MOLECULAR AND BIOCHEMICAL ASPECTS OF GLUCOSE / XYLOSE ISOMERASE FROM *STREPTOMYCES* sp.

THESIS
SUBMITTED TO THE
UNIVERSITY OF PUNE
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
CHEMISTRY (BIOCHEMISTRY)

BY
SNEHALATA ANIL PAWAR

DIVISION OF BIOCHEMICAL SCIENCES
NATIONAL CHEMICAL LABORATORY
PUNE 411 008, INDIA

MARCH 1999

..... TO

MY PARENTS

"I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the sea-shore and diverting myself in now and then finding a smooth pebble or a prettier shell than ordinary whilst the great ocean of truth lay all undiscovered before me."

..... ***Sir Isaac Newton***
quoted in Brewser's Memoirs of Newton

ACKNOWLEDGEMENTS

It is a pleasure to express my sincere thanks to my guide Dr. (Mrs.) Vasanti Deshpande for her kind encouragement, for her insight and perspective of the problem. I express my deep sense of gratitude for her forbearing nature, patience and openmindedness to new ideas, throughout the course of this work.

I express my gratitude towards Dr. Mala Rao for invaluable discussions and timely advice during the course of this work.

I take this oppurtunity to thank in particular Dr. Mohini Ghatge for her collaboration during this work and Dr.V.Chauthaiwale for helpful discussions.

I am highly indebted to Prof. P. Balaram, MBU, Indian Institute of Science, Bangalore for permitting the use of CD facility, Dr. K. N. Ganesh, Head, Organic Synthesis Division, for the use of spectrofluorimeter and Dr. Sourav Pal for the timely help rendered during the making of this thesis.

I thank my labmates Kavita, Aparna, Jui, Neeta, CV and Sudeep for the cordial atmosphere in the lab.

I also thank Devyani and Dr. Mohini for the constant support, help and lighter moments shared together.

I take this oppurtunity to acknowledge Tripathi for his help during the course of this work. Thanks are also due to Narsimha for helpful discussions and to Chandrakumar for his help in the final stages of the thesis.

I thank all my friends at the hostel, especially Meena,Tripathi,Vivek, Thomas, Robert, Ganga, Neelima and Bhavana for the memorable times spent together. I am also thankful to Rangarajan, Neelam, Rajashri, Sami, Venkat, Veena, Amutha, Shamim, Madhura, Vaishali J., Anu, Vaishali D. and Sarita for the funtimes shared together and the incidental help during the course of this work.

I am thankful to our office staff, Usha and Indira for the help rendered during the course of this work.

I also express my thanks to our efficient instrumentation staff, Mr. M .N. Kamthe, Mr. V. Karanjkar for the help rendered on several occassions. I am thankful to our lab attendant Mr. R. Lambhate for his help in routine chores.

I wish to thank my friends Anupama, Snehal and Sandhya for their constant support and encouragement throughout the course of this work.

I fall short of words to express my sincere thanks to my friends Urmila and Ajitha who have made a profound impact on my outlook towards life and science.

I take this oppurtunity to thank my parents-in law for their loving support in this endeavour.

Words fail me, to express my indebtedness to my husband, Anil for his undemanding, unstinting support and tremendous patience during the course of this work. He has always been a source of inspiration, motivation and perseverance to me.

I owe my deepest gratitude to my parents and brother Nitin, without whose enduring support and encouragement, it would not have been possible to embark upon this journey in life.

I am thankful to Dr. P.K. Ranjekar, Head, Division of Biochemical Sciences for allowing me to use the facilities of the Division.

I am thankful to Dr. Paul Ratnasamy, Director, NCL, for granting permission to submit this work in the form of thesis.

The fellowship awarded by the Council of Scientific and Industrial Research is duly acknowledged.

Snehalata Pawar

TABLE OF CONTENTS

DECLARATION

This is to certify that the work incorporated in the thesis entitled " Molecular and Biochemical aspects of Glucose/Xylose Isomerase from *Streptomyces* sp." submitted by Snehalata A. Pawar was carried out under my supervision at Division of Biochemical Sciences, National Chemical Laboratory, Pune. Materials obtained from other sources has been duly acknowledged in the thesis.

Dr. Vasanti V. Deshpande

(Research Guide)

ABBREVIATIONS

ANS	8-Anilino-napthalene-1-sulfonic acid
BSA	Bovine serum albumin
GI	Glucose isomerase
GXI	Glucose/xylose isomerase
GdnHCl	Guanidine hydrochloride
CD	Circular dichroism
CRE	Catabolite repressor element
CR	Catabolite repression
IPTG	Isopropyl thiogalactoside
PEG	Polyethylene glycol
PNK	T4 Polynucleotide kinase
xyl A	Gene encoding xylose isomerase
xyl B	Gene encoding xylulokinase
xyl R	Gene encoding regulatory molecule
xyl T	Gene encoding xylose permease

SYNOPSIS

Glucose/xylose isomerase (E.C. 5.3.1.5) (GXI) catalyzes the conversion of D-xylose and D-glucose to D-xylulose and D-fructose respectively. The enzyme plays an important role in the industrial production of high fructose corn syrup (HFCS) which is used as a sweetener in the food industry. Interconversion of xylose to xylulose by GXI serves a nutritional requirement in saprophytic bacteria and has a potential application in the bioconversion of xylan-containing agricultural biomass to ethanol (Parrish, 1970; Katz et al, 1972).

Streptomyces species are aerobic, Gram positive saprophytic bacteria which typically colonize habitats such as soil. Commercial production of organic acids, antibiotics and industrially important enzymes such as glucose isomerase, xylanase, protease etc. by *Streptomyces* fermentation and the potential for improving the economics of these processes have stimulated interest in the application of novel genetic techniques to *Streptomyces* species. Recombinant DNA technology offers a means to isolate and manipulate the GXI gene to alter the thermostability and substrate specificity to suit its application in HFCS production. The present studies deal with some of the molecular and biochemical aspects of the glucose/xylose isomerase from *Streptomyces* sp. NCIM 2730, an isolate from the National Chemical Laboratory, Pune, India.

The main features of the work are as follows :

1. Molecular cloning, expression and sequence analysis of glucose/xylose isomerase gene
2. Synthesis of recombinant glucose / xylose isomerase : Effect of coordinated additon of specific amino acids
3. Characterization of acid-induced unfolding intermediates of glucose / xylose isomerase

1. Molecular cloning, expression and sequence analysis of Glucose/xylose isomerase

A partial genomic library of *Streptomyces* sp. NCIM 2730, constructed in pUC8 vector was screened for the presence of GXI gene using an 18-mer mixed oligonucleotide probe (5'-GTCACGTTCCACGACGAC3') complementary to a highly conserved six-amino acid sequence of GXI from *Actinomycetes*. Eight clones which hybridized with the radiolabelled oligoprobe showed the ability to complement xylose isomerase-defective *E.coli* mutants. The restriction map of the insert from one (pMSG27) of the eight GXI-positive

clones showing detectable GXI activity, was constructed. The cloned gene has unique restriction sites for *Bgl* II and two restriction sites for *Pvu* II, *Eco*R I and *Sma* I. GXI-deficient strains of *E.coli* were able to utilize xylose as the sole carbon source for their growth upon transformation with pMSG27. *E. coli* JM105 (pMSG27) and *E. coli* JC1553 (pMSG27) were inducible by IPTG suggesting that the expression of the cloned gene was under the control of the *lac* Z promoter. Western blot analysis revealed that, the cloned gene was expressed as a fusion protein of M_r 110. The present chapter deals with the screening of genomic library, restriction mapping of the GXI encoding DNA fragment, characterization of the recombinant gene product and DNA sequence analysis. The partial sequence of the 1.1 kb insert revealed a 50% homology with reported GXIs from bacterial sources.

2. Synthesis of recombinant Glucose/xylose isomerase : Effect of coordinated additon of amino acids

The amplified expression of a recombinant protein is known to lead to an intracellular depletion of specific amino acid pools which in turn may affect the production of the desired protein. In order to counteract and overcome such a situation during the fermentation of the recombinant *E.coli* JM109 (pMSG27) containing the glucose/xylose isomerase (GXI) gene from *Streptomyces* sp. NCIM 2730, the efffect of addition of different amino acids on the specific activity of GXI was studied. The amino acid composition of GXI from *Streptomyces* sp. NCIM 2730 reveals predominance of aspartic acid, glutamic acid and glycine. Therefore, the effect of coordinated addition of the assorted combinations of these three amino acids on the synthesis of recombinant GXI was studied and the results were analyzed using a 2^3 factorial design.

The following conclusions were derived from the analysis of two factor interactions of the three amino acids (a) The interaction between the aspartic and glutamic acid is independent of aspartic acid concentration but is affected by the increasing concentrations of glutamic acid, (b) The effect of aspartic acid concentration is more than that of glycine and (c) During the interaction of glutamic acid and glycine, the effect of glutamic acid is more prominent than that of glycine. The three factor interaction analyses reveal that the effect of the three amino acids is in the order aspartic acid > glutamic acid > glycine.

3. **Characterization of acid-induced unfolding intermediates of Glucose/xylose isomerase.**

Previous studies that have contributed towards our understanding of the protein folding problem have mainly dealt with monomeric proteins. Similar studies on oligomeric proteins are few and provide an insight regarding the tertiary and quaternary interactions. Acid-induced unfolding of the tetrameric glucose/xylose isomerase from *Streptomyces* sp. NCIM 2730 has been investigated using fluorescence quenching, second derivative spectroscopy, hydrophobic dye-binding (ANS) and circular dichroism techniques. It has been shown that GXI undergoes a two-step unfolding transition when the pH is lowered from 7.5 to 2. Based on its increased hydrophobic exposure, increased accessibility to quenching by acrylamide and native-like secondary structure, the intermediate at pH 3.2 was defined as a molten globule state of GXI. The second intermediate at pH 5.0 is suggested to be an intermediate on the N→MG pathway, characterised by its pronounced secondary structure and ANS binding.

LIST OF PUBLICATIONS/ PRESENTATIONS

1 **Bhosale, S. H.** , Rao , M. B. and Deshpande V. V. (1995) Thermostability of high activity alkaline protease from *Conidiobolus coronatus* (NCL 86.8.20) Enz. Microb.Technol.**17,**136-139.

2 **Bhosale, S. H.**, Ghatge, M. S. and Deshpande V. V. (1996) Molecular cloning and expression of the glucose/xylose isomerase gene from *Streptomyces* sp. NCIM 2730 in *Escherichia coli*. FEMS Microbiol. Lett.**145,** 95-100.

3. **Bhosale, S. H.** , Rao , M. B. and Deshpande V. V. (1996) Molecular and industrial aspects of glucose isomerase. Microbiol. Rev. **60,** 280-300.

4. Paul, A, **Bhosale, S. H.**, Maity, T. K. and Deshpande V. V. (1998) Effect of co-ordinated addition of specific amino acids on the synthesis of recombinant glucose isomerase. Enz. Microb.Technol. **23,** 506-510

5. **Pawar, S. A.** and Deshpande V. V. (1998) Abstract entitled "Characterization of acid-induced unfolded intermediates of glucose/xylose isomerase from *Streptomyces* sp. NCIM 2730 " accepted for the Annual Meeting of Society of Biological Chemists, India, held at New Delhi, during Dec17-21, 1998.

CHAPTER 1

General Introduction

GENERAL INTRODUCTION

D-Glucose isomerase (D-xylose ketol isomerase, EC 5.3.1.5) (GI), is an enzyme which catalyzes the reversible isomerization of D-glucose and D-xylose to D-fructose and D-xylulose, respectively (Fig.1.1). *In vivo*, the interconversion of xylose to xylulose by GI serves a nutritional requirement in saprophytic bacteria that thrive on decaying plant material and thus

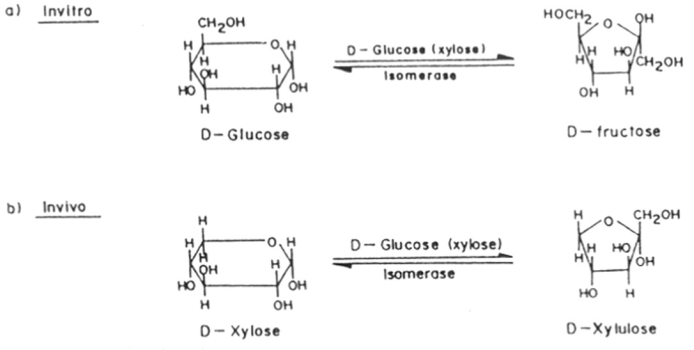


Fig.1.1 Reactions catalyzed by GI.

aids in the bioconversion of hemicellulose to ethanol (Parrish, 1970; Katz *et al*, 1972). Hence the enzyme is known as xylose isomerase (XI). *In vitro,* the enzyme isomerizes glucose to fructose, which finds application in the production of high fructose corn syrup (HFCS) and in commercial parlance is referred to as 'glucose isomerase'.

The origin of today's successful development of HFCS lies in the discovery of D-glucose isomerase. Sucrose derived from sugar beets (40%) and sugarcane (60%) was the main sweetener in the world until 1976. Increasing demands for refined sugar, coupled with its high cost of production and awareness of the adverse effects of sucrose and invert sugar consumption on human health, necessitated the search for acceptable sucrose substitutes. A large number of non-caloric and non-carbohydrate artificial sweeteners such as saccharine, cyclamate, acesulpham-K, aspartame and thaumatin have been discovered and dismissed on the basis of health concerns or other drawbacks. In contrast, high fructose corn syrup (HFCS) has gained wide acceptance as a sweetener in the food industry. HFCS, an equilibrium mixture of glucose and fructose (1:1) is 1.3 times sweeter than sucrose and 1.7 times sweeter

than glucose. HFCS is preferred by the food industry since it does not pose the problem of crystallization as is the case with sucrose. Moreover, D-fructose plays an important role as a diabetic sweetener because it is only slowly reabsorbed by the stomach and does not influence the blood glucose level.

ENZYMATIC *VERSUS* CHEMICAL ISOMERIZATION

Chemical isomerization: The chemical conversion of glucose to fructose constitutes one of a group of reactions collectively known as the Lobry de Bruyn-Alberda van Ekenstein transformation, and is carried out at high temperature and pH. The chemical conversion is non-specific and leads to the formation of non-metabolizable sugars such as psicose and other non-desirable colored products. Moreover, chemically produced fructose possesses off-flavours and reduced sweetness, which cannot be easily remedied.

Enzymatic isomerization: The enzymatic isomerization is carried out under ambient conditions of pH and temperature. This process is highly specific and does not yield side-products and is, therefore, preferred to the chemical isomerization process.

GI is one of the three highest tonnage value enzymes, amylase and protease being the other two. Research on the fundamental as well as applied aspects of GI has received impetus, because of its industrial importance in the production of HFCS and its potential application in the conversion of biomass to ethanol. Several organisms used in the commercial production of HFCS are enlisted in Table 1.1.

Table 1.1 Commercial producers of GI

Organism	Trade name	Manufacturer
Actinoplanes missouriensis	Maxazyme	Gist Brocades & Anheuser-Busch Inc.
Arthrobacter sp.	_	Reynolds Tobacco
Bacillus coagulans	Sweetzyme	Novo-Nordisk
Streptomyces rubiginosus	Optisweet	Miles-Kali-Chemie
	Spezyme	Finnsugar
S. phaeochromogenes	Sweetase	Nagase
S. olivaceus	_	Miles Laboratories Inc

SOURCES OF GI

The presence of GI has been reported in various bacteria (Table 1.2). The enzyme is mainly produced intracellularly, in the exponential phase and is induced by xylose. The existence of extracellular GI has been reported from *Streptomyces glaucescens* (Weber, 1975), *S. flavogriseus* (Chen *et al*, 1979), *Chainia* sp. (Srinivasan *et al*, 1983; Vartak *et al*, 1984) and *Bacillus* sp. (Chauthaiwale & Rao, 1994). Generally, GI is present in bacteria but the occurrence of the enzyme in the yeasts, *Candida utilis* (Wang *et al*, 1980), *Candida boidinii* (Vongsuvanlert *et al*, 1988) and the fungi *Malbranchea pulchella* (Bannerjee *et al*, 1994) and *Neurospora crassa* (Rawat *et al*, 1996) has also been reported. Apart from microbial sources, the existence of GI in barley malt (Bartfay, 1960) and wheat germ (Pubols *et al*, 1963) has been documented.

Table 1.2 GI producing microorganisms

Microorganism	
Actinomyces	*Microellobosporia*
Actinoplanes	*Micromonospora*
Aerobacter	*Mycobacterium*
Ampullariella	*Nocardia*
Arthrobacter	*Pseudonocardia*
Bacillus	*Pseudomonas*
Bifidobacterium	*Sarcina*
Brevibacterium	*Staphylococcus*
Chainia	*Streptococcus*
Clostridium	*Streptomyces*
Corynebacterium	*Streptosporangium*
Escherichia	*Thermopolyspora*
Flavobacterium	*Thermus*
Lactobacillus	*Thermoanaerobacterium*
Leuconostoc	*Xanthomonas*
Microbispora	*Zymomonas mobilis*

IMMOBILIZATION OF GI

GI being an intracellular enzyme requires large amounts of substrate to compensate for the high K_m for glucose and hence a high cost of production. Immobilization of GI offers an excellent oppurtunity for its effective reuse and makes the process for its use cost-effective. A plethora of literature on the immobilization of GI is available. The methods used for

Table 1.3 Immobilized GIs of commercial importance

Microorganism	Trade name	Manufacturer	Immobilization method
Cell-free enzyme			
S. olivochromogenes	G-zyme G-994	CPC (Enzyme biosystems)	Adsorption on anion exchange resin
S. rubiginosus	Spezyme	Genencor International	DEAE-cellulose agglomerated with polystyrene & TiO_2
S. rubiginosus	Optisweet II	Solvay	Adsorption of SiO_2 particles crosslinked with glutaraldehyde
S. olivochromogenes	Ketomax 100	UOP	polyethyleneimine treated alumina with glutaraldehyde cross-linked GI
Whole cells			
A. missouriensis	Maxazyme	IBIS	Cells occluded in gelatin followed by glutaraldehyde
F. arborescens	Takasweet	Solvay	Polyamine glutaraldehyde cross-linked cells extruded & granulated
S. griseofuscus	AGIS-600	Godo-Shusei	Chitosan-treated glutaraldehyde cross-linked cells
S. phaeochromogenes	Sweetase	Nagase	Heat treated cells bound to anion exchange resin
S. murinus & B.coagulans	SweetzymeT	Novo-Nordisk	Glutaraldehyde cross-linked cells extruded

immobilization are namely, whole cell immobilization and cell-free enzyme immobilization. Some of the commercially available immobilized glucose isomerases are described in Table 1.3

PURIFICATION AND PROPERTIES OF GI

Purification of GI

The intracellular GI is extracted from microbial cells by mechanical disruption (such as sonication, grinding or homogenization) or by lysis of the cells using lysozyme, cationic detergents or toluene (Chen, 1980a). Purification of the enzyme has been reported using classical purification methods such as heat treatment, precipitation with ammonium sulphate/ acetone/ Mg^{2+}or Mn^{2+}salts, ion-exchange chromatography, and /or gel filtration (Chen, 1980b).

Properties of GI

The knowledge of specific properties of the enzyme such as its stability, substrate specificity and metal ion requirement, is important to prevent its inactivation and to assess its suitability for the production of HFCS.

Substrate specificity

The enzyme has been reported to utilize D-ribose, L-arabinose, L-rhamnose, D-allose and 2-deoxy-glucose besides its most common substrates *viz.*, D-glucose and D-xylose. Maximum isomerization was obtained with the substrates having hydroxyl groups at carbon 3 and 4 in the equatorial position as in glucose and xylose. The K_m values of the enzyme for D-glucose and D-xylose were in the range of 0.086 to 0.920 M, and 0.005 to 0.093 M, respectively (Chen, 1980b).

Metal ion requirement and inhibitors

The requirement of divalent cation such as Mg^{2+}, Co^{2+} or Mn^{2+} or a combination of these cations for the activity of GI has been reported by several workers. Mg^{2+} is required for catalytic activity, while Co^{2+} is essential for stabilization of the quaternary structure of the enzyme (Callens *et al*, 1988a; Gaikwad *et al*, 1992a). The catalytic activity of GI has been reported to be inhibited by xylitol, arabitol, sorbitol, mannitol, lyxose and Tris and also by the metal ions Ag^{+}, Hg^{2+}, Cu^{2+}, Zn^{2+}, Ni^{2+} and Ca^{2+} (Bucke, 1983; Smith *et al*, 1991).

Subunit structure

The sedimentation constants and molecular weights of GI vary from 7.55 to 11.45 and from 52 to 191 *kDa*, respectively (Chen, 1980b). The subunit structure and amino acid composition of GI revealed that it is a tetramer or a dimer of similar or identical subunits associated by non-covalent bonds, devoid of interchain disulphide bonds. The extracellular GI from *Bacillus* sp. has been reported to be a trimer (Chauthaiwale & Rao, 1994). Basuki *et al* (1992) have reported the existence of isoenzymes of GI from *S. phaeochromogenes*. The isoenzymes differ in their N-terminal amino acids and in the peptide patterns of the digests with trypsin, *Achromobacter* protease I and cyanogen bromide. Each of the isoenzymes was a tetramer of non-identical subunits.

Optimum temperature and pH

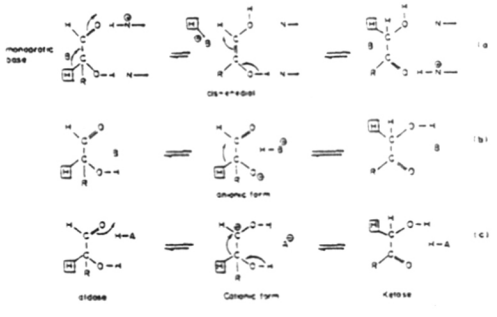
Maximum GI activity has been reported in the temperature range 60 - 80°C, which increased in the presence of Co^{2+}. In contrast to the GI from *Lactobacillus* (Yamanka, 1968) and *Escherichia* (Chen, 1980b), the enzyme from *Streptomyces* (Suekane *et al*, 1975, Basuki *et al*, 1992), *A. missouriensis* (Scallet *et al*, 1974) and *Thermus thermohydrosulfuricum* (Dekker *et al*, 1991a) is stable at high temperatures. The enzyme from various sources exhibited an optimum pH in the range 7.0 to 9.0, except in the case of *Lactobacillus brevis*, wherein a pH optimum of 6.5 was observed (Yamanka, 1968). This is desirable for the commercial application of GI.

MECHANISM OF ACTION OF GI

There is a considerable interest in understanding the catalytic mechanism of GI and has been delineated by chemical modification, X-ray crystallography and isotope exchange studies.

Fig.1.2 Mechanism of action of GI .

a, *cis*-Enediol; b, Proton transfer; c, Hydride shift.



Chemical modification of GI

Chemical modification of amino acid residues using specific chemical reagents serves as a simple means of probing the active site of enzyme. The possible involvement of His in the active site of GI has been shown by studying the effect of the modifier, diethylpyrocarbonate (DEPC) on the activity of the enzyme (Kume & Takahisa, 1983, Gaikwad *et al*, 1988; Vangrysperre *et al*, 1988). His is known to function as a proton abstracting base and assist in hydrogen transfers (Fig 1.2b). Presence of aspartate or glutamate residue in GI was documented by its inactivation by Woodward's Reagent K or guanidine hydrochloride (Vangrysperre *et al*, 1989; Ghatge and Deshpande, 1993). The carboxylate residues have been implicated in the binding of metal ion cofactors (Callens *et al*, 1988b). Chemical modification of protected and unprotected GI and subsequent peptide mapping allowed the identification of an active site region with a consensus sequence consisting of Phe-His-Xaa-Asp-Xaa-Xaa-Pro-Xaa-Gly (Vangrysperre *et al*, 1990). Results of studies on the chemical modification of GI complement the conclusions drawn on the basis of X-ray crystallographic studies described below.

X-ray crystallography

X-ray crystallography yields a detailed account of the three-dimensional structure of the protein and allows actual visualization of the complex between the enzyme and its substrate or inhibitor. GI from different bacterial species has been studied by X-ray crystallography at different levels of resolution, in presence and absence of inhibitors and metal ions, in order to understand and explain the mechanism of action. These studies revealed a similarity in their structures especially at the active site. The structure of GI from *S. rubiginosus* as determined at 4°A (Carell *et al*, 1984) resolution has shown that the enzyme consists of eight α-strands β-helix (α/β) units as found in triose phosphate isomerase. The crystal structure of XI from *S. olivochromogenes* was consistent with a tetrameric structure of the enzyme and also revealed the symmetry of the subunit arrangement. Several internal salt linkages that stabilize the tertiary and quaternary structure of the enzyme were detected by Henrick *et al* (1989). The structure of GI from *S. rubiginosus* in presence of substrate and an active site directed inhibitor at 1.9°A resolution led to the identification of the active site region and two metal-binding sites. The metal ions are important in maintaining the structure of the active site region and one of them binds to C3—O and C5—O of the substrate forming a

six-membered ring. The studies also revealed that His acts as an active site base that abstracts a proton fron the substrate while the water molecule acts as a proton donor or acceptor in the enolization of a ring opened substrate. Analysis of the structure of GI, GI-substrate complex and alkylated GI indicated that the mechanism involves an "open chain" conformation of the substrate and that the intermediate in the isomerization reaction is probably a *cis* enediol (Carrell *et al*, 1989, Blow *et al*, 1990,1992). Studies on X-ray crystallographic structures of metal activated GI from *S. olivochromogenes* indicated that the two metal cofactors bridged through a glutamate residue, promote a hydride shift. Substrate-induced structural changes, occuring prior to the isomerization step were revealed by the differential position of Mg-2 in the enzyme and enzyme-substrate complex. Carrell *et al* (1994) have shown that GI from *S. rubiginosus* binds substrate in a variety of modes depending on the size of the sugar. D-threonohydroxamic acid (THA) a potent inhibitor of GI resembles a putative transition state in the isomerization step of xylose by the enzyme. X-ray crystallographic structure of the GI-THA complex provided evidence for the metal movement during catalysis on deprotonation, followed by the formation of a bridging ligand (Allen *et al*, 1995). These results confirmed the earlier observation that the protonation of the -OH group occurs after ring opening (Allen *et al*, 1994b).

Isotopic exchange

The available crystallographic data for GI rules out a proton-transfer mechanism and suggests a hydride-shift mechanism. However, only structural data is insufficient to conclude the mechanism of action of an enzyme. Uncertainty about a proton-transfer mechanism in GI was prompted by the absence of solvent exchange during investigations on the incorporation of tritiated water into the product (Rose *et al*, 1969). However, the possibility of fast proton transfer in a shielded activity could not be ruled out. Allen *et al* (1994a) have carried out isotope-exchange experiments at high temperature, extreme pH and in presence of Gdn.HCl to investigate the possibility of shielded proton transfer. Their NMR studies coupled with the studies on fluorine-substituted substrate analogs do not support a proton-transfer mechanism for GI.

Recent studies of the wild-type and mutant D-xylose isomerase from *Actinoplanes missouriensis* support the crucial role of the water molecule, Wa-690, Asp-255, and the

adjacent Glu-186 in proton transfer from 2-OH to O-1 of the open and extended aldose substrate (Van Bastelaere *et al*, 1995).

GENETIC REGULATION OF BIOSYNTHESIS OF GI

D-xylose, a major component of plant hemicelluloses, is utilized by bacteria through a pathway involving transport across the cytoplasmic membrane, followed by isomerization to D-xylulose. This pentulose residue is phosphorylated by xylulokinase to yield D-xylulose-5-phosphate which is further metabolized *via* the pentose phosphate and Embden-Meyerhoff pathways (Fig. 1.3).

Genetic organization of *xyl* genes

Genetic mapping and cloning studies on *Salmonella typhimurium* have indicated that the genes responsible for xylose catabolism, are *xyl* T, a gene specifying the transport of xylose across the cell membrane; *xyl* A, the xylose isomerase gene; *xyl* B, the xylulokinase gene and *xyl* R, a regulatory element essential for transcription of *xyl* genes (Shamana & Sanderson, 1979 a, b; Ghangas and Wilson, 1984). These four clustered genes constitute the *xyl* operon (Fig. 1.3). Studies on *E. coli* genome revealed an analogous genetic organization (Malezska *et al*, 1982). Isolation of *E. coli* mutants for *xy l*A, *xyl* B, and *xyl* R (T) coupled with complementation data suggested the order of genes to be *xyl* R (T)-*xyl* A-*xyl* B (Rosenfeld *et al*, 1984). These results strongly suggested a repressor-operator mechanism for the regulation of *xyl* AB expression and postulate a model for coordinate (positive) control of the *xyl* A, *xyl* B and *xyl* T genes by the *xyl* R gene product (David and Weissmeyer, 1970; Rosenfeld *et al*, 1984). In the absence of xylose, the *xyl* R product acts as a repressor, while it acts as an activator in the presence of xylose which is analogous to the *ara* C gene product of the arabinose regulon (Ogden *et al*, 1980).

The regulation of the *xyl* operon in *Bacillus* was studied by integration of a *xyl-lac* Z fusant into the *amy* gene of *B. subtilis* 168 (Gartner *et al*, 1992). The increased expression of the *xyl-lac*Z fusant, when titrated with the *xyl* regulatory DNA in *trans*, suggested a negative regulation of the *xyl* operon in *B. subtilis* contrary to the positive control mechanism described for the *xyl* operon of *E. coli* and *S. typhimurium*. This suggested that regulation occurs at the level of transcription (Kraus *et al*, 1994). The outer 10bp of the *xyl* operator was shown to exhibit palindromic symmetry which was essential for binding the repressor. *Xyl* repressor

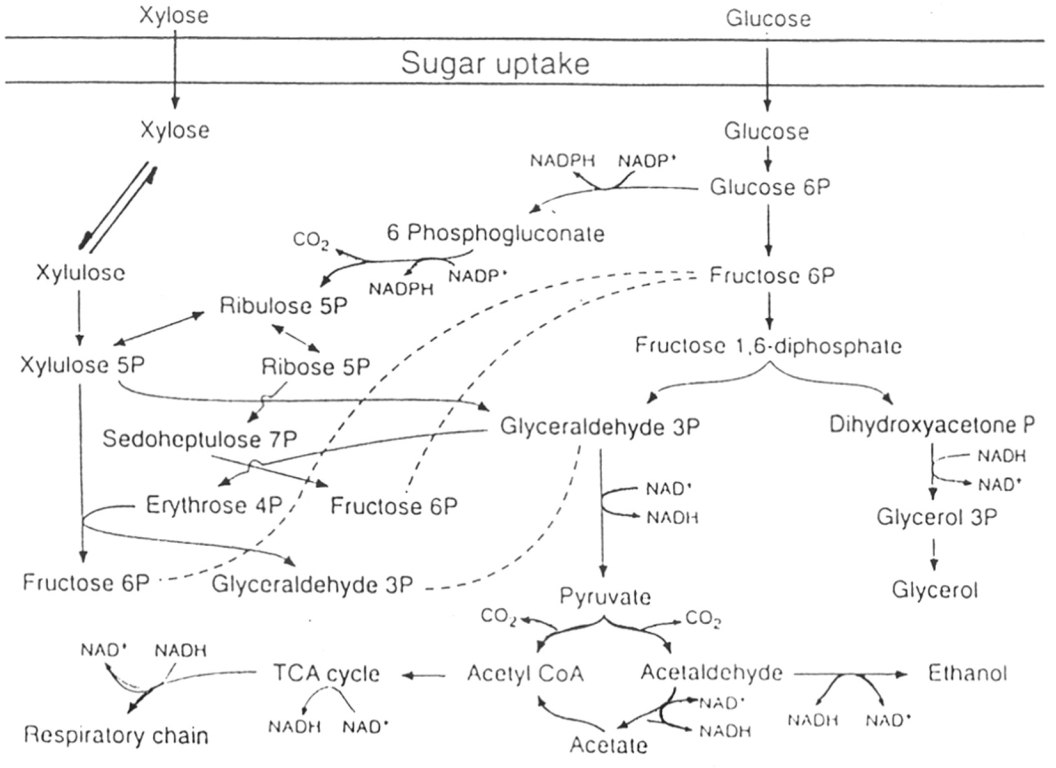


Fig.1.3 Pathway of xylose metabolism in bacteria.

binds to the palindromic *xyl* operator but not to a half site, suggesting that active repressor may be at least a dimer (Gartner *et al*, 1992). A computer analysis of the *xyl* R sequences from *B. subtilis* revealed that operator recognition may be mediated by an α-helix-turn-α-helix motif located between amino acids 29 and 48 of the primary structure (Dodd & Egan, 1990). Kauder *et al* (1993) have shown that an operator binding-negative mutation of *xyl* repressor from *B. subtilis* is *trans* dominant in *B. megaterium*. Arrangement of two operators is also shown to be present in *xyl* regulatory sequences of GI from *B. subtilis* (Dahl *et al*, 1994), *Staphylococcus xylosus* (Sizemore *et al*, 1991), and *Lactobacillus pentosus* (Lokman *et al*, 1991) is assumed to be responsible for efficient regulation in these bacteria.

The *xyl* genes from *S. xylosus* were cloned in *S. carnosus* by complementation to xylose utilization (Sizemore *et al*, 1991). Open reading frames of *xyl* R, *xyl* A and *xyl* B were located with the same polarity. Primer extension analysis revealed that the transcription of *xyl* R was constitutive while that of *xyl* A was xylose-inducible. The two transcriptional units *viz.* *xyl* R and *xyl* A were separated by a transcriptional terminator between the genes and the presence of promoter-like sequences was observed upstream of both transcriptional start sites. *xyl* A and *xyl* B are separated by only five nucleotides between the stop and start codons respectively. This observation together with the absence of terminator-like structures between *xyl* A and *xyl* B strongly suggests that they are co-transcribed. Co-translation of these two genes is also indicated by the presence of a potential Shine-Dalgarno sequence for *xyl* B (AAGGA) which overlaps with the last codons of *xyl* A. These results strongly support a repressor-operator mechanism for the regulation of *xyl* AB expression (Sizemore *et al*, 1993).

Divergent promoters

Studies on the *xyl* A encoding gene of *Streptomyces violaceoniger*, have indicated that the *xyl* A and *xyl* B genes promote transcription in opposite directions (Tiraby *et al*, 1989). The existence of divergent promoters in *Streptomyces* and other prokaryotes has been reported earlier (Beck and Warren, 1988). Sequence analysis has indicated the presence of a third reading frame, which encodes a regulatory protein. The two genes are seperated by a short region of 195 bp which revealed the presence of an element with palindromic symmetry typical of bacterial operators. It is suggested that a regulatory molecule may act within the divergent transcription unit to control the expression of opposite genes and also regulates its own synthesis. Studies by Wong *et al* (1991) have revealed that the genetic organization of

xyl A and *xyl* B genes in *S. rubiginosus* differs considerably from that of the other bacteria. The two genes are transcribed divergently from within a 114 bp sequence separating the two coding regions contrary to the earlier observation that *xyl* A and *xyl* B genes are part of an operon. The transcription initiation sites are 40 and 20 bp upstream of the translation initiation sites of *xyl* A and *xyl* B, respectively. The promoters of the genes share a 33 bp overlapping sequence in the untranscribed region between the two genes. The transcription of *xyl* genes in *S. rubiginosus* was induced by xylose and repressed by glucose. It was believed that the 114 bp nucleotide intergenic region provides the binding site (s) for the regulatory proteins.

Catabolite repression

The expression of the *xyl* operons in *S. typhimurium* and *E. coli* seems to be regulated by a positive control mechanism (Shamana & Sanderson, 1979b) and by catabolite repression exerted by glucose (David & Weissmeyer, 1970). In addition, a regulatory effect of the xylose isomerase itself has been described for *E. coli* (Batt *et al*, 1985).

In *E. coli*, catabolite repression is mediated via transcriptional activation by catabolite gene activator protein (CAP) and cyclic adenosine monophosphate (cAMP). In case of *B. subtilis* the *xyl* operon is negatively regulated by the *xyl* repressor and inducible by xylose (Kreuzer *et al*, 1989). The cAMP-cAMP receptor mediated mechanism as observed in *E. coli* is not functional in the case of *B. subtilis* as evidenced by observations such as (i) cAMP concentration was unaffected by the strength of catabolite repression, (ii) cAMP receptor protein was not detected in Gram positive bacteria, and (iii) catabolite repression in *B. subtilis* was negatively regulated at the transcriptional level (Kraus *et al*, 1994).

In conclusion, the organization of *xyl* A and *xyl* B seems to be highly conserved in all bacteria. These two genes are always adjacent to each other but a closer inspection reveals marked differences in their organization as shown in Fig.1.4. In the case of *B.* subtilis, the *xyl* R gene has a polarity opposite to that of *xyl* A gene, unlike that in *S. xylosus* and *Lactobacillus*. In *Streptomyces*, the *xyl* A and *xyl* B genes are transcribed divergently on different strands whereas in *E. coli*, *Lactobacillus*, and *Bacillus* they are transcribed from the same strand.

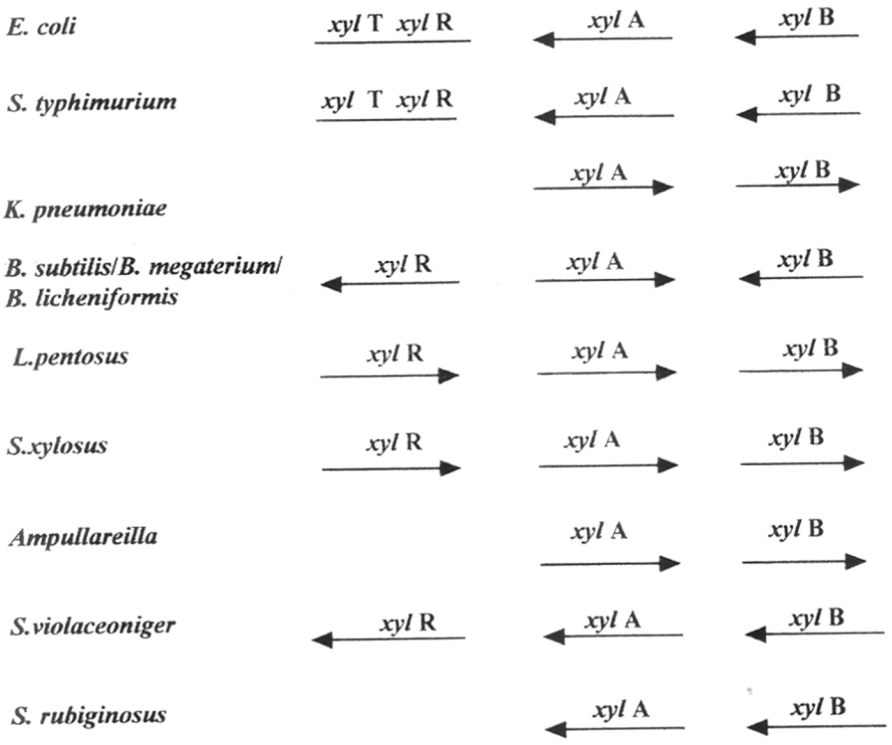


Fig. 1.4 Genetic organization of *xyl* operon in various microorganisms

The analysis of *xyl* genes from a variety of organisms will help to arrive at a consensus opinion about the genetic organization and regulation of *xyl* genes.

STRAIN IMPROVEMENT

Strain improvement is mainly focussed on improvement of yields of GI, constitutive production of the enzyme, elimination of requirement of Co^{2+} ions and the expensive inducer xylose. Several strategies involving conventional mutagenesis and recombinant DNA technology have been adopted for strain improvement to enhance the production of GI. Prior to the advent of recombinant DNA technology, the classical method for enhancing production of the desired product from an organism was to induce mutations and then select for organisms with superior traits. This process was time consuming, laborious and not cost effective, whereby very few traits could be manipulated. In contrast, recombinant DNA technology has

provided a powerful, efficient and rapid means for constructing organisms with specific genetic attributes.

Mutagenesis studies

Several strains of commercial importance were subjected to mutagenesis to produce elevated levels of enzyme or for constitutive production of enzyme. A 60% increase in the enzyme level was obtained by mutagenizing *S. wedmorensis* using ethyleneimine and N-methyl-N-nitro-N-nitrosoguanidine (Bengston & Lamm, 1973). UV-irradiation of *S. olivochromogenes* resulted in a mutant strain with 70% increased activity (Suekane & Iizuka, 1982). A constitutive mutant of *B. coagulans* with twice the activity of the parent was obtained by selecting the mutants on the basis of their resistance to 2-deoxy glucose. Two higher-yielding constitutive mutants showing highest yields on lactose, and one of them showing higher activity on glucose than xylose, have been reported by Lee (1977). A series of constitutive and high glucose isomerase-yielding mutants were isolated by applying multiple UV irradiations to *S. acidodurans* (Bok *et al*, 1984). The mutant SM 805 of *Streptomyces* sp. No.8 developed by NTG and UV treatment resulted in a strain that produced seven-fold more glucose isomerase activity than the wild type strain (Kim *et al*, 1992).

Recombinant DNA technology

An excellent means for isolating and manipulating the gene of a desired protein is possible with the advent of recombinant DNA technology. More than 50% of the industrial enzymes are now produced using genetically engineered microorganisms (Hodgson, 1994). One of the ways for increasing the production of GI is to identify the encoding gene and clone it on a multicopy vector containing strong promoters such as *lac*, *tac* or *PL* promoter. Molecular cloning and expression of GI has been carried out both in homologous as well as heterologous hosts.

Homologous hosts

Several reports describe the homologous cloning of GI from *E. coli* and *Streptomyces*. The cloning and expression of exogenous DNA in a homologous host is advantageous since it allows the easy recognition of the expression signals by the host RNA polymerase. The homologous cloning of GI has been carried out for several microorganisms and is described in Table 1.4

Table 1.4 Homologous expression of *xyl* A

Parent organism	Cloning host	vector	transformant	size of insert	Reference
E.coli	*E.coli*	pBR322	pXI	1.6 kb	Schellenberg *et al*, 1983
E.coli	*E.coli*	YEp 13	pLX7	2.5 kb	Ho & Stevis *et al*, 1983
E.coli	*E.coli*	pMB9	pUC1002	13.3 kb	Wovcha *et al*, 1983
E.coli	*E.coli* MM294	pBR322	pECX14	4.2 kb	Lawlis *et al*, 1984
E.coli	*E.coli*	pKO4	pLX3	1.6 kb	Rosenfeld *et al*, 1984
E.coli	*E.coli*	pKK223	pKKX7	2.5 kb	Batt *et al*, 1986
E.coli	*E.coli*	pSL-1	——	2.396kb	Lastick *et al*, 1986

Table 1.4 Homologous expression of *xyl* A (*contd.*)

Parent organism	Cloning host	vector	transformant	size of insert	Reference
Bacillus licheniformis	*B. subtilis* WH152	pMK4	pWH1451	4.8 kb	Scheler *et al*, 1991a
Bacillus megaterium	*B. subtilis* BR151	pMK4	pIW11	5.8 kb	Scheler *et al*, 1991b
Staphylococcus xylosus	*Staphylococcus carnosus* TM300	pCA43	pXyl12	6.6 kb	Sizemore *et al*, 1991
Streptomyces phaeochromogenes	*Streptomyces lividans* 532S107	pIJ702	pSP107	1.2 kb	Kho *et al*, 1984
Streptomyces violaceoniger	*S. violaceoniger*	pUT206	pUT220	4.0 kb	Marcel *et al*, 1987

E.coli

Ho *et al* (1983) first reported the isolation of GI gene from *E. coli.* Transformation of GI-deficient *E. coli* strain with a plasmid pMB9 bearing a *Hind* III restriction fragment of *E. coli* chromosomal DNA resulted in amplification of D-xylose isomerase and xylulokinase activities (Wovcha *et al*, 1983). The GI gene from *E. coli* was sequenced and was shown to code for GI by purification of the cloned gene product (Schellenberg *et al*, 1984). Briggs *et al* (1984), Lawlis *et al* (1984) and Ueng *et al* (1985) have reported the molecular cloning, sequencing and expression of the GI gene in *E. coli*. Overproduction of GI in *E. coli* has been achieved by several workers (Batt *et al*, 1986; Ho & Stevis, 1985). Ho and Stevis (1985) observed that the hyperexpression of the gene was not accomplished by merely cloning of the gene on a high-copy-number plasmid, since the expression of the gene in *E. coli* is highly regulated through its natural promoter. The fusion of the structural gene with strong promoters such as *lac* or *tac* resulted in a 20-fold overproduction of the enzyme. Ligation of a promoterless DNA fragment containing the *E. coli* gene into a plasmid downstream to a strong *PL* promoter, followed by the transformation of an *E. coli* strain containing a temperature-sensitive repressor, resulted in the overproduction of GI (Lastick *et al*, 1986). Cloning of *xyl* A gene under the control of the *tac* promoter produced GI which accounted for 28% of total cell-protein (Batt *et al*, 1986).

Streptomyces

The *Streptomyces* systems are attractive options for the expression of cloned genes because of their efficient secretion machinery (Gilbert *et al*, 1995).

Homologous cloning of GI from *S. phaeochromogenes* in *S. lividans via* the *Sst* I site of pIJ702 led to a 50 fold increase in GI activity of *S. lividans* which was 2.5 times that of the wild type (Kho, 1984). GI gene from *S. violaceoniger* was cloned using pUT 206 as a cloning vector. Subcloning of the cloned fragment permitted localization of the GI gene on a smaller fragment (2.1 kb) (Marcel *et al*, 1987). Nucleotide sequence and the deduced primary sequence of the gene were determined by Drocourt *et al* (1988). Integration of the *xyl* A gene of *S. violaceoniger* into the chromosome, downstream to the *Streptomyces* promoter (*P*1) using the integration vector pTS55 resulted in the CBS1 strain with 7-fold higher GI activity in the absence of the inducer xylose as compared to the wild type strain induced by xylose (Bejar *et al*, 1994).

Heterologous hosts

GI genes from different organisms have been cloned in *E. coli* which is the most popular host since several cloning vectors have been constructed for its use to meet various specific requirements. Identification of genes in *E. coli* allows their easy sequencing and manipulation by site-directed mutagenesis to produce tailor-made proteins. Table 1.5 enlists the *xyl* A genes of various organisms which have been cloned in *E. coli* and other hosts.

E. coli

A *Bam*H I restricted DNA fragment coding for GI from *B. subtilis* was isolated by the complementation of a xylose isomerase-defective *E. coli* strain. The expression of the gene was shown to be under the control of IS5 which is inserted at 195 bp upstream from the putative ATG initiation codon of the structural gene for GI (Wilhelm & Hollenberg, 1984). The ribosome binding sequence and two hexamer sequences typical of *Bacillus* promoter regions were located in the DNA fragment. *Eco*R I fragments of chromosomal DNA from *B. licheni-formis* were ligated to a vector plasmid pBR322 and used to transform GI-negative mutant of *E. coli* (Shin & Kho, 1985). Stable maintenance of the recombinant plasmid containing GI gene and 20-fold higher activity than the wild type host were achieved. GI gene from a thermophilic *Bacillus* sp. was cloned and expressed in *E. coli*. The GI produced by the recombinant was active at 85°C and was partially purified to yield 49.02 $U.mg^{-1}$ protein which represents the highest ever recorded specific activity for GI (Wuxiang & Jeyaseelan, 1993). Majority (about 94) of the bases in the third wobble position of the codons are G and C in case of GI from *B.subtilis* (Amore & Hollenberg, 1989a). The gene encoding thermostable GI in *Clostridium thermosulfurogenes* was cloned in *E. coli* using a new plate-assay method (Lee *et al*, 1990b). The expression of the protein in *E. coli* was higher (0.46 $U.mg^{-1}$) than that by the parent (0.19$U.mg^{-1}$) and was constitutive. Liu *et al* (1987) have cloned and expressed the D-xylose catabolizing enzymes of *Xanthomonas* in *E. coli*. The *xyl*A gene of several organisms which have been cloned in *E. coli* and other hosts is described in Table 1.6.5

Bacillus

A few reports on heterologous cloning using hosts other than *E. coli* are also available. In case of *E. coli*, there exists the problem of accumulation of recombinant gene product in the periplasm. The ability of *Bacillus* strains to secrete large quantities of protein directly into the

Table 1.5 Cloning and expression of *xyl* A

Parent organism	Cloning host	vector	Transformant	size of insert	Reference
Bacillus subtilis	*E. coli* DG75	YRp7	pIW1	5.8kb	Wilhelm & Hollenberg, 1984
Bacilllus licheniformis ATCC 31667	*E. coli* LE392-6	pBR322	pBGI6	4.1kb	Shin & Kho, 1985
Bacillus subtilis	*E. coli* SP10	pAT153	pTJ1	5.8kb	Jeyaseelan *et al*, 1990
Lactobacillus pentosus	*E. coli*	pUC19	pXH50A	5.0kb	Lokman *et al*, 1991
Lactobacillus brevis	*E. coli*	pUC19		1.9 kb	Bor *et al*, 1992
Klebsiella pneumoniae 1033 strain KAY2026	*E. coli* K12	pUC19	pZY220	6 kb	Feldmann *et al*, 1992
Clostridium thermosulfurogenes	*E. coli* *Bacillus subtilis*	pHSG262 pTB523	pCG138 pMLG1	4.0 kb	Lee *et al*, 1990
Clostridium thermohydrosulfuricum	*E. coli*	pUC119	____	2.2kb	Dekker *et al*, 1991

Table 1.5 Heterologous cloning and expression of *xyl* A (*contd.*)

Parent organism	Cloning host	vector	transformant	size of insert	Reference
Clostridium thermosaccharolyticum	*E.coli* DS941	pUC19	pJA1	4.0kb	Meaden *et al*, 1994
Ampullariella sp. strain 3876	*E.coli*	pUC13	____	2.922 kb	Saari *et al*, 1987
Arthrobacter strain N.R.R.L.3728	*E.coli*	pCG2100	pAXI1	4.8kb	Loviny-Anderton *et al*, 1991
Streptomyces griseofuscus	*E.coli* MN522	pUC13	____	5.7kb	Kikuchi *et al*, 1990
Streptomyces rubiginosus	*E.coli*	pBR322	pTW11	4.0 kb	Wong *et al*, 1991
Thermus thermophilus	*E.coli*	pUC118	pUS12XI	1.8 kb	Dekker *et al*, 1991
Thermotoga neopolitana 5086	*E.coli*	pUC18	pTNE2	2.4 kb	Vielle *et al*, 1995
Thermoanaerobacterium strain JW/SL YS 180	*E.coli*	pBluescript	____	1.317 kb	Liu *et al*, 1996
E.coli	*Zymomonas mobilis* CP4	____	pZB186	7.0 kb	Zhang et al, 1995

growth medium, the ease with which they can be grown and their well proven safety has made them the prime candidates for production of heterologous proteins (Harwood *et al*, 1992).

GI gene from *E. coli* was cloned in *Bacillus* using a bifunctional plasmid. However, the expression of the gene was not observed. Fusion of the *E. coli* structural gene downstream to the promoter of the penicillinase gene from *B. licheniformis* resulted in the functional expression of GI in *Bacillus* (Huang & Ho, 1985). GI gene from *C. thermosulfurogenes* was cloned in *B. subtilis* using an *E. coli-Bacillus* shuttle plasmid pMG1. The expresssion of GI gene in *Bacillus* was constitutive and was higher (1.54 $U.mg^{-1}$) than that produced in *C. thermo-sulfurogenes* (0.29 U.mg-1) (Lee *et al*, 1990b).

Yeast

A wide variety of microorganisms can utilize xylose but none of them can ferment it to ethanol. The main bottleneck lies in the conversion of xylose into xylulose which is usually performed under aerobic conditions as in *Candida utilis*. The pentose utilizing yeasts like *Pachysolen tannophilus* can ferment xylose anaerobically but the rate of fermentation is formidably low and is accompanied by considerable amounts of side products. *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* offer rapid fermentation rate, higher end product yield and increased ethanol tolerance. Transfer of GI genes to these yeasts holds a promise for developing an organism which can ferment xylose directly to ethanol. Some of the organisms from which GI has been introduced into yeast cells are *E. coli*, *Xanthomonas*, *B. subtilis*, *A. missouriensis*, *C. thermosulfurogenes*, *Thermus thermophilus* (Table 1.6).

Most of the attempts of cloning bacterial xylose isomerases in yeast have not been successful. This can be attributed to improper protein folding, post-translational modifications, inter and intra-molecular disulfide bridge formation and the internal pH of yeast which does not allow the assembly of the tetramer in the correct folding pattern.

Plants

The GI gene from *E. coli* has been cloned on a plasmid pBR322 derivative downstream of the nopaline synthetase gene (*nos*) promoter of an *Agrobacterium tumifaciens* plasmid pTiC58 and the construct was transformed into tobacco leaf disks. The transformants expressed GI in transgenic tobacco, thus indicating that the mRNA was successfully translated

Table 1.6 Cloning and expression of *xyl* A gene from bacteria in yeast

Parent organism	Host strain	Description	Reference
E. coli	*Schizosaccharomyces pombe*	Insertion of the XI gene from *E. coli* in the *Pst* I site of plasmid pDB248 when transformed in *S. pombe* resulted in low expression. The lowered expression has been attributed to the effect of yeast proteases and the inhibitory effect of xylitol.	Ueng *et al*, 1985 Chan *et al*, 1989
E. coli	*S. cerevisiae*	The XI gene of *E.coli* was introduced under the control of ADH1 promoter in the yeast expression vector pAAH5. This resulted in a recombinant *S. cerevisiae* expressing XI which was 10^{-3} fold less active as compared to the native enzyme.	Sarthy *et al*, 1987
B. subtilis	*S. cerevisiae*	Fusion of the *B. subtilis* XI gene to the *PDC*1 promoter of 2μm DNA plasmid containing the *leu*2-D gene, resulted in a transformant expressing a catalytically inactive XI, with a polypeptide size similar to the native enzyme.	Amore *et al*, 1989a
A. missouriensis	*S. cerevisiae*	The recombinant containing the *A. missouriensis* XI gene,cloned downstream to the *GAL*1 promoter in the *Hind* III/*Sal* I site of plasmid pBM272 did not result in its expression.	Amore *et al*, 1989b
C. thermosulforogenes	*S. cerevisiae*	The *xyl* A gene of *C. thermosulfurogenes* cloned in *S. cerevisiae* under the control of *ADH*2 promoter, resulted in transformants which were unable to utilize xylose. The presence of *xyl* A transcripts shown by Northern analysis, indicated inability of the yeast to translate the gene.	Moes *et al*, 1996
T. thermophilus	*S.cerevisiae*	The *T. thermophilus xyl* A gene cloned and expressed in *S. cerevisiae* under the control of *PGK*1 promoter, resulted in a recombinant XI showing highest activity at 85°C . The success in this attempt could be explained on the basis of relatedness between the two organisms.	Walfridsson *et al*, 1996

by the plant system (Piruzyan *et al*, 1989). Cloning of the GI gene from *E. coli* in potato (*Solanum tuberosum*) and in tomato (*Lycopersicum esculentum*) has been achieved and the presence of *xyl* gene has been confirmed by the expression of GI activity (Krashininnikova *et al*, 1991; Norova *et al*, 1991).

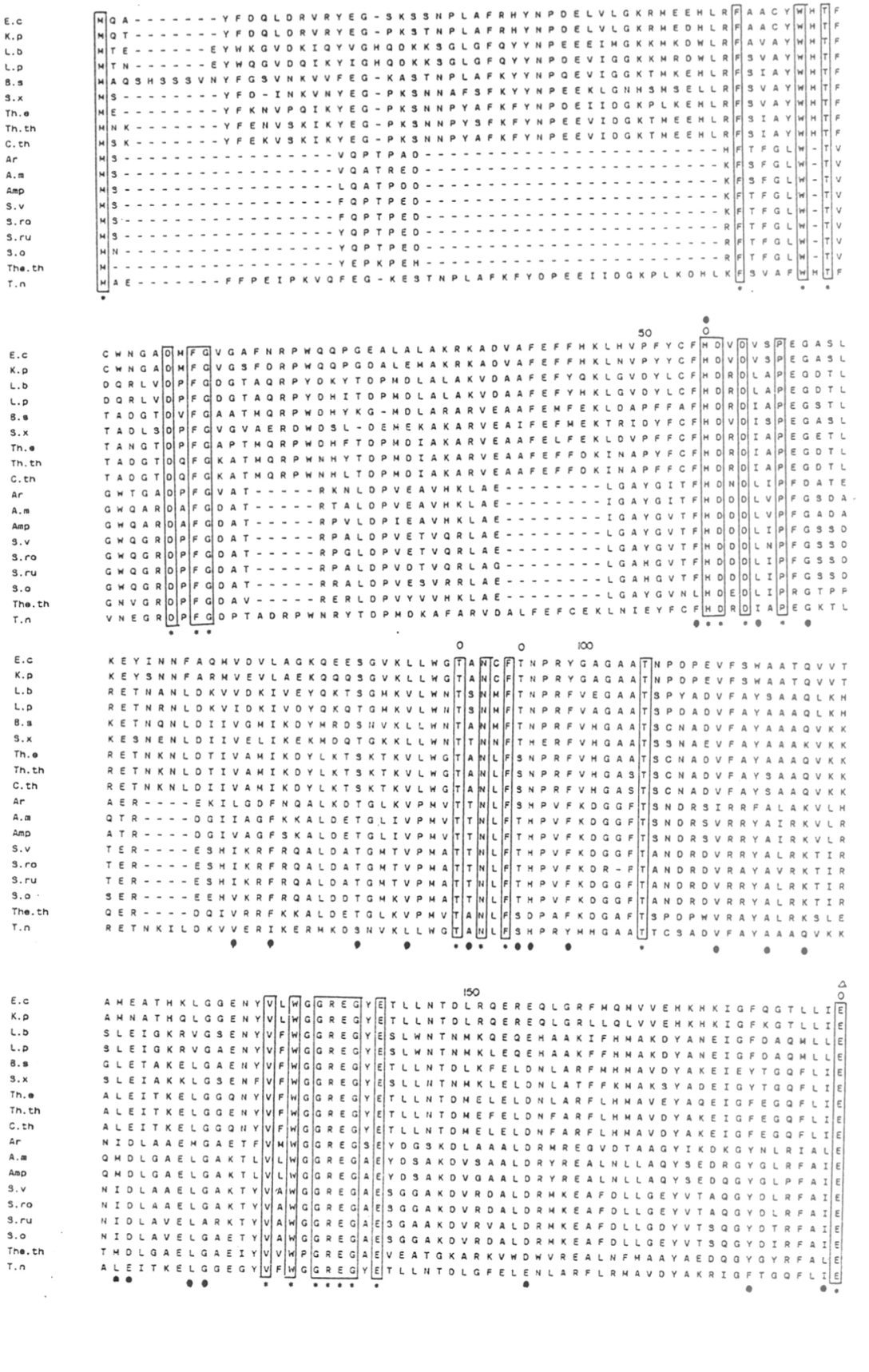
SEQUENCE HOMOLOGY

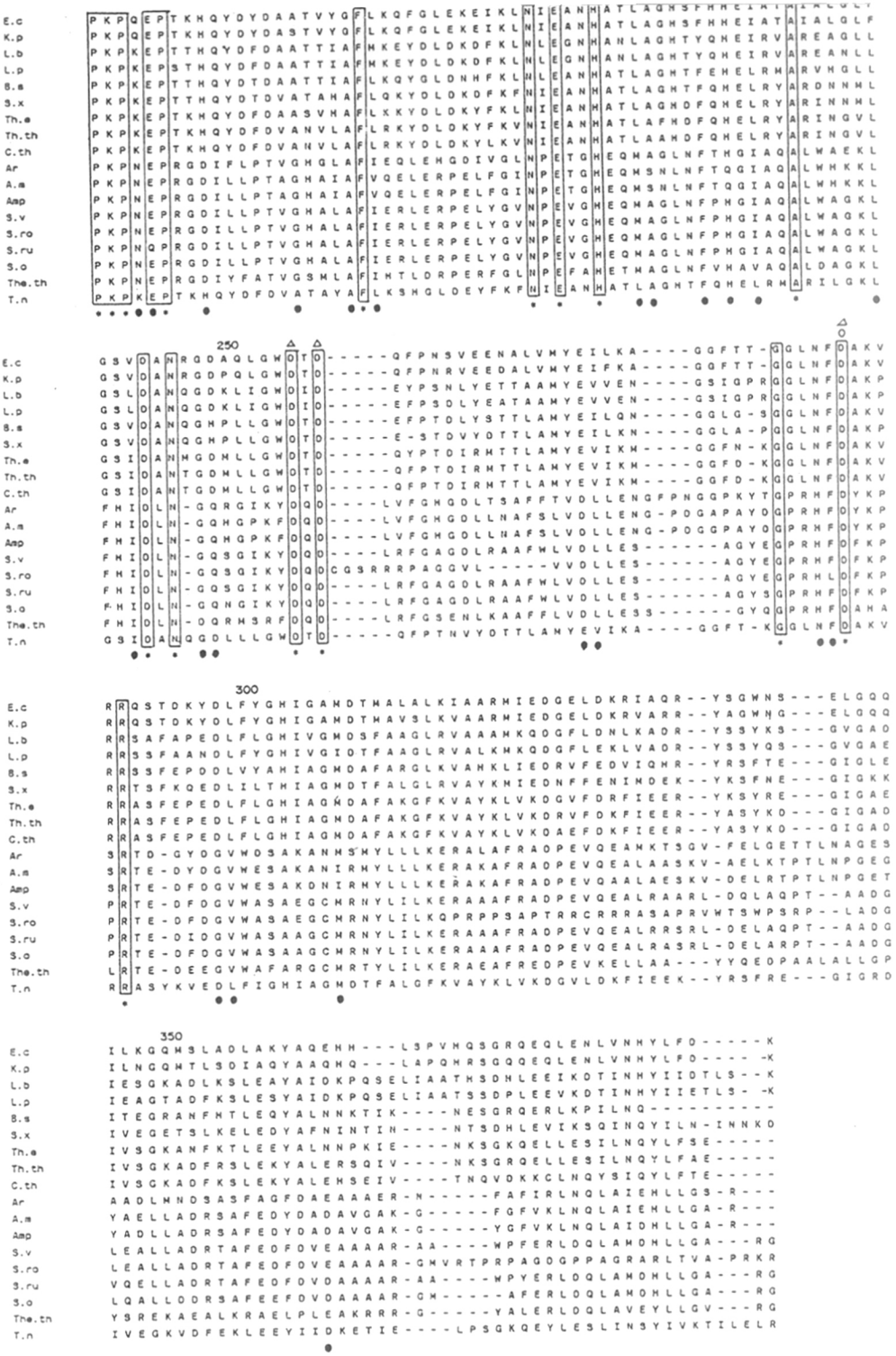
Based on their amino acid sequences (Vangrysperre *et al*, 1988) D-xylose isomerases can be classified into two groups:

a) D-xylose isomerases of *E. coli* and *B. subtilis* and

b) the enzyme from *Actinoplanes*, *Ampullariella* and *Streptomyces* spp.

The enzymes in the second group are less similar and lack a stretch of 30-40 amino acids which are present in the GI from the first group. The enzymes from *E. coli*, *C. thermosulfurogenes*, *L. pentosus* and *B. subtilis* are significantly homologous to each other and consist of approximately 440 amino acids. The GIs from *S. violaceoniger*, *S. griseofuscus*, *A. missouriensis* and *Ampullariella* are homologous and contain 390 amino acids. The existence of two distinct classes of the enzyme *viz.* (i) thermolabile GI of the *B. subtilis* and *E. coli* type and (ii) thermostable GI of the *Streptomyces* type has been reported (Lee *et al*, 1990a,b). The nucleotide sequence of a number of GI genes has been determined and their deduced amino acid sequences are available. The comparison of GI from mesophilic sources such as *Actinomycetes* with thermostable GI from *T. thermophilus* revealed D→E, Q→H and G→P substitutions in the latter organism. Q→H substitution reduces the chances of deamidation of glutamine at higher temperature which is known to be responsible for the irreversible thermoinactivation of the enzyme (Volkin & Klibanov, 1983, 1989). G→P substitutions render rigidity to the polypeptide backbone. GI from *S. violaceoniger* shows 86.8% homology with the enzyme from *S. griseofuscus*. An extensive amino acid homology (93.4%) was observed between GI from *A. missouriensis* and *Ampullariella* sp. The *Streptomyces* enzyme showed significant homology (60%) with the enzyme from the two *Actinomycetes* but poor homology with that from *E. coli* (24.2%) and *B. subtilis* (26.3%) (97). Meaden *et al* (1994) have compared the deduced amino acid sequences of the 19 bacterial GIs obtained from the databases (Swiss-PROT Version 23) and those described by Scheler *et al* (1991a,b). Considering the degree of similarity of the proteins and their statistical significance, it was inferred that all the proteins were homologous and comprised a single family. Based on the


E.c
K.p
L.b
L.p
B.s
S.x
Th.e
Th.th
C.th
Ar
A.m
Amp
S.v
S.ro
S.ru
S.o
The.th
T.n
50
100
150



Homology alignment of the XylA sequences. The XylA protein sequences, deduced from the nucleotide sequences obtained from the EMBL database (updated 1 April 1995), were analyzed with the CLUSTAL V program for multiple alignment. Identical amino acids (*) are boxed, and similar amino acids (.) are indicated below the sequence. Functional amino acid residues involved in catalysis (●), substrate binding (○), metal binding (△), and *cis* peptide linkage (⌐¬) are indicated above the sequences. The numbering of amino acid residues is based on the sequence of GI from *S. rubiginosus* (35). Abbreviations: E.c, *Escherichia coli*; K.p, *Klebsiella pneumoniae*; L.b, *Lactobacillus brevis*; L.p, *Lactobacillus pentosus*; B.s, *Bacillus subtilis*; S.x, *Staphylococcus xylosus*; Th.e, *Thermoanaerobacter ethanolicus*; Th.th, *Thermoanaerobacter thermosulfurogenes*; Ar, *Arthrobacter* sp.; A.m, *Actinoplanes missouriensis*; Amp, *Ampullariella* spp.; S.v, *Streptomyces violaceoniger*; S.ro, *Streptomyces rochei*; S.ru, *Streptomyces rubiginosus*; S.o, *Streptomyces olivochromogenes*; The.th, *Thermus thermophilus*; T.n, *Thermotoga neopolitana*.

G+C content of the DNA and the physico-chemical properties of GI, the proteins were divided into two distinct clusters, *viz*. the proteins from organisms with (i) high and (ii) low G+C DNA content. The first cluster represents GI from four species of *Streptomyces*, *Thermus thermophilus*, *Actinoplanes missouriensis*, *Ampullariella* and *Arthrobacter*. The second cluster includes the enzyme from three species of *Bacillus*, *Staphylococcus xylosus*, two species of *Lactobacillus*, *E. coli* and *Klebsiella pneumoniae*. The distribution of the proteins between the two clusters relates to the phylogenetic relationships of the organisms. Analysis of the aligned sequences revealed two signature sequences *viz*. W[GP]GREG[YSTA]E and [LIVM]EPKPX[EQ]P, which recognised all GI and no other proteins in SWISSPROT (Version 23). We have compiled and performed a multiple alignment of *xyl* A sequences from 18 bacterial sources using the programme Clustal V (Higgins & Sharp, 1988) (Fig.1.5). Inspite of the low homology between the *Streptomyces* and *E. coli* or *Bacillus* enzymes the amino acids involved in the substrate and metal ion binding, as well as in catalysis, are completely conserved. The *cis* peptide linkage between the adjacent glutamic acid and the proline which is responsible for the formation of the rigid structure at the active site is also well conserved in all the GIs studied. Thus the essential structure at the catalytic center of GI appears to be analogous in all the enzymes that were compared. To elucidate the structure-function relationship of the D-xylose isomerases, the *xyl* A sequences from various organisms have been compared. The information on the conserved and homologous regions in the *xyl* A sequences will be a valuable tool for isolating novel GIs with desirable catalytic properties.

GENETIC IMPROVEMENT OF GI USING SITE-DIRECTED MUTAGENESIS

Advances in recombinant DNA techniques allow successful isolation of genes of almost any protein. Engineering proteins by manipulation of their genes is at present a viable approach which complements structure-function studies performed by pre-existing methods and allows production of tailor-made proteins with desirable properties to give a complete insight into the mechanism of the enzyme. Site directed mutagenesis (SDM) of GI has been carried out with several objectives of academic and industrial importance. The studies described below, have contributed substantially to further our knowledge about the molecular mechanism of GI and create new possibilities of producing an enzyme with properties suitable for its biotechnological applications.

Thermal stabilization

Most of the commercial preparations of GI have a temperature optimum of 60-65°C. The thermal inactivation of GI confers a limitation on the operating time of the reactor. Several mechanisms have been reported to be involved in the irreversible inactivation of GI such as irreversible unfolding, glycation, and/or deamidation of Asn or Gln (Volkin & Klibanov, 1983,1989). Under practical conditions, GI is exposed to high sugar concentrations (3M) at higher temperatures (60-65 °C), thus leading to non-enzymatic glycation of Lys residues and subsequent inactivation of GI. Hydrophobic interactions among the aromatic amino acid residues of the active site helps to maintain the association of monomers to active dimers. Strengthening of the interactions at the interface of the active dimers results in enhanced thermostability. Several of these attempts to improve the thermal stability of GI are described in Table 1.7.

Alteration of substrate specificity

GI displays higher affinity for xylose than glucose. However, an increased affinity towards glucose is desirable in view of its application in the production of HFCS. The substrate preference of the thermophilic GI from *C. thermosulfurogenes* was altered by redesigning the amino acids situated in the substrate binding pocket (Table 1.8).

Functional role of essential amino acid residues

SDM studies have been carried out to assess the role of essential amino acids in GI which are tabulated in Table 1.8.

Deciphering the role of metal ions

SDM studies have been employed to dissect the role of metal ions in the thermal stability and catalytic reaction of GI. Substitution of side chains involved in metal binding demonstrated that the metal ions in GI play an essential role in binding and stabilizing the open forms of the substrate and in catalyzing hydride transfer between the C1 and C2 positions (Jenkins *et al*, 1992). Neutron activation analysis and SDM studies of GI from *Streptomyces olivochromogenes* indicated the differential role of two magnesium ions in its activity. Mg-1 bound to Glu180 is essential for isomerization but not for ring opening. Hydroxide coordinated to Mg-2 could act to deprotonate the glucose O2 and protonate O1 in a step

Table 1.7 Role of amino acids involved in thermal stability deduced by SDM studies

Microorganism	Mutation	Inference	Reference
A. missouriensis	Lys -253 → Arg	Half life of the enzyme was increased by 3-fold as compared to the native GI	Quax *et al* (1991)
T. thermosulfurogenes	Trp-139 → Phe,Met,Ala	Increase in catalytic efficiency proportional to the decrease in the side chain of the substituted amino acids.	Meng *et al* (1993)
S. rubiginosus	His-220→Ser, Asn, Gln	Confirmed the role of geometry and binding affinity of metal ion at site 2 in the thermal stability of GI.	Cha *et al* (1994)

Table 1.8 Role of essential amino acids deduced by SDM studies

Microorganism	Amino acid	Inference	Reference
E. coli	His-101 His-271	Acts as a catalytic base mediating the isomerization reaction Imparts thermal stability to the enzyme since it is involved in the binding of active site metal ion	Batt *et al* (1990)
C. thermosulfurogenes	His-101 Trp-139→Phe	Involved in stabilization of the open chain substrate molecule in the active site by hydrogen bond formation with C5 hydroxyl group Reduced the K_m and enhanced the K_{cat} of the mutant thermophilic enzyme toward glucose	Lee *et al*, 1991 Meng *et al*(1991)
A. missouriensis	His-54 His-220 Lys-183 Lys-294 Trp-16 & Trp-137	Governs anomeric specificity of the substrate Involved in metal coordination and positioning of the substrate in the isomerization process Crucial role in the isomerization step by assiting the proton shuttle Indirectly involved in binding the activating cations Maintenance of general architecture of substrate binding site	Lambeir *et al* (1992)
E. coli	Trp-49 & Trp-188 His-101	Stabilizes the transition state during ring opening Mediates intramolecular proton transfer during ring opening	Jamieson & Batt *et al*(1992)

concomitant with the hydride transfer step of isomerization. The metal ion may be involved in the stabilization of the resultant deprotonated O2 (Allen *et al*, 1994b).

Alteration in pH optimum

Commercial application of GI demands an acidic pH optimum to enable starch liquefaction and glucose isomerization to be carried out in a single step. Glu-186 is a conserved residue which is situated near the active site of GI from *A. missouriensis* but does not participate in the substrate or metal ion-binding. The negative charge from this group was removed by its mutation to Gln which lowered its pH optimum from 7.5 to 6.25 and also changed its preference from Mg^{2+} to Mn^{2+} (Tilbeurgh *et al*, 1992). This study adds new information on the catalytic mechanism of aldose-ketose isomerization by GI and demonstrates that a single amino acid substitution is able to shift the pH optimum by more than 1 pH unit.

IDENTIFICATION OF IMPORTANT PROBLEMS AND POTENTIAL FRUITFUL SOLUTIONS

Introduction of enzymatic glucose isomerization for the production of HFCS is beset with several problems. Among the major problems are inactivation of GI at higher temperatures, high pH optima of many of the GI preparations, requirement of Co^{2+} for enzyme activity, lower affinity of GI for glucose than xylose and suboptimal concentrations of the product. Intensive research for overcoming these problems has resulted in the development of substantially improved processes. Nevertheless, there is a scope for further improvement in all the above-mentioned areas to evolve an economicaly feasible commercial process to substitute glucose totally by HFCS.

The ideal GI should possess a lower pH optimum, a higher temperature optimum, a resistance to inhibition by Ca^{2+} and a higher affinity for glucose than presently used enzymes. Introduction of all these properties in a single protein is a herculian task which has hindered the development of an economically feasible commercial process for enzymatic isomerization of glucose to fructose. Advances in recombinant DNA technology and protein engineering open up new and encouraging possibilities of clubbing the desirable properties in a single organism to produce a tailor-made protein. Reduction in the enzyme cost by amplification of the GI gene may cause an increase in fermentation productivity. Isolation of a mutant for the

constitutive production of GI and elimination of the requirement of metal ions will contribute significantly towards the improvement of the existing processes for HFCS production. Combination of saccharification of starch with isomerization of glucose will result in shortening of the reaction time and lead to a major saving in terms of equipment cost. However, the major drawback in the development of the uni-pH process is that the wide difference in optimum reaction conditions for the amylase and glucose isomerase tends to lower the efficiency of a simultaneous system. Efforts to produce thermostable and acid-stable GI with higher affinity for glucose by SDM of GI gene are already under way, for evolving a GI preparation suitable for biotechnological applications.

SCOPE OF PRESENT THESIS

Streptomyces sp. NCIM 2730 is a soil isolate of the National Chemical Laboratory and is a high glucose/xylose isomerase (GXI) yielding strain ($10U.ml^{-1}$). The organism is able to utilize xylan containing agricultural wastes and thus dispenses the use of xylose as an inducer. The production of GXI does not require the addition of cobalt in the fermentation medium. Therefore, this organism is preferred to many others, since it does not pose a threat to the human consumption of HFCS, where lies the ultimate use of this culture. The GXI of *Streptomyces* sp. NCIM 2730 has been the subject of various structure-function studies. Immobilization of the enzyme on the expensive anion exchange resin, Indion 48-R, facilitated its reuse and increased its thermostability (Gaikwad *et al*, 1992b). *Streptomyces* GXI possesses a single active site for isomerization of glucose and xylose, distinct catalytic and co-factor binding sites (Gaikwad *et al*,1989;1992a). Histidine is essential for the binding of the substrate whereas carboxyl groups participate in the metal binding site (Gaikwad *et al*, 1988). Unfolding and refolding of the GXI revealed the occurrence of a molten-globule type intermediate in the folding pathway and indicated that the intact tertiary rather than secondary structure is important for the biological activity of the protein (Ghatge *et al*,1994).

The objective of the present work has been :

1. To identify the GXI encoding gene fragment and to characterize its gene product.
2. To study the effect of the coordinated addition of specific amino acids on the synthesis of the recombinant GXI using factorial design.
3. To gain an insight into the acid-induced unfolding of GXI and to probe the intermediates of the pathway.

CHAPTER 2

Molecular Cloning, Expression And Sequence Analysis Of Glucose / Xylose Isomerase Gene

SUMMARY

A partial genomic library of *Streptomyces* sp. NCIM 2730 constructed in *E. coli* using pUC8 vector, was screened for the presence of GXI gene using an 18-mer mixed oligonucleotide probe complementary to a highly conserved six-amino acid sequence of GXI from *Actinomycetes*. Eight clones which hybridized with the radiolabelled oligoprobe (5'-GTGACCTTCCACGACGAC-3') showed the ability to complement the xylose isomerase-detective *E. coli* mutants. The restriction map of the insert from one (pMSG27) of the eight GXI-positive clones showing detectable GXI activity, was constructed. GXI-deficient strains of *E. coli* were able to utilize xylose as the sole carbon source for their growth upon transformation with pMSG27. *E. coli* JM105 (pMSG27) and *E. coli* JC1553 (pMSG27) were inducible by IPTG suggesting that the expression of the cloned gene was under the control of the *lac* Z promoter. Western blot analysis revealed that, the cloned gene is expressed as a fusion protein of M_r 110. GXI from *Streptomyces* sp. NCIM 2730 shares 55% sequence homology to the reported GXIs from bacterial sources.

INTRODUCTION

Actinomycetes produce the class II type of GI which is thermostable as compared to the class I enzyme produced by *E. coli*, *Bacillus*, *Staphylococcus, Salmonella and Lactobacillus* which is thermolabile but has a pH optimum in an acidic range. There is a tremendous interest in studying the thermostable GI, because of its potential in the production of HFCS which is more efficient at higher temperatures of 60-65°C. Several reports on cloning of GXIs from the various microorganisms are already described in Tables 1.4 and 1.5, of Chapter 1.

Recombinant DNA technology offers a means to isolate and manipulate the gene of a desired protein. The cloning of GXI from several microorganisms has been mainly carried out with the primary aims of :

1. Overexpression of the enzyme by gene dosage effect.
2. Engineering the protein to alter its properties to suit its biotechnological applications.
3. To study the operonic arrangement of the xylose utilization genes.
4. Direct conversion of xylose to ethanol by constructing yeasts containing the *xyl* A gene.

The first step to acheive the desired goals is cloning of the *xyl* A gene in a suitable host. The *xyl* A gene of various organisms has been cloned and isolated using several strategies, for instance immunoscreening, screening with oligoprobes based on either the N-terminal sequence or the sequence corresponding to highly conserved regions of the protein and complementation of the xylose isomerase-negative mutation of the host strain.

Streptomyces sp. NCIM 2730 is one of the highest GXI producers (10 Uml^{-1}) and possesses several novel features such as ability to grow on Co^{2+}-deficient medium containing hemicellulose-rich agricultural wastes, alleviating the problem of health-hazards related to Co^{2+} and thus dispensing the use of xylose as an inducer. The K_m of the enzyme for glucose (200 mM) is higher than that for xylose (4 mM) (Gaikwad, S. M., 1990). The alteration in the substrate specificity would be desirable for the efficient catalysis by the enzyme in industrial processes. As a prelude to protein engineering of GXI, the present study deals with the isolation of the GXI encoding gene from *Streptomyces* sp. NCIM 2730, nucleotide sequence analysis and characterization of the gene product.

MATERIALS AND METHODS

Materials

Acrylamide, agar, ampicillin, boric acid, ficoll, IPTG, X-gal, PEG-8000, polyvinylpyrrolidone, SDS and urea were purchased from Sigma Chemical Co.,(USA). 4-Aminosalicylic acid sodium salt was purchased from Aldrich Chemical Co., (USA). Tri-isopropyl naphthalene sulphonate (sodium salt) was obtained from Eastman Kodak Co., (USA). Media components such as bacto-agar, bacto-tryptone, bacto-peptone, yeast extract and malt extract were obtained from Difco Laboratories, (USA). All restriction enzymes, T_4 DNA ligase and T_4 Polynucleotide kinase were obtained from Amersham (UK), Boehringer Manheim (Germany) and Bangalore Genei (India). Pronase, RNase A, DNase I were purchased from Sigma Chemical Co., (USA). pUC8 and pUC19 were purchased from Pharmacia, (Sweden). The random prime labelling kit was purchased from Boehringer Manheim (Germany). The transfer membranes Hybond-N and nitrocelluose were obtained from Amersham (UK) and Advanced Microdevices (Ambala, India) respectively. The radiolabelled γ^{32}P-dATP, α^{32}P-dCTP and α^{35}S-dATP were supplied by Board of Radiation and Isotope Technology (Dept. of Atomic Energy), BARC, New Mumbai, India. Konika (Japan) X-ray films, commercial developer preparation, (IPC163:Kodak) and sodium thiosulphate were used for autoradiography. All other chemicals used were of analytical grade and were obtained from SISCO Research Laboratories, Qualigens Fine Chemicals, Loba Chemie and E. Merck Laboratories.

Medium composition

Luria-Bertani (LB)		LB- agar
Bacto-tryptone	1.0 %	2% Bacto-agar in LB
Yeast extract	0.5 %	
NaCl	0.5%	
pH 7.0		

IPTG stock solution 200 mg.ml^{-1} dissolved in sterile distilled water

X-gal stock solution 20 mg.ml^{-1} dissolved in dimethyl formamide

Ampicillin stock 100 mg.ml^{-1} dissolved in sterile distilled water

Bacterial strains and plasmids

The *Escherichia coli* strains, plasmids and the *Streptomyces* culture used in the present studies are as listed below

Strain	characteristics	Reference
E. coli JM105	*end*A1, *thi*, *rps*L, *sbc*B15, *hsd*R4, Δ (*lac-pro*AB),[F', *tra*D36, *pro*AB, *lac*$I^qZ\Delta M15$]	Yanisch-Perron *et al*, 1985
E. coli DH5α	F^-, ϕ80d*lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *sd*R17, (r_k^-,m_k^+), *sup*E44, *rel*A1, *deo*R, Δ(*lac*ZYA-*arg*F)U169	Hanahan, 1983
E. coli JC1553	derivative of *E.coli* -K12, containing a *xyl*-7 mutation, the organism is unable to utilize xylose as sole carbon source	*E. coli* Genetic Stock Center, Yale University
Streptomyces sp. NCIM 2730	Soil isolate from the National Chemical Laboratory, Pune, India, posseses xylanase, protease, & GXI activities	National Collection of Industrial Microorganisms, NCL, Pune
pUC8, pUC19	a multicopy vector carrying amp^r marker and displays α-complementation of the lacZ gene in appropriate host	Messing, 1983; Yanisch-Perron *et al*, 1985
pMSG27	a construct of pUC8 containing 1.1kb insert of *Streptomyces* DNA	Present work

Culture maintenance

Streptomyces sp. NCIM 2730 was maintained on MGYP (malt extract, 0.3%; glucose, 1%; yeast extract, 0.3%; and peptone, 0.5%) slants grown at 28°C. The culture was grown in Callens medium for 96 h at 28°C for enzyme production (Callens *et al*, 1986). *E. coli* JM105 was grown on an LB plate/broth at 37°C. *E. coli* recombinants were grown in an LB or M9 medium supplemented with xylose (0.4%) and ampicillin (100 $\mu g.ml^{-1}$) at 37°C with shaking (200 rpm) for 16 h or 24 h respectively). The recombinant *E. coli* cultures were maintained as 15 % glycerol stocks at -70°C.

DNA Isolation Techniques

Genomic DNA isolation

Streptomyces sp. NCIM 2730 genomic DNA was extracted by the method of Kirby (Hopwood *et al*, 1985).

Plasmid DNA isolation

Plasmid DNA isolation was done according to Sambrook *et al* (1989), by the alkaline-lysis method.

Screening of genomic library

The genomic library of *Streptomyces* sp. NCIM 2730 in pUC8 was constructed as described by Ghatge *et al* (1995). DNA from *Streptomyces* sp. NCIM 2730 was extracted by the method of Kirby (Hopwood *et al*, 1985). The partial *Bam*H I digest of genomic DNA in the size range of 2-10 kb obtained by ultracentrifugation in the sucrose density gradient (10-40%), was ligated to dephosphorylated *Bam*H I restricted pUC8 vector. The ligation mixture was used to transform the competent cells of *E. coli* JM105. The recombinants were selected based on blue/white selection resulting due to α-complementation of the *lac* Z gene of the *E.coli* JM105 cells on LB plates containing ampicillin (100 $\mu g.ml^{-1}$), IPTG (4 µl) and X-gal (40µl).

Colony blotting

The transformants were picked up with sterile toothpicks and replica-plated on two LB agar plates containing ampicillin (100 $\mu g.ml^{-1}$), IPTG (4 µl) and X-Gal (40 µl) with the aid of

a numbered grid and incubated at 37°C for 14-16 h. One of the plates served as the master plate and was preserved at 4°C. On the other plate a circular sterile Hybond-N membrane was placed and wetted thoroughly. The membrane was peeled off and the adhering bacterial colonies immediately processed. The membrane with the colony side up was placed on a 3MM Whatman sheet saturated with 10% SDS for 3 min. It was then transferred to a 3MM sheet saturated with denaturing solution (0.5N NaOH, 1.5M NaCl) for 5 min. This was followed by incubation in neutralization solution (1.5M NaCl, 0.5M Tris-HCl, pH 8.8) for 5 min. The membrane was floated on 2X SSC for 5 min, air dried and wrapped in Saran wrap. The DNA was cross-linked to the membrane by exposure to UV for 7 min with the colony side down.

Oligoprobe screening

A mixed 18-mer oligoprobe was synthesized complementary to a six amino-acid stretch of Val-Thr-Phe-His-Asp-Asp which is conserved in GXIs from *Actinomycetes* and related organisms with the sequence 5'-GTGACCTTCCACGACGAC -3'.
C G

End-labelling of the oligoprobe

End-labelling of the oligoprobe was carried out using γ^{32} P-dATP according to Ausubel *et al* (1987) using T_4 polynucleotide kinase.

The labelling reaction comprised of the following:

Oligoprobe (18mer)	100 ng
γ^{32}P-dATP	100 µCi
T_4 Polynucleotide kinase	20 U
Polynucleotide kinase buffer	1 X

Total reaction volume was made upto 50 µl by the addition of sterile deionized water. The reaction was incubated at 37°C for 40 min followed by termination with 20 mM EDTA. The volume of the reaction was made upto 100 µl with TE buffer, pH 8.0 and the radiolabelled reaction mixture was purified using ion-exchange chromatography DE-52 matrix. The labelled oligonucleotides were eluted using TE containing 1M NaCl. The purified fractions containing the labelled oligonucleotide were further quantitated using the liquid scintillation counter (Rack-beta).

Hybridization and washing

Solutions

Denhardt's solution (50X)

Bovine Serum Albumin 1 %

Ficoll 1 %

Polyvinylpyrrolidone 1 %

(the solution was made in sterile deionized water)

20X SSC

Sodium chloride 3 M

Sodium citrate 0.3 M

(pH 7.0 adjusted with 1 M HCl)

20X SSPE

Sodium chloride 3.6 M

Sodium phosphate 0.2 M

EDTA (pH 7.7) 0.02 M

SDS 10 %

Prehybridization solution

SSPE (20X) 5.0 ml

Denhardt's solution (50 X) 0.5 ml

SDS (10%) 0.25 ml

Skimmed milk powder 1.0 %

Hybridization solution

Same as pre-hybridization solution containing end-labelled oligoprobe reaction mixture

Prehybridization was done to block the non-specific sites on the membrane and to avoid non-specific signals on the autoradiograms. Skimmed milk powder acts as a blocking agent. Pre-hybridization was carried out for a minimum period of 3-4 h at 42°C in a shaking water bath. The amount of pre-hybridization solution added was 0.1 $ml.cm^{-2}$. The pre-hybridization solution was removed and replaced by the hybridization solution containing the end-labelled probe. Hybridization was carried out at 42°C for 12-16 h. Washing was done to remove the unbound/unhybridized probe using solutions of different stringencies. The blots were given two washes with 5X SSC containing 0.1% SDS for 15 min at room temperature with gentle shaking. This was followed by a stringent wash at 42°C with the same buffer for 5 min.

Autoradiography

The moist blots were wrapped with Saran wrap and exposed to X-ray film for 24-48 h at -70°C. The X-ray film was developed using commercial developer followed by fixer (2% sodium thiosulphate). The developed film was washed with water and air dried.

Restriction map analysis

The restriction sites on the cloned DNA fragment were mapped using various restriction enzymes to digest the plamid pMSG27 and the pattern was observed by seperating the DNA fragments on agarose gel electrophoresis.

Southern blotting

DNA blots were prepared according to Sambrook *et al* (1989). The DNA blotting technique was originally described by Southern (1975). The DNA was fractionated on a 1% agarose gel, stained with ethidium bromide and photographed to determine the sizes of radioactive signals on the autoradiogram. The gel was subjected to depurination by treatment with 0.25 N HCl for 15 min. The gel was then treated with denaturing solution for 30 min, giving two changes. The gel was rinsed with sterile deionized water and further treated with neutralizing solution (*described earlier*) for 30 min. The gel was transferred to a support of 3MM Whatman sheets presoaked in 10X SSC . The Hybond-N membrane presoaked in sterile deionized water and saturated with the transfer buffer (10X SSC) was placed on the agarose gel taking care to prevent trapping of air bubbles. 3MM Whatman sheet was placed over the membrane loaded with a stack of tissue papers, filter paper stacks, and overlayed with a weight. It was ensured that the 3MM Whatman sheet support on which the gel was placed remains moist by the intermittent addition of the transfer buffer. The capillary blotting was allowed to take place for 24-36 h. The whole assembly was displaced and the membrane was marked for the orientation. It was rinsed in 6X SSC for 5 min. The excess fluid was drained off and the membrane was blotted dry between two Whatman sheets. UV cross-linking of DNA was done by exposure of the DNA side of the membrane to uv light for 15 min.

Random prime labelling

The multi-prime labelling kit from Boehringer Manheim (Germany) was used to label the *Sau*3A I partial digest of genomic DNA.

Preparation of probe

Streptomyces sp. NCIM 2730 genomic DNA (200ng) was subjected to partial digestion by *Sau*3A I and deproteinized by treatment with phenol-chloroform and precipitated with isopropanol. The precipitate of the genomic DNA digest was given a wash with 70% ethanol, dried and redissolved in sterile deionized water.

Labelling reaction

The genomic digest of *Streptomyces* sp. NCIM 2730 DNA was heated to 95°C and snap-chilled on ice. The following reaction components were added in the specified order :

Genomic DNA digest	100 ng
dCTP, dGTP, dTTP	3 µl
hexanucleotide primer	2 µl
α^{32}P-dATP (50 µCi)	5 µl
Reaction buffer	2 µl
Klenow enzyme (10 U)	1 µl

The total reaction volume was made upto 20 µl. The reaction mixture was mixed thoroughly by trituration and incubated at room temperature for 2 h. The reaction was terminated by the addition of 2.5 volumes of chilled ethanol and kept at -70°C for 2-3 h. The DNA pellet obatined was washed with 70% ethanol, warmed to 42°C for 2 min, followed by a spin at 10,000 rpm. The pellet was air-dried and dissolved in TE (pH 8.0).

The pre-hybridization and hybridization was carried out as described before except that the temperature was 65°C. The radiolabelled probe was denatured by boiling for 10 min, before addition to the hybridization solution. Two washes were given to the blot in the following order : 2X SSC, 0.1 % SDS; 1X SSC, 0.1 % SDS and 0.1X SSC, 0.1% SDS. Each wash was for 20 min at room temperature. Autoradiography of the blot was carried out as described earlier.

Complementation studies

E. coli JC1553 obtained from the *E. coli* Genetic Stock Centre containing a *xyl*-7 mutaion which is a genetic lesion is an *E. coli* K-12 derivative, resulting in a xylose isomerase-negative phenotype. The plasmids isolated from the putative GXI-positive clones

were used to transform the competent cells of *E. coli* JC1553. The *E. coli* cells were made competent by the calcium chloride method of Sambrook *et al* (1989). The complementation of the GXI gene in the mutant was monitored by plating out the transformation mixture on Mac Conkey-agar plates containing xylose (0.4%) and ampicillin (100 $\mu g.ml^{-1}$).

Expression studies

Preparation of protein extracts

The recombinants, *E. coli* JM105 (pMSG27), *E. coli* JC1553 (pMSG27) and respective hosts containing pUC8 were grown in 100 ml of M9 medium containing xylose (0.4%) and ampicillin (100 $\mu g.ml^{-1}$), in the presence and absence of IPTG (1.6 mM) at 37°C for 24 h. Cells were harvested by centrifugation, suspended in sodium phosphate buffer (50 mM), pH 7.5 and lysed by sonication (Branson sonicator), four pulses of 30 seconds each. The supernatant obtained after centrifugation at 10,000 rpm at 4°C was heated at 60°C for 5 min to heat inactivate proteases and centrifuged at 10,000 rpm at 4°C.

Enzyme activity

The GXI activity of the supernatant was determined at 70°C for 30 min according to the method of Dische and Borenfreund (1951) as modified by Marshall and Kooi (1957). One unit of enzyme activity was defined as the amount of enzyme which produces 1 µmole of fructose per min under the assay conditions. Protein was estimated by the Bradford method (1976).

Western blotting

In this method, electrophoretically fractionated proteins are transferred from a gel to a solid support and probed with antibodies specific for a particular antigen (in this case the target protein).

Primary antibody

Polyclonal antibodies were raised in rabbit against an electrophoretically homogeneous preparation of GXI from *Streptomyces* sp. NCIM 2730 as described by Ghatge *et al* (1991).

Secondary antibody

Goat anti-rabbit IgG conjugated with alkaline phosphatase (Sigma, USA) was used as secondary antibody.

Solutions :

Towbin buffer

Glycine 39 mM

Tris 48 mM

SDS 0.37 %

Methanol 20 %

pH 7.8 - 8.4

Tris-buffered saline (TBS)

Tris HCl, pH 8.0 100 mM

NaCl 150 mM

TBST

Tween-20 (detergent) in TBS 0.05 %

Blocking solution

BSA in TBST 20 %

Alkaline phosphatase buffer (AP buffer)

TrisHCl, pH 9.5 100 mM

NaCl 100 mM

$MgCl_2$ 5 mM

Nitro-blue tetrazolium salt (NBT)

50 $mg.ml^{-1}$ nitroblue tetrazolium in 70% dimethylformamide

BCIP

50 $mg.ml^{-1}$ 5-bromo-4-chloro-3-indolyl phosphate in 70% dimethylformamide

Ponceau-S

working stock 1:10 diluted in deionized water

The samples for western blotting were prepared by growing the recombinants in M9 medium supplemented with xylose (0.4%) and ampicillin (100 $\mu g.ml^{-1}$). The cell pellet was resuspended in Tris-HCl buffer pH 7.5 (10 mM) and treated with lysozyme (0.5 $mg.ml^{-1}$) at 37°C for 15 min. followed by treatment with DNase I (10 $U.ml^{-1}$) at 30°C for 30 min. The samples were treated with PMSF (2 mM) to inactivate the proteases in the extract. SDS-

PAGE analysis of the cell extract (10 $\mu g.ml^{-1}$) was carried out by Laemmli's method (1970). Western blotting was done according to the method of Towbin *et al* (1979). After electrophoresis, the gel was equilibrated in Towbin buffer for 30 min with gentle shaking. The casette was loaded onto the electro-blotting apparatus along with the negative electrode and the membrane towards the positive electrode. Transfer was carried out in Towbin buffer at a current of 20 mA for 12 h at 4°C. The efficiency of transfer of proteins onto the membrane was monitored by staining with Ponceau-S (a dye that binds to all proteins immobilized on nitrocellulose). The stain was rapidly rinsed out extensively with deionized water, for the subsequent tests. The blot was agitated in TBST buffer for 30 min, followed by a fresh wash of TBST for another 30 min. Non-specific protein binding sites were saturated by incubating the membrane in the blocking solution and kept shaking for 30 min. To the blocking buffer, a 1:1000 dilution of the primary antibody raised against GXI was added and the blot was gently agitated by shaking for 4-6 h. Two washes of TBST were given at intervals of 10 min to remove the unbound antibodies. This was followed by two washes in TBST containing 0.1% BSA for 15 min. The filters were transferred to fresh TBST buffer containing 5 µl of the alkaline phosphatase conjugated anti-rabbit antibodies and agitated for 2 h. The filters were again subjected to washes of TBST for 15 min to remove unbound secondary antibodies. The membrane was blotted dry and incubated in the color development solution (10 ml AP buffer, 66 µl NBT, 33 µl BCIP) for 10-15 min. The reactive areas showed a purple coloration. Further color development was stopped by rinsing the membrane with deionized water. The membrane was either kept moist and photographed or air-dried and stored for record.

Sequence Analysis

DNA sequence analysis was carried out by the dideoxy termination method of Sanger (Sanger *et al*, 1977). In this enzymatic method of sequencing, a specific primer is used and synthesis is initiated at only one site where the oligonucleotide primer anneals to the template. Base specific chain termination results due to incorporation of dideoxynucleoside triphosphates (ddNTPs), since these nucleotide analogues lack the 3'-OH group necessary for DNA chain elongation. Thus a mixture of a small amount of ddNTP with the conventional dNTPs, in the process of DNA synthesis resulted in a competition between extension of the chain and infrequent but specific termination. This gave rise to a population of

oligonucleotides whose lengths were determined by the location of a particular base along the length of template DNAs that differed in length by as little as one nucleotide. When these samples are loaded adjacent on a gel, the order of nucleotides can be read directly from an autoradiographic image of a gel.

Subcloning

The plasmid pMSG27 was subjected to digestion with *Eco*R I, the insert band (0.4 kb) and vector containing 0.25 kb region of the insert were eluted seperately from low melting agarose gel, using the protocol descibed by Sambrook *et al* (1989). The vector containing a portion of the insert was self-ligated. The *Eco*R I inserts were ligated to the EcoR I digested vector. The ligation mixture was transformed into *E. coli* DH5α cells and transformants were selected based on the blue/white selection.

Host strain recommendation

The quality of the template DNA used for sequencing is of importance in its sequenceability. The host strain recommended for the preparation of template DNA is *E. coli* DH5α and yields acceptable sequencing results. The pMSG27 DNA and subcloned plasmids were transformed in *E. coli* DH5α and maintained as a 15% glycerol stock at -70°C. The plasmid was isolated by growing the recombinant in Terrific broth containing ampicillin (100 $\mu g.\ ml^{-1}$) which is an excellent medium for bacterial growth leading to a 4-8-fold increase in the growth of bacteria and in turn to higher plasmid yields.

Purification of plasmid DNA

Purified RNA-free plasmid DNA is a pre-requisite for sequencing. The inclusion of a PEG-precipitation step yields a high quality, super-coiled plasmid DNA, that is free of contaminating chromosomal DNA and RNA. Plasmid DNA isolated from large scale (50 ml) cultures was treated with RNase A at a final concentration of 200 $ng.\mu l^{-1}$ at 37°C for 2h (Sambrook *et al*, 1989). DNA was extracted using phenol-chloroform and precipitated from the aqueous phase with 0.1 volumes of sodium acetate (3M), pH 5.2 and two volumes of absolute ethanol and kept overnight at -20°C. The DNA pellet was washed thoroughly with 70% ethanol and dried under vacuum for 3 min before dissolving in 32 µl of deionized water. The plasmid DNA was precipitated by addition of 1.6 M NaCl and 13% PEG-8000 and kept

on ice to selectively precipitate the supercoiled plasmid DNA. The mixture was immediately centrifuged at 10,000 rpm for 10 min at 4°C. The precipitate was vacuum-dried and redissolved in deionized water. DNA was re-extracted with phenol-chloroform to remove any traces of PEG-8000. It was precipitated again with sodium. acetate and ethanol. The precipitate was washed extensively with 70% ethanol to remove adhering salts and minute traces of PEG-8000, which could interfere with the sequencing reactions. Purity of the plasmid was checked on an agarose gel by electrophoresis.

Sequencing reaction using Sequenase™ version 2.0

The method described in the product instruction mannual (US Biochemicals, USA) was followed and essentially comprised of the following steps:

Denaturation of plasmid DNA

Plasmid DNA was denatured in the presence of 200 μM NaOH, 200 μM EDTA at 37°C for 30 min. This was neutralized with 0.1 volumes of sodium acetate pH 5.2 (3M) and the DNA was precipitated with 2-4 volumes of absolute ethanol at -70°C for 15 min. The precipitate was dried after washing extensively with 70% ethanol.

Annealing of primers to DNA:

Sequencing reactions of pMSG27 and the plasmid from subclones were carried out using the universal Forward primer (-40) (17mer) 5'-GTTTTCCCAGTCACGAC-3' and the Reverse primer (17mer) 5-' CAGGAAACAGCTATGAC-3'.

The dried supercoiled plasmid was dissolved in deionized water and the volume was made upto 10 μl in 1X sequencing buffer. The additions were made in the following order:

Primer	1 μl
Reaction Buffer (5X)	2 μl
DNA	7 μl (approx. 3-5 μg)

The capped tube was warmed to 65°C for 2 min and allowed to cool to room temperature (<35°C) over a period of 30 min and snap-chilled on ice.

Labelling reaction

The labelling mix containing a mixture of all the four deoxynucleoside triphosphates was diluted five-fold. Sequenase version 2.0 enzyme was diluted 1:8 with enzyme dilution

buffer. The following additions were made and the contents were mixed thoroughly by trituration.

Ice-cold annealed DNA mixture	10 µl
DTT	1 µl
Diluted labelling mix	2 µl
α^{35}S-dATP	0.5 µl
Diluted Sequenase polymerase	2 µl

The labelling reaction was incubated at room temperature for 2-5 min.

Termination reactions

In four tubes labelled G, A, T and C, 2.5 µl of the ddGTP termination mix was added and incubated briefly for 1-5 min at 37°C. To each tube containing the termination mix, 3.5 µl of labelling reaction was added and incubated for 5 min at 37°C. The reaction was stopped by addition of 4 µl of Stop solution. The samples were heated to 75°C and chilled on ice, and loaded onto the sequencing gel.

Denaturing polyacrylamide gel electrophoresis

Under optimal conditions, 300 or more bases can be read starting at the bottom of a gel. Factors affecting resolution of a sequence include the quality of reagents used, the polymerization, the temperature of the gel during electrophoresis and proper drying of the gel after electrophoresis.

Gel casting

The thermostatic plate was made hydrophobic by a coat of Repel-silane (Pharmacia LKB, Sweden) whereas the notched plate was made hydrophilic by the application of Bind - silane (Pharmacia LKB, Sweden). The spacers (0.4mm) were positioned lengthwise and clamped.

100 ml of the polymerizing solution contained 1X TBE; 6% Acrylamide; 0.3% *N,N'*-bis-acrylamide; 7M Urea; Formamide; 40%; 25 µl TEMED; and 1ml Ammonium persulphate (10%) (freshly prepared).

The polymerizing gel solution was degassed before casting the gel. The gel casting support was inclined and polymerizing solution was poured at the bottom of the thermostatic

plate, so as to enable sliding of the notched plate which resulted in forming a thin layer of gel till the top of the thermostatic plate. This had to be done taking care to prevent bubble formation in the gel. The support stand was made horizontal as soon as gel casting was completed to prevent leakage and the plates were clamped together. The flush side of the shark tooth comb was inserted between the plates and the glass plates were clamped immediately on both the ends. The gel was prepared 2-20 h prior to use.

Electrophoresis

The gel sandwitch was fixed onto the Macrophor System (Pharmacia LKB, Sweden). A shark-tooth comb was placed and the wells thoroughly washed with 1X TBE. A pre-electrophoresis run was set at a constant power setting of 50 Watts at 50°C for 1 h. Samples denatured by heating at 75°C immediately chilled on ice, were loaded in the order G, A, T and C. Electrophoresis was carried out till the bromophenol blue dye reached the bottom of the gel, after which a second loading of sample was carried out. The run was terminated when the xylene cyanol dye reaches the bottom of the gel in the first lane. This ensured reading of different regions of the same sequence and overlapping them after comparison.

Post electrophoretic procedures

The gel along with the glass plate was washed in 15% methanol containing 10% acetic acid for 1 h for fixation of bands and removal of urea. The gel was dried at 80°C for a minimum of 4-6 h. Radioactive signals were checked using a hand held Geiger-Muller counter. A sheet of X-ray film (XAR) (Kodak,USA) was placed directly over the dried gel under safelight conditions and sandwitched between the two glass plates. This casette was covered with a black cloth and kept in a black box in dark for 48 h. The sequence was read and analyzed using the DNASIS software (Hitachi, Japan).

RESULTS AND DISCUSSION

Screening of the genomic library

The genomic library of *Streptomyces* sp. NCIM 2730 was constructed using the shotgun cloning method as described by Ghatge *et al* (1995), in the *Bam*H I site of the multicopy vector pUC8 using *E. coli* JM105 as a host. The partial genomic library was screened for the presence of GXI encoding gene using a labelled oligonucleotide probe. An 18-mer probe was synthesized complementary to a stretch of six amino acids (Val-Thr-Phe-His-Asp-Asp), which is highly conserved in GXI from *Actinomycetes*. Hybridization with the end-labelled oligonucleotide probe at medium stringency yielded eight putative xylose-utilizing clones (Fig. 2.1 a & b shows representative colony blots of the primary and secondary screening). The putative clones are shown in Fig. 2.2.

Complementation assay

The plasmid DNA from the putative clones, transformed in the xylose isomerase deficient mutants *E.coli* JC1553, resulted in successful complementation of the *xyl*-7 mutation. The xylose utilizing colonies appeared red on Mac Conkey medium containing xylose (0.4%) and ampicillin (100 $\mu g.ml^{-1}$) whereas the colonies containing the vector appeared pale (Fig. 2.3).

Induction of GXI gene

One of the putative clones containing (pMSG27) was characterized further for the expression of GXI. Table 2.1 summarizes the activity of the sonicated cell extracts of *E. coli* JM105 and *E. coli* JC1553 containing pMSG27. The GXI activity of the *E. coli* recombinant

Table 2.1. Effect of inducer on glucose/xylose isomerase activity of the recombinant pMSG27

Organism	Inducer	Total units U	Total protein mg	Specific activity $U.mg^{-1}$
E.coli JM105 (pMSG27)	—	1.590	12.5	0.126
	IPTG	2.320	10.0	0.230
E.coli JC1553 (pMSG27)	—	0.396	11.4	0.346
	IPTG	0.988	7.8	0.126
Streptomyces sp. NCIM 2730	xylose	11.6 $U.ml^{-1}$	3.8 $mg.ml^{-1}$	3.0 $U.mg^{-1}$

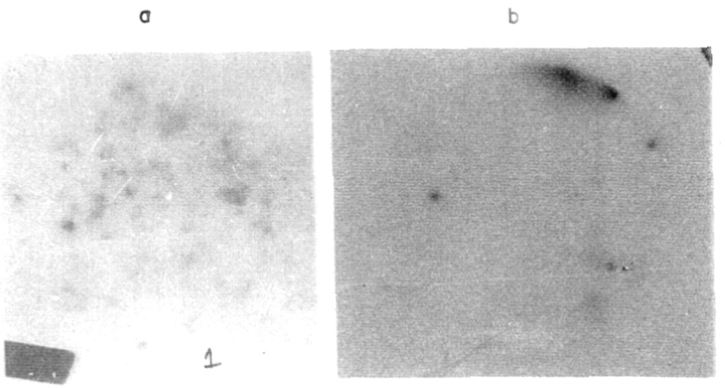


Fig. 2.1 **Representative autoradiogram of screening of partial genomic library of *Streptomyces* sp NCIM 2730 hybridized with γ^{32} P-dATP end-labelled oligoprobe.**

a. Primary screening, b. Secondary screening

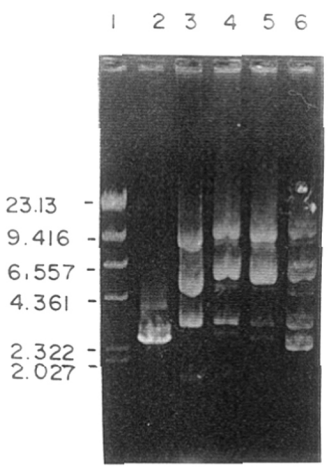


Fig. 2.2 **Agarose gel electrophoretogram of the putative recombinant plasmids.**

lane 1, λ Hind III marker, lane 2, linearised pUC8, lanes 3 - 7, putative plasmids.

was inducible by IPTG (1.6 mM), suggesting that the cloned fragment is in frame with the *lac* Z promoter and the expression of the cloned gene is under the control of the *lac* Z promoter. The expression of *Streptomyces* genes in *E. coli* is a field of tremendous interest. The *Streptomyces* genome has approximately 70 % G+C content, hence the recognition of the *Streptomyces* promoter is the key step in gene expression by a heterologous RNA polymerase. Lower GXI activity in presence of xylose as compared to the GXI produced by the parent organism, indicates either the absence of a functional regulatory gene in the insert or the lack of recognition of *Streptomyces* expression signals by the *E.coli* transcriptional and translational machinery, as GXI expression in *Streptomyces* sp. NCIM 2730 is inducible by xylose. Based on the molecular weight of wild type GXI of 1,60,000, the gene for GXI alongwith its regulatory genes should have a length of greater than 1.2 kb. The insert in pMSG27 is approximately 1.1 kb, hence the possibility of a truncated gene cannot be ruled out, which could also be one of the reasons for the lowered activity.

There have been reports on homologous cloning and overexpression of the *Streptomyces violaceoniger* D-xylose isomerase gene (Marcel *et al*, 1987) and *S. phaeochromogenes* (Kho *et al*, 1984). Although cloning in a homologous host is desirable for efficient expression and secretion of the cloned product, *E. coli* still remains the host of choice as the techniques for gene manipulation such as site directed mutagenesis are readily available for this organism. GI gene from *Streptomyces griseofuscus* S-41 was cloned in *E. coli* NM522 using pUC13 as a vector. DNA sequence of the cloned gene was determined and the amino acid sequence was deduced (Kikuchi *et al*, 1990). The GI gene from *Actinoplanes missouriensis* cloned in *E.coli* has GTG as its initiation codon and encodes a GI monomer of 394 amino acids. Wong *et al* (1991) have reported cloning of the D-xylose isomerase gene from *Streptomyces rubiginosus* in *E.coli* and the transformants were able to utilize D-xylose as sole carbon source.

Western blot analysis of the recombinant GXI

In order to discern the molecular size and the specificity of the recombinant GXI, western blot analysis was carried out using the antibodies raised against the native GXI. This method is useful for the identification and quantitation of specific proteins in complex mixtures. Electrophoretic seperation of proteins under denaturing conditions results in

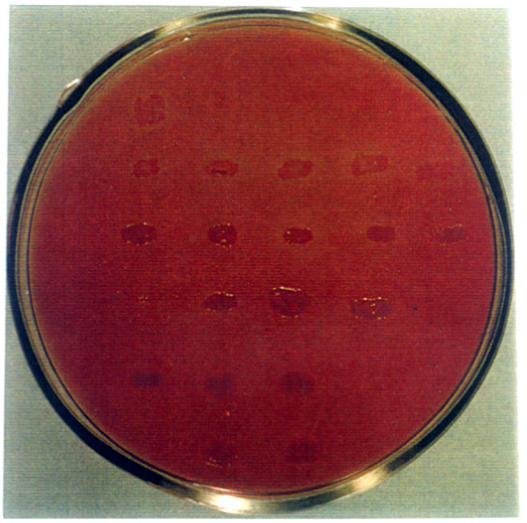


Fig. 2.3 **Complementation of *xyl*$^{-7}$ mutation of GXI-deficient *E.coli***

E.coli JC1553 (pUC8)-pale colonies; *E.coli* JC1553 (pMSG27)- red colonies

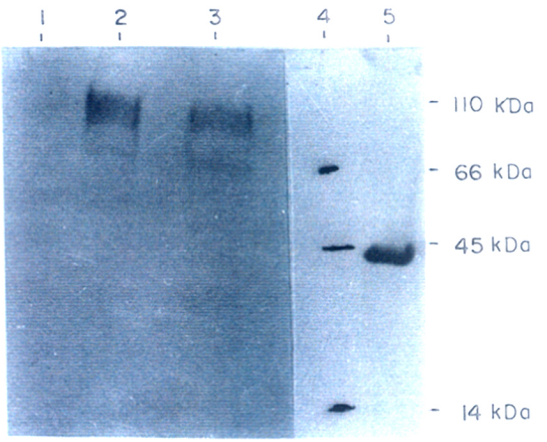


Fig. 2.4 **Western blot analysis of crude extracts of recombinants**

lane 1, *E. coli* JM105 (pUC8); lane 2, *E. coli* JM105 (pMSG27); lane 3, *E. coli* JM 105 (pMSG27) induced with IPTG (1.6mM); lane 4, SDS molecular weight markers; lane 5, purified protein from *Streptomyces* sp. NCIM 2730

elimination of problems of solubilization and co-precipitation of the target protein with adventitious proteins. The western blot pattern of the recombinant GXI revealed positive signals at 110 *kDa* (Fig.2.4). This indicates that the GXI gene is expressed as a fusion protein and possibly as an aggregate. The lower activity of the recombinant GXI can also be attributed to the improper folding of the fusion product.

Restriction mapping of the insert

The putative clone encoding GXI was further characterized by restriction enzyme digestions. Hybridization of pMSG27 with the α^{32}P-dATP labelled genomic DNA of *Streptomyces* sp. NCIM 2730 confirmed the origin of the insert to be from *Streptomyces* sp. NCIM 2730. Southern blotting of the restriction digests of pMSG27 and subsequent hybridization with α^{32}P-dATP labelled genomic DNA of *Streptomyces* sp. is shown in Fig. 2.5. Digestion of pMSG27 with *Bam*H I showed a single fragment of 1.1 kb along with the linearized plasmid at 2.7 kb (Fig. 2.5 a, b lane 7). *Sma* I generated two subfragments of 0.65 kb and 0.2 kb (Fig. 2.5 a, b lane 6) suggesting that there are two internal sites for *Sma* I in the insert. Restriction of the plasmid with *Eco*R I gave rise to an intense band of 0.4 kb, probably corresponding to two fragments of the same size and showed the presence of part of the insert with the linearized vector (Fig. 2.5 a, b lane 5) indicating that there are two internal *Eco*R I sites in the insert. Restriction of pMSG27 by *Pvu* II resulted in generating four fragments (Fig. 2.5 a, b lane 3) indicating that there are two sites for *Pvu* II in the insert. The insert contains a single site for *Kpn* I close to the *Bam*H I site near *Hind* III side of the vector (Fig. 2.5 a, b lane 2). Digestion of pMSG27 with *Bgl* II linearized the plasmid to give a single band at 3.8 kb (Fig. 2.5 a, b lane1) which attributes a unique *Bgl* II site on the insert. Digestion of pMSG27 with a combination of *Bgl* II and *Hind* III showed that the *Bgl* II site is located near the *Bam*H I site on the *Hind* III side of the vector. There were no sites in the insert for *Bam*H I, *Hind* III, *Sal* I, *Pst* I, *Cla* I, *Sph* I, and *Spe* I. Based on the above results the tentative restriction map of pMSG27 is shown in Fig. 2.6.

Sequence analysis

Sequencing of the pMSG27 clone was carried out by Sanger's enzymatic chain termination method (Sanger & Coulson, 1972). By this method it is generally possible to read about 250 bp at a time from each of the sequencing primers. To read the interior sequence

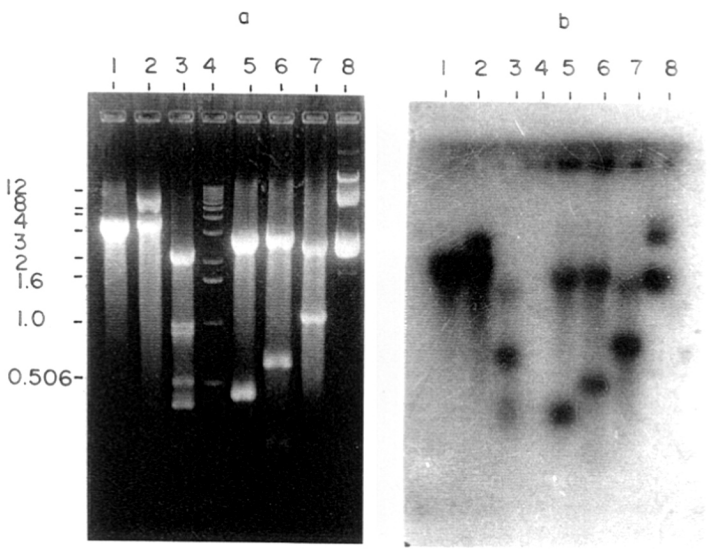


Fig. 2.5 **Southern blot analysis of the *E. coli* recombinant pMSG27 hybridized with α ^{32}P-ATP labelled DNA from *Streptomyces* sp. NCIM 2730 as a probe.**

a. lane 1, pMSG27 restricted with *Bgl* II; lane 2, *Kpn* I; lane 3, *Pvu* II; lane 4, 1 kb DNA ladder ; lane 5,*Eco*R I; lane 6, *Sma* I; lane 7,*Bam*H I, lane 8, unrestricted pMSG27.

b. Autoradiogram of **a**.

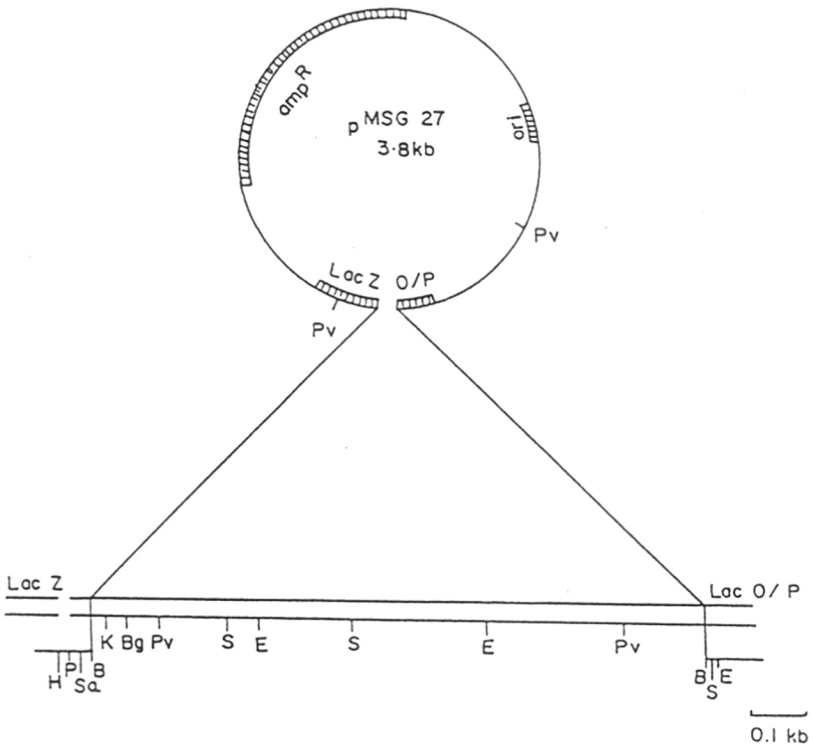


Fig. 2.6 **Restriction map of pMSG27.**

B, *Bam* HI; Bg, *Bgl* II; E, *Eco* RI; H, *Hind* III; P, *Pst* I; Pv, *Pvu* II; S, *Sma* I; Sa, *Sal* I. *K, Kpn* I.

further, the clone PMSG27 was further digested with restriction enzymes and the smaller size fragments were subcloned in pUC19 plasmid to generate the subclones. The subcloning and sequencing strategy of the clone pMSG27 is given in Fig. 2.7

Subcloning

The subclones were generated based on the restriction map of pMSG27. The *Eco*R I digestion yielded (i) the vector band containing 0.25 kb region of the insert, (ii) two fragments which appeared as a doublet band on gel electrophoresis with a size of 0.4 kb.

The pUC8 vector containing a 0.25 kb region of the insert was self ligated and was designated as pMSV1. The double digestion of pMSV1 with *Bam*H I + *Eco*R I yielded an insert of 0.25 kb. The *Eco*R I inserts of 0.4 kb were ligated to the *Eco*R I digested pUC19 vector. The strategy used to differentiate between the subclones containing *Eco*R I insert of the same size, was their digestion pattern with the restriction enzymes *Eco*R I and *Bam*H I (Fig. 2.8). The recombinant plasmid which gave an insert on digestion with *Eco*R I and linearized on digestion with *Bam*H I was designated as pMSV2 (Fig. 2.9). The recombinant plasmid which yielded a 0.4kb insert on digestion with *Eco*R I or *Bam*H I was termed as pMSV3 (Fig. 2.10).

The sequence analysis of the plasmids pMSG27, pMSV1, pMSV2 and pMSV3 revealed an insert of 1.088 kb with an overall mean G+C content of 67.6 % (Fig.2.11-18). The sequence was further verified by restriction digestion analysis of the predicted restriction enzyme sites and restriction fragment lengths and correlates well with the restriction map. In order to determine the uniqueness of the cloned gene, the sequence was subjected to homology search using BLAST (Basic Local Alignment Search Tool) (Altschul et al, 1990) and FASTA search (Pearson & Lipmann, 1988). The blastx search with the non-redundant database revealed a 25% homology to the N-terminal region of the xylulokinase of *Lactobacillus brevis, E. coli, B.subtilis, Tetragenococcus halophilus* 25% homology and 60% homology for a stretch of 16 amino acids for aquaporin of bacteria and *Saccharomyces cerevisiae*. Homology studies with the *xyl* A sequences of *Actinomycetes* revealed a 49-52% homology with the sequence of pMSG27.

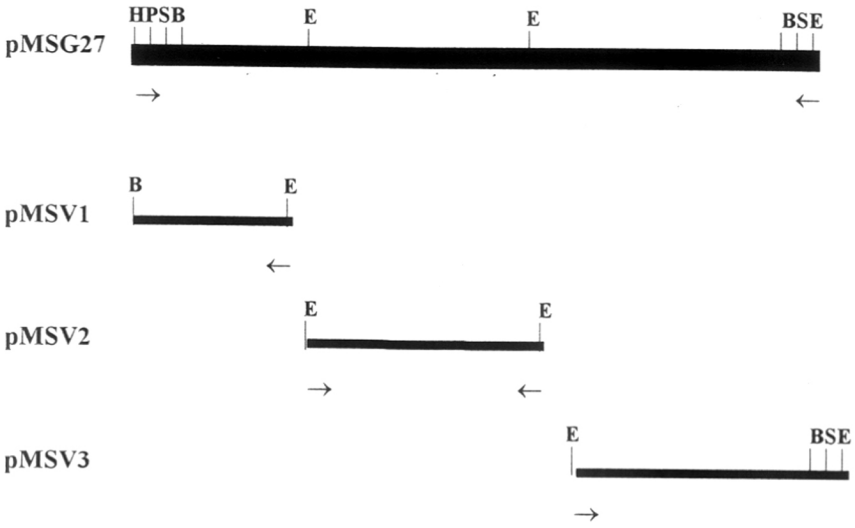


Fig. 2.7 **Subcloning and Sequencing strategy of recombinant pMSG27 clone.**

→ - Forward primer; ← - Reverse primer

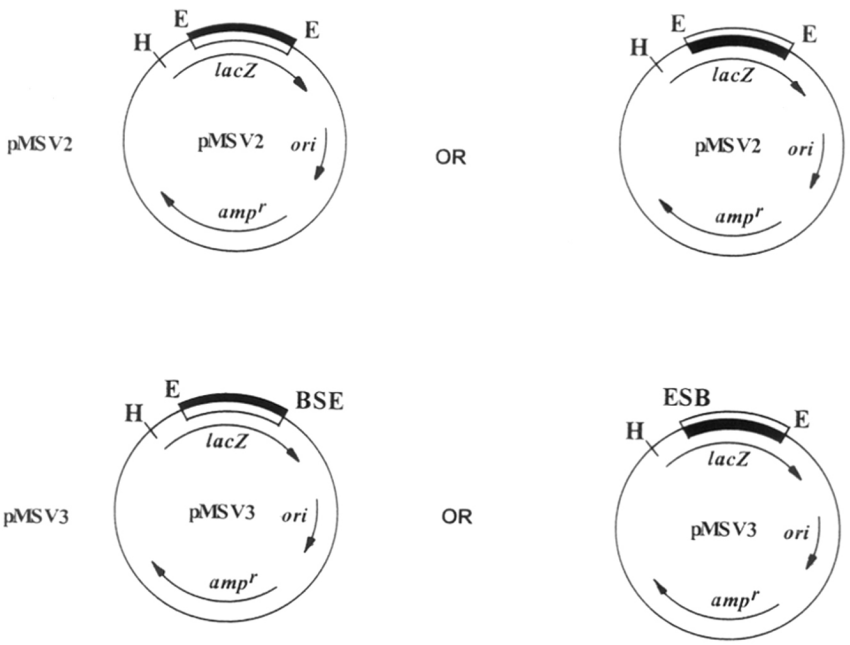


Fig. 2.8 : Possible orientation of *Eco* RI inserts in the subclones pMSV2 and pMSV3

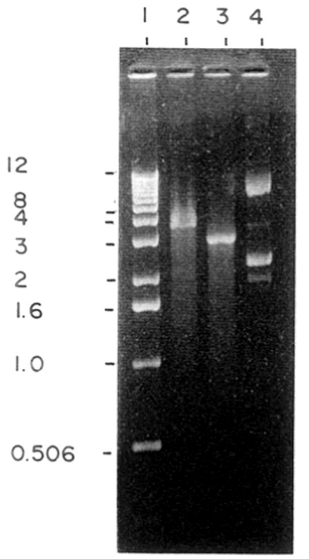


Fig. 2. 9 **Analysis of the recombinant subclone pMSV2.**

lane 1, 1 kb DNA ladder; lane 2, pMSV2 restricted with *Bam* HI; lane 3, *Eco* RI; lane 4, Unrestricted pMSV2 plasmid.

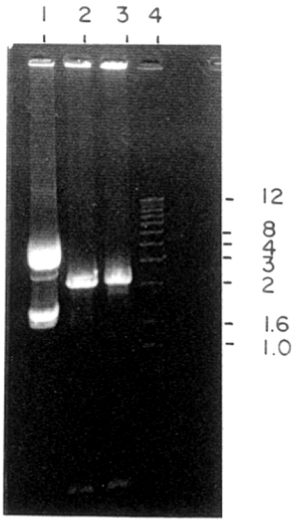


Fig. 2.10 **Analysis of the recombinant subclone pMSV3**

lane 1, Unrestricted pMSV3 plasmid; lane 2, pMSV3 restricted with *Eco* RI; lane 3, *Bam* HI; lane 4, 1 kb DNA ladder.

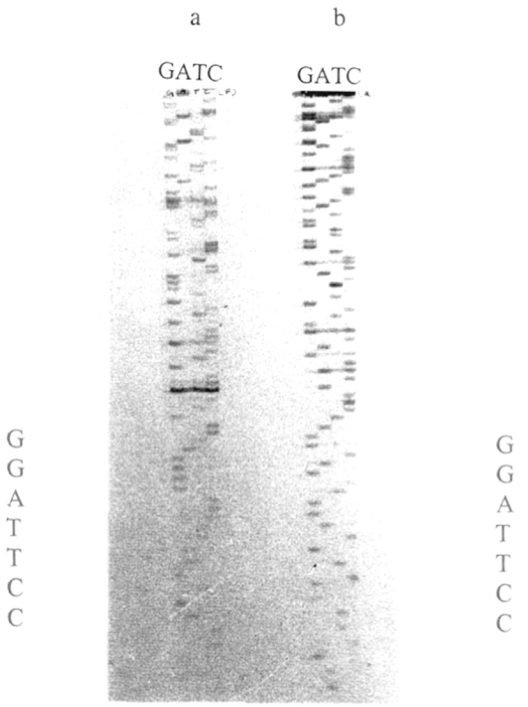


Fig.2.11 **Representative photograph of sequencing autoradiogram of pMSG27**

GTCCCGTCGC CGTCGTGGGC CTGCCCGGTT CCGGGCGGCA GTTCGTGGAC
TTCGGACCGG CGAACTGGCT CATCATCGCC TGGGGCTGGG GTCTCGCCGT
CGTCTTCGGC GTGTACGTGG CCGGCGGCAT CAGCGGCGCC CACCTCAACC
CGGCGGTGAC CCTGGCCTTC ACCGTGCGCA GGGACTTCCC CCTGAAGAAG
GTGCCCGCGT ACTGGCTGGC CCAGCTCTGC GGCGCCTTCG TCGCCGCCGC
GCTCGTCTAC GCCTGCTATC GCTGGGCGAT CGACGCCGCC GACGCGAAGG
CAGGGCTGGC CAGGGCTGGC CCGTGACGAA TCACTGGCCA CCTATTCCAT
CTTCGCGACC TTCCCCGCGG AATACTTCGG GGATTCCTGG TGGGGTCCGC
TGCTCGACCA GATCGTGGGC ACCGGAATTC TCCTGCTCCT GATCTGCGCG
CTCATCGACC TGAAGAATGT CGCGCCCCTG TCGAACTCCA CCCGTTCCTC
ATCCGGCTTG TGGTGTTCTC TGTAAACTTT

Fig. 2.12 **Sequence of pMSG27 read with M13 Universal Forward primer**

CACACGACGG TCGTCTCACG CTGGTTGGTG ATGCCGATCG CCCGCAGGTC
GGAGGCCGTC AGGCCGCCCT CGCGGAGGGC GTTCTGGATC ACCGTGTTGG
TACGCTCCCA GATCTCCACC GGGTCGTGCT CGACCCAGCC GGAACGGGGC
AGGATCTGCT GGTGCTCCAG CTGGTGCTTC GCCACCTCGT TACCGGCGTG
ATCGAAGATC ATGAATCGGG TGCTGGTGGT CCCCTGGTCC ACCGCGCCGA
CGAATTCGGG CATCGCTGCA ACCTTTCCTG AGTGACGGCG GTTCCCGGGG
GCCTAGGCCT CCTCGGTGGC CGGGCCTTCC TCCTGGGCCT CCATCGCCCG
AGTCCTGGCG GTGATGCTCG GCTTGATCAG CGCGTCGTAC ACCAGCACGC
CGACGACACC GCCGATCAGC GGACCGACGA TGGGAATCCA CCAGTAGCCG
CTGAACCACC CCAAAATTTC CCGGAAGGCC AA

Fig. 2.13 **Sequence of pMSG27 read with M13 Reverse primer**

CCCCTTTGAA CGGTGGACCA GGGGACCACC AGCACCCGAT TCATGATCTT
GGTAACGAGG TGGCGAAGCA CCAGCTGGAG CACCAGCAGA TCCTGCCCCG
TTCCGGCTGG GTCGAGCACG ACCCGGTGGA GATCTGGGAG CGTACCAACA
CGGTGATCCA GAACGCCCTC CGCGAGGGCG GCCTGACGGC CTCCGACCTG
CGGGCGATCG GCATCACCAA CCAGCGTGAG ACGACCGTCG TGTG

Fig. 2.14 **Sequence of pMSV1 read with M13 Forward primer**

TCGCTGCGAA CCTTTCCTTT ATTGACCGCG GTTCCCGGGG GCCTAGGCCT
CCTCGGTGGC CGGGCCTTCC TCCTGGGCCT CCATCGCCCG AGTCCTGGCG
GTGATGCTCG GCTTGATCAG CGCGTCGTAC ACCAGCACGC CGACGACACC
GCCGATCAGC GGACCGACGA TGGGAATCCA CCAGTAGCCG CTGAACCACC
CGAACGTGCC CGGGAGCGCG ATGTCGCCCC AGCCCTCGAA GTAGGTGAAC
AGCCGGGGGC CGAAATCGCG CGCGGGATTG ATCGCGTATC CGGCATTCGT
GCCGAAGGTC AGACCGATGG CGACGACCAC CAGGCCGATG AGGAACGGGT
GGAGGTTCGA CAGGGGCGCG ACATTCTTCA GGTCGATGAG CGCGCAGATC
AGGAGCAGGA GAATTC

Fig. 2.15 Sequence of pMSV2 read with M13 Forward primer

GCGCTGGTTG GTGATGCCGA TCGCCCGCAG GTCGGAGGCC GTCAGGCCGC
CCTCGCGGAG GGCGTTCTGG ATCACCGTGT TGGTACGCTC CCAGATCTCC
ACCGGGTCGT GCTCGACCCA GCCGGAACGG GGCAGGATCT GCTGGTGCTC
CAGCTGGTGC TTCGCCACCT CGTTACCGGC GTGATCGAAG ATCATGAATC
GGGTGCTGGT GGTCCCCTGG TCCACCGCGC CGACGAATTC

Fig. 2.16 Sequence of pMSV2 read with M13 Reverse primer

GAATTCCGGT GCCCACGATC TGGTCGACCA TCGGACCCCA CCAGGAATCC
CCGAAGTATT CCGCGGGGAA GGTCGCGAAG ATGGAATAGG TGGCCAGTGA
TTCGTCACGG GCCAGCCCTG CCTTCGCGTC GGCGGCGTCG ATCGCCCAGC
GATAGCAGGC GTAGACGAGC GCGGCGGCGA CGAAGGCGCC GCAGACTGGG
CCAGCCAGTA CGCGGGCACC TTCTTCAGGG GAAGTCCCTG CGCACGGTGA
AGGCCAGGGT CACCGCCGGG TTGAGGTGGG CGCCGCTGAT GCCGCCGGCC
ACGTACACGC CGAAGACGAC GGCGAGACCC CAGGCCCCAG GGCGATGATG
AGCCAGTTCG CCGGTCCGAA GTCCACGAAC TGCCGCCCGG AACCGGGCAG
GCCCACGACG GCGACGGGAC

Fig. 2.17 Sequence of pMSV3 read with M13 Reverse primer

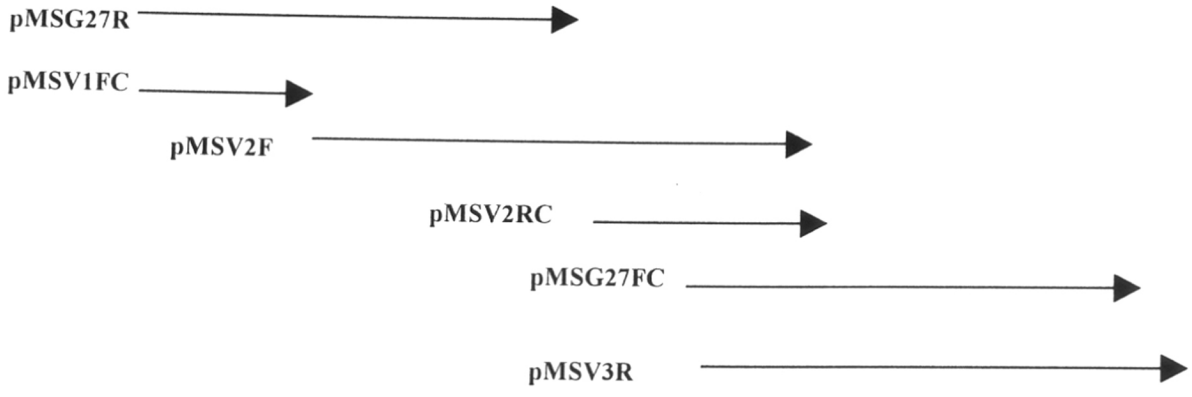


Fig.2.18 Schematic representation of overlapping of subclone sequences

5'-CACACGACGGTCGTCTCACGCTGGTTGGTGATGCCGATCGCCCGCAGGT

CGGAGGCCGTCAGGCCGCCCTCGCGGAGGGCGTTCTGGATCACCGTGTTG

GTACGCTCCCAGATCTCCACCGGGTCGTGCTCGACCCAGCCGGAACGGGG

CAGGATCTGCTGGTGCTCCAGCTGGTGCTTCGCCACCTCGTTACCGGCGT

GATCGAAGATCATGAATCGGGTGCTGGTGGTCCCCTGGTCCACCGCGCCG

ACGAATTCGGGCATCGCTGCAACCTTTCCTGAGTGACGGCGGTTCCCGGG

GGCCTAGGCCTCCTCGGTGGCCGGGCCTTCCTCCTGGGCCTCCATCGCCC

GAGTCCTGGCGGTGATGCTCGGCTTGATCAGCGCGTCGTACACCAGCACG

CCGACGACACCGCCGATCAGCCGGACCGACGATGGGAATCCACCAGTAGC

CGCTGAACCACCCGAACGTGCCCGGGAGCGCGATGTCGCCCCAGCCCTCG

AAGTAGGTGAACAGCCGGGGGCCGAAATCGCGCGCGGGATTGATCGCGTA

TCCGGCATTCGTGCCGAAGGTCAGACCGATGGCGACGACCACCAGGCCGA

TGAGGAACGGGTGGAGGTTCGACAGGGGCGCGACATTCTTCAGGTCGATG

AGCGCGCGCAGGAGCAGGAGAATTCCGGTGCCCACGATCTGGTCGACCAT

CGGACCCCACCAGGAATCCCCGAAGTATTCCGCGGGGAAGGTCGCGAAGA

TGGAATAGGTGGCCAGTGATTCGTCACGGGCCAGCCCTGCCTTCGCGTCG

GCGGCGTCGATCGCCCAGCGATAGCAGGCGTAGACGAGCGCGGCGGCGAC

GAAGGCGCCGCAGAGCTGGGCCAGCCAGTACGCGGGCACCTTCTTCAGGG

GGAAGTCCCTGCGCACGGTGAAGGCCAGGGTCACCGCCGGGTTGAGGTGG

GCGCCGCTGATGCCGCCGGCCACGTACACGCCGAAGACGACGGCGAGACC

CCAGCCCCAGGCGATGATGAGCCAGTTCGCCGGTCCGAAGTCCACGAACT

GCCGCCCGGAACCGGGCAGGCCCACGACGGCGACGGGAC-3'

Fig. 2.19 1088 bp nucleotide sequence of the *Bam*H I fragment of pMSG27

*** INPUT INFORMATION ***

FILE NAME : SFGI.SEQ

SEQUENCE : 1088 BP; 190 A; 358 C; 378 G; 162 T.

*** HOMOLOGY REGIONS REFERENCE *** (NORMAL 1- 1088)

NO TARGET FILE DEFINITION INIT OPT
25 STREDIA.SEQ 122 468
48.3% identity in 979 bp overlap

```
                100        110        120        130        140        150
SFGI.SEQ        GGTACGCTCCCAGATCTCCACCGGGTCGTGCTCGACCCAGCCGGAACGGGGCAGGATCTG
                                                  : :::   :::    ::  :: : :
STREDIA.SEQ     CTCTTCCCAGAAATGTTCCTGCATACTAATTTGTAAACCGCCATGACGAAATAGTCTCAG
                      10         20         30         40         50         60
                160        170        180        190        200        210
SFGI.SEQ        CTGGTGCTCCAGCTGGTGC-TTCGCCACCTCGTTACCGGCGTGATCGAAGATCATGAATC
                : :        ::: :  :: :  :: :::      :::  :: :  : ::: ::: :
STREDIA.SEQ     C-GAGCAAGGAGCCGCGGCATGAGCTACCAGCCCACCCCCGAGGAC-AAGTTCACGTTCG
                      70         80         90        100        110
                220        230        240        250        260        270
SFGI.SEQ        GGGTGCTGGTGGTCCCCTGGTCCACCGCGCCGACGAATTCGGGCATCGCTGCAACCTTTC
                :  ::  :   :::  ::::  ::  :    :::   :::::  :   :  :  :    :
STREDIA.SEQ     GCCTGTGGACCGTCGGCTGG--CAGGGACGGGACCCCTTCGGCGACGCCACCCGCGGCGC
                120        130        140        150        160        170
                280        290        300        310        320        330
SFGI.SEQ        CTGAGTGACGGCGGTTCCCGGGGGCCTAGGCCTCCTCGGTGGCCGGGCCTTCCTCCTGGG
                :   :  :::: :  :::   :::   ::::: ::   : : ::: :  : : ::
STREDIA.SEQ     CCTCGACCCGGCCGAGTCCGTCCGCC---GCCTCGCCGAGCTCGGCGCCCACGGCGTGAC
                 180        190        200        210        220        230
                340        350        360        370        380        390
SFGI.SEQ        CCTCCATCGCCCGAGTCCTGGCGGTGATGCTCGGCTTGATCAGCGCGTCGTACACCAGCA
                  :::: :: :   : :::         :   ::    :  ::::: : ::  : : ::::
STREDIA.SEQ     GTTCCA-CGACGACGACCTCATCCCCTTCGGCGCGACGGACAGCGAG-CGCGC-CGAGCA
                     240        250        260        270        280        290
                400        410        420        430        440        450
SFGI.SEQ        CGCCGA-CGACACCGCCGATCAGCGGACCGACGATGGGAATCCA---CCAGTAGCCGCTG
                :  : : ::   ::::: :   :: :  ::: :: :: ::  :   :: : :  : ::
STREDIA.SEQ     CATCAAGCGGTTCCGCC-AGGGGCTGGACGA-GACCGGCATGAAGGTCCCGATGGCGACC
                       300        310        320        330        340
                   460        470        480        490        500        510
SFGI.SEQ        AACCACCCG--AACGTGCCGGGAGCGCGATGTC-GCCCCAGCCCTCGGAAGTAAGGTGAC
                : : ::: :   ::   :: :: :   : : : ::  :  : :   ::: :  :::::
STREDIA.SEQ     ACCAACCTGTTCACCCACCCGGTGTTTAAGGACGGCGGCTTCACCGCGAACGACCGTGAC
                350        360        370        380        390        400
                      520        530        540        550        560        570
SFGI.SEQ        CAGCCGGGGGCCGAAATCGCGCGCGGGATTGATCGCGTATCCGGCATTCGTGCCGAAGGT
                  ::   :   ::    :: :::: ::     :::    :: :  ::: :   :: :
STREDIA.SEQ     GTGCGCCGTTACG---CCGTGCGCAAGACCATCCGCAACATCGAC-CTCGCGGTCGAGCT
                410        420        430        440        450        460
                      580        590        600        610        620
SFGI.SEQ        CAGACCG-ATGGCGACGACCACCAGGCCGATGAGGAACGGGTGGAGGTTCGACAGGGGCG
                : :  :: :   : ::: :  :  :: :: :  :  :   ::: :  :  ::  : : : ::
STREDIA.SEQ     CGGCGCGCAGACCTACGTCGCCTGGGGCG--GCCGCGAGGGCGCCGAGTC--CGGCGCCG
                   470        480        490        500        510        520
              630        640        650        660        670        680
SFGI.SEQ        CGACATTCTTCAGGTCGATGAGCGCGCGCAGGAGCAGGAGAATTCCGGTGCCCACGATCT
```

*** INPUT INFORMATION ***

FILE NAME : SFGI.SEQ

SEQUENCE : 1088 BP; 190 A; 358 C; 378 G; 162 T.

*** HOMOLOGY REGIONS REFERENCE *** (NORMAL 1- 1088)

NO	TARGET FILE DEFINITION	INIT	OPT
29	SRU.SEQ	94	196

52.5% identity in 259 bp overlap

```
                830        840        850        860        870        880
SFGI.SEQ        AGGCGTAGACGAGCGCGGCGGCGACGAAGGCGCCGCAGAGCTGGGCCAGCCAGTACGCGG
                  ::   : : X:::::::::::: ::   : :::::   : ::    ::  ::::    :  ::
SRU.SEQ         CCCCGAGGCCCAGCGCGGCGGCGGCGTTGTCGCCG--GTGC--CGCACGCCA--CCAGGG
                       10         20         30             40             50
                  890        900        910        920        930        940
SFGI.SEQ        GCACCTTCTTCAGGGGGAAGTCCCTGCGCACGGTGAAGGCCAGGGTCACC--GCCGGGTT
                 : ::::    :  :: :  ::    ::    ::  :: :: :  :: ::::  ::::::
SRU.SEQ         TCCCCTTGGAGAAAGGCAGTTCATGGCTGTCGCGGACGGTCCCGGCCACCTCGCCGGGAC
                    60         70         80         90        100        110
                   950        960                970        980        990
SFGI.SEQ        GAGGTGGGCGCCGCTGATGCCGCCGG-------CCACGTACACGCCGAAGACGACGGCG-
                :       :::  :: :  ::   : ::        :::::: : ::v :^   :   :: ::
SRU.SEQ         GCACCACGCGGGGCAGCAGCGCCGGGTCGAGGCCCACGTGCCCGAGGATCTCCTCGTCGT
                  120        130        140        150        160        170
                 1000       1010       1020       1030       1040       1050
SFGI.SEQ        AGACCCCAGCCCCAGGCGATGATGAGCCAGTTCGCCGGTCCGAAGTCCACGAACTGCCG-
                :  :: : :  :: : :   :  :::: :: :  :::          :: ::  : : ::
SRU.SEQ         ACGCCTCCGTGCCCGAGGCCCACCAGCCGGTGC--CGGAGGCGTCACCGCGGTCGGTCGT
                  180        190        200          210        220        230
                  1060       1070       1080
SFGI.SEQ        CCCGGAACCGGGCAGGCCCACGACGGCGACGGGAC
                 ::    ::::  :: : : ::
SRU.SEQ         GCCCTGCCCGGTGAGCCGCTCGGTGAGGTAGTCGT
                     240        250        260
```

*** INPUT INFORMATION ***

FILE NAME : SFGI.SEQ

SEQUENCE : 1088 BP; 190 A; 358 C; 378 G; 162 T.

*** HOMOLOGY REGIONS REFERENCE *** (NORMAL 1- 1088)

NO	TARGET FILE	DEFINITION	INIT	OPT
47	ACMI.SEQ		86	366

47.8% identity in 776 bp overlap

```
                330        340         350        360        370        380
SFGI.SEQ        TTCCTCCTGGGCCTCCATCGCC--CGAGTCCTGGCGGTGATGCTCGGCTTGATCAGCGCG
                   :::  : :  :::  :: :  : : ::: :  :: : : : :  ::  : :
ACMI.SEQ        CCCGGTGGCGGCGGCGACGGCCGGCGCGGACGGTCGGAGCCGCGCTGGTCGGGCACGGTG
                      10         20         30         40         50         60
                   390        400         410        420        430        440
SFGI.SEQ        TCGTACACCAGCACGCCGACGA-CACCGCCGATCAGCGGACCGAC-GATGGGAATCCACC
                     :::::  ::  ::  :    : :  :    :::::  : : :: ::: :::   :
ACMI.SEQ        CGCAGGACCAG--CGGAGATTATGGGCTCGCACGCTCGGACTGCCGGACGGGGATCGCTC
                      70         80         90        100        110
                  450        460         470        480        490        500
SFGI.SEQ        AGTAGCCGCTGAACCACCCGAACGTGCCGGGAGCG-CGATGTCGCCCCA-GCCCTCGGAA
                  : : :::    :  :: :  :::  ::     : : ::: :   ::: : ::    ::
ACMI.SEQ        GATCG-CGCAACTCAGCCGGGCCGTCACGCCCTTGACAATGCCATACCACGACCAAATAA
               120        130         140        150        160        170
                     510        520         530        540        550
SFGI.SEQ        GTAAGGTGACCAGC-CGGGGGCCGAAATCGCGCGCGGGATTGATCGCGTATCCGG---CA
                 : :    : :  : :::  :  :  : :: :    :  ::   : ::  ::    ::
ACMI.SEQ        TTCAACCAATAAACAAATCGACCGCGTTTCCCGGAGGTAACCGTGTCGCTCCAGGCCACA
                180        190         200        210        220        230
                 560        570         580        590        600        610
SFGI.SEQ        TTCGTGCCGAAGGTC-AGACCGATGGC--GACGACCACCAGGCCGATGAGGAACGGGTGG
                  ::     ::: ::     :: :  :  :::   :  : ::: :  : :  ::: ::
ACMI.SEQ        CCCGATGACAAGTTCTCCTTCGGTCTCTGGACCGTCGGCTGGCAGGCGCGTGACGCGTTC
                240        250         260        270        280        290
                    620        630         640        650        660        670
SFGI.SEQ        AGGTTCGACAGGGGCGCGACATTCTTCAGGTCGAT-GAGCGCGCGCAGGAGCAGGAGAAT
                 :   :: ::   :  ::    :: ::    :::: :::  :: :::  ::: ::   :
ACMI.SEQ        GGTGACGCCACCCGTCCG---GTCCTCGACCCGATCGAGGCCGTGCACAAGCTGGCCGAG
                300        310         320        330        340        350
                    680        690         700        710        720        730
SFGI.SEQ        TCCGGTGCCCACGATCTGGTCGACCATCGGACCCCACCAGGAATCCCCGAAGTATTCCGC
                  ::: ::  ::: :   ::  :  ::    : ::: ::  : ::     :   :  :
ACMI.SEQ        ATCGGCGCGTACG-GCGTCACGTTCCACGACGACGACCTGG--TGCCGTTCGGCGCCGAC
                   360        370         380        390        400        410
                     740        750         760        770        780        790
SFGI.SEQ        GGGGAAGGTCGCGAAGATGGAATAGGTGGCCAGTGATTCGTCACGGGCCAGCCCTGCCT-
                :  :     ::::: :      : :  ::    :      :     : : :: :: ::::
ACMI.SEQ        GCCGCGACCCGCGACGGCATCGTCGCCGGGTTCTCCAAGGCGCTCGACGAGACCGGCCTG
                      420        430         440        450        460        470
                       800        810         820        830        840
SFGI.SEQ        -TCGCGTCGGCGGCGTCGATCGCCCAGCGATAGCAGGC-GTAGACGAGCGCGGCGGCGAC
                 :::  ::  ::  : : :  :: :      ::  : ::   : ::  :::::::  :
ACMI.SEQ        ATCGTCCCGATGGTCACCACCAACCTGTTCACCCACCCGGTGTTCAAGGACGGCGGCTTC
                      480        490         500        510        520        530
             850        860         870        880            890        900
SFGI.SEQ        GAAGGCGCCGCAGAGCTGGGCCAGCCAGTACGCG---GGCACCTTCTTCAGGGGGAAGTC
```

*** INPUT INFORMATION ***

FILE NAME : SFGI.SEQ

SEQUENCE : 1088 BP: 190 A: 358 C: 378 G: 162 T.

*** HOMOLOGY REGIONS REFERENCE *** (NORMAL 1- 1088)

NO	TARGET FILE	DEFINITION	INIT	OPT
48	AMP.SEQ		86	366

47.8% identity in 776 bp overlap

```
                330        340         350        360        370        380
SFGI.SEQ        TTCCTCCTGGGCCTCCATCGCC--CGAGTCCTGGCGGTGATGCTCGGCTTGATCAGCGCG
                          :::  : :  :::  :: :  : : ::: :  :: : : : :  ::  : :
AMP.SEQ         CCCGGTGGCGGCGGCGACGGCCGGCGCGGACGGTCGGAGCCGCGCTGGTCGGGCACGGTG
                        10         20         30         40         50        60
                   390        400         410        420        430        440
SFGI.SEQ        TCGTACACCAGCACGCCGACGA-CACCGCCGATCAGCGGACCGAC-GATGGGAATCCACC
                     :::::  ::  ::  :    : :  :    :::::: : : :: ::: :::   :
AMP.SEQ         CGCAGGACCAG--CGGAGATTATGGGCTCGCACGCTCGGACTGCCGGACGGGGATCGCTC
                       70         80         90        100        110
                  450        460        470        480        490        500
SFGI.SEQ        AGTAGCCGCTGAACCACCCGAACGTGCCGGGAGCG-CGATGTCGCCCCA-GCCCTCGGAA
                  : : :::    :  :: :  :::  ::      : : ::: :   ::: : ::     ::
AMP.SEQ         GATCG-CGCAACTCAGCCGGGCCGTCACGCCCTTGACAATGCCATACCACGACCAAATAA
                120        130        140        150        160        170
                       510        520        530        540        550
SFGI.SEQ        GTAAGGTGACCAGC-CGGGGGCCGAAATCGCGCGCGGGATTGATCGCGTATCCGG---CA
                 : :    :  : :     : :::   :  :  : :: :    :  ::   : ::   ::
AMP.SEQ         TTCAACCAATAAACAAATCGACCGCGTTTCCCGGAGGTAACCGTGTCGCTCCAGGCCACA
                180        190        200        210        220        230
                 560        570        580        590        600        610
SFGI.SEQ        TTCGTGCCGAAGGTC-AGACCGATGGC--GACGACCACCAGGCCGATGAGGAACGGGTGG
                  ::     ::: ::     :: :  :  :::   :  : ::: :  : :  ::: ::
AMP.SEQ         CCCGATGACAAGTTCTCCTTCGGTCTCTGGACCGTCGGCTGGCAGGCGCGTGACGCGTTC
                240        250        260        270        280        290
                    620        630        640        650        660        670
SFGI.SEQ        AGGTTCGACAGGGGCGCGACATTCTTCAGGTCGAT-GAGCGCGCGCAGGAGCAGGAGAAT
                 :   :: ::   :  ::    :: ::    :::: :::  :: :::  ::: ::   :
AMP.SEQ         GGTGACGCCACCCGTCCG---GTCCTCGACCCGATCGAGGCCGTGCACAAGCTGGCCGAG
                300        310        320        330        340        350
                     680        690        700        710        720        730
SFGI.SEQ        TCCGGTGCCCACGATCTGGTCGACCATCGGACCCCACCAGGAATCCCCGAAGTATTCCGC
                  ::: ::  :::  :    ::  :  ::    : ::: ::  : ::    :    :  :
AMP.SEQ         ATCGGCGCGTACG-GCGTCACGTTCCACGACGACGACCTGG--TGCCGTTCGGCGCCGAC
                   360        370        380        390        400        410
                     740        750        760        770        780        790
SFGI.SEQ        GGGGAAGGTCGCGAAGATGGAATAGGTGGCCAGTGATTCGTCACGGGCCAGCCCTGCCT-
                :  :     ::::: :     : :  ::    :     :     : : :: :: ::::
AMP.SEQ         GCCGCGACCCGCGACGGCATCGTCGCCGGGTTCTCCAAGGCGCTCGACGAGACCGGCCTG
                       420        430        440        450        460        470
                       800        810        820        830        840
SFGI.SEQ        -TCGCGTCGGCGGCGTCGATCGCCCAGCGATAGCAGGC-GTAGACGAGCGCGGCGGCGAC
                  :::   ::  ::   : : :  :: :     ::   : ::   : ::  ::::::  :
AMP.SEQ         ATCGTCCCGATGGTCACCACCAACCTGTTCACCCACCCGGTGTTCAAGGACGGCGGCTTC
                       480        490        500        510        520        530
                850        860        870        880        890        900
SFGI.SEQ        GAAGGCGCCGCAGAGCTGGGCCAGCCAGTACGCG---GGCACCTTCTTCAGGGGGAAGTC
```

DNASIS ***** HOMOLOGY SEARCH LIST ***** DATE 03-15-99

*** INPUT INFORMATION ***

FILE NAME : SFGI.SEQ

SEQUENCE : 1088 BP; 190 A; 358 C; 378 G; 162 T.

*** HOMOLOGY REGIONS REFERENCE *** (NORMAL 1- 1088)

NO	TARGET FILE	DEFINITION	INIT	OPT
49	STREPVIO.SEQ		84	276

45.2% identity in 571 bp overlap

```
                520        530        540        550        560        570
SFGI.SEQ        GGGGGCCGAAATCGCGCGCGGGATTGATCGCGTATCCGGCATTCGTGCCGAAGGTCA--G
                     : :   : ::    :   ::: :    : ::  :   :: :   ::: :::
STREPVIO.SEQ    CGTACGGAACAAGGAGCAGCGCCATGAGCTTCCAGCCCACCCCCGAGGACAAGTTCACCT
                       10         20         30         40         50         60
                   580        590        600        610        620        630
SFGI.SEQ        ACCG-ATGGCGACGACCACCAGGCCGATGAGGAACGGGTGGAGGTTCGACAGGGGCGCGA
                 : :  ::  :::   :  : ::: :   ::: ::  ::   :   :: ::   :: :
STREPVIO.SEQ    TCGGTCTGTGGACCGTCGGCTGGCAGGGAAGGGACCCGTTCGGCGACGCCACCCGCCCTG
                       70         80         90        100        110        120
                   640        650        660        670        680        690
SFGI.SEQ        CATTCTTCAGGTCGATGAGCGCGCGCAGGAGCAGGAGAATTCCGGTGCCCACGATCTGGT
                :  ::  :  :  :  :::  :: ::::   : ::   :    :: ::: :::   ::
STREPVIO.SEQ    CCCTCGACCCG--GTCGAGACCGTGCAGCGCCTGGCCGAGCTGGGCGCCTACGGAGTGAC
                       130        140        150        160        170
                   700        710        720        730        740
SFGI.SEQ        CGACCATCGGACCCCACCAGGAATCCCCGAAGTATTCCG---CGGGGAAGGTCGCGAAGA
                :  ::: ::    : ::: :    ::     ::  ::::   : : :   :   :: : :
STREPVIO.SEQ    CTTCCA-CGACGACGACCTGATCCCCTTCGGGTCGTCCGACACCGAGCGCGAGTCGCACA
                  180        190        200        210        220        230
                750        760        770        780        790        800
SFGI.SEQ        TGGAATAGGTGGCCAGTGATTCGTCACGGGCCAGCCCTG-CCTTCGCGTCGGCGGCGTCG
                :  ::  :::  :    :   :   :::   : : : :: :  :  ::  :::  :  ::
STREPVIO.SEQ    T-CAAGCGGTTCCGCCAGGCCCTGGACGCCACTGGCATGACGGTGCCGATGGCCACCACG
                 240        250        260        270        280        290
                810        820        830        840        850        860
SFGI.SEQ        ATCGCCCAGCGATAGCAGGCGTAGACGAGCGCGGCGGCGACGAAGGCGCCGCAGAGCTGG
                : : : :  : :        :::   : ::  ::::::  :   :  : :  :: :  ::
STREPVIO.SEQ    AAC-CTCTTC-ACCCACCCCGTCTTCAAGGACGGCGGTTTCACCGCCAACG-ACCGCGAC
                  300        310        320        330        340        350
                870        880        890        900        910        920
SFGI.SEQ        GCCAGCCAGTACGCGGGCACCTTCTTCAGGGGGAAGTCCCTGCGCACGGTGAAGGCCAGG
                :   :::  ::::::     :        :   : ::   :   :   :::   ::    :
STREPVIO.SEQ    GTGCGCCGCTACGCGCTGCGCAAGACGATCCGCAACATCGACCTGGCGGCCGAGCTGGGC
                     360        370        380        390        400        410
                930        940        950        960        970        980
SFGI.SEQ        GTCACCGC-CGGGTTGAGGTGGGCGCCGCTGATGCCGCCGGCCACGTACACGCCGAAGAC
                : ::   :    :: :    X:::: :  ::: : :::::    ::    :::: : :::
STREPVIO.SEQ    GCCAAGACGTACGTCGCCTGGGGCGGCCGTGAGGGCGCCGAGTCCGGTGGCGCCAAGGAC
                     420        430        440        450        460        470
                990       1000       1010       1020       1030       1040
SFGI.SEQ        GACGGCGAGACCCCAGCCCCAGGCGATGATGAG--CCAGTTCGCCGGTCCGAAGTCCACG
                :   :::X  :::  : ::      : :: : :  : :  ::  :::   : :   :::
STREPVIO.SEQ    GTGCGCGACGCCCTCGACCGCATGAAGGAGGCGTTCGACCTCCTCGGCGAGTACGTCACC
                     480        490        500        510        520        530
                1050       1060       1070       1080
```

The absence of the -10, -35, and ribosome binding site and a complete open reading frame corresponding to the size of GXI supports the earlier studies indicative of a truncated gene fragment. Nevertheless, pMSG27 could be used as a probe to screen for a full length clone with the complete structural gene for further studies.

CHAPTER 3

Synthesis Of Recombinant Glucose/Xylose Isomerase : Effect Of Coordinated Addition Of Specific Amino Acids

SUMMARY

The amplified expression of a recombinant protein is known to lead to an intracellular depletion of specific amino acid pools which in turn may affect the production of the desired protein. In order to counteract and overcome such a situation during the fermentation of the recombinant *E. coli* JM109 (pMSG27) containing the GXI gene from *Streptomyces* sp. NCIM 2730, the effect of addition of different amino acids on the specific activity of recombinant GXI was studied. The amino acid composition of GXI from *Streptomyces* sp. NCIM 2730 revealed a predominance of aspartic acid, glutamic acid and glycine. Therefore, the effect of the coordinated addition of the assorted combinations of these three amino acids on the synthesis of recombinant GXI was studied. The results were analyzed using a 2^3 factorial design. The analysis of two factor interactions of the three amino acids revealed that : (i) The interaction between the aspartic and glutamic acid is independent of aspartic acid concentration but is affected by the increasing concentrations of glutamic acid (ii) The effect of aspartic acid concentration is more than that of glycine and (iii) During the interaction of glutamic acid and glycine, the effect of glutamic acid is more prominent than that of glycine.

The three factor interaction analyses led to the conclusion that the effect of the three amino acids is in the order aspartic acid > glutamic acid > glycine.

INTRODUCTION

Microbes with their ability for rapid multiplication and ease of *in vitro* manipulation have proved to be materials of choice in unravelling the mysteries of bioprocesses. Possibilities to overproduce useful metabolites of value especially by genetically engineered strains make microbes indispensable to progress in biotechnology. The ultimate aim of genetic engineering is to achieve higher production of a desired protein by amplification of the gene. However, the fermentation of genetically engineered microorganisms poses several problems. Proteases constitute a part of the heat shock response and are liable to be induced by stringent responses such as overexpression of cloned gene products (Ramirez *et al*, 1993). The synthesis of foreign proteins imposes a metabolic burden on the host cell and leads to the reduction in cell growth and consequently results in cell death (Zabuskie *et al*, 1986). Another major problem encountered in the overproduction of the recombinant protein is the segreggational and structural instability of the plasmid (Wrigley-Jones *et al*, 1993). Amplified expression of recombinant protein may result in a temporary intracellular depletion of specific amino acid pools and consequently, may stimulate a stress response in the host (Harcum *et al*, 1992).

With the view of overproducing GXI, heterologous cloning of GXI gene was carried out in *E. coli* (*refer Chapter 2*), which is a host of choice, with respect to the ease of performing genetic manipulations. An *E. coli* recombinant PMSG27 harbours a plasmid with a DNA fragment (1.088 kb) containing the GXI gene from *Streptomyces* sp. NCIM 2730. The expression of the gene was under the control of *lac*Z promoter. Amino acid composition of the GXI from *Streptomyces* sp. NCIM 2730 revealed the predominance of aspartic acid, glutamic acid and glycine (Gaikwad,S.M., 1990). This suggested that it is imperative to use these three amino acids as the key precursors during cultivation of the recombinant PMSG27 .

The objective of the present chapter has been to examine the effects of addition of various amino acids individually as well as in assorted combinations of their different concentrations, on the synthesis of GXI with an ultimate aim of enhancing the production of the recombinant protein.

MATERIALS AND METHODS

Recombinant organism

The organism used in the present studies was an *E. coli* JM109 strain bearing pUC8 which harbours a 1.088 kb DNA fragment encoding the GXI gene from *Streptomyces* sp. NCIM 2730, inserted at the *Bam*H I site in the multiple cloning site downstream to the *lacZ* promoter (*refer Chapter 2*).

Growth conditions

Recombinant *E. coli* JM109 (pMSG27) cells grown in LB containing ampicillin (100 $\mu g.ml^{-1}$) for 14h at 37°C were used to inoculate (1% inoculum) eight 500 ml Ehrlenmeyer flasks containing 70 ml of M9 medium (Sambrook *et al*, 1989) containing IPTG (1.66 mM). Xylose (2%) was used as the sole carbon source. Thiamine hydrochloride (2 $\mu g.ml^{-1}$) and ampicillin (50 $\mu g.ml^{-1}$) were added to the medium. The medium was also supplemented with individual amino acids (Table 3.1) as well as with eight assorted combinations of aspartic acid, glutamic acid and glycine, as shown in Table 3.2. The fermentation of PMSG27 was carried out in a medium containing the various combinations of amino acids at 37°C at 200 rpm for 15 h. The samples were withdrawn and assayed for GXI activity.

Enzyme assay and protein estimation

Enzyme assay and protein estimation were done as described in chapter 2.

2^3 factorial design

A general factorial design was performed keeping a fixed number of "levels" for each of a number of variables selected and the experiments were carried out with all possible combinations (Box *et al*, 1978). In the present factorial design the three quantitative variables were the concentration of aspartic acid (T), glutamic acid (C) and glycine (K). The response was assessed by determining the specific activity of the enzyme. The main effects of quantitative variables were calculated by averaging the individual measures of the effect. The individual measure of the effect of changing aspartic acid concentration from 0.04 to 0.1% was calculated by subtracting the specific activity at 0.04% from the specific activity at 0.1% at various combinations of glutamic acid and glycine concentrations. Average of the four values obtained from this data was calculated and the value obtained was the main effect of aspartic

acid concentration. Similarly the main effect of glutamic acid and glycine concentration were calculated. The interaction effects of two variables *i.e.* two factor interactions were calculated at two fixed levels of one factor. The average concentration effect of the other factor was calculated by averaging the individual measure of the effect of changing concentration of the factor from 0.04-0.1% at fixed concentration of the first factor, two values are thus obtained. The average of the two values will give the two factor interaction *viz.*(T x C), (T x K) and (C x K). For calculating three factor interaction, T x C interaction for glycine concentrations of 0.1% and 0.04% were calculated and the difference between these two values was estimated. The difference measures the consistency of the aspartic acid by glutamic acid interaction for the two glycine concentrations. Half of this difference was defined as the three factor interaction, of aspartic acid, glutamic acid and glycine and was denoted as T x C x K interaction. The experiments were carried out in duplicates and the standard error of the effect was calculated statistically.

RESULTS AND DISCUSSION

An elevated transcription rate of the cloned gene may have deleterious effects on the host, since there is an increase in the demand for precursors and metabolites. Introduction of a DNA vector into *E. coli* for the purpose of cloned gene expression can perturb the native cell functions at many levels. The presence of foreign DNA can alter regulation of chromosomal DNA replication, regulation of transcription of chromosomal genes, ribosome functions and RNA turnover, activities of regulatory proteins, chaperone and protease levels, membrane energetics and protein post-translational processing as well as energy and intermediary metabolism of the cells (Bailey, 1993). The cellular metabolism is redirected by the plasmid to channel these precursors into the synthesis of heterologous proteins which may appear "abnormal" and have a dissimilar amino acid composition compared to that of the average cell protein. This may lead to the intracellular depletion of specific amino acids. GXI being rich in aspartic acid, glutamic acid and glycine, the effect of addition of these amino acids to the medium as key precursors on the production of the recombinant enzyme was investigated. Various concentrations of these three amino acids, ranging from 0.04-0.1%, when supplemented separately, were able to boost the specific enzyme activity (Table 3.1).

Table 3.1 Effect of addition of a single amino acid on the synthesis of recombinant glucose isomerase

Amino acid concentration (%)	Specific activity ($nmol.mg^{-1}$)
Aspartic acid	
0.06	6.40
0.1	10.70
Glutamic acid	
0.04	8.30
0.08	2.55
0.1	1.54
Glycine	
0.04	5.10
0.06	4.23
0.1	4.25

Therefore, the interaction effect of the three amino acids on the enzyme activity was carried out using assorted combinations of their various concentrations. The results were analyzed using a 2^3 factorial design method. The 2^3 factorial design method was preferred because of its many advantages over the "One-factor-at-a-time" method, in which the experimental factors are varied one at a time, with the remaining factors held constant. This method was formerly regarded as the only correct approach to conduct research. The method provides an estimate of the effect of a single variable at selected fixed conditions of the other variables. However, for such an estimate to have general relevance, it is necessary to assume that (i) the effect would be the same at other settings of the variables and (ii) the variables act on the response additively. However, if the variables do act additively, the factorial design provides more precision and if the variables do not act additively, the factorial unlike the "One-factor-at-a-time" design can detect and estimate interactions that measure the nonadditivity. To secure the same precision, as in factorial design, for the estimate of the aspartic acid concentration effect, the "One-factor-at-a-time" experiment would need to employ eight runs, four at each level of aspartic acid concentration, with all the observations made at some arbitrarily fixed levels of glutamic acid and glycine concentrations. In a similar manner two further sets of eight runs would be required to study glutamic acid and glycine effect. Thus to obtain estimates of the main effect of three variables with the same precision as is provided by the 2^3 factorial design, the "One-factor-at-a-time" method would require 24 runs which represents a three-fold increased labour over the former method. Moreover, when a more thorough local exploration is needed, 2^3 factorial designs can be suitably augmented to form composite designs. It also forms the basis for two-level fractional factorial designs which are often of great value at an early stage of investigation. Additional advantages of 2^3 factorial design are that the interpretation of the observations produced by the design can proceed largely by using common sense and elementary arithmetic. Due to the above advantages of the 2^3 factorial design method, it was employed in the present studies to find out the effect of assorted combinations of amino acids (aspartic acid, glutamic acid, glycine) on the synthesis of recombinant GXI by *E. coli* JM109 (pMSG27). Table 3.2 shows the effect of supplementation of different concentrations of the three amino acids during the

fermentation of PMSG27 on the synthesis of GXI. Variation in the specific activity implied that there was a differential effect of the type of amino acid and its concentration on the production of the enzyme. The results were analyzed using 2^3 factorial design.

The underlying significance of the numerical interaction values is that the more positive the value the more strength to the effect. Thus, in an interaction of two amino acids with positive and negative individual main effects, a particular amino acid is said to contribute significantly if it is able to ameliorate partially or totally the negative effect of the other amino acid.

Table 3.2 Effect of addition of assorted combinations of amino acids on the synthesis of recombinant glucose isomerase

Test condition	Aspartic acid conc. (%) T	Glutamic acid conc. (%) C	Glycine conc.(%) K	Sp. activity (nmol mg^{-1})
1	0.04	0.04	0.04	18.691
2	0.1	0.04	0.04	32.777
3	0.04	0.1	0.04	19.959
4	0.1	0.1	0.04	26.71
5	0.04	0.04	0.1	8.71
6	0.1	0.04	0.1	9.14
7	0.04	0.1	0.1	7.52
8	0.1	0.1	0.1	10.12

The amino acid is said to have no effect if the total effect of the two-factor interaction is the same as the main effect of the other amino acid.

Table 3.3 shows the effect of the interactions and can be interpreted in the following manner :

Table 3.3 Calculation of the effects of the addition of amino acids using 2^3 factorial design

Effect	Estimates
Average	16.702
Main effects :	
Aspartic acid (T)	5.96
Glutamic acid (C)	-1.25
Glycine (K)	-15.65
Two-factor interactions :	
T x C	-1.2875
T x K	-4.452
C x K	1.1475
Three-factor interactions :	
T x C x K	2.3725

(i) The interaction effect of aspartic acid and glutamic acid (T x C) is -1.2875 and the individual main effects due to aspartic acid and glutamic acid are 5.96 and -1.25 respectively. It can be inferred that during this interaction there is no effect of aspartic acid concentration and only glutamic acid concentration affects the specific activity.

(ii) The interaction effect of aspartic acid and glycine(T x K) is - 4.452 while the values of the main effects due to aspartic acid and glycine are 5.96 and -15.65 respectively. It can be inferred that during this interaction the effect of aspartic acid concentration is more than that of glycine. (iii) The interaction effect of glutamic acid and glycine (C x K) is 1.1475 while the individual main effect due to glutamic acid and glycine are -1.25 and -15.65 respectively, both of which are negative. Therefore, it can be concluded that during this interaction the effect of glutamic acid is more prominent than that of glycine.

(iv) The three factor interaction is 2.3725 compared to the main effects due to aspartic acid, glutamic acid and glycine, with the corresponding values of 5.96, -1.25 and -15.65 respectively. It is evident that the effect of aspartic acid is predominant having positive value and glutamic acid is the next component, which has comparatively less effect on the system.

Fig.3.1 gives the pictorial display of the various interactions calculated by 2^3 factorial design for the three amino acids.

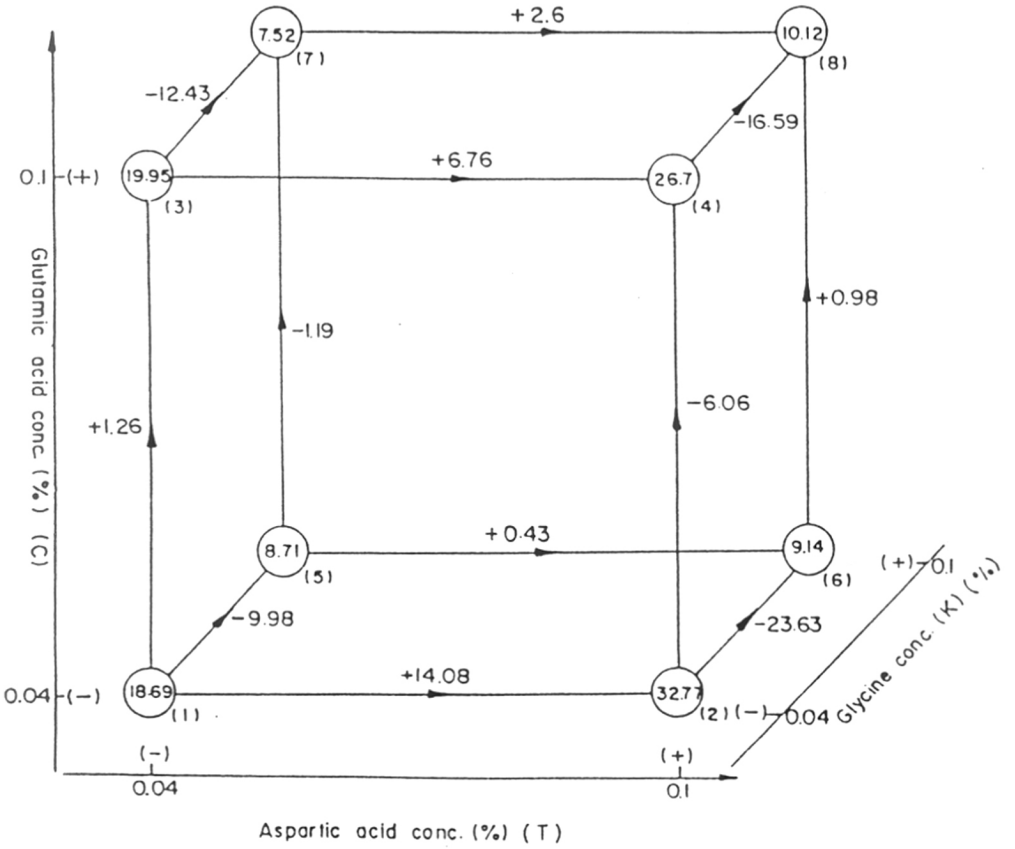


Fig. 3.1 Geometric representation of the interaction of aspartic acid, glutamic acid and glycine in the concentration range of 0.04 - 0.1% of each amino acid

Recombinant DNA technology provides a valuable means to develop biotechnological processes. However, there are a few problems inherent to the cultivation of genetically engineered microorganisms such as the instability of the plasmid encoding the gene for the desired protein and perturbance of the nutrient levels in the cells harbouring the plasmid. The elevated transcriptional rates induced in the recombinants cause a temporary starvation of the specific amino acid pools. In order to overcome these problems it is necessary to study a design for feeding strategies to optimize the fermentation process using a recombinant. In

view of the predominant occurrence of aspartic acid, glutamic acid and glycine in the GXI from the parent organism, it is anticipated that the cloning of this gene in *E. coli* will jeopardize the cellular metabolism of the host. The present studies were directed to analyze the effect of coordinated additon of these amino acids on the synthesis of recombinant GXI using 2^3 factorial design as a tool and to optimize its production on a higher level. The results demonstrate the utility of application of 2^3 factorial design to assess the effect of three variables at two different concentrations by running a reasonably limited number of experiments.

Glutamic acid, aspartic acid and glycine play an important role as they are metabolized to provide the intermediates of the citric acid cycle. Aspartic acid, serves as a precursor for the several amino acids such as arginine, methionine, threonine, isoleucine, lysine and in addition, also serves as a direct precursor in the pyrimidine biosynthesis. In a recombinant organism, where amplified expression of DNA *via* a multicopy plasmid is envisaged, there is an increased demand for the synthesis of nucleic acids which requires an adequate supply of purines and pyrimidines. The observed positive effect of aspartic acid, in the present study, may be attributed to meeting the increased metabolic demand of precursors in the pyrimidine biosynthesis.

CHAPTER 4

Characterization Of Acid-Induced Unfolding Intermediates Of Glucose/Xylose Isomerase

SUMMARY

Previous studies that have contributed towards our understanding of the protein folding problem have mainly dealt with monomeric proteins. Similar studies on oligomeric proteins are few and provide an insight regarding the tertiary and quaternary interactions. Acid-induced unfolding of the tetrameric glucose/xylose isomerase from *Streptomyces* sp. NCIM 2730 has been investigated using fluorescence quenching, second derivative spectroscopy, hydrophobic dye-binding (ANS) and circular dichroism techniques. It has been shown that GXI undergoes a two-step unfolding transition when the pH is lowered from 7.5 to 2. Based on its increased hydrophobic exposure, increased accessibility to quenching by acrylamide and native-like secondary structure the intermediate at pH 3.2 was defined as a molten globule state of GXI. The second intermediate at pH 5.0 is suggested to be an intermediate on the N→MG pathway, characterised by its pronounced secondary structure and ANS binding.

INTRODUCTION

The acquisition of a native, biologically functional conformation by a linear polypeptide chain has been called 'the second translation of the genetic code' (Anfinsen, 1973). Numerous *in vitro* protein folding experiments have demonstrated that proteins successfully achieve their correct native structures in the absence of cellular factors and without input of energy. With the advent of recombinant DNA technology, the major problem with overexpression of proteins is their improper folding and formation of inclusion bodies. The protein folding problem is of increasing practical concern for the biotechnology industry and for interpreting the growing database of DNA sequences (Inouye & Shinde, 1997). Protein folding is currently one of the most intensely investigated areas of structural biology.

The *in vitro* process of refolding differs considerably from the *in vivo* folding process. The release of the nascent polypeptide chain as it emerges from the ribosome during translation or translocation and its assembly into oligomeric complexes, is subject to a cellular milieu of proteins and thus to many oppurtunities for inappropriate associations. Although the proposal made by Anfinsen and co-workers has been the central dogma of protein folding, additional factors that enhance protein folding efficiencies such as peptidyl disulfide isomerase (Fischer *et al*, 1984), peptidyl prolyl isomerase (Freedman, 1991) and molecular chaperones (Ellis *et al*, 1991) have been demonstrated. The cell has evolved molecular chaperones and enzymes responsible for preventing misassociation and misfolding. In contrast to the *in vivo* process, the *in vitro* process is in a more dilute environment and the studies mainly emphasize on the refolding from the denatured states.

The protein folding problem essentially comprises of two aspects : *viz.* to determine what particular three-dimensional structure is specified by a given amino acid sequence and the sequence of events by which a protein folds to its native conformation. It was thought that the protein searches through all possible conformations to attain the most stable spatial configuration (Levinthal, 1968). Protein folding is a very rapid process, for instance a protein like ribonuclease takes less than 1 sec to acquire the three-dimensional conformation, hence the assumption made by Levinthal does not conform with the *in vivo* rate of folding, since on a biological time scale, the time taken by the molecule to acquire its most stable native conformation would be much larger (Jaenicke, 1987). This is the Levinthal paradox. The

"framework" model assumes 'local' interactions among the neighbours in the amino acids, the interactions that form helices and turns are the main determinants of protein structure (Kim & Baldwin, 1982).

Physical basis of the stability of the folded conformations

Interactions among spatially neighbouring amino acids can be divided into local interactions among monomers that are close together in the polypeptide chain and non-local interactions among monomers that are widely seperated in the sequence (Chan & Dill, 1991). The physical basis of the stability of the folded state is attributed to the local interactions like hydrogen bonds, van Der Waals interactions, electrostatic interactions and non-local interactions such as hydrophobic interactions. The hydrophobic interactions seem to be the most critcal aspect for stability of the normal folded state (Jaenicke, 1987).

Kinetics vs thermodynamics of protein folding

Until recently it has been generally accepted that the observed tertiary structure of a protein is thermodynamically and not kinetically controlled. Anfinsen's hypothesis states that the three dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups and temperature) is the one in which the Gibbs free energy of the whole system is lowest and the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment. This was further confirmed by Kim & Baldwin (1990) and Dill (1990). The hypothesis remains true even in case of chaperone- assisted folding, since the chaperone has no bearing on the rate of folding, it only navigates the protein to the sequential energetically stable conformation (Baker & Agard *et al*, 1994).

Stable conformations of proteins

The three stable conformations in which proteins are known to exist are the native, unfolded and the molten globule state. The native state is defined by its minimum free energy state whereas the unfolded state is achieved in the presence of a denaturant resulting in a polypeptide chain with all its side-chains exposed to solvent. In literature, the term denaturation has been used to refer to any changes from the biologically active, native state, including irreversible alterations such as covalent modifications and aggregation. Globular proteins are transformed from their native states into different non-native states under various

denaturing conditions, such as extreme pH, high denaturant concentrations and elevated temperatures (Tanford, 1971). The studies have mostly been directed to study conformational changes, that are in principle reversible. The transition of a globular protein from its native state to its denatured state has been usually considered to be highly cooperative (Dobson *et al*, 1992). Earlier the presence of only two stable states that is N and U was thought to exist, but the presence of intermediates on the pathway from N→U state has been shown based on spectroscopic evidences. One such intermediate state is the molten globule state.

A variety of proteins has been observed to exist in stable conformations that are neither fully folded or unfolded, and are termed as the 'molten globule state.' The molten globule has been rightly called the 'molecular state' of the decade. The term 'molten globule' has been coined by Ohgushi & Wada (1983) for an intermediate of α-lactalbumin (Kuwajima *et al*, 1976; 1977; Ptitsyn *et al*, 1990) which retained secondary structure and lacked the tertiary structure. The MG state thus represents a third conformational state and possesses several characteristic features such as (i) high content of native-like secondary structure, (ii) overall compactness with buried but highly mobile aromatic side-chains; (iii) exposure of hydrophobic surface, as indicated by dye binding ability and susceptibility to aggregation, (iv) lack of cooperative interactions, (v) rapid equilibration with the fully unfolded states, (vi) increase in the hydrodynamic volume as compared to the folded state.

Several proteins have shown the existence of this intermediate state. The MG state, being transient in nature, is difficult to detect but has been located under favourable circumstances. The MG state has been characterized under unphysiological conditions such as low pH, high temperature, high pressure, and presence of denaturing agents (Dolgikh *et al*, 1983; Goto & Fink, 1989; Holzman *et al*, 1990). The presence of kinetic intermediates resembling the MG state, has been shown for a number of proteins suggesting that it may be an essential step on the protein folding pathway (Elove *et al*, 1992; Goldberg *et al*, 1990; Radford *et al*, 1992). Different proteins behave differently on acid-denaturation (Fink *et al*, 1990). Some do not unfold at the lowest pH values, some undergo transition to a compact molten globule state A and a third type first unfolds to an extended conformation and then undergoes transition to a compact state A on addition of anions (Goto *et al*, 1990; Stigter *et al*, 1991).

Interest in the MG state has been due to its physiological role in its being recognized by the chaperone Gro EL, during insertion of the protein into the cell membrane (Martin *et al*, 1991; Van der Goot *et al*, 1991). Recent studies show that defective protein folding can be a cause of serious diseases. Potentially thermodynamic destabilization of the native or an intermediate state, alteration of the folding kinetics, prolonged or inappropriate associations with molecular chaperones or folding enzymes, preferential formation of off-pathway or toxic conformations or folding in an improper compartment could all lead to a loss in functional protein and thus resulting in a defective phenotype (Thomas *et al*, 1995). Cystic fibrosis, hypercholesterolemia, Maple syrup disease, Tay Sacch's disease, scrapie, Alzheimers are a few examples of consequences of improper protein folding.

Previous studies with monomeric proteins have contributed towards our understanding of the protein folding problem (Redfield *et al*, 1994; Fillipis *et al*, 1996; Fink *et al*, 1994; Khurana *et al*, 1994). Similar studies on oligomeric proteins are few and provide an insight regarding the tertiary and quaternary interactions. In the case of oligomeric proteins like alkaline phosphatase, glucose/xylose isomerase and phosphoglucose isomerase, partially folded dimeric intermediates were detected (Blackburn *et al*, 1981; Rao *et al*, 1991; Ghatge *et al*, 1994).

Many studies of large proteins consisting of multiple domains and subunits have demonstrated that the individual domains and subunits fold atleast approximately to their final conformations and then associate (Privalov, 1982; Jaenicke, 1984). Multimeric proteins are prone to aggregation upon perturbation of subunit interactions (Garel, 1992; Jaenicke, 1987). Except for the β_2 subunit of tryptophan synthase (Goldberg *et al*, 1990) and aspartate Aminotransferase (Leistler *et al*, 1992), all proteins that have been investigated so far are small single domain proteins.

The present work on GXI was initiated with a view to characterize the unfolding pathway of GXI and compare it with the monomeric α/β barrel proteins. 22 proteins are known to have the α/β barrel fold (Farber, 1993). The use of the α/β barrel domain as a building block for creating new proteins is the important question being currently addressed (Lesk *et al*, 1989; Pickett *et al*, 1992). These studies would help in defining the rules, which would then allow prediction of α/β barrel structures from amino acid sequences. Folding

studies have been carried out extensively with monomeric barrel structures. In contrast, there are no detailed reports on unfolding of multimeric barrel proteins. Most of the work on GXI has been focussed on the mechanistic aspects of its isomerization reaction, its genetic engineering and site-directed mutagenesis for improved productivity and properties (*refer chapter 1*). Studies in the present chapter discuss the molten globule-like intermediate of the tetrameric enzyme glucose/xylose isomerase from *Streptomyces* sp. NCIM 2730 during its acid-induced unfolding.

MATERIALS AND METHODS

Chemicals

High purity grade guanidine hydrochloride (GdnHCl), 1-anilino-8-naphthalene sulfonate (ANS), acrylamide and potassium iodide (KI) were obtained from Sigma Chemical Co., St. Louis, MO, (USA). All other reagents used were of analytical grade.

Buffers and Solutions

The buffers used for the spectroscopic measurements at different pH values were glycine-HCl (pH 2 and 3), sodium acetate (pH 4 and 5), sodium phosphate (pH 6, 7 and 7.5) at a concentration of 50 mM. Unfolding conditions were provided by GdnHCl (6M). The concentration of stock solution of GdnHCl (8 M) was determined by the measurement of the refractive index (Pace, 1986). ANS concentration was calculated spectrophotometrically using an extinction coefficient of 5000 M^{-1} cm^{-1} at 350 nm (Stryer, 1965). All the solutions were prepared in deionised water and filtered through a 0.45 μm filter.

Data analysis

All data were analyzed using the ORIGIN software provided by MicroCal.

Purification of GXI

Streptomyces sp. NCIM 2730, is an isolate of the National Chemical Laboratory, Pune, India. The organism was grown for 96 h at 28°C on Callens medium (Callens *et al*, 1986) containing xylose (1%) as the inducer. The cells were lysed by sonication and the extract was heated at 60°C to inactivate the proteases. The enzyme was concentrated by ammonium sulfate precipitation (0.9 saturation), purified to homogeneity by preparative gel electrophoresis followed by DEAE-cellulose treatment to remove acrylamide impurities (Gaikwad *et* al, 1989). The purified enzyme showed a single band at 40 kDa on sodium dodecyl sulfate polyacrylamide gel electrophoresison silver staining. The GXI activity was determined by the method of Dische & Borenfreund (1951) as modified by Marshall & Kooi (1957). One unit of enzyme activity was defined as the amount of enzyme which produces one micromole of fructose per min under the assay conditions. Protein concentration was estimated colorimetrically by Bradford's method (1976).

Fluorescence studies

Steady state fluorescence was recorded on a Perkin-Elmer Luminiscence spectrofluorimeter LS50B. The spectrofluorimeter was attached to a Julabo water bath to maintain the temperature of the samples and of the cuvette chamber in the range of 30 to 60°C. Intrinsic tryptophan fluorescence spectra were recorded by exciting the samples at 295nm with excitation and emission slit widths set at 5.0nm. The emission spectra were recorded in the range of 300-500 nm. Base-line corrections were done with buffer without protein in all cases.

Binding of ANS

Binding of ANS to GXI was studied by excitation of the dye at 375 nm and the relative emission spectra were recorded from 400 to 500 nm.

Quenching studies

Fluorescence quenching experiments were carried out by the addition of small aliquots of acrylamide stock solution (5 M) to the protein solution (0.625 μM) previously incubated at pH 3.2 or 7.5 at 25°C for 60 min and the fluorescence intensities were determined (λ_{ex}. 295 nm; λ_{em} 344 nm). Correction for the inner filter effect due to quencher was made using the formula

$$F_{\text{corrected}} = F_{\text{observed}} \text{ x antilog } [A_{ex}+A_{em}/2],$$

where A_{ex} is the absorption at the excitation wavelength and A_{em} is the absorption at the emission wavelength (Eftink & Ghiron, 1981). The quenching data was analyzed according to the Stern-Volmer equation

$$F_o / F = 1 + K_{sv} [Q],$$

where F_o and F are the fluorescence intensities in the absence and presence of a quencher [Q] and K_{sv} is the collisional quenching constant (Eftink & Ghiron, 1981; Lehrer, 1971). For proteins containing more that one fluorescing tryptophan residue differing in their accessiblity to the quencher, the Stern-Volmer plot will be non linear and hence a modified Stern-Volmer equation has been applied

$$F_o / F_o - F = 1 / K_{sv(eff)} [Q] f_{a(eff)} + 1/f_{a(eff)}$$

where F_0 is the fluorescence intensity in the absence of quencher, F is the difference in the fluorescence intensities in the absence and presence of the quencher [Q] and $f_{a\ (eff)}$ is the fraction of tryptophan residues accessible to the quencher (Lehrer, 1971).

Circular Dichroism

Circular Dichroism studies were performed on a Jasco Model J500A spectropolarimeter equipped with a DPN-1 data processor. Secondary structure of the GXI (0.625 μM) was monitored in the far-UV region (200 to 250 nm) using a path length of 0.1 cm. The tertiary structure of the GXI (1.25 μM) was monitored in the near-UV (250 to 320 nm) region using 0.5 cm path length cell. Each spectrum was recorded as an average of 4 scans. All measurements were made at 30°C. The molar ellipticity (θ) was calculated using the formula,

$$\theta = \frac{\theta_{observed} \times \text{molecular weight}}{10 \times l \times c},$$

where l is the length of the light path in cm and c is the concentration in gm ml^{-1} (Yang *et al*, 1986).

Second derivative absorption spectrum

Second derivative absorption spectroscopy was used to determine the changes in the tyrosine and tryptophanyl environment of the different intermediate states of GXI. Absorption of the protein (0.1 $mg.ml^{-1}$) was measured in the range 260-300 nm.

Light scattering studies

Rayleigh light scattering experiments were carried out with the spectrofluorimeter to follow protein aggregation. Both excitation and emission spectra were set at 400nm and the time dependent change in scattering intensity was followed.

Results and Discussion

Spectroscopic analysis of GXI

The pH dependent changes in GXI were followed by using fluorescence and circular dichroism studies. The intrinsic fluorophore tryptophan (Trp) is highly sensitive to the polarity of its surrounding environment (Lakowicz, 1983). The GXI from *Streptomyces* sp. NCIM 2730 contains 10 Trp residues per subunit (Gaikwad, 1990). The intrinsic fluorescence of GXI (λ_{ex} 295 nm) showed an emission maximum (λ_{em}) at 344 nm indicating the solvent exposure of some Trp residues at pH 7.5 (Fig. 4.1).

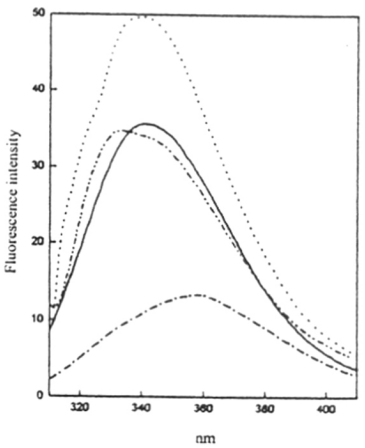


Fig. 4.1 Spectroscopic characterization of GXI.

GXI (λ_{ex} 295nm) at pH 7.5, (—); pH 5.0; (....) , pH 3.2; (- -), 6M GdnHCl, pH 7.5, (-··-)

At pH 5.0, GXI shows an enhanced fluorescence intensity at 340 nm, which is indicative of dissociation of the tetramers to monomers, resulting in an aggregated species which has been later confirmed by light scattering studies. GXI at pH 3.2 shows a blue-shifted emission maximum at 333 nm, indicative of burial of the Trp residues in a hydrophobic environment. Unfolding of GXI occurs in presence of 6M guanidine hydrochloride (Gdn HCl) indicated by the red shifted fluorescence at 357 nm as a result of solvent exposure of the tryptophan residues. The fluorescence patterns of GXI revealed that the enzyme exists in two different states at pH 3.2 and pH 5.0 distinct from the native (N) and unfolded (U) states.

In order to investigate the temperature dependence of pH-induced unfolding, the fluorescence of GXI was recorded at different pH in the temperature range of 30-60°C. The ratio of the relative fluorescence intensity of GXI at 330 and 350 nm was measured to detect small changes in emission maximum. A decrease in $F_{330/350}$ is indicative of a red shift whereas increase in the ratio signifies a blue shift (Jiang & London, 1990). The decrease in the $F_{330/350}$ ratio at pH 7.5 with an increase in temperature indicated gradual unfolding of the enzyme (Fig. 4.2). $F_{330/350}$ of GXI does not change significantly at 30°C between pH 3.0 and 7.5. However,

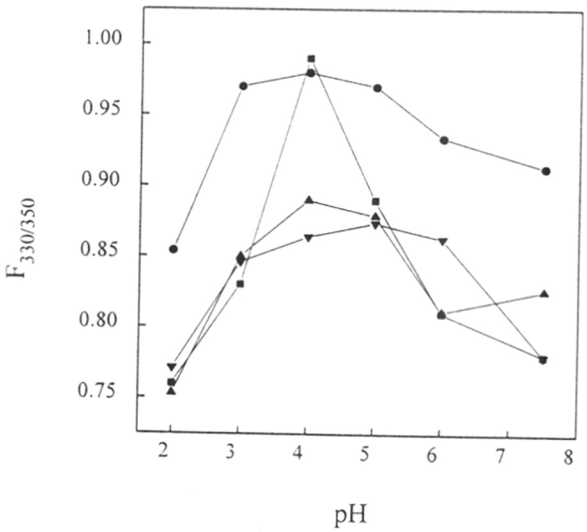


Fig. 4.2 Effect of pH on the fluorescence emission ratio $F_{330/350}$ of GXI at different temperatures.

30°C(•), 40°C (▲), 50°C (▼) and 60°C (■)

there is a decrease in the ratio at 40 and 50°C, suggesting that at higher temperatures there is a change in the protein tertiary structure resulting in the exposure of the buried Trps to the polar solvent. Further increase in temperature results in drastic changes in the ratio as seen by the sudden red shift at pH 6.0 followed by the blue shift at pH 4.0. It is important to note that at pH 3.0 and 5.0, there is no significant change in the $F_{330/350}$ ratio in the temperature range of

40- 60°C confirming that at these pH values, the enzyme exists in stable intermediate states in the N ⇔ U transition.

The changes in secondary and tertiary structure of GXI were followed as a function of pH by far and near-UV circular dichroism studies, respectively. The CD spectrum of native GXI at pH 7.5, in the far-UV region (200-250 nm) exhibited a strong negative ellipticity in the region 215-222 nm and a weaker one at 210 nm, characteristic of a protein having an α-helix. The changes in the negative band in the far-UV region were monitored for GXI at different pH (Fig 4.3 A). GXI at pH 5.0 showed a pronounced ellipticity and an altered secondary structure. At pH 3.2 the enzyme exhibited a slightly decreased ellipticity as compared to that at pH 7.5, and a non-native secondary structure. Incubation of GXI with 6M Gdn HCl resulted in disruption of the secondary structure indicative of unfolding of GXI.

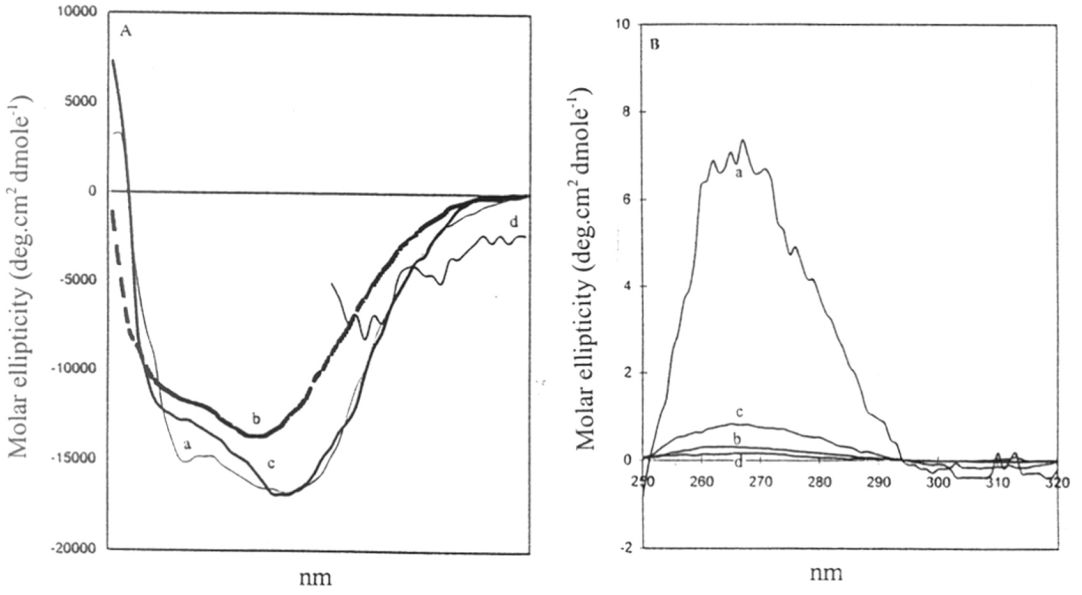


Fig. 4.3 Circular dichroism spectra of GXI

A, Far-UV; **B**, near-UV

a, pH 7.5; b, pH 3.2; c, pH 5.0; d, 6M Gdn HCl, pH 7.5.

It was observed that ellipticity in the near-UV region decreased with decreasing pH. The near-UV CD spectrum of GXI exhibited a fine structure characteristic of a folded protein with a positive ellipticity, contributed by its aromatic residues (Fig. 4.3 B). However, there was a drastic decrease in the ellipticity from pH 7.5 to 3.2, indicating unfolding of the tertiary structure upon acidification. It has been reported by Strickland (1974) that addition of acid induces a looser and more flexible environment around the aromatic residues. GXI showed a loss in tertiary structure in presence of 6M Gdn HCl.

Binding of ANS to GXI

In the present studies, the fluorophore ANS was used to determine the relative amount of exposed hydrophobic surfaces in the folding intermediates of GXI. ANS is not fluorescent in aqueous solution (λ_{em} 520 nm). However, in presence of proteins containing hydrophobic pockets its emission maximum shifts to a shorter wavelength and the emission intensity is enhanced. Binding of ANS has been widely used to detect the formation of molten globule-like intermediates in the folding pathways of several proteins (Semisotnov *et al*, 1991; Stryer, 1965). This property of ANS was used to differentiate between the conformations of GXI at pH 3.2 and pH 5.0 and to locate the possible molten globule state of GXI.

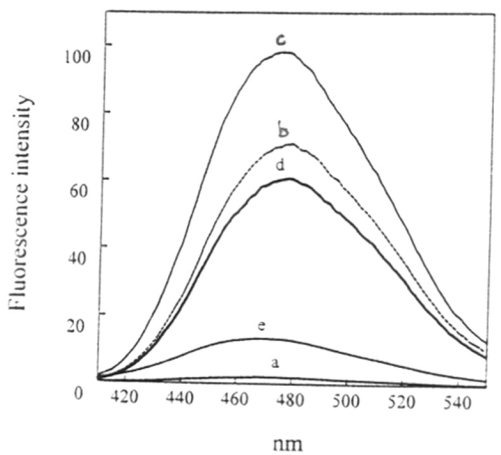


Fig. 4.4 ANS fluorescence of GXI as a function of pH.

GXI in presence of ANS (λ_{ex} 375nm, λ_{em} 470) at a, pH 7.5; b, pH 5.0; c, pH 3.2; d, pH 2.0 and e, 6M Gdn HCl.

The fluorescence intensity of ANS in the presence of GXI at pH 3.2 is higher than that observed at pH 2.0, 5.0, 7.5 and in presence of 6M Gdn HCl. Based on its maximum fluorescence upon ANS binding, the MG state for GXI has been detected more accurately at pH 3.2 (Fig.4.4). The pH 5.0 state also showed substantial ANS binding as compared to the native and unfolded state, indicative of exposure of hydrophobic regions.

Fluorescence quenching

In order to probe changes in the Trp environments and their exposure to solvent during the pH perturbed unfolding, fluorescence quenching studies were undertaken. The structural transitions at pH 3.2 and 7.5 were detected by studying the fluorescence quenching of the two pH states using the polar quencher. Acrylamide is an effective quencher of Trp fluorescence and because of its polar nature, it does not quench completely buried Trp residues, but quenches the surface exposed and partially buried Trp residues (Eftink & Ghiron, 1976). Fig. 4.5a shows a linear Stern-Volmer plot for intrinsic protein fluorescence quenching by acrylamide at pH 7.5 and 3.2.

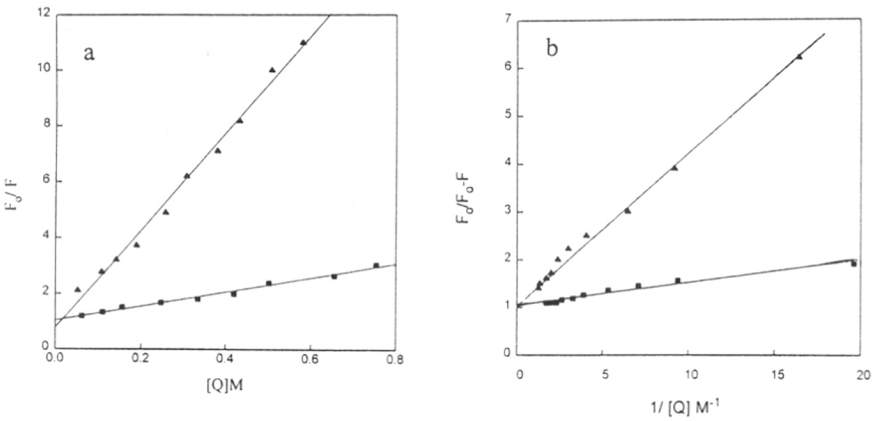


Fig. 4.5 Fluorescence quenching of GXI by acrylamide.

a. Stern-Volmer plot for pH 3.2 (▲) and pH 7.5 (■)

b. Modified Stern-Volmer plot pH 3.2 (■) and pH 7.5 (▲)

The quenching constant (K_{sv} values) calculated for the pH 7.5 and pH 3.2 were found to be 2.28 and 16.3 M^{-1} respectively. The modified Stern-Volmer plot (Fig. 4.5b) permits the estimation of the fraction of the total Trp residues quenched. At pH 3.2, the fractional accessibility (f_a) is 1 indicating that all the Trps are accessible to the quencher (Fig. 4.5b). The pH 7.5 quenching plot also shows a value f_a of 1, with a different quenching constant, confirming the differential environments of Trp residues under the two pH conditions. Since GXI is a heterogeneously emitting system, no quantitative interpretation of the data can be given. The conclusion that can be drawn is that compared to the native enzyme, the Trp residues in the MG state showed increased accessibility to the quenching agent.

Second derivative absorption spectra

Conformational and microenvironmental changes of side-chain residues which occur during the unfolding of proteins can be detected by the second derivative absorption spectroscopy. It has been proved to be an effective analytical tool to resolve the complex protein absorption spectrum of Phe, Tyr and Trp residues (Ichikawa *et al*, 1981). In the second derivative spectra such milieu-dependent alterations are reflected in the changes of the band amplitude of respective amino acids. The peak to trough distances in the 280-295 nm region expressed as *a/b* ratio are related to the polarity of the medium in which Tyr residues are embedded (Ragone *et al*, 1984). In order to determine the changes occuring in the microenvironment of Tyr residues, second derivative spectra of acid unfolded intermediates of GXI was studied (Fig. 4.6). The second derivative spectrum for the model compound N-AcTrpNH$_2$ shows two maxima centered around 287 and 295 nm and two minima at 283 and 290.5 nm, the position of which is marginally affected by solvent polarity (Ragone *et al*, 1984). The peak to peak distance between the maximum at 287 nm and the minimum at 283 nm is indicated as *a* and the peak to peak distance between the maximum at 295 nm and the minimum at 290.5 nm is indicated as *b*. Since GXI is known to contain 30 Tyr residues per mole (Gaikwad, 1990), it was of interest to determine any changes in the absorption spectrum in the various pH states . There was not much change in the position of maxima and minima at pH 3.2, 5.0, 7.5 and in the unfolded state. However, there was a substantial difference in the ratio of *a/b* obtained at pH 7.5, 5.0, 3.2 and in the unfolded state. The *a/b* ratio for GXI at pH 7.5 was found to be 0.67, whereas that at pH 3.2 and pH 5.0 was found to be 0.63 and 0.65

respectively. The low ratio of *a/b* observed for pH 3.2 was indicative of an alteration in the environment of the Tyr residues, resulting in the burial of these residues in a hydrophobic

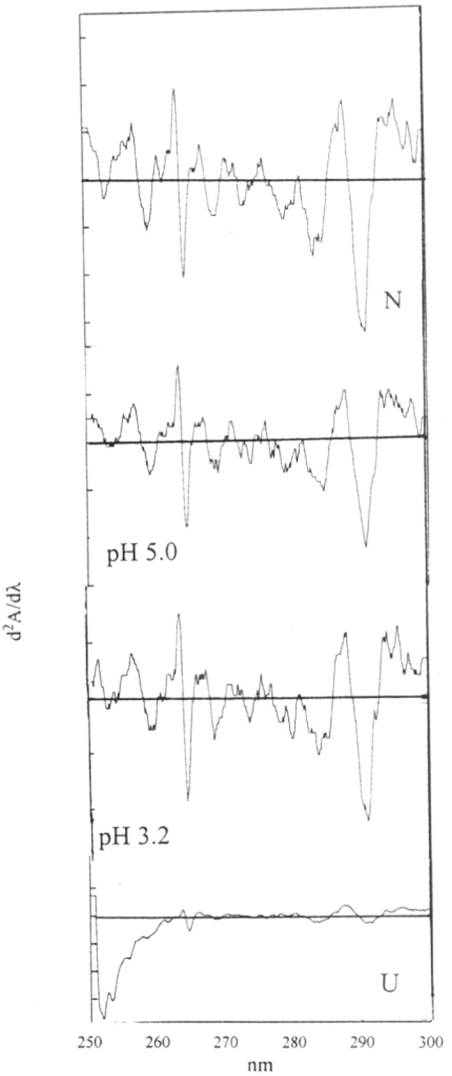


Fig. 4.6 Second derivative spectra of GXI under various pH conditions

environment. In presence of 6M Gdn HCl, GXI showed a ratio of 0.96, as a result of the solvent exposure of the Tyr residues in the unfolded state.

Scattering studies

The molten globule has a propensity for aggregation due to exposure of hydrophobic regions. Rayleigh light scattering studies were undertaken to determine whether GXI at pH 3.2 exists in an aggregated state (Fig. 4.7).

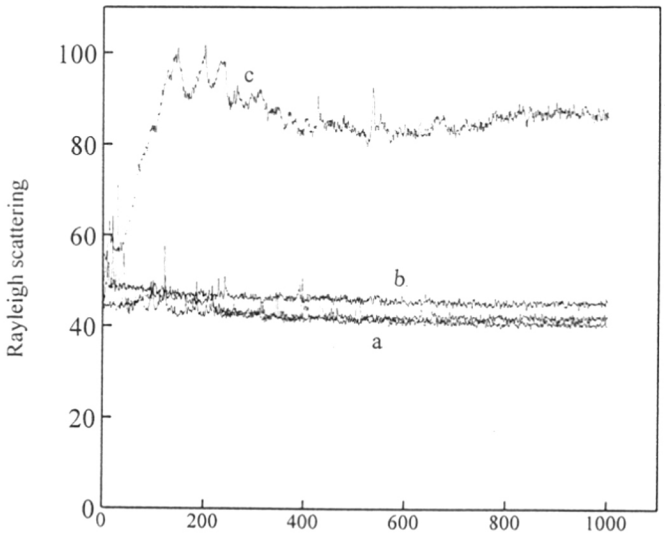


Fig. 4.7 Scattering studies of GXI

a, pH 7.5, b, pH 3.2, c, pH 5.0

No increase in the light scattering value indicated the absence of aggregation at pH 3.2. However GXI at pH 5.0 showed a rapid increase in scattering, which may be indicative of dissociation of the tetramers into monomers followed by aggregation. This also may be the plausible reason for the increase in fluorescence observed at this pH.

The molten globule state of proteins is characterized by an ordered secondary structure similar to the native state and by a disordered tertiary structure. In this conformation the protein molecule has a compact structure but lacks the dense packing of side chains (Uversky & Ptitsyn, 1994). A decrease in the molar ellipticity reveals that the secondary structure is retained in the MG state, but it is different from that in the native state. The loss of tertiary structure is a characteristic of the molten globule state. The acid-induced molten globule state of GXI has a looser and more flexible environment around the aromatic residues.

Goto *et al* (1990) have proposed that upon acid-titration, the intramolecular charge repulsions are the driving force for partial unfolding of the protein molecule. The electrostatic interactions of the protons in the aqueous medium with the positively charged centres on the protein decrease the internal repulsive forces and favour the intrinsic hydrophobic interactions of proteins, which may be responsible for the blue shifted fluorescence of the acid-induced molten globule of GXI. A similar blue-shifted fluorescence has been reported for the molten globule states of bovine growth hormone (Goto & Fink, 1989) and interferon-γ (Nandi, 1997). The presence of anions has been reported to be essential for the formation of A state of β-lactamase and cytochrome c (Goto *et al*, 1990). Proteins such as α-lactalbumin and creatinase do not require a high salt concentration for the A-state formation (Kuwajima et al, 1985, Schumann & Jaenicke, 1993). Similarly the molten globule of GXI also does not require the presence of salt.

The present studies describe the existence of an acid molten globule state at pH 3.2 which shows a characteristic secondary structure, disrupted tertiary structure, binding with ANS, an altered microenvironment of Trp and Tyr residues as shown by acrylamide quenching and second derivative spectroscopy. Light scattering studies revealed non-formation of aggregates of the molten globule state.

Fluorescence as well as CD data support the involvement of an intermediate state at pH 5.0. The pH 5.0 state bears resemblance to a pre-molten globule state which lies on the N→MG pathway, characterized by a native-like secondary structure and an altered tertiary structure indicative of an altered microenviroment of the Trp and Tyr residues as compared to the native and MG state at pH 3.2. Both ANS binding and light scattering data suggest the exposure of hydrophobic pockets in this state. These studies are in agreement with the pre-molten globule state described for the α-lactalbumin (Kaul & Lala, 1992; Gussakovsky & Haas, 1995).

In conclusion, the present studies reveal the presence of a molten globule state of GXI at pH 3.2 and an intermediate analogous to the pre-molten globule state localized on the N→MG pathway.

BIBLIOGRAPHY

Allen, K. N., Lavie, A., Farber, G. K., Glasfeld, A., Petsko, G. A. and Ringe, D. (1994a) *Biochemistry* **33,** 1481-1487.

Allen, K. N., Lavie, A., Glasfeld, A., Tanada, T. N., Gerrity, D. P., Carlson, S. C., Farber, G. K., Petsko, G. A. and Ringe, D. (1994b) *Biochemistry* **33,** 1488-1494.

Allen, K. N., Lavie, A., Petsko, G. A and Ringe, D. (1995) *Biochemistry* **34,** 3742-3749.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403-410.

Amore, R. and Hollenberg C. P. (1989a) *Nucleic Acids Res.* **17,** 7515.

Amore, R., Wilhelm, M. and Hollenberg C. P. (1989b) *Appl. Microbiol. Biotechnol.* **30,** 351-357.

Anfinsen, C. B. (1973) *Science* **181,** 223-230.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. K. and Strutki, K. A. (1987) In : *Current Protocols in Molecular Biology*, John Wiley & Sons, 6.4.1- 6.4.7.

Bailey, J.E. (1993) *Advances in Biochem.Engg. Biotechnol.* **48,** 30-52.

Bannerjee, S., Archana, S. and Satyanarayana, T. (1994) *Curr. Microbiol.* **29,** 349-352.

Bartfay, J. (1960) *Nature* (London) **185,** 924.

Basuki, W., Iizuka, M., Ito, K., Furuichi, K., and Minamiura, N. (1992) *Biosci. Biotech. Biochem.* **56,** 180-185.

Batt, C. A., Jamieson, A. C. and Vandeyar, M. A. (1990) *Proc. Natl. Acad. Sci.*, USA **87,** 618-622.

Batt, C. A., O'Neill, E., Novak, S. R., Ko, J. and Sinskey, A.. (1986) *Biotechnol. Progress* **2,** 140-144.

Beck, C.F. and Warren, R. A. J. (1988) *Microbiol. Rev.* **52,**318-326.

Bejar, S., Belghith, K., Gargouri, R. and Ellouz, R. (1994) *Biotech. Lett.* **16,**1259-1264

Bengston, B. L. and Lamm, W. R. (1973) *Fr. Pat* 2172882.

Blackburn, M. N. and Noltmann, E. N. (1981) *Arch. Biochem. Biophys.* **212,** 162-169.

Blow, D. M. and Collyer, C. A. (1990) 6th International symposium. *Genetics of Industrial Microorganisms*, Strassberg, France, 12-16 August.

Blow, D. M., Collyer, C. A.. Goldberg, J. D. and Smart, O. S. (1992) *Faraday Discuss.* **93,** 67-73.

Bok, S. K., Seidman, W. and Wopat, P. W. (1984) *Appl. Environ. Microbiol.* **47,** 1213-1215.

Bor, Y., Moraes, C., Lee, S., Crossby, W. L., Sinskey, A. J. and Batt, C. A. (1992) *Gene* **114,** 127-131.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. In : Statistics for experimenters. - An Introduction to Design, data, analysis and model building, Chapter 10 : Factorial designs at two levels (1978) 306-325.

Bradford, M. M. (1976) *Anal.Biochem.* **72,** 248-254.

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

Buchner, J., Renner, M., Lilie, H., Hinz, H. J., Jaenicke, R., Kiefhaber, T. and Rudolph, R. (1991) *Biochemistry* **30,** 6922-6929.

Bucke, C. (1981) In : *Topics in enzyme and fermentation biotechnology* Vol. 1, Industrial Glucose Isomerase, p. 147-171. Ed. A.. Wiseman, Ellis Horwood, Chichester, U.K.

Bucke, C. (1983) Glucose transforming enzymes, 93-127. In : *Microbial enzymes and biotechnology.* Ed. W. Fogarty. Applied Science Publishers, London.

Callens, M., Kersters-Hilderson, H., Van Opstal, O. and Debruyne, C. K. (1986) *Enzyme. Microb. Technol.* **8,** 696-700.

Callens, M., Kersters-Hilderson, H., Vangrysperre, W. and Debruyne, C. K. (1988a) *Enzyme. Microb.Technol.* **10,** 695-700.

Callens, M., Tomme, P., Kersters-Hilderson, H., Cornelis, R., Vangrysperre, W. and deBruyne, C. K. (1988b) *Biochem. J.* **250,** 285-290.

Carell, H. L., Hoeir, H. and Glusker, J. P. (1994) *Acta. Cryst.* **50,** 113-123

Carell, H. L., Rubin, B. H., Hurley, T. J. and Glusker, J. P. (1984) *J. Biol. Chem.* **259,** 3230-3236.

Carrell, H. L., Glusker, J. P., Burger, V., Manfre, F., Tritsch, D. and Biellman, J. F. (1989) *Proc. Natl. Acad.. Sci,* USA **86,** 4440-4444.

Carrell, H.L., Hoer, H. and Glusker, J.P. (1994) *Acta Crystallogr. Sect.* **D50,** 113-123.

Cha, J., Cho, Y., Whitaker, R.. D., Carrell, H. L. ,Glusker, J. P., Karplus, P. A. and Batt, C. A. (1994) *J. Biol. Chem.***269,** 2687-2694.

Chan, E., Ueng, P. D. and Chen, L. F. (1989) *Appl. Microbiol. Biotechnol.* **31,** 524-528.

Chauthaiwale, J. V. and Rao, M. B. (1994) *Appl. Environ. Microbiol.* **60,** 4495-4499.

Chen, W. P. (1980a) *Process Biochemistry* **15,** 30-35.

Chen, W. P. (1980b) *Process Biochemistry* **15,** 36-41.

Chen, W. P., Anderson, A. W. and Han, Y. W. (1979) *Appl. Environ. Microbiol.* **37,** 324-331.

Collyer, C. A. and Blow, D. M. (1990) *Proc. Natl. Acad. Sci*, USA **87,** 1362-1366.

Collyer, C. A., Hendrick, K. and Blow, D. M. (1990) *J. Mol. Biol.* **212,** 211-235.

Dahl, M. K., Degenkolb, J. and Hillen, W. (1994) *J .Mol. Biol.* **243,** 413-424.

David, J. D. and Weissmeyer, H. (1970) *Biochem. Biophys. Acta* **201,** 497-499.

Dekker, K., Yamagata, H., Sakaguchi, K. and Udaka, S. (1991a) *Agric. Biol. Chem.***55,** 221-227.

Dekker, K., Yamagata, H., Sakaguchi, K. and Udaka, S. (1991b) *J. Bacteriol.* **173,** 3078-3083.

Dische, Z. and Borenfreund, E. (1951) *J. Biol. Chem.* **192,** 583-587.

Dobson,C. M. (1992) *Curr.Opin.Struct.Biol.* **2,** 6-12.

Dodd, I. B. and Egan, J. B. (1990) *Nucleic Acids Res.* **18,** 5019-5026.

Dolgikh, D. A., Kolomieto, A. P., Bolotina, I. A. and Ptitsyn O. B. (1984) *FEBS Lett.* **165,** 88.

Drocourt, D., Bejar, S., Calmels, T., Reynes, J. P. and Tiraby, G. (1988) *Nucleic Acids Res.* **16,** 337

Eftink, M. R. and Ghiron, C. A. (1981) *Anal. Biochem.* **114,** 199-227.

Eftink, M. R., and Ghiron, C. A. (1976) *Biochemistry* **15,** 672-680.

Ellis, R.J. and van der Vies, S. M. (1991) *Ann. Rev. Biochem.***60,** 321-347.

Elove, G. A., Chaffotte, A. F., Roder, H.and Goldberg, M. E. (1992) *Biochemistry* **31,** 6876-6883.

Ewbank, J. J. and Creighton, T. E. (1991) *Nature* **350,** 518-52.

Farber, G. K., Glasfeld, A., Tiraby, G., Ringe, D. and Petsko, G. A.(1989) *Biochemistry* **28,** 7289-7297.

Farber, G. K., Petsko, G. A. and Ringe, D. (1987) *Protein.Engg.* **1,** 459-466.

Farber , G.K. and Petsko, G.A. (1990) *Trends Biochem. Sci.* 15, 228-234.

Feldman, S. D., Sahm, H. and Sprenger, G. A. (1992) *Mol.Gen.Genet.* **234,** 201-210.

Fillipis V. D, Polverino de Laureto, P., Toniutti, N. and Fontana, A. (1996) *Biochemistry* **35,** 11503-11511.

Fink, A. L., Calciano, L. J., Goto, Y., Kurotso, T. and Palleros, D. R. (1994) *Biochemistry* **33,** 12504-12511.

Fischer, G., Bang, H. and Mech, C. (1984) *Biomed. Biochim. Acta* **10,** 1101-1111.

Freedman, R.B. In Conformation and Forces in Protein Folding (Nall, B.T., Dill, R.A. eds.) AAAS, Washington, D.C. (1991) 204-214.

Gaikwad, S.M. (1990) *Ph.D thesis, University of Pune.*

Gaikwad, S. M., More, M. W., Vartak, H. G. and Deshpande, V. V. (1988) *Biochem. Biophys. Res. Commun.* **155,** 270-277.

Gaikwad, S. M., Pawar H. S., Vartak H. G. and Deshpande, V. V. (1989) *Biochem Biophys.Res.Commun.* **159,** 457-463.

Gaikwad, S. M., Rao, M. B. and Deshpande, V. V. (1992a) *Enz. Microb. Technol.* **14,** 317-320.

Gaikwad, S. M. and Deshpande, V. V. (1992b) *Enzyme. Microb. Technol.* **14,** 855-858.

Gaikwad, S. M., Rao, M. and Deshpande, V. (1992c) *Enzyme. Microb. Technol.* **15,**155-157.

Gartner, D., Degenkolb, J., Ripperger, J. A. E., Allmansberger, R. and Hillen, W. (1992) *Mol. Gen. Genet.* **232,** 415-422.

Gartner, D., Geissendorfer, M. and Hillen, W. (1988) *J. Bacteriol.* **170,** 3102-3109.

Ghangas, G. S. and Wilson, D. B. (1984) *J. Bacteriol.* **157,** 158-164.

Ghatge M. S., Deshpande, V. V. and Ranjekar P. K. (1995) *W. J .Microbiol. Biotechnol.*, **11,** 144.

Ghatge, M. S. and Deshpande, V. V. (1993) *Biochem. Biophys. Res.Commun.* **193,** 979-984.

Gavel et al 1992

Ghatge, M. S., Mawal, Y., Gaikwad, S. M. and Deshpande, V. (1991) *Appl. Biochem. Biotechnol.* **31,**11-20

Ghatge, M. S., Phadatare, S. U., Bodhe, A. M. and Deshpande,V. V. (1994) *Enz. Microb. Technol.* **16,** 323-327.

Ghatge, M., Mawal, Y., Gaikwad S. and Deshpande V. V. (1995) *Appl. Biochem. Biotechnol.* **31,** 11-20.

Gilbert, M., Morosoli, R., Shareck, F. and Kluepfel, D. (1995) *Crit. Rev. Biotechnol.* **15,** 13-39.

Glasfeld, A., Farber, G. K., Ringe, D., Marcel, T., Drocourt, D, Tiraby, G. and Petsko, G. A. (1988) *J. Biol. Chem.* **263,** 29, 14612-14613.

Goldberg, M. E., Semisotnov, G. V., Friguet, B., Kuwajima, K., Ptitsyn, O. B., and Sugai, S. (1990) *FEBS Lett.* **223,** 327-329.

Goto, Y. and Fink, A. L. (1989) *Biochemistry* **28,** 945-952.

Goto, Y., Calciano, L. J., and Fink, A. L. (1990) *Proc. Natl. Acad. Sci, USA.* **87,** 573-577.

Goto, Y.,Takahashi, N. and Fink A. L. (1990) *Biochemistry* **29,** 3480-3488.

Griko, Yu. V., Venyaminov, S. Yu. and Privalov, P. L. (1989) FEBS *Lett.* **244,** 276-278.

Gussakovsky, E. E. and Haas, E. (1995) *Protein Science* **4,** 2319-2326.

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

Harcum, S. W., Ramirez, D. M. and Bentley, W. E. (1992) *Appl. Biochem. Biotechnol.* **34/35,** 161-173.

Harwood, C. R. (1992) *Trends in Biotechnol.* **10,** 247-256.

Henrick, K., Collyer, C. A. and Blow, D. M. (1989) *J.Mol.Biol.* **208,** 129-157.

Higgins, D. M. and Sharp, P. M. (1988) *Gene* **73,** 237-244.

Ho, N. W. Y., and Stevis, P. E. (1985) *Enzyme. Microb. Technol.* **7,** 592-596.

Ho, N. W. Y., Rosenfeld, S., Stevis, P. and Tsao, G. T. (1983) *Enzyme. Microb. Technol.* **5,** 417-420

Hodgson, J. (1994) *Bio/technology* **12,** 789-790.

Holzman T. F., Dougherty, J. J., Brems D. N. and MacKenzie, N. E. (1990) *Biochemistry* **29,** 1255-1261.

Hopwood, D. A., Bibb, M. J., Chatter, K. F., Kieser,T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P. and Schremp, H. (1985). *In* : Genetic manipulation of *Streptomyces*; a laboratory mannual, The John Innes Foundation,Norwich.

Huang, J. J. and Ho, N. W. Y. (1985) *Biochem. Biophys. Res. Commun.* **126,** 1154-1160.

Jacob, S., Allmansberger, R., Gartner, D. and Hillen, W. (1991) *Mol. Gen. Genet.* **229,** 189-196.

Jaenicke, R. (1984) *Angew. Chem. Int. Ed. Engl.* **23,** 395-413.

Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol.* **49,** 117-237.

Jaenicke, R. (1991) *Biochemistry* **30,** 3147- 3161.

Jamieson, A.C. and Batt, C. A. (1992) *Protein. Eng.* **5,** 235-240.

Jeng, M. F. and Englander, S. W. (1991) *J. Mol. Biol.* **221,** 1045-1061.

Jenkins, J., Janin, J., Rey, F., Chiadmi, M., Tilbergh, H., Lasters, I., DeMaeyer, M., Belle, D. V., Wodak, S. J., Lauwerays, M., Stanssens, P., Mrabet, N. T., Snauwaert, J., Matthyssens, G. and Lambeir, A. M. (1992) *Biochemistry* **31,** 5449-5458.

Jiang , J. X. and London, E. R. (1990) *J. Biol.Chem*, **265,** 8636-8641.

Katz, E., Ehrenthal, I. and Scallet, B. L. (1972) *US Patent* **3,**690,948

Kauder, C., Allmansberger, R., Gartner, D. Schmiedel, D. and Hillen, W. (1993) *FEMS Microbiol. Lett.* **109,** 81-84.

Kaul, P and. Lala, A. K. (1992) *J. Biol. Chem.* **267,** 19914-19918.

Kho, Y. H. (1984) *Kor. J. Appl. Microbiol. Bioeng.* **12,** 253-259.

Khurana, R. and Udgaonkar, J. B. (1994) *Biochemistry* **33,** 106-115.

Kikuchi, T., Itoh, Y., Kasumi, T, Fukazawa, C. (1990) *Agric. Biol.Chem.* **54,** 2469-2472

Kim, H. R. and Pyong, Su O. (1992) *J. Microbiol. Biotechnol.* **2,** 78-84.

Kim, P. S. and Baldwin R. L. (1982) *Annu. Rev. Biochem.* **51,** 459-489.

Kim, P. S. and Baldwin R. L. (1990) *Annu. Rev. Biochem* **59,** 631-660.

Krashininnikova, L. V., Rassadina, G. V., Kirsanova, S. V., Khromova, S. N., Yusibov, V. M., Pak, C., Andrianov, V. M. and Piruzyan, E. S. (1991) *Mol.Gen.Mikrobiol.Virusol.* **11,** 17-20 (Russ.)

Kraus A., Hueck , C., Gartner, D. and Hillen, W. (1994) *J. Bacteriol.* **176,** 1738-1745.

Tohekawa et al 1981

Kreuzer, P., Gartner, D., Allmansberger, R. and Hillen, W. (1989) *J. Bacteriol.* **171,** 3840-3845.

Kume, T. and Takahisa, M. (1983) *Agric. Biol. Chem.* **47,** 359-363.

Kuwajima, K., Nitta, K., Yoneyama, M. and Sugai,S. (1976) *J. Mol. Biol.***106,** 359-373.

Kuwajima, K. (1989) *Proteins : Struct.Funct.Genet.* **6,** 87-103.

Laemmli, U. K. (1970) *Nature* **227,** 680-685.

Lakowicz, J. R. (1983) in *Principles of Fluorescence Spectroscopy*, Plenum Press, New York. 42

Lambeir, A., Lauwereys, M., Stanssens, P., Mrabet, N. T., Snauwaert, J., Tilbergh, H. V., Matthyssens, G., Lasters, I., Maeyer, M. D., Wodak, S. J., Jenkins, J., Chiadmi, M. and Janin, J. (1992) *Biochemistry* **31,** 5459-5466.

Lastick, S. M., Mohagheghi, A., Tucker, M. P. and Grohmann, K. (1990) *Appl. Biotechnol. Bioeng.* **24/25,** 431-439.

Lastick, S. M., Tucker, M. V., Mackedomski, V. and Grohman, K. (1986) *Biotechnol. Lett.* **8,** 1-6.

Lastick, S. M., Tucker, M. Y., Beyett, J. R., Noll, G. R. and Grohmann, K. (1989) *Appl. Microbiol. Biotechnol.* **30,** 574-579.

Lavie, A., Allen, K. N., Petsko, G. A and Ringe, D. (1994) *Biochemistry* **33,** 5469-5480.

Lawlis, V. B., Dennis, M. S., Chen, E. Y., Smith, D. H. and Henner, D. J. (1984) *Appl. Environ. Microbiol.* **47,** 15-21.

Lee, C. K. (1977) *U.S. Patent*, 4061539

Lee, C., Bagdasarian, M., Meng, M. and Zeikus, J. G. (1990a) *J. Biol. Chem.* **265,** 19082-19090.

Lee, C., Bhatnagar, L., Saha, B. C., Lee, Y., Takagi, M., Imanaka T., Bagdasarian, M. and Zeikus, J. G. (1990b) *Appl. Environ. Microbiol.* **56,** 2638-2643.

Lee, C., Saha, B. C and Zeikus, J. G. (1990c) *Appl. Environ. Microbiol.* **56,** 2895-2901.

Lehrer, S. S. (1971) *Biochemistry* **10,** 3254-3263.

Leistler, B., Herold, M.and Kirschner, K. (1992) *Eur. J. Biochem.***205, 603-611**

Lesk , A. M., Braenden, C. I. and Chothia, C. (1989) *Proteins : Struct. Funct. Genet.***5,** 139-148.

Levinthal, C. (1968) *J. Chim. Phys.* **65,** 44-45.

Liu, C. Q., Lucas, R. J., Daly, J. M., Ricard, P. A. D. and Dann, J. M. (1987) *J. Biotechnol.* **6,** 159-165.

Liu, S.Y., Weigel, J. W. and Gheradim, F.C. (1996) *J. Bacteriol.* **178,** 3938-3943.

Lokman, B. C., van Santen, P., Verdoes, J. C., Kruse, J., Leer R. J., Posno, M and Pouwels, P. H. (1991) *Mol. Gen.. Genet.* **230,**161-169.

Loviny-Anderton, T., Shaw, P., Shin, M. and. Hartley, B. S. (1991) *Biochem. J.* **2,** 263-271.

Maleszka, R., Wang, P. Y and Schneider, H. (1982) *Can. J. Biochem.* **60,** 144-151.

Marcel, T., Drocourt, D. and Tiraby, G. (1987) *Mol. Gen. Genet.* **208,** 121-126.

Marshall, R. O. and Kooi, E. R. (1957) *Science* **125,** 648-649.

Martin J., Langer, T., Boteva, R., Schamel, A., Horwich A. L.,and Hartl, F. U. (1991) *Nature* **352,** 36-42.

Meaden, P. G., Andrise-Opoku, J., Reizer, J., Reizer, A., Lanceman, Y. A.,. Martin, M. F and Mitchell, W. J. (1994) *Gene* **141,** 97-101.

Meng, M., Bagdasarian, M. and Zeikus, J. G. (1993) *Bio/technology* **11,** 1157-1161.

Meng, M., Lee, M., Bagdasarian, M. and. Zeikus, J. G. (1991) *Proc. Natl. Acad. Sci.,* USA. **88,** 4015-4019.

Messing, J. (1983) *Met. Enzymol.* **101,** 20-78

Miles Laboratories Inc. (1974) *U.K. Patent*, 1376787.

Moes, C. J., Pretorius, I. S. and van Zyl, W. H. (1996) *Biotechnol.Lett.* **18,** 269-274.

Nandi, P. K. (1998) *Int. J. Biol. Macromol.***22,** 23-31.

Norova, G. E., Kaliev, A. B., Andrianov,V. M. and Piruzyan, E. S. (1991) *Bioteknologiya* **2,** 16-18.

Ogden, S., Haggerty, D., Stoner, D., Kolodrubetz, D. and Schleif, R. (1980) *Proc. Natl. Acad. Sci.*, USA. **77,** 3346-3350.

Ohgushi, M. and Wada, A. (1983) *FEBS Lett.* **164,** 21-24.

Pace, N. C. (1986) *Mtds.in Enzymol.* **131,** 266-279.

Parrish, F. W. (1970) *US Patent* **3,** 514,327

Pickett , S. D., Mansoor, S., Sternberg, A. and Michael, J. E. (1992) *J. Mol. Biol.* **228,** 170-187.

Piruzyan, E., Andrianov, S., Yushibov, V. M. and Mett, V. L. (1989) *Dokl. Akad. Nauk. USSSR.* (Genet) **305,** 729-731(Russian)

Privalov , P.L. (1982) *Adv. Protein Chem.* **35,** 1-104.

Ptitsyn O. B. (1987) *J. Protein. Chem.* **6,** 273-293.

Ptitsyn, O.B. (1995) *Trends Biochem. Sci.* **20,** 376-379.

Ptitsyn,O. B. (1995) *Adv. Protein Chem.* **47,** 83-229.

Pubols, M. H., Zahnley, J. C. and Axelrod, B. (1963) *Plant Physiol.* **38,** 457-461.

Quax, W. J., Mrabet, N. T., Luiten, R. G., Schuurhuizen, P. W., Stanssens, P. and Lasters, I. (1991) *Bio/Technol.* **9,** 738-742.

Radford, S. E., Dobson, C. M. and Evans, P. A. (1992) *Nature* **358,** 302-307.

Ramirez, D. M. and Bentley, W. E. (1993) *Biotechnol. Bioengg.* **41,** 557-565.

Rao, N. M., and Nagaraj, R. (1991) *J. Biol. Chem.* **266,** 5018-5024.

Rawat, U. B., Phadtare, S. U., Deshpande, V. V. and Rao, M. B. (1996) *Biotech. Lett.* **18,** 1267-1270.

Redfield, C., Smith, R. A. G. and Dobson, C. M. (1994) *Nature Struct.Biol..* **1,** 23-29.

Rose, I. A., Conell, E. L. and Mortlock, R. P. (1969) *Biochem. Biophys. Acta* **178,** 376

Rosenfeld, S. A., Stevis, P. E. and Ho, N. W. Y. (1984) *Mol. Gen. Genet.* **194,** 410-415.

Saari, G. C., Kumar, A. A., Kamasaki, G. H., Insley, M. Y. and O'Hara, P. J. (1987) *J. Bacteriol.* **169,** 612-618

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning : A Laboratory Mannual* (2nd Edn.) Cold Spring Harbor, NY.

Sanger, F. Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci., USA* **74,** 5463-5467.

Sarthy, A. V., McConaughty, B. L., Lobo, Z., Sundstrom, J. A., Furlong, C. E., and Hall, B. D. (1987) *Appl. Environ. Microbiol.* **53,** 1996-2000.

Scallet, B. L., Shieh, K., Ehrenthal, I. and Slapshak, L. (1974) *Starke*, **26,** 406.

Scheler, A., Rygus, T., Allmansberger, R. and Hillen, W. (1991a) *Arch. Microbiol.* **155,** 526-534.

Shinde, U. and Inouye, M. (1995) *Intramolecular chaperones and protein folding,* 1-6, Springer-Verlag, Heidelberg, Germany .

Schumann, R. and Jainicke, R. (1993) *Eur. J. Biochem.* **213,** 1225-1233.

Scheler, A., Rygus, T., Allmansberger, R. and Hillen, W. (1991b) *Arch. Microbiol.* **155,** 535-542.

Schellenberg, G. D., Sarthy, A., Larson, A. E., Backer, M. P., Crabb, J. W., Lidstrom, M., Hall, B. D. and Furlong, C. E. (1984) *J. Biol.Chem.* **259,** 6826-6832.

Schruender, J., Gunge, N. and Meinhardt, F. (1997) *Curr. Microbiol.* **35,** 129.

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N. and Gilmanshin, R. I. (1991) *Biopolymers* **31,** 119-128.

Shamana, D. K. and Sanderson, K. E. (1979a) *J. Bacteriol.* **139,** 64-70.

Shamana, D. K. and Sanderson, K. E. (1979b) *J. Bacteriol.* **139:**71-79.

Shin, M. and Kho, Y. H. (1985) *Kor. Journ. Microbiol.* **23:** 138-146.

Sizemore, C., Buchner, E., Rygus, T., Witke, C., Gotz, F. and Hillen, W. (1991) *Mol. Gen. Genet.* **227,** 377-384.

Sizemore, C., Geidorfer, W. and Hillen, W. (1993) *FEMS Microbiol. Lett.* **107,** 303-306.

Smith, C. A., Rangarajan, M. and Hartley, B. S. (1991) *Biochem. J.* **277,** 255-261.

Southern, E.M. (1975) *J. Mol. Biol.* **98,** 503.

Srinivasan, M. C., Vartak, H. G., Powar, V. K. and Khire, J. M. (1983) *Biotechnol. Lett.* **5:** 611-614.

Stern, O. and Volmer, M. (1919) *Phys.Z.* **20,** 183-193.

Strickland, E. H. (1974) *CRC Crit.Rev.Biochem.* **3,** 113-175.

Stigter, D., Alonso, D.O.V. and Dill, K.A. (1991) *Proc. Natl.Acad.Sci.* **88,** 4176-4180.

Stryer, L. (1965) *J. Mol. Biol.* **13,** 482-495.

Suekane, M. and Iizuka, H. (1982) *Z. Aug. Mikrobiol.* **22,** 577.

Suekane, M., Tamura, M. and Tomimura, C. (1975) *Agric. Biol. Chem.* **42,** 909.

Tanford, C. (1968) *Adv. Protein.Chem.* **23,** 121-282.

Tanford, C. (1970) *Adv. Protein.Chem.* **24,** 1-95.

Thomas P. J.,Qu, B. H., and Pedersen P. L. (1995) *Trends Biotechnol.Sci.* **20,** 456-459.

Tilbeurgh, H. V., Jenkins, J., Chiadmi, M., Janin, J., Wodak, S. J., Mrabet, N. T and Lambier, A. M. (1992) *Biochemistry* **31,** 5467-5471.

Tiraby, G., Bejar, S., Drocourt, D., Reynes, J. P., Sicard, P. J., Farber, G. K., Glasfeld, A. Ringe, D. and Petsko, G. A. (1989). In: Genetics & Molecular biology of Indusrial microorganisms. ed.C. L. Hershberger, S. W. Queener and G. Hegeman.119-126.

Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl.Acad.Sci.* **76,** 4350.

Van der Goot F. G., Gonzalles-Manes, J. M., Lakey, J. H., and Pattus, F. (1991) *Nature* **359,** 408-410.

Ueng, P. P., Volpp, K. J., Tucker, J. V., Gong, C. S. and Chen, L. F. (1985) *Biotechnol. Lett.* **7,** 153-158.

Uversky, V. N. and Ptitsyn, O. B. (1994) *Biochemistry* **33,** 2782-2791.

Uversky, V. N. and Ptitsyn, O. B. (1996) *J. Mol. Biol.* **255,** 215-228.

Van Bastelaere, P. B. M., Kirsters-Hilderson, H. L. M. and Lambeir, A. M. (1995) *Biochem. J.* **307,** 135-142.

Vangrysperre, W., Callens, M., Kersters-Hilderson, H. and de Bruyne, C. K. (1988) *Biochem. J.* **250,** 153-160.

Vangrysperre, W., Callens, P., Kersters-Hilderson, H. and deBruyne, C.K (1989) *Biochem. J.* **260,** 163-169.

Vangrysperre, W., van Damme, J., Vandekerchkhove, J., de Bruyne, C. K., Cornelis, R. and Kersters-Hilderson , H. (1990) *Biochem.. J.* **265,** 699-705.

Vartak, H. G., Srinivasan, M. C., Powar, V. K., Rele, M. V and Khire., J. M. (1984) *Biotechnol. Lett.* **6,** 493-494.

Vieille, C., Hess, M. J., Kelly, R. M. and Zeikus, J. G. (1995) *Appl. Environ. Microbiol.* **61,** 1867-1875

Volkin, D. B. and Klibanov , A. M. (1983) *Science* **219,** 722-727

Volkin, D. B. and Klibanov, A. M. (1989) *Biotechnol.Bioeng.* **33,** 1104-1111.

Vongsuvanlert, V. and Tani, Y. (1988) *Agric. Biol. Chem.* **52,** 7,1817-24.

Walfridsson, M., Bao, X., Anderlund, M., Lilius, G., Bulow, L., and Hahn-Hagerdal, B. (1996). *Biochem. Biophys .Acta* **62,** 4648-4651.

Wang, P. Y., Johnson, B. F. and Schneider, H. (1980) *Biotechnol. Lett* .**2,** 273-278.

Wang, P.Y., Shopsis, C. and Schneider, H. (1980) *Biochem. Biophys. Res. Commun.* **94,**

Ward, L. D., Matthews, J. M., Zhang, J-G, and Simpson, R. J. (1995) *Biochemistry* **34,** 11652-11659.

Weber, P. (1975) *United Kingdom Patent*, 1410579

Wilhelm, M. and Hollenberg, C. P. (1984) *EMBO J.* **3,** 2555-2560.

Wiseman, A. ed. (1975). In :Handbook of Enzyme Biotechnology. Chichester : Ellis Horwood Ltd.

Wong, H. C., Ting, Y., Lin, H. C., Reichert, F., Myambo, K., Watt, K. W. K., Toy, P. L, and Drummond, R. J. (1991) *J. Bacteriol.* **173,** 6849-6858.

Wovcha, M. G., Steuerwald, D. L. and Brooks, K. E. (1983) *Appl. Environ. Microbiol.* **45,** 1402-1404.

Wrigley-Jones, C. A., Richards, H., Thomas, C. R. and Ward, J. (1993) *Biotechnol. Bioeng.* **41,** 148-155.

Wuxiang, L. and Jeyaseelan, K. (1993) *Biotechnol.Lett.* **15,** 1101-1106.

Yamanaka, K. (1968) *Biophys. Biochem. Acta* **151,** 670.

Yang, J. T., Wu, C-S. C. and Martinez, H. M. (1986) *Methods Enzymol.* **130,** 208-255.

Yanisch-Perron, C., Viera, J. and Messing, J. (1985) *Gene* **33,** 103.

Zabuskie, D. W. and Arcuri, E. J. (1986) *Enz. Microb. Technol.* **8,** 706-717.

Zerovnik, E., Jerala, R., Kroon-Zitko, L., Turk, V. and Lohner, K. (1997) *Eur. J. Biochem.* **245,** 364-372.

binds to the palindromic *xyl* operator but not to a half site, suggesting that active repressor may be at least a dimer (Gartner *et al*, 1992). A computer analysis of the *xyl* R sequences from *B. subtilis* revealed that operator recognition may be mediated by an α-helix-turn-α-helix motif located between amino acids 29 and 48 of the primary structure (Dodd & Egan, 1990). Kauder *et al* (1993) have shown that an operator binding-negative mutation of *xyl* repressor from *B. subtilis* is *trans* dominant in *B. megaterium*. Arrangement of two operators is also shown to be present in *xyl* regulatory sequences of GI from *B. subtilis* (Dahl *et al*, 1994), *Staphylococcus xylosus* (Sizemore *et al*, 1991), and *Lactobacillus pentosus* (Lokman *et al*, 1991) is assumed to be responsible for efficient regulation in these bacteria.

The *xyl* genes from *S. xylosus* were cloned in *S. carnosus* by complementation to xylose utilization (Sizemore *et al*, 1991). Open reading frames of *xyl* R, *xyl* A and *xyl* B were located with the same polarity. Primer extension analysis revealed that the transcription of *xyl* R was constitutive while that of *xyl* A was xylose-inducible. The two transcriptional units *viz.* *xyl* R and *xyl* A were separated by a transcriptional terminator between the genes and the presence of promoter-like sequences was observed upstream of both transcriptional start sites. *xyl* A and *xyl* B are separated by only five nucleotides between the stop and start codons respectively. This observation together with the absence of terminator-like structures between *xyl* A and *xyl* B strongly suggests that they are co-transcribed. Co-translation of these two genes is also indicated by the presence of a potential Shine-Dalgarno sequence for *xyl* B (AAGGA) which overlaps with the last codons of *xyl* A. These results strongly support a repressor-operator mechanism for the regulation of *xyl* AB expression (Sizemore *et al*, 1993).

Divergent promoters

Studies on the *xyl* A encoding gene of *Streptomyces violaceoniger*, have indicated that the *xyl* A and *xyl* B genes promote transcription in opposite directions (Tiraby *et al*, 1989). The existence of divergent promoters in *Streptomyces* and other prokaryotes has been reported earlier (Beck and Warren, 1988). Sequence analysis has indicated the presence of a third reading frame, which encodes a regulatory protein. The two genes are seperated by a short region of 195 bp which revealed the presence of an element with palindromic symmetry typical of bacterial operators. It is suggested that a regulatory molecule may act within the divergent transcription unit to control the expression of opposite genes and also regulates its own synthesis. Studies by Wong *et al* (1991) have revealed that the genetic organization of