Bacterial Media

Super Broth: 32 g Bacto tryptone, 20 g yeast extract, 5 g sodium chloride and 5 ml of 1 N NaOH were dissolved in a liter of H₂O. Then the pH was adjusted to 7.0 with 1 N NaOH. The medium was autoclaved at 121°c and 15 lbs/in² of pressure for 15 min. Finally, 10 ml of 20% glucose (autoclaved) was added to the above medium.

2xYT: 16 g Bacto tryptone, 10 g yeast extract and 5 g NaCl per liter H₂O

Minimal Agar Plates: 15 g agar, 10.5 g K₂HPO₄, 4.5 g KH₂PO₄, 1 g (NH₄)₂SO₄, 0.5 g sodium citrate.2H₂O, 1 ml of autoclaved 20% solution of MgSO₄.7H₂O, 0.5 ml of .22 μm filter sterilized 1% stock solution of B1 (thiamine hydrochloride) and 10 ml of autoclaved 20% solution of glucose

The salt solution and the agar were prepared and autoclaved separately. The salts were prepared in more concentrated form. It is convenient to autoclave 15 g agar in 800 ml H₂O and the salts (at 5x normal strength) in 200 ml H₂O, or the agar in 500 ml H₂O and the salts (at 2x normal strength) in 500 ml H₂O.

Normally, 40μg/ml of the D,L-amino acids (or 20μg/ml of the L-amino acids) are sufficient.
Agar Plates: Agar was made by dissolving 15 g of Bacto-agar in 500 ml of distilled H₂O. A 2x concentrated medium was prepared, also in 500 ml. Agar and medium were autoclaved in separate flasks and then mixed. The mixture was poured into sterile plastic petri dishes after cooling to about 50°C. Antibiotics were added to the medium just before pouring.

**F-TOP Agar:** 8 g of agar and 8 g of NaCl were dissolved in one liter of distilled water and autoclaved. Then 2 ml of sterile 1 M CaCl₂ and 5 ml of sterile 20% glucose were added to the F-TOP agar which was kept molten at 45°C.

**2-Yeast Media**
If used in plates, add 20 g agar.

**YPD Medium:** 1% yeast extract, 2% peptone, 2% glucose

**SD Medium:** 0.67% yeast nitrogen base and 2% glucose

**SA Medium:** 0.67% yeast nitrogen base and 0.6% potassium acetate

**SAG Medium:** 0.67% yeast nitrogen base, 0.6% potassium acetate and 0.4% glyoxylic acid (sodium salts)

**SC Medium:** 0.67% yeast nitrogen base, 2% glucose and 0.2% drop-out mix
(see appendix 1)

**Minimal Sporulation Medium:** 1% potassium acetate and 2% agar

**5-Fluoro-orotic Acid Medium:** 6.7 g yeast nitrogen base (0.67%), 2 g drop-out mix minus uracil (0.2%), 20 g glucose (2%), 50 mg uracil (50 μg/ml) and 1 g 5-FOA (0.1%) were dissolved in 500 ml of distilled water and filter sterilized using
a 0.22 μm filter. 20 g of agar was separately dissolved in 500 ml of water and autoclaved. The two solutions were mixed after cooling the agar to ~ 80°C and poured.

**B-Buffers**

**TE Buffer:** 10 mM Tris-HCl pH 8.0, 1 mM EDTA

**TBE Buffer:** 50 mM Tris-HCl pH 8.0, 50 mM boric acid, 1 mM EDTA

**5x DNA Loading Buffer:** 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol

**MC Buffer:** 0.1 M MgSO₄, 0.005 M CaCl₂

**1 x SDS Gel Loading Buffer:** 50 mM Tris-HCl pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol

**SDS Overlay Buffer:** isobutanol

**SDS Electrophoresis Buffer:** 25 mM Tris, pH 8.3, 250 mM glycine, 0.1% SDS

**Breaking Buffer:** 100 mM Tris-HCl pH 8.0, 1 mM dithiothreitol, 20% glycerol

**Z Buffer:** 16.1 g Na₂HPO₄, 7H₂O, 5.5 g NaH₂PO₄.H₂O, 0.75 g KCl, 0.246 g MgSO₄, 7H₂O and 2.7 ml β-mercaptoethanol were dissolved in 1 liter water. The pH was adjusted to 7.0 and the buffer was stored at 4°C for future use.

**8x Binding Buffer:** 40 mM imidazole, 4M NaCl, 160 mM Tris-HCl pH 7.9

**8x Wash Buffer:** 480 mM imidazole, 4M NaCl, 160 mM Tris-HCl pH 7.9

**4x Elution Buffer:** 4M imidazole, 2M NaCl, 80 mM Tris-HCl pH 7.9
**4x Strip Buffer**: 400 mM EDTA, 2 M NaCl, 80 mM Tris-HCl pH 7.9

**8x Charge Buffer**: 400 mM NiSO₄

**Prehybridization Buffer**: 1 g of dextran sulphate, Na salt (MW ~ 500,000) and 1 ml of 10% SDS were dissolved in 8.46 ml water.

### C- Stock Solutions and Reagents

#### (I) Antibiotic stock solutions

**Ampicillin stock**: 100 mg/ml ampicillin stock (1000x) was made by dissolving 1 g of ampicillin in 10 ml sterilized distilled water. This stock solution was dispensed in 1 ml aliquots and kept at -20°C until needed. The final concentration of the ampicillin in rich media is 100 μg/ml and in minimal media is 75 μg/ml.

**Chloramphenicol stock**: 34 mg/ml chloramphenicol stock was made by dissolving 340 mg chloramphenicol in 10 ml 100% ethanol (kept at -20°C). The final concentration of chloramphenicol used on media was 12.5 μg/ml.

**Kanamycin stock**: A stock of 20 mg/ml (1000x) was made by dissolving 200 mg of kanamycin in 10 ml sterilized water and dispensed in 1 ml aliquots (kept at -20°C). The final concentration of this antibiotic on media was 20 μg/ml.

#### (II) Amino acids stock solutions:

**L-Histidine stock**: 10 mg/ml stock solution was made by dissolving 100 mg of L-histidine - HCl in 10 ml distilled water. The histidine stock was then autoclaved for 15 min at 121°C and 15 lbs/in² of pressure. The final concentration on media was 20 μg/ml.
L-Leucine stock: 10 mg/ml stock was prepared and autoclaved. The final concentration used on media was 100 μg/ml.

L-Glycine stock solutions: A 10 mg/ml glycine stock solution was prepared and autoclaved. This stock solution was used on bacterial minimal plates to give a final glycine concentration of 20 μg/ml. Another glycine stock solution (750 mg/ml, 1000x) was made and used on yeast media to give a final concentration of 750 μg/ml or 10 mM.

L-Serine stock solution: A stock of 750 mg/ml L-serine was prepared and used on yeast media to give a final concentration of 750 μg/ml.

Amino acid stock solutions were cooled after being autoclaved and stored at -20°C.

Base stock solutions:

Adenine sulfate stock: A stock of 2 mg/ml was prepared and autoclaved. This stock was used on yeast media to achieve a final concentration of 20 μg/ml.

Uracil stock: a stock of 2 mg/ml was prepared and autoclaved. The final concentration on yeast media was kept at 20 μg/ml.

Solutions

IPTG stock: A stock solution of 1M IPTG was made in H₂O and was used in induction experiments at a final concentration of 1mM.
**PMSF stock:** A 40 mM PMSF stock was made in 100% isopropanol and was stored in -20°C to be used in β-galactosidase assay. PMSF is toxic and should be handled with care.

**ONPG stock:** A 4 mg/ml stock in Z buffer was prepared and was kept at -20°C to be used in β-galactosidase assay.

**Na₂CO₃ stock:** A 1M stock was prepared in water and was kept at room temperature to be used as a stop buffer for the β-galactosidase assay.

**Dithiothreitol stock:** A 1M DTT was prepared in H₂O.

**CaCl₂ solution:** 1 ml stock solution of CaCl₂ was made in pure water and kept in 10 ml aliquots at -20°C. From this stock solution 0.1 ml working solution is made and autoclaved to be used in the preparation of bacterial competent cells.

**20X SSC Solution:** 3M NaCl (175.35 g) and 0.3M sodium citrate (88.23 g) in 1 liter of H₂O

**SDS-PAGE staining solution:** 0.25 g of Coomassie Blue R250 was dissolved in a mixture of 90 ml of methanol: water (1:1) and 10 ml of glacial acetic acid. The solution was then filtered through a Whatman No. 1 filter to remove any particulate matter.

**SDS-PAGE destaining solution:** The destaining solution contains the same components as the staining solution except the Coomassie Blue.

**SDS-PAGE Solution A:** 28 g acrylamide and 1 g N,N'-methylene bisacrylamide were dissolved in 90 ml of distilled water and the volume was adjusted to 100 ml.
**SDS-PAGE Solution B:** 36.3 g of Tris was dissolved in 180 ml of water and the pH was adjusted to 8.8. The final volume was adjusted to 200 ml.

**SDS-PAGE Solution C:** 21.29 g of acrylamide and 0.56 g of bisacrylamide were dissolved in 90 ml water and the final volume was brought to 100 ml.

**SDS-PAGE Solution D:** 12.1 g of Tris was dissolved in 180 ml water and the pH was adjusted to 6.8. The final volume was adjusted to 200 ml.

**Reagents**

$[^{32}P] \alpha$-ATP (10 mCi / mmol) was purchased from DuPont Canada Ltd.

Nucleotides, sodium glyoxylate, acetaldehyde and alcohol dehydrogenase were purchased from Sigma Chemical Company. Agar was obtained from Difco Laboratories. X-ray films, developer and fixer were obtained from Kodak Canada Inc. Restriction enzymes and ligase were purchased from Pharmacia, Boehringer Mannheim Canada Ltd., New England Biolabs (NEB) and Bethesda Research Laboratories (BRL).

**Cell Strains and Plasmids:** A number of bacterial and yeast cell strains were used in this study which can be seen in Table 4 and Table 5 respectively. The salient points about the bacterial strains used are as follows:

- JF1754 is a highly transformable cell strain which was used throughout this project.

- BL 21 strain is a lambda DE3 lysogen which carries a chromosomal copy of the T7 RNA polymerase. This strain was used to establish the DNA construct
(having a pET 15b vector where the gene of interest under the control of a T7 lac promoter) for the purpose of gene expression.

- The GS459 is a bacterial strain (provided kindly by Dr. G.V. Stauffer) whose SHM gene is disrupted due to the insertion of a transposable element (Tn5 or kanamycin transposon). This strain was used as a donor strain in a P1cm transduction experiment.

Table 6 depicts all the vectors which were used in this study. The salient points regarding the vectors used in this study are as follows:

- pET 15b vector is an expression plasmid (Fig. 8) where the gene of interest is put under the control of a strong T7 promoter and the resultant protein is expressed with a hexahistidine tag at its N-terminal and a thrombin cleavage site (Leu-Val-Pro-arg-Gly-Ser) at its C-terminal.

- pJM1308 plasmid is a shuttle vector (Fig. 9) in which the promoter and the 5' coding region of the GLY1 gene is fused to lac Z reporter gene.

- Yep24-GLY1 is a shuttle vector (Fig. 10) in which the 5' and 3' non-coding flanking regions and the coding region of the GLY1 gene is cloned into an Xhol site. This plasmid (having a URA3 marker) was used as a template for amplification of the GLY1 gene and also for its overexpression in the YM 14 yeast strain.

- pJM1223 is the GLY1 knock-out plasmid (Fig. 11) where part of the GLY1 5' non-coding flanking region, the promoter region (0.8 kb) and the first three amino acids of GLY1 are replaced by a URA3 marker (1.1 kb).
Fig. 8 Schematic diagram of pET 15b and its cloning/expression regions
Fig. 9 Schematic Diagram of pJM1308 used in β-Galactosidase in Yeast
**Fig. 10 Schematic Diagram of Yep24-GLY1 Construct**

Yeast Episomal Plasmid 24 (7769 bps) is a cloning vehicle used for introduction of genes, in this case the GLY1 gene, into *S. cerevisiae* at a high copy number. The plasmid can replicate in both *E. coli* (due to the presence of the pBR322 origin of replication) and in yeast (due to the presence of the replication determinant of the yeast 2 micron circle plasmid). It carries the ampicillin resistance gene (AP) of pBR322 for use as a selectable marker in *E. coli* and the URA3 gene for use as a selectable marker in yeast. The tetracyclin resistance gene (TC) of pBR322 is present but is separated from its promoter by the URA3 sequence. The Yep24 plasmid was digested with *SalI* and *XhoI* linkers were engineered in it. Then an eight-kb yeast genomic fragment containing a 3-kb *XhoI/SalI* GLY1 fragment was subcloned in it.
Fig. 11 Schematic diagram of GLY1 knock-out plasmid used in gene disruptional studies
**TABLE 4**

**Bacterial Strains Used in This Study**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Sources of References</th>
</tr>
</thead>
<tbody>
<tr>
<td>JF 1754</td>
<td>( hsdR ) lac gal metB leuB hisB</td>
<td>McNeil et al., 1981</td>
</tr>
<tr>
<td>GS459</td>
<td>serA glyA:: Tn5 lysA thi-1 lac gal 6 melA xyl mtl-2 tonA ( \lambda ): supE 44</td>
<td>Stauffer et al., 1981</td>
</tr>
<tr>
<td>BL21</td>
<td>( F ) omp( T ) hsd ( S_\beta ) (( r_\beta^+ m_\beta^- )) gal dcm lon</td>
<td>Grodberg et al., 1988</td>
</tr>
<tr>
<td>Strain</td>
<td>Genotype</td>
<td>Sources or References</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>YM09</td>
<td>$\text{Mat}^\alpha \text{ ura3-1 trp-1}$</td>
<td>McNeil et al., 1994</td>
</tr>
<tr>
<td></td>
<td>ade2-1 $\text{his 3-11,15}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>leu2-3,112 $\text{can1-100}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>shm1:: $\text{HIS3}$ shm2:: $\text{LEU2}$</td>
<td></td>
</tr>
<tr>
<td>YM11</td>
<td>$\text{Mat}^\alpha \text{ ura3-1 trp1-1 ade2-1}$</td>
<td>McNeil et al., 1994</td>
</tr>
<tr>
<td></td>
<td>$\text{his3-11,15 leu2-3, 112}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{can1-100 shm1:: HIS 3}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{gly1:: URA3}$</td>
<td></td>
</tr>
<tr>
<td>YM13</td>
<td>$\text{Mat}^\alpha \text{ ura3-1 trp1-1 ade2-1}$</td>
<td>McNeil et al., 1994</td>
</tr>
<tr>
<td></td>
<td>$\text{his3-11,15 leu2-2,112}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{can1-100 shm1::HIS3}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{shm2::LEU2 gly1::URA3}$</td>
<td></td>
</tr>
<tr>
<td>YM14</td>
<td>$\text{MAT}^\alpha \text{ his-3,11,15 leu2-3,112}$</td>
<td>McNeil et al., 1994</td>
</tr>
<tr>
<td></td>
<td>$\text{ura3-52 ser1}$</td>
<td></td>
</tr>
<tr>
<td>YM22</td>
<td>$\text{Mat a his3-11, 15 leu2-3,112}$</td>
<td>McNeil et al., 1996</td>
</tr>
<tr>
<td></td>
<td>$\text{ura3}$</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 6

Plasmids Used in This Study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>References &amp; Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET 15b</td>
<td>Expression vector</td>
<td>Novagen Ltd.</td>
</tr>
<tr>
<td>pJM1308</td>
<td>GLY1-lac fusion</td>
<td>McNeil (1996)*</td>
</tr>
<tr>
<td>Yep24-GLY1</td>
<td>GLY1 shuttle vector</td>
<td>McNeil (1994)*</td>
</tr>
<tr>
<td>pJM1223</td>
<td>GLY1 knockout plasmid</td>
<td>McNeil (1994)*</td>
</tr>
</tbody>
</table>

* Kindly provided by Dr. J.B. McNeil

Oligonucleotides: Two oligonucleotides, p5'N and p3'B, (24 mer each) were synthesized by Toronto Biotech Services on an ABI DNA synthesizer. Table 7 depicts the sequences of these oligonucleotides and particular details about them.
## TABLE 7

**Single-Stranded Oligonucleotides Used in This Study**

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>p5′N′</td>
<td>5′GGG GTT CAT ATG ACT GAA TTC GAA3′</td>
<td>GLY1 PCR</td>
</tr>
<tr>
<td>p3′B</td>
<td>5′GGG GTT GGA TCC TCA GTA TT GTA3′</td>
<td>GLY1 PCR</td>
</tr>
<tr>
<td>p10</td>
<td>5′CCC CCC GGA TCC ATG CCT TAC ACT C3′</td>
<td>SHM2 PCR</td>
</tr>
<tr>
<td>p20</td>
<td>5′CCC CCC GGA TCC TTA CA CAG CCA ATG3′</td>
<td>SHM2 PCR</td>
</tr>
</tbody>
</table>

1-The underlined sequences are 6 extra nucleotides added for the proper functioning of the Taq DNA polymerase. The *BamHI* sites and *Ndel* sites are printed in **BOLD** and **BOLD-ITALIC** respectively. The initiation codon, ATG, and the termination codons, TTA & TCA, are double underlined with an arrowhead and cross on their tops respectively.
Methods

Routine Procedures

A- Isolation of Plasmid DNA (mini-preps):

Bacteria containing a target plasmid were grown in 2 ml of 2x YT + 100 μg/ml ampicillin with shaking in a 37°C shaker bath for 7 hrs to overnight. The cells (1.5 ml) were centrifuged for 60 sec in a microfuge and the supernatant was discarded. The cell pellet was resuspended in 100 μl of GET (50 mM Glucose, 10 mM EDTA, 25 mM Tris-HCl pH 8.0) by gentle vortexing and kept on ice for 5 min. A 200 μl volume of freshly prepared alkaline-SDS (0.2 N NaOH, 0.1% SDS) was added and the samples mixed by gentle inversion. After 5 min on ice, 150 μl of ice-cold potassium acetate pH 5.2 (3 ml glacial acetic acid, 7 ml distilled water, 15 ml 5M potassium acetate) was added to precipitate genomic DNA. The tubes were mixed by inversion and incubated on ice for 15 min. The tubes were then centrifuged for 10 min and the supernatants were carefully removed and placed in new tubes. DNA was precipitated by addition of 1 ml ice-cold 95% ethanol and incubated at -70°C for 20 min. The tubes were then centrifuged for 10 min and the supernatants removed by aspiration. The DNA pellet was resuspended in 200 μl of distilled water and extracted with an equal volume of phenol: chloroform (1:1) followed by extraction by an equal volume of chloroform. The DNA was then precipitated by addition of 2 volumes of 95%
ethanol and incubation at -70°C for 20 min. Following centrifugation, the DNA pellet was rinsed with 70% ethanol, dried under vacuum and resuspended in 50 μl of sterile distilled water. The DNA (1-2 μg) was then stored at -20°C until needed.

B- Restriction Digests:

Restriction endonuclease digestion of DNA (0.5-1 μg) was carried out in 10-20 μl volumes using buffers recommended by the manufacturers of the restriction enzyme. The DNAs were usually digested at 37°C for 2 hrs. DNA fragments were electrophoresed on 0.8% agarose gels for separation based on size.

C- Agarose Gel Electrophoresis

Plasmid DNA and DNA fragments from restriction digests were electrophoresed on agarose gels for separation based on size and visualized with ethidium bromide under ultraviolet (UV) light. Agarose gels (0.8%) were prepared by placing 0.48 g of agarose in 60 ml of TBE buffer (50 mM Tris-HCl pH 8.0, 50 mM boric acid, 1 mM EDTA) and heating in a microwave oven for 2-3 min. The agarose mixture was then allowed to cool down, and while still liquid, cast into a plastic mold whose ends were closed with masking tape. A plastic comb was placed at one end of the gel and the agarose
was allowed to set for 30 min. at room temperature. Once the gel had set, the plastic comb and masking tape were removed and the gel placed in a buffer chamber. DNA samples were mixed with loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol and 30% glycerol in sterilized distilled H₂O) and placed in the wells formed by the comb. The gel was covered with TBE buffer and current was applied through the gel. The anode was placed at the opposite end from the sample wells. The DNA samples were electrophoresed for 45-60 min at 100 Volts. DNA fragments were visualized by soaking the gel in 0.5 μg/ml of ethidium bromide for 15 min and then viewing on a UV transilluminator.

**D- Ligations:** Ligations were carried out with 100 ng each of vector DNA in a 20-μL volume containing 50 mM Tris-HCl, pH 7.4, 10 mM MgCl₂, 1 mM ATP, 1 mM DTT and 1 unit of T4 DNA ligase. The mixture was incubated overnight at 16°C and was used to transform competent cells.

**E- Preparation of fresh competent *E. coli* cells using the calcium chloride method:** A single colony (2-3 mm in diameter) from a plate freshly grown for 16-20 hrs at 37°C was picked and used to inoculate 2 ml of 2x YT broth. The cells were incubated in a 37°C shaking water bath overnight. From the overnight culture, 2 ml was used to inoculate 100 ml of 2x YT broth. The culture, in a 500 ml flask, was incubated for about 3 hrs at 37°C in a shaking water bath. Once the culture became turbid (OD of about 0.8 at 600 nm), the cells were taken out
of the shaking bath and aseptically transferred to two disposable ice-cold 50-ml polypropylene tubes. The culture was allowed to cool to 0°C by storing the tubes on ice for 10 min. After this period, the cells were spun at 5000 rpm for 10 min at 4°C in a Sorvall GS3 rotor. The media was then decanted from the cell pellets. The cell pellets in each tube were resuspended in 0.1 ml CaCl₂ and stored in ice for 10 min. After this period of time, the cells were recovered by centrifugation at 5000 rpm for 10 min at 4°C in the same rotor. The supernatant was then decanted completely and the cell pellets in each tube were resuspended in 2 ml of ice-cold 0.1 ml CaCl₂. At this point, the cells were either dispensed into aliquots that could be frozen (in 10% DMSO or 15% glycerol) or the cells can be stored on ice in 100 mM CaCl₂ for up to 3 weeks.

F- Transformation of plasmid DNA into competent cells: Plasmid DNA was routinely used to transform bacteria using the prepared CaCl₂ treated competent cells. A 200 µl volume of competent cells was placed in an ice-cold polypropylene tube with 100 ng of DNA and incubated on ice for 30 min. The cells were then heat-shocked for 90 sec at 42°C and placed on ice for 2 min. An aliquot (100-200 µl) of transformed cells was plated directly onto 2x YT (plus the appropriate antibiotics) plates and incubated overnight at 37°C.

Cell Storage: Bacterial and yeast cells were routinely frozen for storage. 0.9 ml of a growing culture of bacterial cells was taken and mixed with 0.1 ml of DMSO. The mixture was stored at -70°C. For yeast storage, a few large single colonies
from a two-day old YPD plate were taken and resuspended into 1 ml of sterilized 15% glycerol and stored at -70°C.

G- SDS-PAGE:

(1) Preparation of the Gel Cast: The glass plates were washed with a detergent several times and rinsed with distilled water and finally with 95% ethanol. Two spacers were placed between the glass plates and the plates were mounted onto the SDS-PAGE cast. A piece of paraffin paper was placed at the bottom of the glass plate to avoid any leakage.

(2) Preparation of the 13% Resolving Gel: In a small beaker, 3.1 ml of solution A, 1.65 ml solution B, 1.8 ml water, 70 µl 10% SDS and 50 µl ammonium persulfate (APS) were mixed thoroughly. To this mixture, 5 µl TEMED (N,N,N',N' tetramethylethylene diamine) was added and mixed. This mixture was poured between the glass plates and sufficient space was left for the stacking gel (the length of the Teflon comb plus 1 cm). Using a Pasteur pipette, 50 µl isobutanol was gently overlaid on the top of the resolving gel. The gel was allowed to sit at room temperature for 30 min. After this period of time, the overlay was poured off and the gel washed several times with distilled water.
(3) Preparation of the Protein Samples: The protein samples were mixed with 15-20 μl 1x gel loading buffer and boiled for 3 min to denature the proteins.

(4) Preparation of the 5% Stacking Gel: In a small beaker, 0.6 ml of solution C, 0.75 ml of solution D, 1.1 ml water, 25 μl 10% SDS and 25 μl 10% APS were mixed and to this mixture 5 μl TEMED was added. This mixture was poured on the top of the resolving gel and the Teflon comb was placed in the stacking gel solution without trapping any air bubbles. The gel was allowed to sit at the room temperature at a vertical position for 30 min. After the polymerization was complete, the Teflon comb is removed from the gel while the whole cast was held under a gentle stream of distilled water to avoid polymerization of unpolymerized materials in the wells. The whole cast was then put into the SDS-PAGE tank and the protein samples (15 μl) were loaded dry. After sample loading was complete, the wells were topped up with SDS-PAGE running buffer and then the upper and the lower reservoirs were filled with the running buffer (Any unused well was loaded with 1x gel loading buffer).

(5) SDS-PAGE Staining: The gels were routinely removed from the cast, placed in staining solution and microwaved at the “HIGH” setting for 60 sec. Then, the gels were shaken on a shaking rotator at room temperature for 5 min.

(6) SDS-PAGE Destaining: Gels were washed thoroughly in distilled water and placed in the destaining solution. The gels were then heated in a microwave at
the "HIGH" setting for 60 sec and were shaken at room temperature for another 5 min, following which the bands were visualized on a neon light box.
**H- PCR Protocol for the Amplification of the GLY1 Gene:** In a 0.5 ml thin-walled polypropylene tube (on ice), the following reaction components were added together to reach a total reaction volume of 100 µl:

<table>
<thead>
<tr>
<th>Reaction Components</th>
<th>Concentration of the Stock Solution to be added</th>
<th>Volume in µl</th>
<th>Final Concentration of the Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoclaved distilled deionized water</td>
<td>-</td>
<td>72.4</td>
<td>-</td>
</tr>
<tr>
<td>Taq Reaction Buffer</td>
<td>10x</td>
<td>10</td>
<td>x</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>25mM</td>
<td>0.6</td>
<td>1.5mM</td>
</tr>
<tr>
<td>dNTPs</td>
<td>5 mM</td>
<td>4</td>
<td>200 µM</td>
</tr>
<tr>
<td>p5'N</td>
<td>20 µM</td>
<td>5</td>
<td>1 µM</td>
</tr>
<tr>
<td>p3'B</td>
<td>20 µM</td>
<td>5</td>
<td>1 µM</td>
</tr>
<tr>
<td>DNA Template</td>
<td>50 µg / ml</td>
<td>2</td>
<td>100 ng</td>
</tr>
<tr>
<td>Paraffin Oil</td>
<td>70 µl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taq DNA polymerase</td>
<td>2000 units / ml</td>
<td>1</td>
<td>2 units</td>
</tr>
</tbody>
</table>
I- PCR Programs: PCR amplification of the target gene (GLY1 in this case) was achieved using the following programs on the Perkin Elmer DNA Thermal Cycler 480 (These files were linked together):

1. 60 sec at 94°C
2. 25 cycles of: 60 sec at 94°C (melting stage), 60 sec at 55°C (annealing stage) and 120 sec at 72°C (extension stage)
3. 10 min at 72°C (final extension)

J- Protein Assay: The protein assay was accomplished using the Bradford assay protocol. The Bradford reagent was diluted fivefold in distilled water. The diluted reagent was then filtered through Whatman 540 paper or equivalent paper. 10-20 μl cell extract or the BSA standard (0.1-1 mg/ml dilutions) was then added to 1 ml of the diluted reagent and mixed well. The OD of the blue colour formed (for cell extracts and the standard) was then measured at 595 nm and using the data from the BSA standard, a standard curve was made.

K- Protein Purification Using IMAC: BL 21 cells were transformed with a construct carrying a gene of interest. A transformant was picked and streaked out for single colonies and then a single colony was picked to inoculate 2 ml 2x YT + Amp100 broth. The culture was incubated at 37°C overnight in a shaking bath. The following morning, 0.5 ml of the overnight culture was used to inoculate 105 ml of the 2x YT+Amp100 broth. The culture was incubated under
the above conditions until OD$_{600}$ was reached 0.6 (3-4 hrs). From the culture, 5 ml was removed to be used as an uninduced control and IPTG was then added to the rest of the culture to a final concentration of 1 mM. The cells (test and control) were then incubated at 30³C for 3 hrs. After this period of time, the cells were removed and placed on ice. The cells then were washed three times with ice-cold sterilized water (5000 x g for 5 min). The cell pellet was then resuspended in 4 ml 1x Binding Buffer containing 1 mM phenylmethylsulfonylfluoride (PMSF) and 10 mM benzamidine. The cells were sonicated in an ultrasonic disintegrator operating between 18,000 and 20,000 Hz while on ice for 10 min (one minute burst time and cooling periods in between). The cell lysate was then centrifuged at 39,000 x g for 20 min to remove debris. The supernatant was then filtered through a 0.45 µm filter (to prevent the clogging of the resin used in IMAC) and was kept on ice until the His-Bind resin was charged using the following procedures (from this point on, all the manipulations were performed in the cold room):

The resin was gently inverted back and forth and 2.5 ml of it was delivered into a polypropylene column. When the level of the resin storage buffer (20% ethanol) was dropped to the top of the column bed, the following sequences of washes to charge and equilibrate the column (where 1 volume is equivalent to the settled bed volume):

→ 3 volumes sterile deionized water

→ 5 volumes 1x Charge Buffer
Once the Binding Buffer was dropped to the top of the column bed, the filtered cell extract was delivered into the column. The extract was allowed to drain away over the resin (FT). The column was then washed with 10 volumes of 1x Binding Buffer (W1). After this wash, the column was washed with 6 volumes of 1x Wash Buffer containing 40 mM imidazole (W2) (Different imidazole concentration for Wash Buffer can be made by simply mixing calculated volumes of Elute Buffer and Binding Buffer). A third wash (6 volumes) was also applied in which the concentration of imidazole was 60 mM (W3). The protein was finally eluted with 15 volumes of Elute Buffer containing 0.5 M imidazole (E). The purified protein was then dialyzed against 2 l 83 mM potassium phosphate buffer, pH 7.0/100μM PLP while stirring in the cold room overnight in dialysis tubing with an exclusion limit of 6,000 or less. The dialyzed protein was then concentrated against solid polyethylene glycol (PEG, 15,000-20,000 molecular weight) packed outside the dialysis tubing.

**L- P1cm Transduction:** P1cm is a bacteriophage with a hexagonal head and complex tail (Zinsser et al., 1980) that has quite a different mechanism for its maintenance. This virus resembles plasmids in the respect that its genome does not actually become integrated into the host chromosome, but instead replicates in the cytoplasm as circular DNA molecules. Although the plasmid prophage is not physically connected to the host DNA, phage DNA replication is closely
coordinated with cell division, since only one copy of the phage is present per host chromosome.

(1) **Preparation of Lysogen:** 0.05ml of P1cm was streaked in a horizontal line on a 2x YT plate containing 12.5 μg/ml chloramphenicol. The phage streak was allowed to dry, following which the plate was streaked perpendicularly with GS 459 *E.coli* strain (taken from an overnight culture plate containing 20 μg/ml kanamycin). The plate was incubated at 30°C overnight. The next morning the chlR cells (infected cells) were selected and subcultured on another fresh chloramphenicol plate and incubated at 30°C overnight.

(2) **Preparation of Lysate:** 1 ml superbroth was inoculated with a single colony of the chloramphenicol resistant GS 459 and incubated at 30°C overnight in a shaking water bath. After this period of time, 0.1 ml of the overnight culture was used to inoculate 5 ml of superbroth. The culture was incubated at 30°C in a shaking water bath for 2 hrs during which time the culture became cloudy. The cell culture was transferred to a 42°C shaking water bath and shaken for 30 min. Then the culture was transferred back to the 30°C shaking water bath until it became clear, indicating cell lysis (sometimes addition of 0.5 ml chloroform helped lyse the infected cells). The culture was vortexed briefly and was centrifuged at top speed for 15 min in a clinical centrifuge. The supernatant containing the lysate was then filtered through a 0.22 μm filter and the filtrate was stored at 4°C until needed.
(3) Transduction of BL 21: The recipient cells (BL 21) were grown overnight in 2x YT broth. Five hundred µl of the overnight culture was used to inoculate 5 ml 2x YT broth. The new culture was incubated at 30°C in a shaking water bath for 60 min. The culture was then centrifuged at 5,000 rpm for 10 min and the cell pellet was resuspended in 5 ml MC Buffer. The cells were placed in a 30°C shaking water bath for another 15 min. Meanwhile, the phage lysate was diluted 10-fold in the MC Buffer (10⁻¹ and 10⁻²). The rest of the experiment was conducted according to the table below:

<table>
<thead>
<tr>
<th>Tube #</th>
<th>Cells</th>
<th>Lysate</th>
<th>MC Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (cell control)</td>
<td>0.1 ml</td>
<td>-</td>
<td>0.1 ml</td>
</tr>
<tr>
<td>2</td>
<td>0.1 ml</td>
<td>0.1 ml of 1/100</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>0.1 ml</td>
<td>0.1 ml of 1/10</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>0.1 ml</td>
<td>0.1 ml of undiluted</td>
<td>-</td>
</tr>
<tr>
<td>5 (lysate control)</td>
<td>-</td>
<td>0.1 ml of undiluted</td>
<td>0.1 ml</td>
</tr>
</tbody>
</table>

These tubes were incubated in a 37°C shaking water bath for 20 min. After this period of time, 0.2 ml 1 M sodium citrate was added to each tube followed by 2.5 ml of F-TOP agar (45°C) and the mixture was plated onto 2x YT+kanamycin plates. The plates were then incubated at 30°C overnight. The next morning, transductants were streaked onto kanamycin containing 2x YT plates and incubated at 42°C to isolate the chi⁸ cells.
M- Threonine Aldolase Assays:

M1- Coupled Threonine Aldolase Assay:

Threonine aldolase activity was measured by adding an excess of alcohol dehydrogenase [EC 1.1.1.1] to remove the acetaldehyde produced and by following the oxidation of NADH.

\[
\text{Threonine} \rightarrow \text{glycine} + \text{acetaldehyde}
\]

\[
\text{Acetaldehyde} + \text{NADH} \rightarrow \text{ethanol} + \text{NAD}^+
\]

In a total volume of 1.5 ml, the standard assay system contained 83 mM potassium phosphate buffer, pH7.0, 0.3 \( \mu \)mol NADH, 18.57 units alcohol dehydrogenase, 20 \( \mu \)mol L-threonine, 100 \( \mu \)mol pyridoxal phosphate (PLP) and 750 \( \mu \)g purified protein (in the case of test sample, 750 \( \mu \)g purified GLY1 protein was used and in the case of positive control, 750 \( \mu \)g purified SHMTc was utilized) (Dainty, 1970).

The reaction was carried out under aerobic conditions in Beckman cuvettes where potassium phosphate buffer, PLP and purified protein were added together and used to blank the Beckman DU 600 spectrophotometer at 340 nm. To this mixture NADH was added, and it was monitored for 5 min for any change (decrease or increase) in OD readings. Then, threonine was added and the changes in OD readings were monitored for 15 min (for the purpose of negative control, the assay mixture was set up using the purified GLY1 protein,
but without any threonine substrate). Finally, alcohol dehydrogenase was added and the OD was monitored for another 10 min.

**M2- Direct Threonine Aldolase Assay:**

In this assay, the threonine aldolase activity of a respective protein was measured directly by detecting and quantitating the glycine end product of the reaction (in an amino acid analyzer) in which the protein of interest is presented with threonine as substrate. The assay protocol is as follows:

The crude extract of *S. cerevisiae* harvested during the logarithmic growth phase was incubated with the reaction mixture (80 mM threonine, 25 mM PLP, 100 mM HEPES/NaOH pH 7.0 and 0.5-1.5 mg/ml protein from crude extract) at 28 °C in a final volume of 0.5 ml. After 2 hrs, the reaction was stopped by addition of 0.5 ml 10% TCA. Denatured protein was subsequently removed by centrifugation and glycine in the supernatant was determined using an amino acid analyzer (precolumn derivatization of amino acids with o-phtalaldehyde and a fluorometric detection system) (unpublished data from a German group).

**N- Assay of β-Galactosidase in Yeast:**

A 5 ml culture of appropriate yeast cells was grown to a density of 1 x 10^7 - 2 x 10^7 cells/ml in an appropriate medium (which selects for the the presence of the plasmid of interest) at 30°C in a shaking water bath. Once such a cell density was reached, the cell culture was transferred onto ice and the cells were
harvested by centrifugation at 2,000 rpm for 5 min in a clinical centrifuge in a cold room. (NOTE: From this point onward, all the manipulations were performed on ice). After the centrifugation was complete, the supernatant was discarded and the cells were resuspended in 250 μl Breaking Buffer. Acid washed glass beads (450-600 μm) were added to the cells just below the meniscus of the liquid. Then, 12.5 μl PMSF stock solution was added to the cells. The mixture was vortexed six times at top speed in 15-second bursts and chilled on ice in between. To this mixture, another 250 μl Breaking Buffer was added and mixed well. The liquid phase was withdrawn using a 1-ml pipettor and transferred to a new tube. The crude extract was spun at the top speed of a microcentrifuge and then the clarified extract was transferred to a new tube. From this crude extract, 50 μl was taken and added to 0.9 ml of Z Buffer and the volume was adjusted to 1 ml with Breaking Buffer. This mixture was incubated at 28 °C for 5 min. The actual β-galactosidase reaction was initiated by adding 0.2 ml ONPG stock solution. The time of ONPG addition was recorded precisely. The incubation at 28 °C was continued until a pale yellow colour appeared. Then, the reaction was terminated by the addition of 0.5 ml Na₂CO₃ stock solution. The timing of this addition was also recorded precisely. The optical density of the yellow colour produced in this reaction was then measured at 420 nm. The amount of protein in the crude extract was also determined using the Bradford Assay.
Having all the necessary data, the specific activity of the extract was then calculated using the formula below:

\[
\text{OD}_{420} \times 1.7 \\
0.0045 \times \text{protein} \times \text{extract volume} \times \text{time}
\]

where \( \text{OD}_{420} \) is the optical density of the product, o-nitrophenol, at 420 nm. The factor 1.7 corrects for the reaction volume. The factor 0.0045 is the optical density of 1 nmol/ml of o-nitrophenol. Protein concentration is expressed as mg/ml. Extract volume is the volume assayed in ml. The time is in min. Specific activity is expressed as nmole/minute/mg protein.

**O- Preparation of ssDNA for Yeast Transformation:** 100 mg of Sigma D1626 type II sodium salt salmon testes DNA was dissolved in 10 ml of distilled water in a beaker. The DNA solution was stirred (using a magnetic bar) at 4°C overnight in order to reach a uniform solution. This DNA solution (while kept on ice) was then sonicated (18,000-20,000 Hz) for 7 min to obtain DNA fragments of ~ 7 kb average size. The DNA was then extracted once with buffer saturated phenol, once with phenol/chloroform and finally with an equal vol chloroform. The DNA was precipitated with 2 vol ice-cold ethanol and then rinsed with 70% ethanol. The DNA was allowed to dry and then dissolved in autoclaved distilled water at
10mg/ml concentration and kept in the cold-room overnight. The next morning, the OD<sub>260</sub> of the DNA solution was measured.

For the purpose of yeast transformation, an aliquot of this DNA was taken and boiled for 10 min and quickly cooled on ice-water.

**P- Yeast Transformation:** Yeast cells were grown in 10 mls of YPD broth in a 30°C shaking water bath to a density of ~2x10<sup>7</sup> cells/ml. The cells were centrifuged at 2,000 rpm for 5 min in a clinical centrifuge. The cells were then washed three times with 0.1 M LiOAc, TE (pH 7.5). Cells were resuspended in 1/100 volume of the same buffer and transferred into a 10-ml plastic tube. The tube was placed on a roller drum at 30°C for 60 min. After this period, 1-10 µg plasmid DNA and 10 µl 5mg/ml sheared deproteinized carrier DNA (single stranded salmon testes DNA) were added to 0.1 ml of cells and incubated at 30°C for 30 min. Then, 0.7 ml of 40% PEG 3500, 0.1 M LiOAc, TE (pH 7.5) was added to the cell mixture and incubation continued at the same temperature for another 60 min. The cells were then transferred to a 42°C water bath for 5 min (heat shock stage). Cells were spun at 2,000 rpm for 5 min and washed once in 1 ml TE (pH 7.5). Finally, cells were resuspended in 0.2 ml sterile deionized water and plated on selective plates. Plates were let dry and were incubated at 30°C for 2-3 days.

**Q- Isolation of Yeast Genomic DNA:** A single yeast colony from a putative gene-disrupted mutant was used to inoculate 2 ml YPD broth. The culture was
incubated at 30°C in a shaking water bath overnight. The next morning, 1.5 ml of the overnight culture was spun down in a small polypropylene tube at top speed in a microfuge for 30 sec. The pellet was resuspended in 200 μl TE (50 mM/10 mM). To this suspension was added acid-washed glass beads occupying 0.5 X volume. The tube was vortexed for 30 sec, in 4 separate intervals followed by phenol/chloroform treatment of the extract and subsequent precipitation of yeast genomic DNA with ice-cold 95% ethanol. The precipitate was centrifuged at the top speed of a microcentrifuge for 10 min. The DNA pellet was then rinsed with 75% ethanol and air-dried.

**R-Gene disruption:** Gene disruptions were carried out to obtain null mutants at the GLY1 locus. The GLY1 gene disruption was performed using a GLY1 knock-out plasmid (pJM1223) in which part of the 5' non coding region, promoter region and the first three amino acids of the gene had been replaced by a URA3 marker. This plasmid was digested with SalI and BamH1 to transform the recipient yeast strains to uracil prototrophy.

**S- Southern Blot Analysis:** The yeast genomic DNA obtained from the putative mutants was digested with HindIII in a total volume of 50 μl containing the appropriate 10x BRL reaction buffer and 50 units of enzyme. The digestion mixture was incubated at 37°C for 4 hrs to overnight. The digested DNA was then subjected to Southern analysis to verify the putative gene-disruptions. The
HindIII-digested genomic DNA was electrophoresed in a 0.8% agarose gel at 40V overnight to separate the DNA fragments. The DNA was then denatured under alkaline conditions (0.4N NaOH) and transferred overnight in 10x SSC to a nylon membrane (Gene Screen Plus; DuPont) by the method of Southern (Southern, 1975). DNA probes were labeled with [α-32P] dATP by the random priming method (Feinberg and Vogelstein, 1983) and the yeast genomic DNA was probed with 6 kb HindIII pET15b-GLY1 digested fragment. Prehybridization of the membrane-DNA was carried out for 60 min at 65°C in an oven armed with a rotating device. Then the probe solution was added to the prehybridization buffer and kept at 65°C overnight. After this period of time the membrane was washed twice with excess 2x SSC for 5 min at room temperature, then with 2x SSC, 1% SDS at 60°C for 30 min, and finally, washed with 0.1x SSC at room temperature for 30 min. After this wash, the membrane was rinsed with excess distilled water, wrapped in plastic wrap and exposed to the Kodak film for 1-2 hrs.

NOTE: Apparently no UV crosslinking of DNA to the membrane is required for the Gene Screen Plus membrane; however, UV crosslinking of the DNA to the membrane for a short period of time (90sec) would be of no harm.

**T- Yeast Mating:** Yeast strains of opposite mating type from fresh YPD plates were cross streaked onto another fresh YPD plate and incubated at 30°C for 24 hrs. After this time, the resultant diploids were removed from the YPD plate and
streaked onto a minimal sporulation plate with the necessary supplements. The minimal sporulation plate was incubated at 30°C for 3-5 days, and the formation of spores confirmed. Then, a loopful of asci was taken and transferred into a 1.5-ml polypropylene tube. The asci were then washed three times with sterile distilled water and finally resuspended in 0.9 ml sterile water. Ether (0.1 ml) was added to this mixture and vortexed for 60 sec to get rid of diploids. The spores were then washed again three times with sterile water and treated with 10% glusulase (10,000 units/g of lytic activity) for 30 min at 30°C (the separation of spores from the ascus walls was confirmed microscopically). Spores were sonicated for 25 sec to completely separate them from each other. The spore mixture was directly plated onto YPD plates containing 10 mM glycine and incubated for 2 days at 30°C. After this period, the single colonies were transferred onto the selective plates to check for the desired phenotypes.

**Tetrad Analysis:** The diploids formed on the minimal sporulation plates were taken and washed with sterile distilled water three times. Then, the spore pellet was resuspended in 20 μl of a solution containing 10 mg/ml zymolase and incubated at 30°C for 2 min. The zymolase action was stopped by adding 180 μl 1M sterile sorbitol and then the spores were blotted onto a slab of YPD agar (4% agar). The blot was let dry slightly to avoid movement of spores under the microscope stage. Using a micro-dissection needle, the spores were pried apart and seeded at the known co-ordinates onto the agar slab. Once the dissection
was over, the agar slab was transferred onto a YPD plate and incubated at 30°C for 2 days. The resultant haploids were then checked for the desired phenotypes.
Results

A- The coding region of the GLY1 gene was successfully amplified

Having the GLY1 sequence (McNeil et al., 1994) at hand (Fig. 12), two primers, p5’N and p3’B (Table 7), were designed and made. Using these primers along with the Yep24-GLY1 as template and utilizing the PCR protocol presented in the Materials & Methods section, a reaction mixture was made. This GLY1 PCR reaction mixture was put through four different PCR programs differing in their time spans of their DNA premelting stages (2 min vs. 1 minute) and also in annealing temperatures (50°C vs. 55°C). The results of these PCRs (shown in fig. 13) clearly demonstrate that all programs worked well; however, an annealing temperature of 55°C appeared to be optimal. In figure 13, lane #1 corresponds to 1 kb markers, lanes 2&7 represent positive controls (SHM2) (composed of two sets of different primers (p10 and p30 and SHM2 template with 1 minute premelting stage) amplified at two different annealing temperature, 50°C and 55°C, respectively. Lanes 3&8 represent negative controls which are the same as positive controls except the fact that the DNA template is excluded from the reaction mixture. Lanes 4&9 represent the GLY1 amplification at 1 minute premelting stage and 50°C & 55°C annealing temperatures respectively. Lanes 5&10 represent the GLY1 PCR at 2 min premelting stage and 50°C & 55°C.
Fig. 12. Nucleotide sequence of the *S. cerevisiae* GLY1 gene and flanking regions

\[-143\] TTA CTG ATT TTA TCT TGA AGG CGG GAG GAA TTT TTT TTA CTA
\[-114\] GAA ATT AGG AGG CAA GCG TCC AGG CAT AGC TCA TCT CAT GCT
\[-128\] AAC TCT CTA ATG CTT TCA TCA TAT TTA CCA CTA TTA TTA
\[-119\] CAA TTT TCT ATT CTT TTA ACA CTA TGT CAA TTA TTA TTA
\[-117\] TTA TTA ATT CTT TTA TTA TTA TTA TTA TTA TTA TTA

The nucleotide sequence of the *S. cerevisiae* GLY1 gene and flanking regions is shown. The sequence is presented as a series of nucleotide triplets, with the start and stop codons indicated (e.g., ATG for start and TAA for stop). The sequence is transcribed from the 5' to 3' direction and includes the coding region as well as the flanking regions on both sides. This type of information is crucial for understanding the genetic structure and function of the GLY1 gene in *S. cerevisiae*. The sequence is essential for bioinformatic analysis, genetic engineering, and functional studies in yeast biology.
**Fig 12.** Nucleotide sequence of the *S.cerevisiae GLY1* gene and flanking regions. The deduced amino acid sequence is shown above the nucleotide sequence. The predicted termination codon is indicated with a cross. A putative TATA box sequence is underlined. The two hydrophobic stretches of amino acids forming the transmembrane domains are indicated in Bold letters. The most hydrophilic amino acid stretch is indicated in Bold-Italic letters with dotted underlinings.
Fig. 13 The GLY1 PCR Amplification

This figure depicts a 0.8% agarose gel electrophoresis of the GLY1 PCR products along with positive (SHM2) and negative controls.

In this figure:
Lane 1: the 1 kb DNA markers
Lane 2: the positive control (SHM2) at 1' premelting and 50°C annealing temperature
Lane 3: the negative control (the same as the positive control but with no template added) at 1' premelting and 50°C annealing temperature
Lane 4: the GLY1 PCR products at 1' premelting and 50°C annealing temperature
Lane 5: the GLY1 PCR products at 2' premelting and 50°C annealing temperature
Lane 6: loaded with the DNA loading buffer
Lane 7: the positive control at 1' premelting and 55°C annealing temperature
Lane 8: the negative control at 1' premelting and 55°C annealing temperature
Lane 9: the GLY1 PCR products at 1' premelting and 55°C annealing temperature
Lane 10: the GLY1 PCR products at 2' premelting and 55°C annealing temperature
annealing temperatures (the GLY1 amplification at 2 min premelting and 50°C annealing did not work out).

**B- The GLY1 amplicons with 5' Ndel site and 3' BamHI site were subcloned into pET 15b expression vector**

The GLY1 amplicons were placed onto commercial spin columns in order to separate the paraffin oil and unused nucleotides from the GLY1 amplicons. The pET 15b and the GLY1 amplicons were then digested sequentially with Ndel and BamHI. The digested inserts of ~1.2 kb (GLY1 PCR products) and the digested pET 15b were run on a 0.8% low melting agarose gel and purified through spin columns. The purified inserts and vectors were then ligated in a 20 µl total reaction vol at a 5:1 ratio (2 µg inserts: 400 ng vectors) (Fig. 14). The ligation mix was incubated at 16°C overnight and then transformed into JF1754 cells. The plasmid DNAs of the resultant transformants were isolated and digested with HindIII to check for the desired construct, pET 15b-GLY1 (Fig. 15). One transformant contained the recombinant desired since HindIII digestion of pET 15b-GLY1 construct (PG), according to the restriction maps of the plasmid and the gene, would give rise to two fragments of ~6 kb and ~1 kb (lane 2 in Fig.15). Lane 3 represents the HindIII digested pET 15b which was used as a control.
Fig. 14 Subcloning of the GLY1 coding region in pET 15b
This figure depicts a 0.8% agarose gel electrophoresis of the *Hind III* digestion products. Lane 1 represents the 1 kb DNA markers and lanes 2 & 3 represent *Hind III* digestion of a putative PG construct and pET 15b respectively.
C- The GLY1 protein was successfully expressed in BL 21 cells

The pET 15b-GLY1 construct (where the GLY1 gene is put under the control of a T7 lac promoter) was used to transform the BL 21 lysogens (which carry a chromosomal copy of the T7 RNA polymerase). Induction of the T7 lac promoter by a lac inducer such as IPTG in BL 21 results in the expression of the GLY1 protein with a N-terminal hexahistidine tag (which can be advantageous when it comes to the protein purification since such proteins can be purified rather easily on a nickel column using immobilized ion metal affinity chromatography), and a thrombin cleavage site following the hexahistidine tag (which is advantageous when one wishes to obtain the purified protein in its native state with no histidine tag attached). Figure 16 depicts the results of the GLY1 protein expression. In this figure, lane 1 represents the BioRad protein markers. Lane 2 represents the IPTG induced (3 hrs) BL 21 transformed by pET 15b-GLY1 constructs (PG). Lane 3 represents the uninduced PG-transformed BL 21 cells. Lanes 4&5 represent induced (4 hrs) and uninduced PG-transformed BL 21 cells. Lanes 6&7 show induced and uninduced pET 15b-transformed BL 21 cells respectively. Finally, lanes 8&9 represent the IPTG-induced and uninduced BL 21 cells. As can be seem from figure 16, in lane 5 where no induction has occurred, there is indication of protein expression which can be attributed to the basal level expression of the GLY1 gene due to leakiness of the lac promoter.
Fig. 16  Induction of the GLY1 Gene Analyzed by 13% SDS-PAGE

In this figure:
Lane 1: the BioRad protein markers
Lanes 2 & 4: the IPTG induced PG- transformed BL 21 cells for 3 & 4 h respectively
Lanes 3 & 5: the uninduced PG BL 21 cells for 3 & 4 h respectively
Lane 6: the IPTG induced pET 15b transformed BL 21 cells for 3 h
Lane 7: the uninduced pET 15b BL 21 cells for 3 h
Lane 8: the IPTG induced BL 21 cells for 3 h
Lane 9: the uninduced BL 21 cells for 3 h
D- Most of the GLY1 protein expressed in BL 21 cells is produced in an insoluble state in the form of inclusion bodies

Eukaryotic genes expressed in a bacterial system can be toxic to the bacterium. Bacterial cells can counteract this toxicity by simply surrounding the toxic protein in sacs with inverted plasma membranes which are called Inclusion Bodies. One advantage of inclusion bodies is the fact that they are easily isolated by centrifugation to yield highly concentrated and relatively pure protein. With many proteins it has been possible to solubilize the inclusion bodies with urea or guanidine-HCl and then refold by slowly diluting or dialyzing the denaturant at low protein concentration (Ausubel et al., 1989). Even when inclusion bodies are formed, some portion of the target proteins usually remains soluble within the cell.

Figure 17 shows that most of the GLY1 protein has entered the insoluble phase. In this figure, lane 1 corresponds to the BioRad protein markers, lane 2 indicates the IPTG-induced supernatant, lane 3 represents the uninduced supernatant; lane 4 shows the IPTG induced pellet and lane 5 represents the uninduced pellet.

E- Construction of *glyA*− BL 21 by P1cm transduction of GS 459 E.coli cell strain as donor and BL 21 As recipient

In order to confirm the functionality (glycine biosynthesis) of the GLY1 gene in the pET 15b-GLY1 construct, this construct had to be used to transform
Fig. 17  The major portion of the GLY1 protein expressed in BL 21 cells is in the form of inclusion bodies.

As analyzed on a 13% SDS-PAGE, in this figure:
Lane 1: the BioRad protein markers
Lane 2: the IPTG induced supematant phase of the lysate
Lane 3: the uninduced supematant phase of the lysate
Lane 4: the IPTG induced pellet phase of the lysate
Lane 5: the uninduced pellet phase of the lysate
a glycine auxotrophic E.coli mutant to look for glycine complementation in that strain (it is shown that the majority of yeast genes are not split and thus can complement the respective auxotrophies in the E.coli mutants. HIS3, LEU2 and TRP1 are among these genes). The only glycine auxotrophic mutant at hand was the E.coli strain GS 459; however, this strain could not be used for this purpose since it lacks a chromosomal copy of T7 RNA polymerase which is required for expression of the GLY1 gene from the PG construct. For this reason a generalized P1cm transduction was performed in which the strain GS 459 was the donor strain that contributed the glyA::Tn5 element, and BL 21 strain acted as a recipient cell where the integration of glyA::Tn5 at the SHM locus would lead to glycine auxotrophy in this strain (Fig.18). The resultant glyA⁻ BL 21 strains were plated on kanamycin-containing minimal plates with and without glycine and incubated at 37°C overnight. No growth was observed on the plate lacking glycine and significant growth was observed on the plate with glycine supplements (20 µg/ml), indicating that the SHM gene is disrupted in this strain (Table 8). Similarly, the GS 459 strain was unable to grow on kanamycin-containing minimal plates with no glycine supplements, but conversely, the wild type BL 21 could grow on minimal plates without any glycine supplements.
Fig. 18  Generalized Transduction of BL 21 E.coli Strain with the P1cm Lysate of GS 459 E.coli Containing the glyA:: Tn5 Element
The glycine auxotrophy of glyA^{-} BL 21 could be weakly overcome provided these cells are transformed with the PG construct and induced by IPTG

The glyA^{-} BL 21 was transformed with the following constructs: ① pET 15b-GLY1 (PG), ② pET 15b-SHM2 (PS) and ③ pET 15b. The transformants were then plated onto ampicillin-containing minimal plates in the presence and absence of IPTG & glycine and onto ampicillin-containing 2x YT plates (Table 9).

As can be seen from table 9, pET 15b transformed cells did not grow on minimal plates with no glycine supplements. PS transformed cells grew on all plates after 24 hrs incubation at 37°C indicating that the yeast cytoplasmic SHMT can complement the glycine auxotrophy of this strain even at the basal level of expression, as happened on the plates with no IPTG. PG-transformed cells were able to grow on the rich plate and on the minimal plates with IPTG or glycine supplements. The growth of PG-transformed cells on the minimal plates with the IPTG occurred after 3 d incubation, thereby pointing to a rather weak glycine complementation.
TABLE 8

Growth of Glycine Auxotrophic *E.coli* Mutants on Different Media

<table>
<thead>
<tr>
<th>Strain</th>
<th>Media</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M''+Kan^b</td>
</tr>
<tr>
<td>GS 459</td>
<td>-</td>
</tr>
<tr>
<td>glyA*BL 21</td>
<td>-</td>
</tr>
</tbody>
</table>

The glycine auxotrophic mutants, GS 459 and glyA* BL 21, were streaked on the above media and incubated in a 37°C incubator for 24 h. After this period of time, the presence or absence of the growth on the above media was recorded for each mutant.

a: Minimal Bacterial Medium  
b: 20 μg/ml kanamycin  
c: No growth after 24 h  
d: Growth after 24 h
### TABLE 9

**Glycine Complementation of glyA' BL 21 by Utilization of Different Constructs**

<table>
<thead>
<tr>
<th>Cell Strain</th>
<th>Media</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M^3+Amp75</td>
</tr>
<tr>
<td></td>
<td>+IPTG</td>
</tr>
<tr>
<td></td>
<td>2xYT+Amp75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>pET 15b BL21*</th>
<th>PS* BL21*</th>
<th>PG' BL21*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- c</td>
<td>+++ d</td>
<td>++ g</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+ g</td>
<td>+++</td>
</tr>
</tbody>
</table>

The BL 21 cell strains were transformed with pET 15b, pET 15b-SHM2 and pET 15b-GLY1 plasmid DNAs. The resultant transformants were plated on the above media and kept in a 37°C incubator for up to three days. The presence or absence of growth was recorded for each transformant.

a: Minimal bacterial medium  
b: 75 μg/ml ampicillin  
c: No growth  
d: Growth after 24 h  
e: pET 15b-SHM2 transformed BL 21*  
f: pET 15b-GLY1 transformed BL 21*  
g: Growth after 3 days  
*: glycine auxotrophic BL 21
The GLY1 protein was expressed in glyA- BL 21 and purified using Immobilized Ion Affinity Chromatography on the HIS-BIND resin. PG transformed glyA- BL 21 cells were IPTG-induced and prepared for the IMAC purification procedure (the purification of the GLY1 protein from this background has the merit of the fact that if this protein is to be used for enzyme assays such as threonine aldolase assay, the chance of GLY1 protein getting contaminated with the bacterial SHMT which is known to have threonine aldolase activity is practically nil). This procedure takes advantage of the fact that a protein with a N-terminal hexahistidine tag can be bound by a Ni\(^{+2}\)-charged resin (and after proper washes to get rid of protein impurities) and can be released by a competing agent (0.5M imidazole). Figure 19 shows the results of the GLY1 protein expression and purification.

In this figure, lane 1 represents the BRL protein markers, lane 2 represents the IPTG induced whole cell extract, lane 3 indicates the uninduced whole cell extract, and lanes 4 -10 correspond to pellet (induced), supernatant (induced), flow through (FT), wash 1 (W1), wash 2 (W2), wash 3 (W3) and eluent (E) respectively. The purified protein (soluble) was then dialyzed against 83 mM potassium phosphate and concentrated to a final concentration of 7.5 μg/μl.
**Fig. 19** GLY1 protein expression and purification from the glyA\textsuperscript{-} BL 21 background analyzed by 13% SDS-PAGE

In this figure:

- **Lane 1**: BRL protein markers
- **Lane 2**: IPTG induced whole cell extract of PG transformed glyA\textsuperscript{-} BL 21 cells
- **Lane 3**: uninduced PG-glyA\textsuperscript{-} BL 21 whole cell extract
- **Lane 4**: IPTG induced pellet phase of the lysate
- **Lane 5**: IPTG induced supernatant phase of the lysate
- **Lane 6**: flow through phase of the GLY1 protein purification process
- **Lanes 7- 9**: wash stages of 1 to 3
- **Lane 10**: eluent.
H- Anti-GLY1 antibody could not be raised against the purified GLY1 protein

The GLY1 protein was purified, concentrated and used to inject the lab animals (mice) to raise antibodies (This work was done in collaboration with Dr. Ron Pearlman, Biology Department, York University). However, no antibody was obtained from the immunization of mice with this protein.

I- Yeast cytoplasmic SHMT was also expressed in glyA− BL 21 and purified on nickel column utilizing IMAC

The yeast cytoplasmic SHMT was expressed in glycine auxotrophic BL 21 cell strains containing the pET 15b-SHM2 construct (Fig. 20) and purified with the nickel column in a manner similar to that used for the GLY1 protein (Fig. 21).

In figure 20, lane 9 presents the IPTG-induced whole cell extract and lane 10 corresponds to uninduced whole cell extract.

In figure 21, lane 1 represents the BRL protein markers and lane 9 represents the product of the elution stage of SHMTc IMAC. This purified protein was dialyzed against 83 mM potassium phosphate buffer overnight and was then concentrated against PEG (MW ~ 15,000-20,000) to a final concentration of about 7.5 μg/μl. This protein was used as a positive control in the threonine aldolase assay since it contains such an activity.
Overexpressed SHMTc in glyA- BL21

Fig. 20  13% SDS-PAGE analysis of the Overexpressed SHMTc in glyA- BL21 Cell Strain Transformed with pET 15b-SH2.

In this figure, Lane 9 represents the IPTG- induced whole cell extract for a period of 3 hours and Lane 10 represents the uninduced whole cell extract for the same period of time.
Fig. 21 13% SDS-PAGE analysis of Purified Yeast Cytoplasmic SHMT

In this figure, Lane 1 represents the BRL protein markers. Lane 7 represents the flow through phase of the purification process. Lane 8 represents the first wash. Lane 9 represents the product of the second wash which was used for enzyme assays since it seemed pure enough when compared to the eluent at Lane 10.
J- Threonine Aldolase Assays of GLY1 Protein

J1- Coupled Threonine Aldolase Assay

It has been shown that the GLY1 protein functions in glycine synthesis (McNeil et al., 1994). There are a number of biochemical reactions by which glycine can be made. Among these reactions, that involving threonine aldolase seems to be a candidate since aldolytic activity on threonine brings about its conversion to glycine and acetaldehyde.

This assay (that is in fact a coupled assay in which the acetaldehyde produced along with NADH are acted upon by alcohol dehydrogenase to produce ethanol and NAD⁺, and the reduction in OD₃₄₀ measured) was performed on the purified GLY1 protein and SHMTc (both from yeast) under aerobic condition.

The results of this assay can be seen in figure 22. As can be seen, no threonine aldolase activity was observed using the purified GLY1 protein and no activity could also be shown for the positive control (SHMTc) which is known to have threonine aldolase activity. A negative control was also included in this assay where the assay mixture included the purified GLY1 protein, but with no threonine substrate added.

The results of this assay should be considered inconclusive for a number of reasons which will be addressed in the “Discussion” section.
Fig. 22 Coupled Threonine Aldolase Assay with the Purified SHMTc and GLY1 Proteins. Experimental details are given in the Materials and Methods section.

- SHMTc (positive control)
- GLY1 (test)
- No Threonine Substrate (negative control)
J2- Direct Threonine Aldolase Assay of the GLY1 Protein (Bognar et al., unpublished data)

The results of the direct threonine aldolase activity is presented in the Table below:

**TABLE 10**

**Direct Threonine Aldolase Assay**

<table>
<thead>
<tr>
<th>Yeast Strain</th>
<th>Thr. Ald. sp. ac&lt;sup&gt;a&lt;/sup&gt;.</th>
</tr>
</thead>
<tbody>
<tr>
<td>YM 14</td>
<td>879</td>
</tr>
<tr>
<td>YM 13</td>
<td>34</td>
</tr>
<tr>
<td>YM 13 Transformed with Yep24-GLY1</td>
<td>1730</td>
</tr>
</tbody>
</table>

Crude extracts of YM 14 (wild type at the GLY1 locus), YM 13 (disrupted at the GLY1 locus) and Yep24-GLY1 transformed YM 13 were obtained and subjected to the direct threonine aldolase assay (outlined in Materials and Methods section).

<sup>a</sup>-specific activity in nanomoles per hour per mg of protein
According to the results outlined in table 10, the threonine aldolase activity in YM13, where \textit{SHM1, SHM2} and \textit{GLY1} genes are disrupted, is low. However, overexpression of the \textit{GLY1} gene in the above strain from the Yep24-
\textit{GLY1} construct has led to a significant increase in threonine aldolase activity thereby, pointing to the conclusion that the \textit{GLY1} protein is indeed threonine aldolase.

\textbf{K- Glyoxylate anaplerotic pathway is functional in yeast and can provide the} \textbf{ser1 mutant yeast strain with serine when grown on a non-fermentable carbon source without serine supplements}

In yeast serine can be made from the glycolytic intermediates (Fig. 3) and the serine produced can be converted to glycine and 5,10-methylene THF through the action of cytoplasmic SHMT. In the \textit{ser1} yeast strain, YM 14, serine can not be provided to the cell from the glycolytic intermediates due to a point mutation at the \textit{SER 1} locus coding for phosphoserine aminotransferase. These cells are unable to grow on the minimal media where the sole carbon source is glucose. However, there is a glucose repressible pathway in yeast called the \textit{glyoxylate shunt} which is induced when cells are grown on the minimal media with non-fermentable carbon sources such as ethanol or potassium acetate (Fig. 3). This pathway provides the yeast cells with \textit{glycine} which in turn can be converted into serine due to the action of mitochondrial SHMT. Consequently, this is another biochemical pathway by which glycine can be made.
The growth of the ser1 yeast strain was studied on glucose minimal medium with and without serine supplements (Fig. 23). No growth was obtained on minimal medium without serine supplements.

The ser1 yeast strain was able to grow slowly on the minimal medium with potassium acetate, but the growth rate was augmented when sodium glyoxylate was added to the medium (Fig. 24).

L. Overexpression of the GLY1 Gene and Relief of Serine Auxotrophy in a ser1 yeast strain

It was postulated that the GLY1 gene might code for an aminotransferase which presumably functions in the glyoxylate anaplerotic pathway by transaminating glyoxylate to glycine using alanine as an amino group donor. It was reasoned that if this is the case, the overexpression of this gene on non-fermentable carbon sources would lead to the relief of serine auxotrophy due to the overproduction of glycine which can subsequently converted into serine. The results obtained from two separate growth studies suggest otherwise.

Figure 25 shows that the overproduction of the GLY1 gene in a ser1 strain led to the relief of serine auxotrophy on glucose minimal media, whereas such a phenomenon did not occur on the media with non-fermentable carbon sources (Fig. 26), ruling out the possibility of the GLY1 gene coding for an alanine-glyoxylate aminotransferase.
Fig. 23 Growth of ser1 YM 14 on SD (SYNTHETIC DEFINED) Medium

- SD +HIS +LEU +URA
- SD +HIS +LEU +SER +URA
Fig. 24 Growth of \textit{ser1} YM14 on SA (acetate) and SAG (acetate+sodium glyoxylate) Media

![Graph showing growth of ser1 YM14 on SA and SAG media over time.](image-url)
Fig.25  Growth of Wild Type YM14 and Yep24-GLY1 Transformed YM14 on SD (Synthetic Defined) with Histidine, Leucine and Uracil Supplements

- Wild Type YM14
- Yep24-GLY1 TRANSFORMED YM14
Fig. 26 Growth of Wild Type YM14 and Yep24- GLY1
Transformed YM14 ON SA + Histidine
+ Leucine + Uracil

WILD TYPE YM14
Yep24-GLY1 YM14
Another group of investigators in Germany (unpublished data), using alanine-glyoxylate aminotransferase assay, have shown that in a GLY1 knock-out strain there is comparable aminotransferase activity to that of the wild type strain at the GLY1 locus, and that this activity is still inducible by growth on ethanol. The results of these studies also point to the fact that the GLY1 protein does not have AGAT activity.

M- Genetic Probing of the Possible Regulation of the GLY1 Gene Expression by Threonine and Glycine in a Yeast Setting

To probe the biochemical regulation of the GLY1 protein, a genetic approach was utilized in the yeast setting. This approach took advantage of a shuttle vector, pJM1308, where the promoter and part of the 5’ coding region of the GLY1 gene is fused to the lac Z gene (reporter gene). This plasmid was used to transform the yeast strain YM 22 to leucine prototrophy. The transformants and the wild type cell strains were then grown on SD minimal with threonine (10 mM final concentration), glycine (10 mM), both or none. The cells were then harvested and prepared for β-galactosidase assay. The results of this assay are shown in figure 27. As can be seen, there is no β-galactosidase activity for the wild type strain in the presence of threonine, which is as it should be since Baker’s yeast does not have β-galactosidase. However, in pJM1308-transformed yeast cells the presence of threonine brings about a 1.5 x increase in β-galactosidase activity in comparison to β-galactosidase activity of the
Fig. 27 β-Galactosidase Activity in Yeast

β-Galactosidase activity in nmols/minute/mg protein

Yeast Cells

1: wild type YM 22
2: wild type YM 22 + threonine
3: wild type YM 22 + glycine
4: wild type YM 22 + threonine + glycine
5: p1308 YM 22
6: p1308 YM 22 + threonine
7: p1308 YM 22 + glycine
8: p1308 YM 22 + threonine + glycine
transformants grown in the absence of threonine. This increase, however, is not significant enough to prove that threonine up-regulates the GLY1 gene expression. The results of the same assay with and without glycine supplements indicated that glycine could not possibly be involved in the down-regulation of the GLY1 gene expression.

**N- Disruption Studies of the GLY1 Gene**

Since it was hypothesized that the GLY1 gene might be involved in the glyoxylate anaplerotic shunt, it was reasoned that disruption of this gene would terminate the ability of a ser1 yeast strain to grow without serine on media in which the sole carbon sources were non-fermentable. To achieve this, GLY1 gene disruptional studies were embarked upon utilizing two strategies:

1. **Gene Replacement Strategy**
2. **Mating Strategy**

**N1- No GLY1 knock-out mutants were obtained from the ser1 yeast strain, YM 14, using the gene replacement technique**

Two yeast strains, YM 14 and YM 09, were made competent and were transformed with *BamH*1/SA1 pJM1223 plasmid to uracil prototrophy. The uracil prototrophic mutants of both strains were then subjected to Southern blot analysis where the DNA probe was the *Hind*III digest of the PG construct
carrying the first 504 nucleotides of the GLY1 coding region (Fig. 28). In this figure, the first 6 lanes depict the GLY1 knock-outs of the YM 09 strain. As can be seen, there is a single band in each of these lanes which corresponds the positive control band (YM 11) in lane 14. Wild type YM 09 was loaded in lane 7. Lanes 8-13 represent Southern analysis of the YM 14 uracil prototrophs indicating absolutely no difference between these mutants and the wild type YM 14 loaded in the last lane (pointing out that these mutants were still wild type at the GLY1 locus). However, there was some confusion on the Southern blot analysis results of the YM 14 uracil prototrophs and the wild type YM 14 since there should have been only one single band corresponding to the lower band of the two in each lane. Another Southern blot analysis with a freshly prepared probe was done on the above mutants (plus two more uracil prototrophic YM 14 mutants from the same experiment) and the results of which are shown in figure 29. As can be seen from this figure, the Southern results of all the putative GLY1 knock-outs corresponds to the wild type YM 14 (lane 10), pointing to the fact that these mutants are still wild type at the GLY1 locus. Lane 9 shows the single band for the positive control (YM 11).

Notwithstanding these observations, it should be noted that GLY1 gene disruption has been achieved by another group of investigators led by Dr. D. Appling (unpublished data) using the pJM1223 knock-out plasmid and the yeast strain DAY4: a ura3-52 trp1 leu2 his3 his4 ser1-171. The only phenotype
Fig. 28 Southern Blot Analysis of the Uracil Prototrophic Yeast Haploids of YM 09 and YM 14 Obtained Through the Gene Replacement Strategy Using pJM1223 GLY1 Knock-Out Plasmid and Utilizing a 8 kb Hind III PG Digest as the Probe

Lanes 1-6 represent the GLY1 knock-outs of YM 09 and lanes 8-13 represent the uracil prototrophic YM 14 mutants. The wild type YM 09 and wild type YM 14 are loaded in lanes 7 and 15 respectively. The positive control, YM 11, is loaded in lane 14.
Fig. 29 Southern Blot Analysis of the Uracil Prototrophic YM 14 Haploids Obtained from the Gene Replacement Experiment with the pJM1223 Knock-Out Plasmid and Utilizing a 6 kb HindIII PG Digest as the Probe

Lanes 1-6 represent the results of Southern blot of the 6 uracil prototrophic YM 14 haploids and lanes 7-8 represent the Southern analysis of two more of such haploids obtained from the same experiment. The positive control (YM 11) and the wild type YM 14 are in lanes 9 and 10 respectively.
depicted by the DAY4 disrupted at the GLY1 locus is that formate could no longer bring about a slow growth in the absence of serine

**N2: No true GLY1 knock-out mutants of ser1 YM 14 were obtained using mating strategy; however, a diploid auxotrophic for serine and prototrophic for uracil was obtained during these mating experiments**

YM 14 and YM 11 were mated. The resultants diploids were selected and sporulated on the sporulation media with leucine and histidine supplements. The spores were germinated and plated onto selective media. Two selective media were routinely used for this purpose: 1- SC minus uracil; 2- SC minus serine. No true GLY1 knock-out haploids were obtained using this technique; however, a number of cells were obtained which were serine auxotrophic and uracil prototrophic. These mutants were subjected to Southern analysis, with the results shown in figure 30. As can be seen, these mutants show two bands in each lane which correspond to the bands shown by the negative control (YM 14) and positive control (YM 11) loaded in the last two lanes respectively. These results indicate that these cells are in fact diploids which have a copy of their GLY1 gene disrupted and the other copy intact, which is why two bands corresponding to the parental bands appear in their lanes.
Fig. 30 Southern Blot Analysis of the Uracil Prototrophic YM 14 cells Obtained by Mating Strategy Using a 6 kb Hind III PG Digest as the Probe

Lanes 1-12 (excluding lane# 8) represent the Southern analysis of the uracil prototrophic YM 14 cells which were proved to be some sorts of strange YM 14 serine auxotrophic diploids in which a copy of the GLY1 gene is intact and the other copy is disrupted. Lanes 13 and 14 represent the positive and negative controls respectively.
O- Tetrad analysis of the ser1/ser1;gly1::URA3/GLY1 YM 14 diploids shows that the GLY1 disruption in the ser1 background of the above strain might be lethal.

The above diploids were subjected to a number of tetrad analyses in which, from four spores of each ascus, 2 died and 2 survived (Figure 31 shows the results of one of these tetrad analyses where four asci were dissected and two spores out of four from each ascus survived). The survivors were all uracil auxotrophs, showing that no uracil prototrophs had survived. The results of one of these tetrad analysis is depicted in Table 11. Five asci (20 spores) were dissected and seeded at the known coordinates. The survivors were taken on different selective plates to check for different phenotypes. As can be seen from this Figure, all the survivors grew onto the 5-FOA plates which selects for uracil auxotrophs (the parental YM 14, grew on this medium) and none grew on the SC-URA. These results indicate that not even one true GLY1 knock-out haploid had survived thereby, pointing to a lethality hypothesis.
This figure illustrates the results of dissection studies of four asci and the follow up on the survival of the resultant haploids obtained from each ascus.

1- Ascus #1 where 2 spores out of four survived
2- Ascus #2 where 2 spores out of four survived
3- Ascus #3 where 2 spores out of four survived
4- Ascus #4 where 2 spores out of four survived
<table>
<thead>
<tr>
<th>CELLS</th>
<th>SC-URA</th>
<th>5-FOA</th>
<th>SC-SER</th>
<th>SC-HIS</th>
<th>SC-TRP</th>
<th>SPORU</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROGENY #1</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #2</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #4</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #5</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #6</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #7</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #8</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>PROGENY #9</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>YM11</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>NS</td>
</tr>
<tr>
<td>YM14</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>DIPLOIDS</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>S</td>
</tr>
</tbody>
</table>

**TABLE 11**

GROWTH OF THE HAPLOID PROGENY AFTER TETRAD DISSECTION ON DIFFERENT SELECTIVE PLATES

**SYMBOLES:**

- **SC-URA**: SYNTHETIC COMPLETE MINUS URACIL
- **5-FOA**: 5-FLUORO-OROTIC ACID MEDIUM
- **SC-SER**: SYNTHETIC COMPLETE MINUS SERINE
- **SC-HIS**: SYNTHETIC COMPLETE MINUS HISTIDINE
- **SC-TRP**: SYNTHETIC COMPLETE MINUS TRYPTOPHAN
- **SPORU**: SPORULATION PLATES
- **NS**: NO SPORES OBSERVED
- **S**: SPORES OBSERVED
- **(+)**: GROWTH OBSERVED
- **(-)**: NO GROWTH OBSERVED
P- The GLY1 gene lacks the Gcn4p consensus sequence

The regulation of genes encoding amino acid biosynthetic enzymes in budding yeast involves a number of control mechanisms. Most of these genes are subject to a cross-pathway regulatory system known as general amino acid control that increases their expression under conditions of amino acid starvation. Gcn4p is the direct positive regulator of gene expression in this system. Starvation for any one of 11 amino acids (histidine, arginine, lysine, isoleucine, leucine, serine, phenylalanine, valine, tryptophan, methionine and proline) will cause enzyme derepression mediated by the general control system.

A short nucleotide sequence located upstream of the start sites of transcription of the amino acid biosynthetic genes mediates their transcriptional activation by Gcn4p in response to amino acid starvation. This sequence, approximately 12 base pairs in length and containing the highly conserved hexanucleotide core 5' -TGACTC- 3' is the binding site for Gcn4p. The consensus sequence proposed for Gcn4p binding site is 5' -RRTGACTCATTT 3', where R designates a purine nucleotide.

The GLY1 sequence was searched using Vector NTI program for the following consensus sequences:

1. GCN→ 5' -TGACTC- 3'
2. GCN1→ 5' -ATGACTCCT- 3'
3. GCN2→ 5' -ATGACCCTT- 3'
4. GCN3→ 5' -ATGAGTCTT- 3'
5. GCN4→5' -ATGACTCAT- 3'

6. GCN5→5' -ATGACGTCA- 3'

Figure 32 represents the results of this search. There were no such sites on the 5' non-coding region of GLY1 gene, whereas a number of the above sites were detected by this program on the 5' non-coding region of the SHM2 gene, which is involved in serine biosynthesis.
FIG. 32 Vector NTI Search of the GLY1 and SHM2 Genes for the Gcn4p Consensus Sequences. In this figure, the GCN consensus sequences are as follows:

- **GCN**: 5'-TGACTC-3'
- **GCN1**: 5'-ATGACTCCT-3'
- **GCN2**: 5'-ATGACCGCTT-3'
- **GCN3**: 5'-ATGACGTCA-3'

No such sequences were found in the GLY1 gene, whereas the above sequences were detected in the SHM2 gene sequence.
Q- Hydrophobicity analysis of the GLY1 protein

The GLY1 protein sequence was put into a number of computer programs to characterize the identity and the site of its localization.

GLY1 Protein sequence was put into a program named “TMPRED” at the url location of: http://ulrec3.unil.ch/cgi-bin/TMPRED-form-parser. The results of this analysis can be seen in Appendix 2 where the program strongly suggests that the GLY1 protein has two transmembrane domains: (1) starting at the amino acid number 59 and ending at position 79 in an inside to outside orientation(2) starting at the position 98 and ending at the amino acid number 118 in an outside to inside orientation. According to the results of this analysis, GLY1 protein fits the description of the complex multitopic proteins where both N-terminal and the C-terminal ends of the protein come to rest in the cytoplasm.

The amino acid sequence of the GLY1 protein was also put through another program named “PSORT” at the url location of: http://psort.nibb.ac.jp/cgi-bin/okumur. The results of this analysis can be seen in Appendix 1 where the results suggest that the GLY1 protein has a cytoplasmic distribution. The amino acid sequences of the yeast GCV1, SHM1, and SHM2 were also put through the PSORT (as controls to check the accuracy of the predictions by the program), the results of which can be found in Appendices 3, 4, 5 and 6. This program correctly assigned the GCV1 as mitochondrial, SHM1 as mitochondrial and SHM2 as cytoplasmic proteins. However, Gly1 protein does not seem to have either the N-glycosylation sites (CHO-Asp-x-Ser/Thr,
where “x” can be any residue except proline (Bause, 1983) or O-glycosylation sites (-Gly-x-Hyl-x-Arg- where x is any amino acid and Hyl is hydroxy lysine) (Jentoft, 1990) on its sequence which makes it unlikely that this protein is a glycoprotein.
Discussion and Conclusions

The GLY1 gene in *S.cerevisiae* codes for the GLY1 protein (387 amino acids long and with molecular weight of ~ 42 kDa) which functions in glycine biosynthesis.

The characterization of the GLY1 gene product was started off by subcloning the coding region of this gene in the pET 15b expression vector, purifying the protein product and finally using this product for different biochemical assays.

The GLY1 gene was successfully subcloned into the pET 15b, which allowed the GLY1 protein to be expressed with a N-terminal hexahistidine tag that permitted the easy purification of this protein using IMAC. However, before the GLY1 protein could be used for any biochemical assays, the functionality of this protein (glycine biogenesis) had to be confirmed. The PG construct was expressed in glycine auxotrophic BL 21 *E.coli* strains and complemented the glycine requirement of these cells rather weakly (after 3 days of incubation), in comparison to the positive control where glycine complementation occurred after 24 hrs of incubation in the above cells transformed with the PS construct. This rather weak glycine complementation by the PG construct left questionable the functionality of the GLY1 protein expressed from this construct. Despite this problem, the biochemical characterization of the GLY1 protein was continued by setting up a threonine aldolase assay, since it has been known that glycine can be made from threonine by a threonine aldolase. The threonine aldolase assay
used was, in fact, a coupled assay in which the acetaldehyde produced during the aldolytic cleavage of threonine was acted upon by the alcohol dehydrogenase to produce ethanol and NAD$^+$, then the reduction in extinction coefficient at 340 nm was linked to the aldolytic activity of this protein. No threonine aldolytic activity could be shown for the GLY1 protein or the SHMTc which is known to have such an activity. The failure of this assay to show threonine aldolase activity, at least for the positive control was due to the experimental set up of this assay. This assay had to be performed under anoxic condition (under N$_2$) due to the volatile nature of the reaction products (acetaldehyde). However, due the lack of proper equipment, this assay was performed under room condition with open test tubes. That was why it was not possible to show threonine aldolase activity for the positive control (SHMTc expressed as a N-terminal HIS tagged protein similar to that of the GLY1 protein) even though its activity was certain, due to its prompt glycine complementation in the glycine auxotrophic BL 21 E.coli mutants. It was also not possible to show threonine aldolase activity for the GLY1 protein due to the same shortcomings of the experimental set up.

The results of direct threonine aldolase assay done by Bognar and coworkers indicate, however, that the GLY1 protein is, in fact, threonine aldolase since threonine aldolase activity is lost where the GLY1 gene gets disrupted, and is significantly elevated where the GLY1 gene gets overexpressed from a shuttle vector in a gly1 background.
A genetic approach to GLY1 characterization was based on the fact that this protein might be part of the glyoxylate anaplerotic pathway (which is glucose repressible), thereby furnishing the yeast cells with glycine. In the pursuit of this approach, the ser1 yeast strain, YM 14, was selected and studied. The major source of serine in yeast comes from glycolytic intermediates. In the above yeast strain, this pathway of serine genesis is blocked due to a point mutation in one of the key enzymes (SER1, phosphoserine aminotransferase) which makes the growth of this strain on yeast minimal medium with glucose as a sole carbon source and without serine impossible. However, this yeast strain is able to grow on non-fermentable carbon sources without serine supplements, implying the viability of the glyoxylate shunt pathway. Consequently, it was reasoned that if the GLY1 gene were part of this pathway, the overexpression of this gene would enhance the growth rate on non-fermentable carbon sources dramatically. The GLY1 gene was overexpressed in the ser1 yeast strain, YM 14, on fermentable and non-fermentable containing minimal media. The overexpression of this gene on glucose minimal medium without the serine supplement relieved the serine auxotrophy of this strain, whereas its overexpression on the other medium failed to complement that auxotrophy. These results suggest that this gene can not be part of the glyoxylate shunt.

The possibility of up-regulation and down-regulation of the GLY1 gene expression by threonine and glycine respectively was tested genetically in a yeast setting. These experiments took advantage of the yeast shuttle vector
pJM1308 where the \textit{GLY1} gene (promoter and part of the 5' coding region) was fused to a reporter gene (lac \textit{Z}). If threonine is the substrate for the \textit{GLY1} gene product, so a yeast strain (YM 22) transformed with the above vector could show at least several fold increase of \(\beta\)-galactosidase activity when grown in the presence of threonine than in its absence. The \(\beta\)-galactosidase activity in the pJM1308-transformed YM 22 was \(1.5 \times\) more in the presence of threonine than in its absence. However, this increase was not large enough to be interpreted as a positive result in order to draw any conclusion on the threonine induced up-regulation of the \textit{GLY1} gene expression and threonine being its substrate. The \(\beta\)-galactosidase readings of the above experiment in the presence and absence of glycine are also of no help to conclude that the presence of end-product (glycine) has any negative feed back effect on the \textit{GLY1} gene expression since the \(\beta\)-galactosidase reading with glycine supplement was only slightly lower than that of the same reading with no glycine supplement.

To characterize the possible aminotransferase activity of the \textit{GLY1} protein, a genetic approach was used in a yeast setting. This approach was based on the reasoning that if the \textit{GLY1} gene is part of the glyoxylate shunt, so its disruption would annul the ability of the \textit{ser1} yeast strain to grow on non-fermentable carbon sources without serine.

The disruption of the \textit{GLY1} gene was attempted using gene replacement and mating strategies using the \textit{ser1} YM 14 (\textit{ser1 his3 leu2 ura3}) strain. It was not possible to obtain any true \textit{GLY1} knock-out mutants. However, Dr. Appling
and coworkers were able to knock out the GLY1 gene in the ser1 yeast strain DAY4 (ser1 ura3 trp1 leu2 his4), where the only observable phenotype was the failure of formate supplement to confer a slow growth upon the DAY4 yeast strain in the absence of serine. The only explanation that can be offered at this point on the discrepancy of the GLY1 gene disruption results using YM 14 and DAY 4 is the fact that the ser1 mutation in DAY 4 might be leaky allowing the GLY1 gene disruption without any lethality outcome to be achieved.

The failure to obtain GLY1 knock-outs pointed to the possibility that such a disruption in the ser1 background of the YM 14 strain might be lethal. This notion was put into test by performing tetrad analysis on ser1/ser1&GLY1gly1::URA3 YM 14 diploids. The results of tetrad analysis suggested that the lethality theory might be correct since only two out of four spores survived and all survivors were uracil auxotrophs.

The GLY1 protein was also used to stimulate laboratory animals to raise antibody against this protein. However, no such Ab could be raised which further hampered achieving one of goals of this project; ie. finding the location of this protein using immunofluorescence microscopy. To compensate for these shortcomings at this front, another approach was utilized in order to pinpoint the localization site of the GLY1 protein in yeast cells. In this approach, the GLY1 amino acid sequence was put through a number of computer programs searching for protein motifs such as transmembrane domains and signal sequences. The results of these protein motif searches suggested that the
GLY1 protein is a complex membrane protein (where its N-terminal and C-terminal ends come to lie in the cytosol) and it has a cytoplasmic distribution.
Future Directions

1- Raising Antibody against the GLY1 Protein.

GLY1 protein can be purified in the form of inclusion bodies and be used with fresh and proper adjuvants to stimulate Ab production in mice. However, if all attempts at raising antibodies against the native GLY1 protein failed, the choice of synthetic peptides (corresponding to the most hydrophilic regions of the GLY1 protein) could also be considered since such peptides have been used with a great degree of success in raising cross reacting antibodies.

The GLY1 protein sequence was put into the MacVector analysis program and the results shown in appendix 7 were obtained. According to these results, the GLY1 aa sequence has a 18-residue long hydrophilic region (ie., polar amino acids extending from residue 231 to 248) which is the best candidate for synthesis of synthetic peptides. Once the synthetic peptide is made, it can be conjugated to a carrier molecule such as keyhole limpet haemocyanin (KLH) for antibody production.
2- A More Sensitive Threonine Aldolase Assay

Since the results of the threonine aldolase assay of the GLY1 protein done during the course of this work were inconclusive, a more sensitive threonine aldolase assay would definitely be the order of the day. This assay should utilize the yeast extract where the GLY1 gene has been overexpressed from the Yep24-GLY1 shuttle vector. The final reading of this assay should be based on detection of glycine production using an amino acid analyzer.

3- The Lethality Hypothesis

It was shown during the course of this work that disruption of the GLY1 gene in the ser1 background of YM 14 was lethal. This lethality could be confirmed by a simple experiment where the ser1/ser1; GLY1/gly::URA3 diploids are to be transformed with a GLY1 expression shuttle vector and a tetrad dissection to be performed. The survival of all 4 spores of each ascus would be a confirming result since it would show that the expression of the GLY1 gene has salvaged the spores that would die as it happened when the diploids with no GLY1 expression vector were sporulated and dissected.

If this lethality scenario is correct, then the GLY1 gene will be inserted into a yeast vector in which its expression is under the control of GAL1 promoter. This allows the gene to be expressed when the yeast host is grown on galactose but not when it is grown on glucose. This plasmid will be used to transform a
ser1 yeast strain. The chromosomal GLY1 gene will then be disrupted under conditions in which the cells are grown on galactose. This resultant strain will be conditional lethal when grown on glucose. This strain could be used to characterize the ser1/gly1 mutants and to screen other genes involved in this pathways.

Three separate screens will be used to find such genes:

1) Screen for mutations which abolish the lethal phenotype.

The yeast will be mutagenized with ethyl methane sulfonate (EMS) and switched to glucose medium. Cells which grow under these conditions may have acquired a mutation which overcomes the lethality of ser1/gly1 combinational mutations. This could be analyzed with a lethal gene trap which looks for recessive mutations. Mutant strains in which ser1 gly1 is no longer lethal would be transformed with the GLY1 ADE2 plasmid followed by transformation with a gene library. Those which lost the lethal phenotype through recessive mutations would again be lethal as judged by their inability to segregate the GLY1 plasmid and produce red colonies. If the mutation is dominant, a gene library could be constructed from the mutant strain and used to transform the conditional lethal strains. The responsible gene could be cloned from transformants that can grow on glucose.

2 A library of the yeast genome in the shuttle vector Yep 13 would be used to transform the conditional lethal ser1/gly1 strain to select for genes that could be
overcome the lethal phenotype through multicopy expression. The surpressing gene(s) would then be subcloned.

3: The GAL1 dependent GLY1 construct would be replaced with a construct containing the GLY1 gene on a plasmid which also contained the ADE2 gene. The whole strain would be gly1/ade2 with the GLY1 and ADE2 genes expressed from a plasmid. The yeast would be mutagenized and grown without selection for the plasmid auxotrophic markers. Most yeast would be able to segregate the plasmid and produce red ade2 derivatives. Mutants that remained white, however, would have a mutation that is lethal in a gly1 background (synthetic lethal phenotype). The SER1 gene should be identified by this screen and other genes also involved in this pathway could be detected.

4- Transmembrane Protein Identity Confirmation of the GLY1 Protein

It is predicted in appendix 2 that the GLY1 protein has 2 transmembrane domains in its primary structure. The amino acid sequences of the transmembrane domains can be substituted with structurally similar amino acids without loss of any enzymatic activity.
APPENDIX 1

Drop-out mix:

Drop-out mix is a combination of the following ingredients minus the appropriate supplement. It should be mixed very thoroughly by turning end-over-end for at least 15 min; adding a couple of clean marbles helps.

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine</td>
<td>0.5g</td>
<td>Leucine</td>
<td>10.0g</td>
</tr>
<tr>
<td>Alanine</td>
<td>2.0g</td>
<td>Lysine</td>
<td>2.0g</td>
</tr>
<tr>
<td>Arginine</td>
<td>2.0g</td>
<td>Methionine</td>
<td>2.0g</td>
</tr>
<tr>
<td>Asparagine</td>
<td>2.0g</td>
<td>para-Aminobenzoic acid</td>
<td>0.2g</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>2.0g</td>
<td>Phenylalanine</td>
<td>2.0g</td>
</tr>
<tr>
<td>Cysteine</td>
<td>2.0g</td>
<td>Proline</td>
<td>2.0g</td>
</tr>
<tr>
<td>Glutamine</td>
<td>2.0g</td>
<td>Serine</td>
<td>2.0g</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>2.0g</td>
<td>Threonine</td>
<td>2.0g</td>
</tr>
<tr>
<td>Glycine</td>
<td>2.0g</td>
<td>Tryptophan</td>
<td>2.0g</td>
</tr>
<tr>
<td>Histidine</td>
<td>2.0g</td>
<td>Tyrosine</td>
<td>2.0g</td>
</tr>
<tr>
<td>Inositol</td>
<td>2.0g</td>
<td>Uracil</td>
<td>2.0g</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>2.0g</td>
<td>Valine</td>
<td>2.0g</td>
</tr>
</tbody>
</table>
Appendix 2

Computer Prediction for Possible Transmembrane Helices in GLY1 Protein

Sequence: MTE...YKY length: 387
Prediction parameters: TM-helix length between 17 and 33

1.) Possible transmembrane helices
The sequence positions in brackets denominate the core region. Only scores above 500 are considered significant.

Inside to outside helices: 3 found
from to score center
59 (59) 79 (79) 546 69
103 (103) 121 (121) 879 113
206 (206) 224 (224) 266 214

Outside to inside helices: 3 found
from to score center
59 (59) 79 (79) 377 69
98 (98) 118 (118) 604 108
207 (209) 230 (225) 717 217

2.) Table of correspondences
Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices. Helices shown in brackets are considered insignificant.
A "+" symbol indicates a preference of this orientation.
A "++" symbol indicates a strong preference of this orientation.

inside->outside | outside->inside
59- 79 (21) 546 + | ( 59- 79 (21) 377 )
103-121 (19) 879 ++ | 98-118 (21) 604
( 206-224 (19) 266 ) | 207-230 (24) 717 ++

3.) Suggested models for transmembrane topology
These suggestions are purely speculative and should be used with
EXTREME CAUTION since they are based on the assumption that all transmembrane helices have been found.
In most cases, the Correspondence Table shown above or the prediction plot that is also created should be used for the topology assignment of unknown proteins.

2 possible models considered, only significant TM-segments used

*** the models differ in the number of TM-helices ! ***

-----> STRONGLY preferred model: N-terminus inside
2 strong transmembrane helices, total score :  1150
# from  to length score orientation
1  59  79 (21)  546 i-o
2  98  118 (21)  604 o-i

-----> alternative model
1 strong transmembrane helices, total score :  604
# from  to length score orientation
1  98  118 (21)  604 o-i
Computer Program Prediction for the Localization Site of the GLY1 Protein

<TITLE> Nakai server </TITLE>
ORIGIN yeast
BEGIN
>GLY1 PROTEIN SEQUENCE

MTEFELPPKY ITAANDLRSID TFTTPTAEMM EAAEAASIGD AVYGEDVDTV
RLEQTVARMA GKEAGLFCVS GTLSNQIAIR THLMQPPYSI LCDYRAHYVT
HEAAGLIALS QAMVVPVVP SNGDYLLELI KSHYVPDDGD IHHGAPTRLS
LENTLHIGIVY PLEELVRIKA WCMENGLKLH CDGARIWNAA AQSGVPLKQY
GEIDISISIC LSKSMSGAPIG SVLGNLKFV KKHFRKQQQ GGGIRQSGMM
ARIALVIIINN DWSQGLYSH SLAHELAEYC EAKGIPLESAP ADTNVFNL
KAARMDPDLV VKKGLKYNVK LMGGVFSFHV QVTRDTEKV KLAISEAFDY
AKEHPFDCNG PTQIYRSEST EVDDGNAIR EIKTYKY

<H2>Result Information</H2>

PSORT --- Prediction of Protein Localization Sites
version 6.4(WWW)

GLY1 387 Residues

Species classification: 3

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
count: 1
Position of the most N-terminal TMS: 103 at i=1
<a href="http://psort.nibb.ac.jp/helpwww.html#mtop">MTOP: membrane topology (Hartmann et al.)</a>

I(middle): 110 Charge difference(C-N): -2.5
<a href="http://psort.nibb.ac.jp/helpwww.html#esig">McG: Examining signal sequence (McGeoch)</a>

Length of UR: 6
Peak Value of UR: 0.36
Net Charge of CR: -1
Discriminant Score: -14.51

<a href="http://psort.nibb.ac.jp/helpwww.html#esig">GvH: Examining signal sequence (von Heijne)</a>

Signal Score (-3.5): -4.22
Possible cleavage site: 44

>>> Seems to have no N-terminal signal seq.

Amino Acid Composition of Predicted Mature Form:
calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
<a href="http://psort.nibb.ac.jp/helpwww.html#ealom">ALOM: finding transmembrane regions (Klein et al.)</a>
count: 0 value: -1.70 threshold: -2.0
PERIPHERAL Likelihood = -1.70
modified ALOM score: -0.56
(2) or uncleavable?

Gavel: Examining the boundary of mitochondrial targeting seq.<a>

Uncleavable? Ipos set to: 12

Discrimination of mitochondrial target seq.:</a>

negative (-3.45)

*** Reasoning Step: 2

HDEL Count: 0

Checking apolar signal for intramitochondrial sorting<a>

Mitochondrial matrix? Score: 0.10

SKL motif (signal for peroxisomal protein):</a>

pos: -1(387), count: 0

Amino Acid Composition Tendency for Peroxisome: -1.63

Peroxisomal proteins? Status: negative

Amino acid composition tendency for vacuolar proteins</a>

Score: 0.62 Status: notcl

Checking the amount of Basic Residues (nucleus)

Checking the 4 residue pattern for Nuclear Targeting<a>

Checking the 7 residue pattern for Nuclear Targeting<a>

Checking the Robbins & Dingwall consensus (nucleus)</a>

Checking the RNA binding motif (nucleus or cytoplasm)

Nuclear Signal Status: negative (0.00)

Checking CaaX motif..</a>

Checking N-myristoylation..</a>

Checking CaaX motif..</a>

Cytoplasmic protein? Score1: 0.650

----- - Final Results ----

cytoplasm — Certainty= 0.650(Affirmative) < succ>

mitochondrial matrix space — Certainty= 0.100(Affirmative) < succ>

endoplasmic reticulum (membrane) — Certainty= 0.000(Not Clear) < succ>

endoplasmic reticulum (lumen) — Certainty= 0.000(Not Clear) < succ>

---- The End ----

APPENDIX 4

Computer Program Prediction for the Localization Site of the GCV1 Protein

<TITLE> Nakai server </TITLE>
ORIGIN yeast
BEGIN
>GCV1 PROTEIN SEQUENCE

MSIIKKIVFK RFNSTLKKTA LHDLHVSLLG TMVPYAGYSM PVLYKGQTHI
ESHNWTRRTNA GLFDVSMHLQ SKLSGPHSVK FLORVTPTDF NALPVGSGLL
SVLLNQPQVV VDDTIITKEN DENEFYIVTN AGCAERDTEF FHDELQNGST
LDCQWKIEG RSSLALQGPK AKDVKLEPILS KTAPGKDLKE LFFGQREFA
LKDGSVLQIA RGGYTGDGEF EISIANEKA VFAEQLLANP VMKPIGLAAR
DSLREAGMM LCYHELDES TIPEEAALNWW ISKSRRDLVD QKYWFNGYAK
IMDQLNNKTY SKVRVGFKYL KKGPAARNGV KIFLPDAETE VGLVTSGSAS
PTLNINIQG AYVQKGYHKK GTKLLVQVRN KFYPIELAKM PLVPThYKQ

PSORT --- Prediction of Protein Localization Sites
version 6.4(WWW)

GCV1 400 Residues

Species classification: 3

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
count: 0
<a href="http://psort.nibb.ac.jp/helpwww.html#esig">McG: Examining signal sequence (McGeoch)</a>
Length of UR: 5
Peak Value of UR: 0.61
Net Charge of CR: 4
Discriminant Score: -9.73
<a href="http://psort.nibb.ac.jp/helpwww.html#esig">GvH: Examining signal sequence (von Heijne)</a>
Signal Score (-3.5): -5.91
Possible cleavage site: 39

>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B
<a href="http://psort.nibb.ac.jp/helpwww.html#ealom">ALOM: finding transmembrane regions (Klein et al.)</a>
count: 0 value: 5.89 threshold: -2.0
PERIPHERAL Likelihood = 5.89
modified ALOM score: -2.08
<a href="http://psort.nibb.ac.jp/helpwww.html#mit">Gavel: Examining the boundary of mitochondrial targeting seq.</a>
**<R-2> motif at: 86**
QRVTPT

**Discrimination of mitochondrial target seq.:**
- positive (2.31)
- Rule: mitochondrial protein
- Rule: mitochondrial protein
- Rule: mitochondrial protein
- Rule: mitochondrial protein

***Reasoning Step: 2***

**HDEL Count: 1**

- Checking apolar signal for intramitochondrial sorting
  - (Gavel position 86) from: 26 to: 43 Score: 4.0
  - Mitochondrial matrix? Score: 0.45

- SKL motif (signal for peroxisomal protein)
  - pos: 71(400), count: 1 SKL
  - SKL score (peroxisome): 0.3
  - Amino Acid Composition Tendency for Peroxisome: 1.01
  - Peroxisomal proteins? Status: notc1
  - AAC score (peroxisome): 0.201

- Amino acid composition tendency for vacuolar proteins
  - Score: -0.73 Status: negative

- Checking the amount of Basic Residues (nucleus)

- Checking the 4 residue pattern for Nuclear Targeting
- Checking the 7 residue pattern for Nuclear Targeting

- Checking the Robbins & Dingwall consensus (nucleus)

- Checking the RNA binding motif (nucleus or cytoplasm)

- Nuclear Signal Status: negative (0.00)

- Checking CaaX motif
- Checking N-myristoylation
- Checking CaaX motif

--- **Final Results** ---

- mitochondrial matrix space --- Certainty= 0.633(Affirmative) < succ
- microbody (peroxisome) --- Certainty= 0.441(Affirmative) < succ
- mitochondrial inner membrane --- Certainty= 0.331(Affirmative) < succ
- mitochondrial intermembrane space --- Certainty= 0.331(Affirmative) < succ

--- The End ---
APPENDIX 5

Computer Program Prediction for the Localization Site of the SHMTm Protein

<TITLE> Nakai server </TITLE>
ORIGIN yeast
BEGIN
>SHMTm PROTEIN SEQUENCE

MFPRASALAK CMATVHRGL LTSGAQLSLS KPVSEGDPEM FDILQGQERHR
QKHSITLIPS ENFTSKAVMD LSGSELQNYF SEGYPGERRY GNNEIDKSE
SLCOARALEL YGLDPKWXGV NVQPLSGAPA NLVYVSAIMN VGERLMGLDL
PDGGHLSHGY QLKSSTPISF ISKYGQSMPY HVDTTGLID YDNLQLVLA
FRPKVIVAGT SAYSRLIDYA RFKEISQGCAY ALMSDMAHISGLVAANVP
SPFHEHSDIVT TTHKSLRGP RGAMIFRKG IKSVTKIUGE IPYELEKKIN
FSVFPGHQGPPHNTITAMAVALQAMSPEFKEYQQKIDVNSKWFAQEILT
KMGYKLVSGGGTDLHIVLDLGTQVDGARVETILSALNIAANKNTIFPDK
SALFPGLRI GTPAMITRGFGREEFSQVAKYIDSASVLAENLKTLEPTTK
LDARSRLNEF KKLNSESVEAALSGEISKWSVGGYPVPGDII

<H2>Result Information</H2>

PSORT -- Prediction of Protein Localization Sites
version 6.4(WWW)

SHMTm 490 Residues

Species classification: 3

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
count: 0
<a href="http://psort.nibb.ac.jp/helpwww.html#esig">McG: Examining signal sequence (McGeoch)</a>
Length of UR: 5
Peak Value of UR: 0.45
Net Charge of CR: 2
Discriminant Score: -12.20

<a href="http://psort.nibb.ac.jp/helpwww.html#esig">GvH: Examining signal sequence (von Heijne)</a>
Signal Score (-3.5): -2.27
Possible cleavage site: 27

>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B

<a href="http://psort.nibb.ac.jp/helpwww.html#ealom">ALOM: finding transmembrane regions (Klein et al.)</a>
count: 0 value: 2.76 threshold: -2.0
PERIPHERAL Likelihood = 2.76
modified ALOM score: -1.45

- Gavel: Examining the boundary of mitochondrial targeting seq.</a>
  - R-2 motif at: 20
  - RRGLLT

Discrimination of mitochondrial target seq.:</a> positive (2.69)
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein
Rule: mitochondrial protein

*** Reasoning Step: 2

HDEL Count: 0

Checking apolar signal for intramitochondrial sorting</a> (Gavel position 20) from: 5 to: 9 Score: 3.5
Mitochondrial matrix? Score: 0.49

SKL motif (signal for peroxisomal protein):
  - pos: 455(490), count: 2 SRL
  - SKL score (peroxisome): 0.3
Amino Acid Composition Tendency for Peroxisome: 0.14
Peroxisomal proteins? Status: notclr
  - AAC score (peroxisome): 0.085

Checking the amount of Basic Residues (nucleus)

Checking the 4 residue pattern for Nuclear Targeting</a>
Checking the 7 residue pattern for Nuclear Targeting</a>
Checking the Robbins & Dingwall consensus (nucleus)</a>

Checking the RNA binding motif (nucleus or cytoplasm)
  - RNP motif found: pos. 417 (6) RGFGREEF

Nuclear Signal Status: negative (0.07)

Checking CaaX motif..</a>
Checking N-myristoylation..</a>

----- Final Results -----

mitochondrial matrix space --- Certainty= 0.681(Affirmative) < succ>
mitochondrial inner membrane --- Certainty= 0.369(Affirmative) < succ>
mitochondrial intermembrane space --- Certainty= 0.369(Affirmative) < succ>
mitochondrial outer membrane --- Certainty= 0.369(Affirmative) < succ>

---- The End ----
Computer Program Prediction for the Localization Site of SHMTc Protein

<TITLE> Nakai server </TITLE>
ORIGIN yeast
BEGIN
>SHMTc PROTEIN SEQUENCE

MPYTLSDAHH KLITSHLVDT DPEVDSIIKD EIERQKHSID LIASENFTST
SVFDALGTPL SNKYSEGYPG ARYYGGNEHI DRMEILCQQR ALKAFLHTPD
KWGVNVTQLS GSPANLQVYQ AIMKPHERLM GLYPDGGLH SHYTATENRK
ISAVSTYFES FPYRVNPETG IIDDYTLEKN AILYRPKVLV AGTSAVCRKI
DYKRMREID KGCAYLMVDM AHSGLIAAG VIPSPFEYAD IVTTTTHKSL
RGPRGAMIFF RRVGVRSPNK TGKEVLYDLE NPIIFSVPFG HQGGPHHNTI
AALATALKQA ATPEFKEYQT QVLLKNAKALE SEFKNLYLRL VSNGTDSHVM
LVSLREKGVG GARVEYICEK INIALNKNSI PGSALSALPG GVRIGAPAMT
TRGMEEDDFR RIVSYNKAV EFAQQVQQKL PKDACRLKDF KAKVDEGSVD
LTWKKKIEYD WAGEYPLAV

PSORT --- Prediction of Protein Localization Sites
version 6.4(WWW)

SHMTc 469 Residues

Species classification: 3

*** Reasoning Step: 1

Preliminary Calculation of ALOM (threshold: 0.5)
  count: 1
  Position of the most N-terminal TMS: 216 at i=1
<a href="http://psort.nibb.ac.jp/helpwww.html#mtop">MTOP: membrane topology (Hartmann et al.)</a>
  |(middle): 223 Charge difference(C-N): 1.0
<a href="http://psort.nibb.ac.jp/helpwww.html#esig">McG: Examining signal sequence (McGeoch)</a>
  Length of UR: 4
  Peak Value of UR: 1.01
  Net Charge of CR: 2
  Discriminant Score: -8.57
<a href="http://psort.nibb.ac.jp/helpwww.html#esig">GvH: Examining signal sequence (von Heijne)</a>
  Signal Score (-3.5): -4.45
  Possible cleavage site: 20

>>> Seems to have no N-terminal signal seq.
Amino Acid Composition of Predicted Mature Form:
calculated from 1
ALOM new cnt: 0 ** thrshld changed to -2
Cleavable signal was detected in ALOM?: 0B

\(<a href="http://psort.nibb.ac.jp/helpwww.html#alom">ALOM: finding transmembrane regions (Klein et al.)</a>\>
- count: 0 value: -1.86 threshold: -2.0
- PERIPHERAL Likelihood = -1.86
- modified ALOM score: -0.53

(6) or uncleavable?

\(<a href="http://psort.nibb.ac.jp/helpwww.html#mit">Gavel: Examining the boundary of mitochondrial targeting seq.</a>\>
- <R0> motif at: 6
- Uncleavable? lpos set to: 16

\(<a href="http://psort.nibb.ac.jp/helpwww.html#mit">Discrimination of mitochondrial target seq.</a>\>
- negative (-3.35)

--- Reasoning Step: 2 ---

HDEL Count: 0

\(<a href="http://psort.nibb.ac.jp/helpwww.html#mit">Checking apolar signal for intramitochondrial sorting</a>\>
- Mitochondrial matrix? Score: 0.10

\(<a href="http://psort.nibb.ac.jp/helpwww.html#pox">SKL motif (signal for peroxisomal protein)</a>\>
- pos: 15(469), count: 1
- SKL score (peroxisome): 0.3

Amino Acid Composition Tendency for Peroxisome: -0.57

Peroxisomal proteins? Status: not clr

\(<a href="http://psort.nibb.ac.jp/helpwww.html#lys">Amino acid composition tendency for vacuolar proteins</a>\>
- Score: 0.16 Status: not clr

Checking the amount of Basic Residues (nucleus)

\(<a href="http://psort.nibb.ac.jp/helpwww.html#nuc">Checking the 4 residue pattern for Nuclear Targeting</a>\>

\(<a href="http://psort.nibb.ac.jp/helpwww.html#nuc">Checking the 7 residue pattern for Nuclear Targeting</a>\>
- Checking the Robbins & Dingwall consensus (nucleus)

Checking the RNA binding motif (nucleus or cytoplasm)

Nuclear Signal Status: negative (0.00)

\(<a href="http://psort.nibb.ac.jp/helpwww.html#anc">Checking CaaX motif</a>\>

Cytoplasmic protein? Score1: 0.450

--- Final Results ---

| Cytoplasm — Certainty = 0.450 (Affirmative) < succ |
| Microbody (peroxisome) — Certainty = 0.300 (Affirmative) < succ |
| Mitochondrial matrix space — Certainty = 0.100 (Affirmative) < succ |
| Endoplasmic reticulum (membrane) — Certainty = 0.000 (Not Clear) < succ |

--- The End ---
REFERENCES


Schirch, V. 1984. in Folates and Pterins: Chemistry and Biochemistry of Folates. 1: 399-432


Szoka, P.R., Schreiber, A.B., Chan, H. & Murphy, J. 1986. DNA 5:11-20.


