INVESTIGATION INTO THE STRUCTURE AND ACTIVITY OF CONJUGATED BILE SALT HYDROLASE AND RELATED PENICILLIN ACYLASE

R. SURESH KUMAR

**NATIONAL CHEMICAL LABORATORY
PUNE – 411008
NOVEMBER 2006**

Investigation into the Structure and Activity of Conjugated Bile Salt Hydrolase and Related Penicillin Acylase

Thesis Submitted to

University of Pune

For the Degree of

Doctor of Philosophy

In

Biotechnology

By

R. Suresh Kumar

Division of Biochemical Sciences

National Chemical Laboratory

Pune – 411008

November 2006

DECLARATION

I hereby declare that the thesis entitled "**Investigation into the Structure and Activity of Conjugated Bile Salt Hydrolase and Related Penicillin Acylase"** submitted for Ph.D. degree to the University of Pune has not been submitted by me to any other university for a degree or diploma.

R. Suresh Kumar   Date:

Biochemical Sciences Division,

National Chemical Laboratory,

Dr. Homi Bhabha Road,

Pune- 411008. India

CERTIFICATE

Certified that the work incorporated in this thesis entitled, "**Investigation into the Structure and Activity of Conjugated Bile Salt Hydrolase and Related Penicillin Acylase"** submitted by Mr. R. Suresh Kumar was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.

(Dr. C. G. Suresh)   **Date:**

Research Guide

Dedicated to Amma and Appa

ACKNOWLEDGEMENTS

I would like to thank my present and past lab members Narashima Rao, Anu, Manish, Priya, Satya, Uma, Poorva, Raamesh, Nitin, Nishant and Urvashi. Many thanks to Sharath, Sachin, Ambrish, Atul, Sridevi and Ashwini for being fun and supportive. I like to thank Marek, Chandra, Tony, Nicolas, Annabelle, Sally, Shirly, Jullia, Simon Grist, Elena and Simon Charnock. It was a pleasure to work with them and the other people of YSBL, York.

I also want to thank Raju, Rohini, Ramchander, Bhushan, Shashidhar, Sathish, Nagaraj, Gowda, Anil, Suhas, Ajay, Nitin, Anish and Atul for their timely help and for all the interesting discussions. I would also like to thank other members of our division for their commiseration and friendship.

Special thanks to my hostelmates, Selva, Sankar, Elan, Thiru, Vijay, Malli, Victor, Marivel, Eshwar, Murugan, Deveraj, Kannan, Mari anna, Prathap, Nagender, Subu. My heart is filled with gratitude to all those wonderful people who have been my friends. A very special thank you goes out to Mathi, Yogi, Dhanda, Senthil, Siva, Chakra and Mohan.

I need to acknowledge Indira, Mari, Satyali, Caroline, Louise for the help in administrative work and Mr. Trehan, Karanjekar and Tim for technical assistance.

Finally, I would like to thank my parents and family without whose support, whatever I have accomplished would not have been possible.

I gratefully acknowledge Council of Scientific and Industrial Research, New Delhi for the award of CSIR Junior and Senior Research Fellowship.

R. Suresh Kumar

Contents	Page no

Contents	Page no

Contents	Page no

3. Crystal structure analysis of *Bl*BSH

Contents	Page no

4. Active site characterization and substrate specificity of *Bl*BSH

5. Stability and denaturation studies on *Bl*BSH

Contents	Page no

6. Biochemical characterization of *Kc*PGA and study of its post-translational processing

7. Crystallization and crystal structure studies of PGA

Contents	**Page no**

LIST OF ABBREVIATIONS

6-APA	:	6- aminopenicillanic acid
7-ADCA	:	7-aminodeacetoxycephalosporanic acid
*Af*PGA	:	*Alcaligence faecalis* Penicillin G Acylase
AGA	:	Aspartylglucosaminidase
AS	:	Ammonium Sulfate
AmoRe	:	Automated Molecular Replacement
APA	:	L-Aminopeptidase-D-Ala-amidase
ATCC	:	American Type Culture Collection
ASAH	:	N-acylsphingosine amidohydrolase
AU	:	Absorption Unit
BPB	:	Bromo-phenol-blue
*Bl*BSH	:	*Bifidobacterium longum* Bile Salt hydrolase
BOG	:	β-octyl-glucosidase
*Bsp*PVA	:	*Bacillus sphaecus* Penicillin V Acylase
BSH	:	Bile Salt Hydrolase
CA	:	Cholic acid
CCD	:	Charge-Coupled Device
CCP4	:	Colloborative Computational Project No. 4
CD	:	Circular dichroism
CDCA	:	Chenodeoxycholic acid
CHO	:	Chinese hamster ovary
CPA	:	Cephalosprin Acylase
CPB	:	Citrate-Phosphate buffer
CpC	:	Cephalosporin C
*Cp*BSH	:	*Clostrodium perfigence* Bile Salt Hydrolase
DCA	:	Deoxycholic acid
Dcase	:	N-carbamyl-D-aminoacid amidohydrolase
DLS	:	Dynamic light scattering
DMF	:	Dimethyl formamide
DMSO	:	Dimethyl Sulfoxide
DTNB	:	5,5'-dithiobis-(2-nitrobenzoic acid
DTT	:	Dthiothreitol
*Ec*PGA	:	*Escherichia coli* Penicillin G Acylase
EDTA	:	Ethylenediaminetetraacetic acid
EHC	:	Enterohepatic cycle
ESRF	:	European Synchrotron Radiation Facility, France
FXR	:	Farnesoid X receptor
GA	:	Glycosylasperaginase
GACA	:	glutaryl 7-aminocephalosporanic acid acylase
GAT	:	Glutamine amido transferase
GCA	:	Glycocholic acid
GGT	:	Gamma-glutamyl transferase
GCDCA	:	Glycochenodeoxycholic acid
GDCA	:	Glycodeoxycholic acid

GIT	:	Gastrointestinal tract
GPS	:	Glutaminase domain of glucosamine-6-phosphate synthase
HEPES	:	N-(2-hydroxyethyl)-piperazine-N'-2-ethanesulfonic acid
HNBB	:	2-hydroxy 5-nitrobenzylbromide
HXT	:	1,2,6-Hexane triol
IBAT	:	Ileum bile acid transporter
IEF	:	Isoelectric focusing
IPTG	:	Isopropyl- -D-thiogalactopyranoside
ITC	:	Isothermal titration calorimetry
IU	:	International Unit
*Kc*PGA	:	*Kluyvera citrophila* Penicillin G Acylase
LB	:	Luria-Bertani
LCA	:	Lithocholic acid
L	:	litre
m	:	metre
Mb	:	Mega base
mol	:	mole
M	:	molar
mm	:	milli meter
MAD	:	multi-wavelength anomalous dispersion
MAP	:	methionine aminopeptidase
min	:	Minute
ML	:	Maltose
MOE	:	Molecular operating environment
MOPS	:	3-(N-morpholino) propanesulfonic acid
MR	:	molecular replacement
M-RIP	:	myosin phosphatase-Rho interacting protein
M.W./Mr	:	molecular weight
NAI	:	N-acetylimidazole
NBS	:	N-bromo succinimide
nm	:	Nano metre
NCS	:	non-crystallographic symmetry
NEM	:	N-ethylmalamide
NMR	:	Nuclear magnetic resonance
Ntn	:	N-terminal nucleophile
OA	:	Ornithine acetyltransferase
OD	:	Optical density
PAA	:	Phenoxy Acetic Acid
Pas	:	Penicillin acylases
pac	:	Penicillin G Acylase gene
PAGE	:	poly-acrylamide gel electrophoresis
pepDA	:	dipeptidase A
PB	:	Phosphate Buffer
PCMB	:	p-chloromereuribenzoate
PCR	:	polymerized chain reaction
PDB	:	Protein Data Bank

PEG	:	Polyethylene Glycol
PGA	:	Penicillin G acylase
PG	:	phenylglyoxal
PH domain	:	Pleckstrin Homology domain
PI	:	iso-electric point
PMSF	:	Phenylmethanesulfonyl fluoride
PSI-BLAST	:	Position-Specific Iterated - Basic Local Alignment Search Tool
POAA	:	phenoxyacetic acid
PVA	:	penicillin V acylase
PGA	:	penicillin G acylase
penV	:	Penicilin V (Phenoxymethyl penicillanic acid)
penG	:	Penicillin G (Benzyl penicillanic acid)
pNPG	:	p-nitrophenyl glyoxal
PRO	:	proteasome
pre	:	precursor
r.m.s.d	:	Root Mean Square Deviation
rpm	:	revolution per minute
sec	:	second
SAH	:	S-adenosyl-homocysteine
SCB	:	Sodium Cacodylate Buffer,
SDS	:	sodium dodecyl sulfate
TCA	:	Taurocholic acid
TCDCA	:	Taurochenodeoxycholic acid
TDCA	:	Taurodeoxycholic acid
TMAO	:	Trimethylamine N-oxide
TNBS	:	2,4,6-trinitrobenzenesulfonic acid
Tris	:	tris-hydroxymethyl amino methane
U	:	Unit
Vm	:	Matthew's number
K_m	:	Michaelis-Menten constant
k_{cat}	:	Turn over number
K_i	:	Inhibition constant
Å	:	Angstrom
	:	alpha
(h,k,l)	:	phase angle
F(hkl)	:	structure factor
	:	sigma
	:	summation
∫	:	integration
kDa	:	kilo Dalton
ºC	:	degree centigrade
µg	:	microgram
µL	:	microlitre
µM	:	micromolar
h	:	hour

ABSTRACT

The Penicillin acylases are used in the commercial production of 6-amino penicillanic acid (6-APA), the starting compound for the synthesis of semi-synthetic penicillins. The enzyme bile salt hydrolase (BSH) is related to penicillin V acylase and has potential application in lowering the serum cholesterol level in hypercholestremic humans or to prevent hypercholesterolemia in individuals with normal cholesterol.

We report in this thesis the extensive studies carried out on a bile salt hydrolase (Cholylglycin hydrolase, EC 3.5.1.24) from *Bifidobacterium longum* (*Bl*BSH) and two penicillin G acylases (PGAs) (penicillin amidohydrolase, EC 3.5.1.11) from *Kluyvera citrophila* (*Kc*PGA) and *Alcaligence faecalis* (*Af*PGA). These enzymes could be placed in N-terminal nuleophile (Ntn) hydrolase superfamily. Ntn hydrolases are a class of enzymes that share a common fold of - core structure possessing a N-terminal catalytic nucleophile residue. In BSH the N-terminal nucleophile is cysteine while in PGAs it is serine. The Ntn hydrolases invariably produce their active form from the corresponding precursor by an intra-molecular autocatalytic cleavage to create a free amino group at the nucleophile residue.

Bile salt hydrolase (BSH) catalyses the hydrolysis of glycine or taurine conjugated bile salt into the corresponding amino acid residues and the deconjugated bile salt. The ability to deconjugate bile salts is widespread among the members of the genus *Bifidobacterium*. The deconjugation of six different salts by all *bifidobacteria* strains that inhabit animal gut is one way of transforming the bile salt in intestine. In recent years therapeutic use of bile salt deconjugation by lactic acid bacteria to lower serum cholesterol level is being tried.

Penicillin acylases, a subclass of β-lactam antibiotic acylase superfamily, catalyze the selective hydrolysis of relatively stable amide bond in penicillins and some cephalosporins while leaving the labile β-lactam ring intact. The capacity of penicillin G acylases to catalyze the acylation of amino group of key intermediates can be used in the environment-friendly synthesize of semi-synthetic β-lactam antibiotics. PGAs are also useful for other applications such as peptide synthesis, removal of protecting groups and separation of racemic mixtures of certain compounds.

Biochemical and biophysical characterization, including the three-dimensional structure determination have been carried out on BSH from *B. longum*. Structural studies have been conducted on PGA from *A. faecalis*. The autocatalytic processing of *Kc*PGA has been investigated. The objective of this study is to achieve a better understanding of the mechanism of action and evolutionary relationships of these enzymes, which in turn is expected to help in protein engineering them for further applications. Results from the studies reported here have helped us to compare the structural, mechanistic and evolutionary relationships of industrially and therapeutically important enzymes.

Chapter one: Introduction

This chapter presents the literature survey of the research reported on bile salt hydrolases, penicillin V acylases and penicillin G acylases. The sources of BSH known, the role of BSH in physiology of gastrointestinal system, possible physiological effect of BSH producing bacteria on host and their resultant clinical importance, negative impact on host health, BSH activity and its implication for BSH producing bacterial cell are described. Also presented is the research on penicillin acylases and a brief account of all known proteins that belong to the Ntn hydrolase family.

Chapter Two: Materials and methods

This chapter describes in detail the materials and methods used for the preparation of BSH and PGA clones, their purification, crystallization, X-ray data collection, data processing, structure determination, structure refinement, analysis of the refined structure biochemical/biophysical experiments and kinetic studies.

The proteins were purified by column chromatography using Octyl-sepharose, Q-sepharose and hydroxyl chromatography. Hanging-drop vapour-diffusion method was used for crystallizing BSH and PVA proteins. The X-ray data were collected using Raxis IV^{++} detector mounted on a Rigaku rotating anode at NCL, Pune. India and on beamline at ESRF, Grenoble, France. The X-ray images were processed using DENZO and SCALEPACK programs in HKL suit. The crystal structures were determined using molecular replacement technique implemented in AMoRe and PHASER. The REFMAC5 program was used for structure refinement in cycles of refinement followed by model

fitting using QUANTA and COOT. The CCP4 suit programs were used for most of the calculations. MOE was used for molecular modelling and docking studies.

Chapter Three: Crystal structure analysis of *Bl*BSH

This chapter describes the results of crystallization experiments and techniques used to improve the quality of *Bl*BSH crystals. The details of cryoprotectant screening and strategy followed for getting enzyme-substrate complex are discussed. Several sets of X-ray data of wild type and mutant *Bl*BSH crystals in various space groups were collected. The structure of wild type *Bl*BSH, determined using molecular replacement method using PVA co-ordinates (PDB: 3pva), was refined at resolutions 2.25 Å (space group $P6_122$) and 2.5 Å (space group $P3_22_1$). These structures were compared with the reported structures of PVA from *Bacillus sphaericus* (*Bsp*PVA) and BSH from *Clostridium perfringens* (*Cp*BSH).

Chapter Four: Active site characterization and substrate specificity of *Bl*BSH

This chapter describes the results of experiments carried out to assess the role of selected residues in catalysis and substrate binding by chemically modifying them and measuring the effect on enzyme activity. Two sets of substrate molecules were tested for activity. One set contained the original substrates, the bile salts, which showed variations in activity depending on the type of conjugated amino acid moiety (glycine or taurine) and also depending on the position of –OH group in the cholesterol ring. The second set of compounds was basically the inhibitors such as penicillin V. The inhibition of the bile salt hydrolysis by these compounds was confirmed by inhibition kinetic studies. The effective inhibition was measured using inhibition constant K*i*. Most of the inhibitors showed competitive inhibition pointing to the possibility that they bind at the same site as substrates.

Since cysteine was involved in activity we used cysteine-specific chemical modifiers and found complete loss of activity. The cysteine modification was hampered in the presence of substrates and inhibitors that again indicated that inhibitors, like substrates, were binding at the active site and blocking it. Further studies were conducted on a representative compound of each type and the parameters of thermodynamics and

energetics of binding were calculated using fluorescence studies and isothermal calorimetry. To identify the molecular interactions involved in binding *insilico* experiments such as docking and molecular dynamics simulations were attempted.

Chapter Five: Stability and denaturation studies on *Bl*BSH

Conformational transitions and functional stability of the bile salt hydrolase from *Bifidobacterium longum* (*Bl*BSH) were studied in thermal, chemical and pH mediated denaturation conditions using biophysical techniques such as fluorescence and CD spectroscopy. The major findings are i) Thermal and Gdn-HCl mediated denaturation of *Bl*BSH is a multistep process of inactivation and unfolding. Enzyme activity seems sensitive to even minor conformational changes at the active site. ii) Thermal denaturation as such did not result in any insoluble protein aggregate. However, on treating with 0.25 to 1M Gdn-HCl the enzyme showed increasing aggregation at temperatures 40-55°C indicating more complex structural changes taking place in the presence of chemical denaturants. iii) The enzyme secondary structure was turning more compact at extremely low pHs (1-3 units). The perturbation in the tertiary structure at the above pH was detected through freshly formed solvent exposed hydrophobic patches on the enzyme. We show that these changes can be interpreted as due to formation of an acid induced molten globule-like state. This is the first time any such studies on the conformational changes of a member of Ntn-hydrolase family are conducted. Since all the members of these family share common mechanistic and structural features, the study reported here has wider implications to structure, function and stability of this class of enzymes in general.

Chapter Six: Biochemical characterization of *Kc*PGA and study of its post-translational processing

This chapter discusses preliminary biochemical characterization and active site residue modifications of PGA from *K. citrophila*. During purification of *Kc*PGA the presence of its 96 KD precursor was constantly observed in the preparation, almost in equal amounts as the matured enzyme. The presence of this precursor was confirmed by chromatographic and light scattering experiments. Various conditions of pH, temperature

and time intervals were tried for achieving highest *invitro* processing. An account of the efforts to clone and over express this enzyme for using in structural studies is included in this chapter.

Chapter Seven: Crystallization and crystal structure studies of PGA

This chapter describes the crystallization and structural studies attempted on *Kc*PGA and on PGA from *A. faecalis*. We were successful in collecting diffraction data at a resolution of 3.5 Å from tetragonal crystals of *Af*PGA. The crystal structure was determined using molecular replacement method using co-ordinates of *E. coli* PGA (PDB:1gk9). Although the resolution was modest we could identify features like the presence of a disulphide bridge in the structure that would impart more stability to this PGA.

Chapter Eight: Comparison of Ntn-hydrolases including Ntn-hydrolase domains

To compare the Ntn-hydrolase superfamily of proteins we divided them into three categories based on the type of N-terminal nucleophile residue, which is a cysteine, serine or a threonine. An extensive sequence comparison and analysis was carried out in each category separately. Many related proteins from eukaryotes in the database were identified in serine and cysteine groups. In the category of threonine as the N-terminal nucleophile residue two distinct groups could be identified based on the closeness of amino acid sequences. Thus, through careful sequence comparison we not only could identify new, but distantly related, Ntn-hydrolase members or domains but also could place in this family some of the un-annotated proteins in the database.

CHAPTER 1

INTRODUCTION

Digestion of food material in our gastrointestinal system requires many enzymes and associated digestive secretion. Among them bile is a predominant digestive secretion that plays a major role in the emulsification and solubilization of lipids. It has the ability to affect the phospholipids and proteins of cell membranes and disrupt cellular homeostasis. Human gastrointestinal tract is a natural habitat for many microflora. However, the ability to tolerate bile is an important factor for the survival of these organisms, pathogens or commensals, and for their subsequent colonization of gastrointestinal tract. Bile acid constitutes approximately 50% of the organic components of bile. Bile acids are synthesized de novo in the liver from cholesterol by multi-enzyme process and conjugated as N-acyl amidase with amino acids. The indigenous intestinal bacteria extensively modify them, the major transformation is the deconjugation (Hill, 1995).

1. 1. Conjugated Bile Salt Hydrolase

Bile salt hydrolase (BSH) (Cholylglycine hydrolase EC 3.5.1.24) catalyses the hydrolysis of amide bond in conjugated bile salts resulting in the release of free amino acids (Figure 1.1). This deconjugation is a biologically very crucial reaction among the bacterial alterations of bile acids in humans and animals.

1.2. Sources of conjugated bile salt hydrolase

BSH activity is found primarily in organisms from the gastrointestinal tracts of mammals and also from fermented milk preparation and vegetables (Mann, 1974). Most of these organisms were isolated and BSH activity was characterized (Table 1.1).

BSHs are generally intracellular enzymes that are oxygen insensitive, their activity is coupled to biomass production and they are not regulated by bile salts (Christiaens et al., 1992 and De Smet et al., 1995). The extracellular BSH of *Clostridium perfringens* does not require an inducer (Kishinaka et al., 1994), whereas the BSH of *L. acidophilus* that requires low oxidation–reduction potential (Gilliland et al., 1977) and

the BSH of *L. johnsonni* strain 100–100 both of which are inducible by bile (Lundeen et al., 1992a).
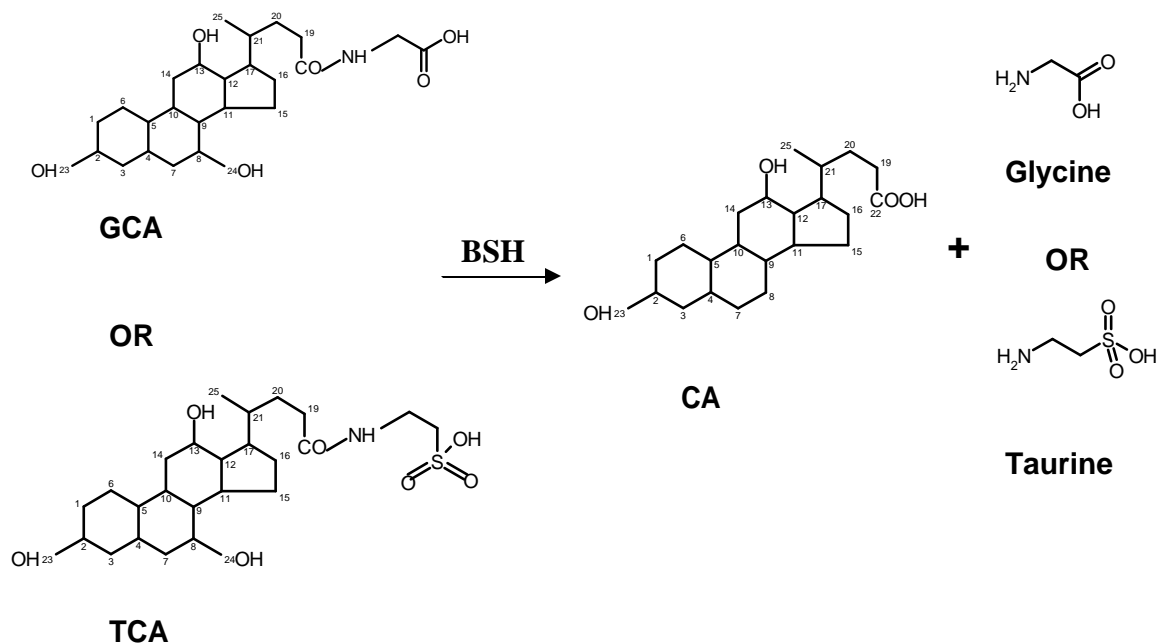
Figure 1.1: The reaction catalyzed by BSH. Glycocholic acid (GCA) or Taurocholic acid (TCA) is hydrolsed into Cholic acid (CA) and glycine or taurine

Table 1.1 Sources characterized for producing BSH with corresponding gene bank accession number (wherever available) and their references are listed.

S. No	Source	Accession No	Reference
1	*Bacteroides fragilis* subsp. *fragilis*	gi:60494473	Stellwang et al., 1976
2	*Clostridium perfringens*	gi:1705662	Gopal-Srivastava & Hylemon, 1988; Archer et al., 1982
3	*Enterococcus faecium*	gi:30171803	Wijaya et al., 2004
4	*Xanthomonas maltophilia*	-	Dean et al., 2002
5	*Listeria monocytogenes*	gi:16411537	Dussurget et al., 2002; Sue et al., 2003
6	*Lactobacillus acidophilus*	gi:58337369	McAuliffe et al., 2005

S. No	Source	Accession No	Reference
7	*Lactobacillus plantarum* 80	gi:84314036	Christiaens et al., 1992; Lundeen & Savage, 1990
8	Bacteroides	gi:29339396	Kawamoto et al., 1989; Stellwag & Hylemon, 1976; Masuda, 1981
9	*Lactobacillus casei*	gi:12802349	Lundeen et al., 1990; Pereira et al., 2003
10	*Enterococcus*	gi:84372150	Franz et al., 2001; Knarreborg et al., 2002; Wijaya et al., 2004
11	*Bifidobacterium adolescentis*	gi:62546323	Kim et al., 2005
12	*Bifidobacterium animalis*	gi:46486762	Lepercq et al., 2005
13	*Bifidobacterium bifidum*	gi:83630914	Kim et al., 2004a & b
14	*Bifidobacterium longum*	gi:23465372	Grill et al., 1995a; Tanaka et al., 2000
15	*Bifidobacterium infantis*	-	Liong et al., 2005
16	*Bifidobacterium breve*	gi:83630916	Liong et al., 2005
17	*Lactobacillus reuteri*	gi:81427616	De Boever et al., 2000; Tannock et al., 1989
18	*Lactobacillus amylovora*	gi:10342233	Grill et al., 2000a

1.2.1 *Bifidobacterium* and its beneficiary role to humans

Vast and diverse communities of microbes that are indispensable to its function colonize the human gastrointestinal tract (GIT) (Marteau et al., 2001). These microbes have evolved in concert with their host to occupy specific regions and niches in the GIT (Mitvedt & Norman, 1967 & 1968). A balanced, complex microflora is necessary for normal digestion and to maintain the homeostasis of intestinal ecosystem (Buddington et al., 1996). There are 32 species of bifidobacteria; a few have been isolated from human

vagina and oral cavity, but a vast majority is from mammalian GITs (Takahashi et al., 1991; Bateup et al., 1995).

The genus *Bifidobacterium* has been reported to possess higher BSH activity than any other bacterial groups (Grill et al., 1995b). Discovered at the start of the last century, *Bifidibacteria* are considered a key commensal in human-microbe interactions, and are believed to play a pivotal role in maintaining a healthy gastrointestinal tract. *Bifidobacterium* are characterized as Gram-positive, non-spore forming, obligate anaerobes and non-motile, which sometimes could be observed in the form of club-shaped or spatulate often branched (Figure 1.2) (Ventura et al., 2004). Among the 32 species of *Bifidobacteria* (Kaufmann et al., 1997), *Bifidobacterium longum* strain has been the most studied one. *Bifidobacterium bifidum* is another major bifidobacterial species commonly detected in human faeces (Matsuki et al., 2003).
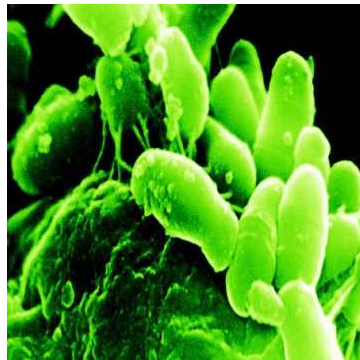
Figure 1.2: Morphology of *Bifidobacterium longum*

Attempts have been made in recent years to define beneficial effects of *Bifidobacterium* strains (Leahy et al., 2005; Picard et al., 2005) which are listed here: (1) Use of *Bifidobacterium* to combat diarrhea. (2) Relief in case of lactose intolerance (Jiang et al., 1996) (3) Resistance to microbial infection (Asahara et al., 2004) (4) Cancer prevention (Reddy & Rivenson 1993; Hirayama & Rafter 2000) (5) Treatment of inflammatory bowel disease (Sartor 1997) (6) Alleviation of constipation (7) Role in immune function (Medici et al., 2004; Marteau & Rambaud 1993).

While *B. infantis, B. brevi,* and *B. longum* are the largest bacterial groups in the intestines of infants, *Bifidobacteria* are said to be only the 3rd or 4th largest group of bacteria in adults (and about 3-6% of adult faecal flora). The number of these bacteria

actually decline in the human body with age (Figure 1.3). In infants who are breast-fed, *Bifidobacteria* constitute about 90% of their intestinal bacteria; *Bifidobacteria* as well as other beneficial bacteria can be found in fermented dairy foods, especially in yogurt (Marteau et al., 2001).
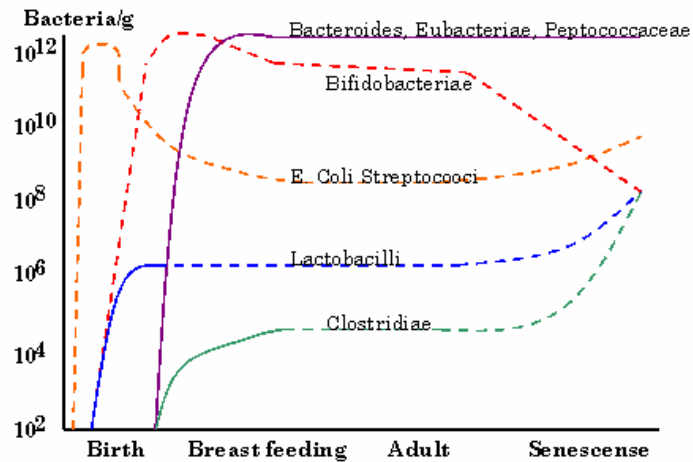
Figure 1.3: Presence of microflora in intestinal track of humans at different stages of life cycle

Presently there are three genome sequences of *Bifidobacterium* available in public databases. *Bifidobacterium longum* NCC2705 (Genbank accession no: AE014295) *B. longum* DJO10A (Genbank accession no.: AF148138) *B. breve* UCC2003 (Genbank accession no: AY488174). There are reports of BSH enzymes being purified from five strains of *bifidobacteria*, and they are identified as three different types based on their biochemical characteristics. The 2.26 Mb genome sequenced from an infant-derived strain of *Bifidobacterium longum*, had been identified as consisting of 1,730 possible coding regions organized in a circular chromosome of 60% GC content (Schell et al, 2002). Analysis using bioinformatics tools revealed several physiological traits that partially explain the successful adaptation of these bacteria to the colon (Kim et al., 2004a & b).

1.3. The role of bile salt hydrolase in the physiology of gastrointestinal system

The major *in vivo* role of bile is to act as a biological detergent to emulsify and solubilize fats. This also confers potent antimicrobial properties to bile and gives it an important role in body's physicochemical defense system. To understand the role of bile

26

salt hydrolase in gastrointestinal system we need to acquaint with the bile acid circulation in the gastrointestinal system.

1.3.1. Enterohepatic cycle of bile acid

Bile salts are produced de novo in the liver from cholesterol by multi enzyme process. Bile acids contain a perhydrocyclopentano phenanthrene steroid nucleus that consists of three six-membered rings fused to a fourth five-membered ring. Before secretion of bile acid the steroid nucleus is conjugated with an amide bond at the carboxyl C-24 position to one of two amino acids, glycine or taurine. These conjugates are secreted via the common bile duct into the duodenum (Johnson, 1998). Because of their amphipathic nature, the conjugates form spontaneous micelles that trap dietary cholesterol and fats and facilitate their absorption by the intestinal epithelium. A sodium-dependent transporter actively transports the bile acids through the epithelium and into the bloodstream. A brief sketch of the enterohepatic cycle (EHC) is shown (Figure 1.4).

1.3.2. Conjugation

The terminal step in bile acid synthesis is the addition of glycine or taurine, through an amide linkage to carbon 24. The reaction is catalyzed by the amino acid N-acyltransferase enzyme (Johnson et al., 1990). The main deciding factor for the ratio of glycine to taurine conjugated bile acids in humans is the relative abundance of glycine and taurine and may have no functional or regulatory consequences. Diets rich in taurine-containing foods such as meat and seafood will increase tauro-conjugation (Usman & Hosono, 1999). Bile acids conjugated with other amino acids such as leucine or lysine are rapidly hydrolysed by pancreatic carboxypeptidases (Huijghebaert & Hofmann, 1986a & b).

Conjugated bile salts are impermeable to cell membranes because of their enhanced solubility and amphipathicity as a result of conjugation. The oxygen atoms bonded to sulfur of taurocholic acid or to the terminal carbon of glycocholic acid will be ionized at physiological pH. The ionization of oxygen atom together with the planar structure of the bile acid and the hydroxyl groups present in the rings render it highly amphipathic. Conjugation of cholic acid with glycine reduces pKa of cholic acid from 6.4 to 4.4 units ensuring that the bile acid gets completely ionized and becomes highly soluble (Hofmann, 1992).

Bile acids due to their amphipathic nature have detergent action on particles of dietary fat by which the fat globules break down or be emulsified into minute, microscopic droplets. For lipase action emulsification is an essential step, which results in easy accessibility for enzyme action by increasing the surface area of lipid molecule. Bile acids also function as ''lipid carriers'' by solubilizing lipids by forming micelles and thus allowing their transport in an aqueous environment. This is critical for the absorption of fats and fat-soluble vitamins by intestinal ville.

The conjugated bile acids are not absorbed during digestion in the proximal small intestine because they are relatively bulky, ionized molecules and resistant to deamidation by pancreatic and mucosal carboxypeptidases (Hofmann, 1992). Instead, the conjugated bile acids pass to the distal ileum, where they are efficiently absorbed by an active transport system, which is called as ileum bile acid transporter (IBAT) and members of the ABC family of transporters. A small fraction of the bile acid escapes absorption from the ileum and is excreted in the faeces.

1.3.3. Deconjugation of bile acids

The hydrolysis of conjugated bile acids by bacterial enzyme BSH (Drasar et al., 1966) in distal ileum results in the production of free bile acids, which bind with lower affinity to IBAT transporter. This leads to majority of deconjugated bile acids passing into large intestine or cecum. The most important metabolic transformation occurs here, when the free bile acids undergo 7-dehydroxylation. This reaction results in the conversion of cholic acid or chenodeoxycholic acid to deoxycholic acid and lithocholic acid, respectively (Figure 1.4). The enterohepatic re-circulation of bile acids is partially dependent on the recognition of their conjugated form by active transport sites in the terminal ileum (Van den Berg et al., 1990). Bacterial BSH may be conceptualized as being in competition with these active transport sites.

It must be emphasized that the physical and chemical properties of bile acids drastically change on hyderolysis. The solubility decreases, particularly at low pH (Klaver & van der Meer 1993). Free bile acids (deconjugated) are poorer detergents than their conjugates and are far less efficient in forming mixed micelles (Macdonald et al., 1983).
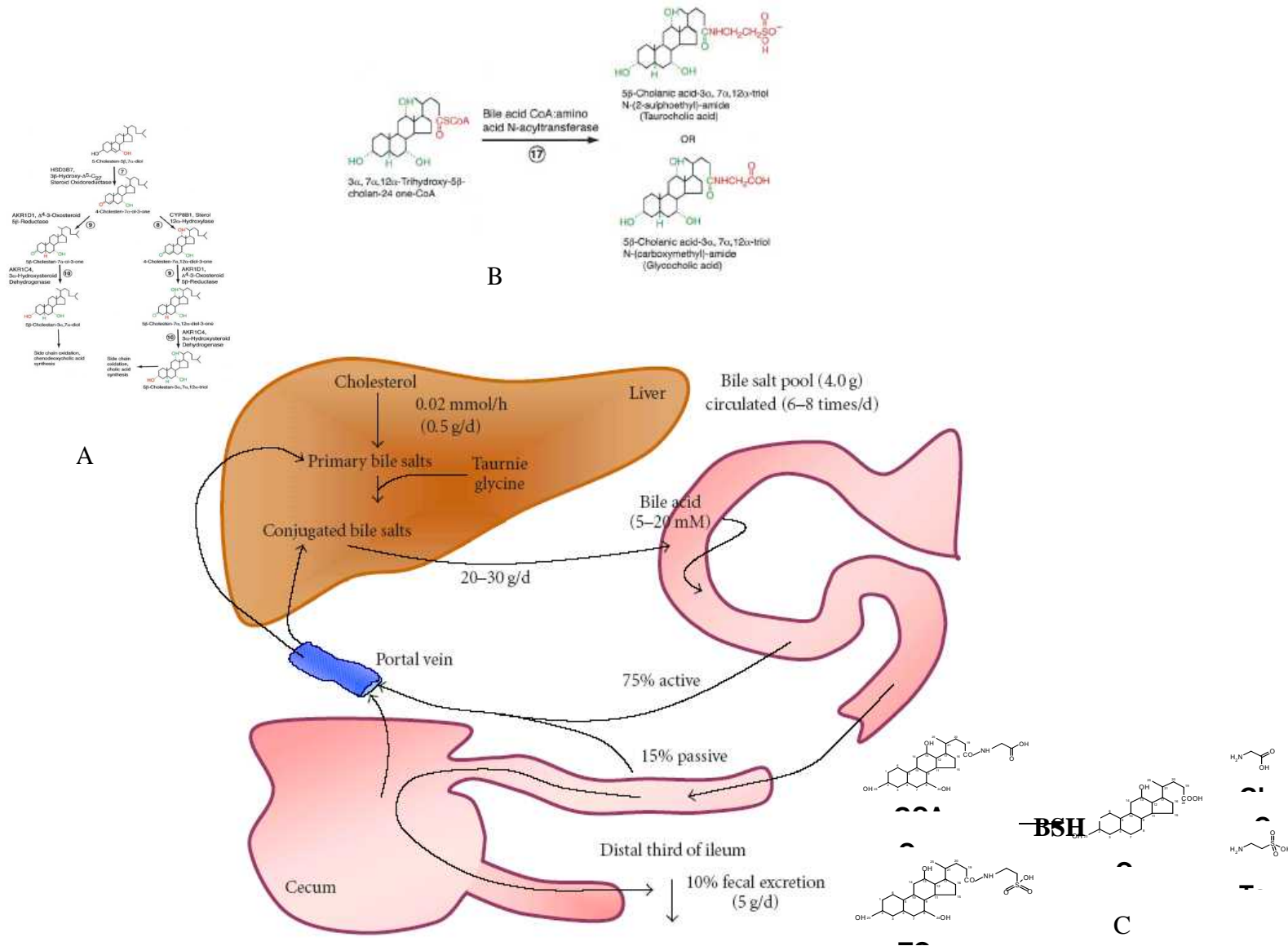
Figure 1.4 : Bile salt metabolisam in mammals. A) bile acid synthesis from cholesterol B) conjugation of bile acids in liver C) deconjugation of bile acids at terminal ileum (Adopted from Jones et al., 2004)

In humans, conjugated bile salts are produced from cholesterol and taurine or glycine at a rate of 0.02 mmol/h in the liver (Hofmann, 1999). During normal entero-hepatic circulation, the average bile salt pool of 4.0 g is secreted into the duodenum twice during each meal, or an average of 6-8 times per day (Hofmann, 1989a) for lipid digestion. Daily secretions of bile acid approach 20-30 g (Hofmann, 1989b) or 20-60 mmol (Bongaerts et al., 2000a) sustaining an intestinal bile salt concentration of 5-20mM (Hofmann, 1989b). During intestinal transit, 90-95% of the secreted bile salts are reabsorbed in the terminal ileum and are returned to the liver via the portal vein. About 75% of the bile acid secretion is reabsorbed in the conjugated form in an active sodium dependent pathway (Hofmann, 1989b). The remaining 25% is hydrolyzed during intestinal transit and only 15% is reabsorbed in a passive way (Hofmann, 1989b). Intestinal absorption of bile acid pool is about 95% efficient, so that faecal loss of bile acids is in the range of 0.3-0.6 g per day. This faecal loss is compensated by an equal daily synthesis of bile acids by the liver, thus the entero-hepatic circulation effectively balances the bile salt pool.

The common belief that a detergent is needed for the absorption of lipids and fat-soluble nutrients is challenged by a recent discovery by three independent groups (Parks et al., 1999; Makishma et al., 1999; Wang et al., 1999). The member of the nuclear hormone receptor super-family, farnesoid X receptor (FXR), functions as a specific receptor for a wide variety of bile acids. Since other members of this family include the receptors for steroids and thyroid hormone, it is suggested that bile acids could be thought of as acting like hormones (Goodwin et al., 2000).

Cholestatic liver injury is the result of the induction of hepatocyte apoptosis by toxic bile salts such as glycochenodeoxycholate (GCDC). GCDC-induced hepatocyte apoptosis involves ligand-independent oligomerization of Fas, recruitment of FADD, activation of caspase 8, and subsequent activation of effector proteases, including downstream caspases and cathepsin B (William et al., 1999).

The liver-derived McNtcp.24 cells transport bile acids and show distinctive responses to two classes of conjugated bile acids. Whereas taurine-conjugated bile acids are non-toxic, glycine-conjugated bile acids efficiently induce apoptosis. Studies were conducted to determine whether the differential sensitivity is limited to cells that

normally transport bile acids and if bile acid-binding proteins could reduce bile acid-mediated apoptosis (Torchia et al., 2001).

Even though it is well known that BSH activity is widely spread over many enteric bacteria, the specific functions of this enzyme neither in the bacterium that produces it nor in the mammalian host infested by some of these bacteria are still unknown. The precise physiological impact exerted by the enzymatic products on mammalian cells, although not fully understood at this stage, will be of considerable pharmacological interest.

1.4. Possible physiological effect of BSH producing bacteria on host and its Clinical Importance

1. Reduction of serum cholesterol level

Cardiovascular disease is the most important cause of death in the developed countries and it is strongly linked with hypercholesterolemia (Levine et al., 1995; Pekkanen et al., 1990). Therefore, decreasing serum cholesterol level is very important towards preventing cardiovascular diseases. For each 1 mM higher than the normal level of cholesterol the risk of coronary heart disease is approximately 35% greater, while coronary death is 45% higher (Gallo, 1983). Cellular cholesterol homeostasis is very important for the prevention of cardiovascular disease.

The cholesterol concentration in plasma can be regulated by reducing the biosynthesis of cholesterol from saturated fat, by removal of cholesterol from the circulation, by preventing absorption of dietary cholesterol or by excretion of cholesterol via bile and faeces. Thus, the deconjugation of bile salts in the small intestine and their resultant excretion can result in reduction of serum cholesterol. To replenish the bile salts it would require utilization of some cholesterol in the body. Free bile salts are less soluble than conjugated bile salts, resulting in lower absorption of de-conjugated ones in the intestinal lumen. Therefore, utilizing the bile salt hydrolase activity of probiotics offers a potential ''biological'' alternative to pharmaceutical interventions in the treatment of hypercholesterolaemia. Pharmacologic agents that effectively reduce cholesterol levels such as fibrates, nicotinic acid, bile acid sequestrants, and statins are available for the treatment of high cholesterol levels (Hay et al., 1999). But they are expensive and known to have severe side effects (Bliiznakov, 2002; Kolata et al., 2001). Control of cholesterol

through oral live bacterial cell therapy is based on the demonstration that naturally occurring bacteria such as *Lactobacillus acidophilus, Lactobacillus bulgaricus, and Lactobacillus reuteri* can significantly lower serum cholesterol levels due to BSH activities (De Smet et al., 1994; Anderson & Gilliland, 1999; De Smet et al., 1998). It has also been reported that it reduces serum cholesterol levels by 22% to 33% (Jones et al., 2004; Lim et al., 2004). Even a small reduction in serum cholesterol of 1% can reduce risk of coronary heart disease by 2–3% (Manson et al., 1992; Gilliland et al., 1985).

The microencapsulated *Lactobacillus plantarum* 80 (pCBH1) cells can efficiently break down and remove bile acids which is the basis of their application in lowering blood serum cholesterol (Jones et al., 2004; Canzi et al., 2000; du Toit et al., 1998). In this way, interruption of EHC can result in an increased bile acid synthesis of up to 15-fold, from 0.02 to 0.3 mmol/h (Hofmann, 1999). One can measure the effects of different dosages of microencapsulated LP80 (pCBH1) on blood serum cholesterol levels through interruption of the EHC in this way.

Different hypotheses have been advanced to explain the reduction of cholesterol by probiotics, including assimilation (Gilliland et al., 1985; 1990), co-precipitation (Klaver & Van der Meer, 1993; Ling et al., 1992) and enzymatic hydrolysis of conjugated bile acid (Corzo & Gilliland, 1999a & b).

Co-precipitation of cholesterol with cholic acid was observed on deconjugation of both sodium glycocholate and sodium taurocholate, and by all *bifidobacterial* strains studied. More cholesterol was precipitated with cholic acid when sodium glycocholate was used instead of sodium taurocholate (Gnnewald et al., 1982). Increased cholesterol co-precipitation with deconjugated bile salt was observed with decreasing pH levels. Although the potentially positive aspects of probiotic BSH activity have been discussed, other possible negative concerns on BSH activity have become evident recently.

1.5. Suspected adverse impact of deconjugation on host health

Deconjugated bile salts are thought to be a cause of gallstones (Thomas et al., 2000). They are shown to cause depression of growth in chickens due to the poor intake of lipid in the small intestine (Knarreborg et al., 2002) and are suspected to cause colorectal cancer (Singh et al., 1997). Furthermore, it has been thought that the virulence factor of virulent strains in *Listeria monocytogenes* is contributed by the BSH activity

(Dussurget et al., 2002). Many reports are available about the adverse effects of bile salt metabolism, such as extensive deconjugation of bile salts in the small bowel in humans leading to steatorrhoea and gall stone formation (Takahashi et al., 1998)

In healthy persons, the primary bile acids, cholic acid and chenodeoxycholic acid, account for 60%–70% of the total serum bile acids, but in short small bowel (SSB) children with intractable diarrhea, they account for more than 95% (Ohkohchi et al., 1997). In these children, the (taurine- and glycine-) conjugated bile acids account for only 10% of the total serum bile acids. Because of the massive amount of *lactobacilli* present in the intestines of SSB patients the conjugated bile acids present in them will be rapidly deconjugated (Bongaerts et al., 1995; Tannock et al., 1994). Thus large amounts of lipophilic nutrients fail to get reabsorbed, and will be excreted in the faeces (sometimes observed as steatorrhea). This situation leads to steatorrheal loss of nutritional energy, may be an important cause for growth retardation in these children (Dunne, 2001). Antibiotic-mediated replacement of *lactobacilli* is not the preferred therapeutic method because secondary pathogenic bacteria will probably create greater problems of infection. The bicarbonate meant to neutralize the low intestinal pH can stop growth of *lactobacilli* and stimulate the growth of some unwanted bacteria (Kim, 1996a & b). So the alternate approach is to selectively target BSH using its inhibitors.

Clinical studies have shown that the patients with colorectal cancer have higher levels of bile acids in colonic lumen, and animal models confirm the potential role of bile acids as tumor promoters (Dayal & Ertel, 1997). However, the molecular mechanisms of tumor-promoting action of bile acids remain unclear (Cohen & Raicht, 1981; Breuer & Goebell, 1985; Reddy & Rivenson, 1993). Some of the studies have shown that deoxycholic acid-induced tyrosine phosphorylation stabilizes and translocates β-catenin into the nucleus and stimulates urokinase-type plasminogen activator (uPA), its receptor (uPAR) and cyclin D1 expression (Pai et al., 2004). Tyrosine phosphorylation dissociates -catenin from E-cadherin and thus induces loss of cell adhesion. uPA/uPAR-mediated proteolytic degradation of extracellular matrix accompanied by cyclin D1/uPA/uPAR-induced cell proliferation enhances colon cancer growth and progression (Figure 1.5).
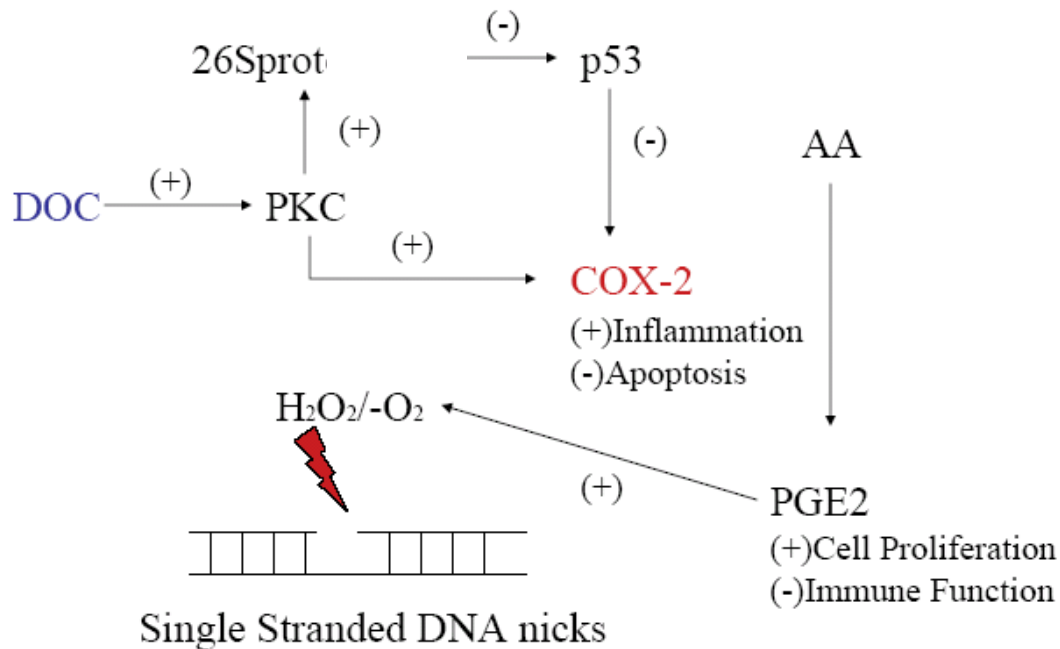
Figure 1.5: Pathway of bile acid induced cell proliferation (26Sprot- 26S proteasome, DOC- deoxycholic acid, PKC- protein kinase C, COX-2 – Cyclo-oxygenase 2, PGE2- Prostaglandin E2)

Primary bile acids were studied as possible colon tumor promoters or inhibitors in a rat model of chemically induced colon cancer. Feeding cholic acid increased the number of animals with tumors and the number of tumors per tumor-bearing animal. The secondary bile acids lithocholic acid (LCA) and deoxycholic acid (DCA) have both demonstrated co-carcinogenic and co-mutagenic activity. Elevations in the ratio of LCA relative to DCA (LCA:DCA) have been implicated in the pathogenesis of both breast and colorectal cancer. Owen, (1985) concluded that the LCA:DCA ratio may be an important discriminating marker in susceptibility to colo-rectoral cancer (CRC). Their study revealed that in 75% of patients with CRC this secondary bile acid ratio was greater than 1.0, whereas in 85% of healthy controls, the ratio was less than 1.0. Patients with breast cancer also tend to demonstrate a higher LCA:DCA ratio compared to healthy controls (Owen et al., 1987).

Several studies have pointed to the activation of the caspase cascade by toxic bile acids in hepatocytes. Members of the caspase family of cysteine proteases have been implicated in the initiation and execution phase of apoptosis (Faubion et al., 1999).

34

Activation of caspase 8 was detected in Chinese hamster ovary (CHO) cells after treatment with either taurine- or glycine- conjugated bile acids, but absent in bile acid-treated CHO cells. Analysis of genomic DNA isolated from treated cells showed TCA and GCDCA induced DNA laddering in CHO cells, indicating that both classes of conjugated bile acids induce apoptosis in this non-hepatic derived cell line. Because the bile acid transporters cause bile acid uptake activity in bile acid receptor-over-expressed (BRO) cells and CHO cells were not the same; it was possible that the sensitivity of CHO cells to taurine-conjugated bile acids was a consequence of non-liver derived cell model. These results demonstrate that the sensitivity of this cell line to taurine-and glycine-conjugated bile acids does not dependent on a specific kind of bile acid transporter (Torchia et al., 2001).

The activation of a PI3K-dependent pathway by taurine-conjugated bile acids appears to decide the fate of BRO cells (Figure 1.5). PI3K activity in response to TCA or GCDCA was not evident in CHO cells, suggesting that the inability to activate survival pathways by taurine-conjugated bile acids commit these cells to apoptosis. These findings also indicate that liver derived cells posses a mechanism that respond to presence of taurine-conjugated bile acids leading to the activation of PI3K (Qiao et al., 2000).

Currently, there is great interest in medical treatment of gallstones using bile acids. A disadvantage of the currently available bile acids used in this therapy is their products of bacterial metabolism that may be toxic or inactive. If these compounds retain their physiologic activity (being of comparable polarity as the glycine and taurine conjugates of CDCA and UDCA), they could act as good detergents and form miscelles with phospholipids to solubilize cholesterol; thus a reduced bacterial destruction will permit use of smaller doses and improved safety (Masclee et al., 1989).

However, it remains to be confirmed whether the BSH activity of probiotics is really beneficial or detrimental to the host. Once this is clarified, it could be used as one of the criteria for the selection of probiotics. Future genetic analyses perhaps can concentrate on expanding the information available on the molecular mechanisms by which bacteria regulate the *bsh* gene in the gastrointestinal conditions and try to answer why the gastrointestinal microbes have evolved to harbor the BSH activity. This may lead

to a better understanding of the interaction between animal hosts and microbes in their gastrointestinal tracts.

1.6. Bile salt hydrolase activity and its implication for BSH-producing bacterial cell

The most essential question still remains unanswered about the advantages and evolutionary importance of *bsh* gene for the bacteria expressing it in the human gastrointestinal tract. Many studies were conducted and several hypotheses have been put forward over the last several years to answer this.

1.6.1. Taurine as an electron acceptor

Some species of bacteria that are able to deconjugate bile salts may be able to use the product amino acid, taurine, as an electron acceptor. Van Eldere et al. (1996) and Huijghebaert et al. (1982) reported that certain strains of *Clostridia* utilized the enzymatically released taurine as an electron acceptor.

1.6.2. Bile salt tolerance and detoxification phenomenon

De Smet et al. (1995) suggested that deconjugation might be a detoxification reaction also, essential for the survival of *Lactobacillus* cell. They assume that the protonated form of the bile salt exhibits toxicity through mechanisms same as that of organic acids, i.e. intracellular acidification. The primary mechanism to prevent intracellular acidification depends upon proton expulsion from the cytoplasm by H^+-ATPases at the expense of ATP (Kobayashi, 1985). The bile salt hydrolase enzyme converts conjugated bile salts into their deconjugated counterparts, which are weaker acids. The latter may then recapture the co-transported proton and prevent in this way the excessive expenditure of ATP to maintain pH homeostasis. Nevertheless, De Smet et al. (1995) observed a high die-off of the BSH-active *Lactobacillus plantarum 80* strain when glycodeoxycholic acid was added to the cell suspensions. It was therefore suggested that a detoxification of the deconjugated bile salt had to follow the deconjugation step.

The BSHs decrease toxicity of conjugated bile acids for bacteria. Compared to their conjugated counterparts, deconjugated bile acids have decreased solubility and diminished detergent activity and may, therefore, be less toxic to bacteria in the intestine (De Boever & Verstraete, 1999). It has also been suggested that BSHs are detergent shock proteins that protect the bacteria that produce them from the toxicity of bile acids in the gastrointestinal tract.

Numerous investigators have disproved also the above hypothesis. Taranto et al. (1996) did not find a relationship between the ability of *E. faecium* strains to grow in bile (0.3% oxgall) and their ability to hydrolyse bile salts (glycocholic acid or taurocholic acid). Similar studies by Gopal et al. (1996) (6 strains of *L. acidophilus* and 8 strains of *Bifidobacterium* spp.), Usman & Hosono (1999) (28 strains of *L. gasseri*), Moser & Savage (2001) (49 strains of numerous *Lactobacillus* spp.) and Ahn et al. (2003) (5 *L. acidophilus* strains) did not correlate bile tolerance with BSH activity. However, studies by three independent groups using wild-type and *bsh* mutant pairs provide a link between bile salt hydrolysis and bile tolerance (Dussurget et al., 2002). *L. plantarum bsh* mutant created by De Smet et al. (1995) showed a strong pH-dependent inhibition due to the presence of glycodeoxycholic acid but not due to taurodeoxycholic acid. As the protonated (non-dissociated) form of bile salts are thought to exhibit toxicity through intracellular acidification in a manner similar to organic acids, BSH positive cells may protect themselves through the formation of weaker deconjugated counterparts.

This could help counteract the drop in pH by recapturing and exporting the co-transported proton (Figure 1.6). These investigations with *bsh* mutants strongly suggest a role for BSH enzyme in bile tolerance, and particularly tolerance to glyco-conjugated bile salts. It is possible that the contradictory studies may have used inappropriate experimental conditions; for example, using tauro-conjugated bile acids to detect BSH activity even though the majority of BSHs show a preference for glyco-conjugated bile acids. Also the bile tolerance of strains was often assayed by growth in lower concentration of bile where BSH activity may not be important. The deconjugated bile acids resulting from bile salt hydrolysis have greater inhibitory effects on bacteria than conjugated bile acids in vitro. However, it is possible that they are precipitated at low pHs in the intestine caused by the fermentation products of lactic acid bacteria. The toxic effect of deconjugated bile acids on *L. plantarum* could be partially alleviated by the addition of a metabolizable carbon source, and suggested that BSH active strains might be capable of detoxifying bile acids by an energy-dependent process (De Smet et al., 1995; Gunn, 2000). It has also been suggested that BSH positive strains may be associated with 7 -dehydroxylating bacteria that will dehydroxylate unconjugated bile

acids (De Boever et al., 1999). Dehydroxylated molecules have low solubility and precipitate at moderately acidic pHs.
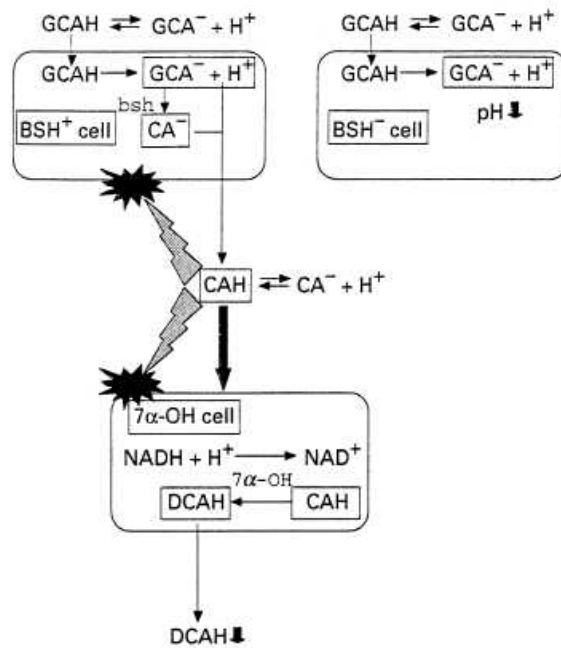
Figure 1.6: Schematic representation of the hypothesis that bile salt hydrolysis is a detoxification mechanism. (Adopted from Begley et al., 2005)

A role for BSH in colonization of the gastrointestinal tract was investigated in *L. monocytogenes* (Dussurget et al., 2002). A confirmed *bsh* mutant reduced bacterial faecal carriage after oral infection of guinea pigs. It was also observed that intestinal multiplication of the *Listeria* could be increased by transforming cells with an extra copy of the gene on a plasmid, and also that BSH activity was found only in pathogenic species of *Listeria*, further confirming the role of BSH in persistence of *L. monocytogenes* within the gastrointestinal tract. It is likely that future investigations will reveal a similar role for BSHs in other organisms. In fact, it has been noted that BSH activity is found primarily in organisms isolated from the gastro intestinal tracts of mammals (*Bifidobacterium* spp., *L. acidophilus, L. gasseri, L. johnsonni* and *L. plantarum*), while organisms isolated from environments where bile salts are absent such as fermented milk preparations and vegetables (*L. lactis, L. del bruecki, L. helveticus, Strep. thermophilus*) do not exhibit BSH activity. Interestingly, the complete genome sequences of intestinal probiotic bacteria have revealed that some strains may possess

more than one *bsh* gene e.g. *L. johnsonii* NCC533 has three and *L. plantarum* WCFS1 has four (Kleerebezem et al., 2003), pointing to the importance of BSH for intestinal survival.

1.6.3. The energy production

Recently, Kurdi et al. (2000) proposed one possible beneficial effect of BSH activity in *Bifidobacteria* upon investigation of the cholic acid transport and accumulation in some intestinal *Bifidobacterium* strains, including two *B. bifidum* strains. They observed that cholic acid, the main free bile acid produced by BSH activity in the intestine, could accumulate inside the bacterial cell in the intestine so long as the bacteria were energized. The entrapment of free bile acids by *Bifidobacteria* can contribute to the decreased production of secondary bile acids, which are considered cytotoxic and precarcinogenic. The enzyme responsible for this undesirable reaction, 7 -dehydroxylase, has been found in *Clostridium* and *Eubacterium* species but not in lactic acid bacteria or *Bifidobacteria* (Begley et al., 2005). Since the hydrolytic products of BSH have properties that cause them toxic to cells and capable of damaging the membrane of mammalian cells, the accumulation of free bile acids within BSH-producing bacteria could counteract the toxic nature of deconjugated bile salts to host cells. On the other hand, it has yet to be determined what advantage BSH activity provides to *Bifidobacterium* strains. So far, there is no report of the ability of *Bifidobacteria* to produce energy from bile salts or of any additional catabolic pathways of the steroid ring structure of bile salts in *Bifidobacteria*.

1.6.4. As virulence factor

The presence of this enzyme perceived as a positive impact on host health till recently has been questioned with the identification of a BSH producing pathogenic *L. monocytogenes* not considered a member of the normal enteric flora (Dussurget et al., 2002). However, the G + C content of the *bsh* gene in *L. monocytogenes* is lower than that of the neighbouring genes which suggests that it may have been acquired from a low G + C content bacteria such as *lactobacilli* (Dussurget et al., 2002).

Virulence genes defined as being part of the σB regulation include *bsh* as well as two genes from the internalin family (*inlA* and *inlB*). Although σB-dependent transcription of *inlA* and *bsh* has also been independently verified by reverse

transcription-PCR and while RACE-PCR confirmed the presence of functional σB promoters for these genes, it is important to note that both genes also appear to be transcribed from σB-independent promoters. Interestingly, PrfA, a positive transcriptional regulator of virulence gene expression, regulates both *bsh* and *inlA*. While PrfA binding has been shown to activate transcription by binding to the promoter region, none of the PrfA binding boxes upstream of the σB-dependent virulence genes described here overlap with the σB consensus promoter region.

PrfA, the transcriptional activator of known *L. monocytogenes* virulence genes, positively regulates the *bsh* gene. Moreover, BSH activity increases at low oxygen concentration. Deletion of *bsh* results in decreased resistance to bile *in vitro*, reduced bacterial faecal carriage after oral infection of the guinea-pigs, reduced virulence and liver colonization after intravenous inoculation of mice. Taken together, these results demonstrate that BSH is a novel PrfA-regulated *L. monocytogenes* virulence factor involved in the intestinal and hepatic phases of listeriosis (Dussurget et al., 2002).

1.6.5. Incorporation of cholesterol in membrane

It has been proposed that BSHs facilitate incorporation of cholesterol or bile into bacterial membranes (Dambekodi & Gilliland, 1998). This incorporation may increase the tensile strength of the membranes or may change their fluidity or charge in a way that could affect sensitivity to -defensins and other host defense molecules (Elkins & Mullis, 2004; Elkins & Savage, 2003). Another suggestion is that this adaptation could strongly select commensals with BSHs, while inhibiting pathogens or other transients lacking BSHs (Pridmore et al., 2004).

1.6.6. Advantages in interspecies competition

Gilliland & Speck (1977) have suggested that deconjugation of bile salts by BSH may enhance antagonist action of autochthonous microorganisms. The deconjugation of bile salts resulted in the formation of more toxic compounds such as cholic acid and deoxycholic acid. The strain with BSH gene has advantage of self-defense mechanism against these toxic compounds.

1.6.7. Role of BSH in stress tolerance

The alternative sigma factor B contributes to *L. monocytogenes* survival under extreme conditions. *opuCA, lmo1421* and *bsh* were identified as putative B-dependent

genes based on the presence of a predicted B-dependent promoter sequence upstream of each gene. *opuCA* and *lmo1421* encode known and putative compatible solute transporter proteins. In conjunction with recent findings that indicate a role for *opuCA* and *bsh* in *L. monocytogenes* virulence, the data presented here provide further evidence of specific B-mediated contributions to both environmental stress resistance and intra-host survival in *L. monocytogenes* (Dussurget et al., 2002). This is consistent with transcriptional studies and confirms that B plays a greater role in *bsh* regulation than does PrfA under the conditions tested.

1.6.8. Other functions

The transcription of *B. longum bsh* gene is coupled to a homolog of *glnE* that encodes a glutamine synthetase adenyltransferase that is part of the nitrogen regulation cascade (Tanaka et al., 2000). However, other studies have shown that *Lactobacillus* do not utilize the steroid moiety of bile salts as cellular precursors suggesting that this is not a universal function of BSH.

1.7 Studies of conjugated bile salt hydrolase and related Ntn hydrolases

Christians et al., (1992) first carried out the cloning of *bsh* gene from the genomic library of *Lactobacillus plantrum* 80. Comparison of BSH promoter from different *bifidobacterium* species can be useful in extending our knowledge on the genetic diversity of the genus *Bifidobacterium* (Table 1.2) and may provide a new genetic marker for their phylogenetic study. Further, studies on the genetic analysis of the *bsh* gene in the *Bifidobacterium* would be of great importance in gaining a better understanding of the physiological role of BSH activity for the animal host and for the BSH-producing microorganisms in the gastrointestinal tracts of the host (Leer et al., 1993).

Table 1.2. Comparison of bsh promotor sequence among three *bifidobacteria* species.

Species[a]	Score[b]	Promoter sequences
B. adolescentis	0.94	GGGATTTTTTCATTTTCAACGACTCTTGGCCATTTTTGGCCACCCTTCATG (N90) **ATG**
B. bifidum	0.39	AACTTAAGATTTTCCCAGCAAGTGACGCTTACCATGAACACGCAAGCAAG (N32) **ATG**
B. longum	0.19	AACCCAGAGTTCTCCCAGCGGCCGATGGGTATCATGAACGTGCAAGCCAA (N25) **ATG**
Consensus[c]		TTGACA ⟶ TG TATAAT ⟶ ↑
		-35 -10 TSP

[a]Sequence obtained from *B.adolenscentilis* ATCC 15705,*B. bifidum* ATCC 11863 (Kim et al. 2004b), and *B.longum* SBT2928 (Tanaka et al., 2000).

[b]Score indicates the fitness value to the consensus promoter sequence when predicted by using a Nural Network Promoter Prediction (NNPP).

[c]The consensus -10 and -35 motifs of the E. coli are presented. The upward arrow indictes the putative transcriptional start point (TSP).

1.7.1. Conjugated bile salt hydrolase as member of the Ntn Hydrolase family

B. longum BSH exhibits high levels of sequence homology with BSH of various *lactobacilli* and *C. perfringens*, and also with Penicillin V acylase (PVA) of *Bacillus sphaericus* (Suresh et al., 1999). Based on its structural similarity to PVA, the BSH enzyme has been proposed to be a new member of the N-terminal nucleophile (Ntn) hydrolase superfamily with Cys as the N-terminal residue.

1.7.2. N- terminal nucleophile hydrolase (Ntn hydrolases) family

The members of the Ntn hydrolase superfamily have an N-terminal nuclephilic residue, Ser, Thr or Cys whose side chain O or S atom act as nucleophile and its own free - amino group acts as base in catalysis (Brannigan et al., 1995). In the BSH enzyme discussed here the Cys residue at the N-terminus serves as the nucleophile and proton donor in the catalytic process. The representative fold of this superfamily is composed of a four-layered catalytically active   core structure. This core structure consists of two antiparallel -sheets packed against each other, and they are sandwiched by a layer each of antiparallel -helices on either side (Brannigan et al., 1995). The comparison of structures between members of the family shows great variability in the number of secondary structural elements and in details of their arrangement (Oinonen & Rouvinen,

2000). Another reported characteristic of these enzymes is that the respective precursor undergoes an intramolecular autocatalytic cleavage to remove a peptide to generate the active enzyme.

1.7.3. Members of the Ntn hydrolase family

The hydrolases and references mainly to their three-dimensional structures are listed below according to their importance and utility in research reported in the thesis:

- Penicillin V acylase from *Bacillus sphaericus* (*Bsp*PVA) (Suresh *et al.*, 1999)
- Bile salt hyrolase from *Clostridium perfringens* (*Cp*BSH) (Rossocha et al., 2005)
- Penicillin G acylase from *E. coli* (*Ec*PGA) (Duggleby *et al.*, 1995), *Providencia rettgeri* (McDonough et al., 1999)
- Aspartylglucosaminidase (AGU) from human (Oinonen *et al.*, 1995; Peräkylä & Rouvinen,1996; Aronson et al., 1996 and 1999; Peräkylä & Kollman, 1997; Saarela et al., 2004) Lupinus (Michalska et al., 2006)
- Glycosylasperaginase (GA) from *Flavobacterium meningosepticum* (Guan et al., 1998; Xu et al., 1999, Hejazi et al., 2002)
- Glutamine phosphoribosyl-pyrophosphate amidotransferase (GAT) from *Bacillus subtilis* (Smith *et al.*, 1994), *E. coli (*Muchmore et al., 1998)
- L-Aminopeptidase-D-Ala-amidase (APA) from *Ochrobactrum anthropi* (Bompard-Gilles *et al.,* 2000, Polderman-Tijmes et al., 2002a & b)
- N-carbamyl-D-aminoacid amidohydrolase (DCase) from *Agrobacterium sp.* (Nakai et al., 2000)
- Glutaminase domain of glucosamine-6-phosphate synthase (GPS) from *E. coli* (Isupov *et al.*, 1996; Heuvel et al., 2004).
- Ornithine acetyltransferase (OA) (Abadjieva et al., 2000; Elkins et al., 2005)
- U34 peptidase (Pei & Grishin, 2003)
- MTH1020 conserved protein from *Methanobacterium thermoautotrophicum* (Saridakis et al., 2002)
- Cephalosporin acylase (CPA) from *Pseudomonas sp.* (Kim *et al.*, 2000; Kim et al., 2006).
- Glutaryl 7-aminocephalosporanic acid acylase from *E. coli* (Lee *et al.*, 2000)

- Heat shock locus V (HslV) from *E. coli* (Bochtler, et al., 1997)
- Gamma-glutamyl transferase (GGT) from *E. coli* (Okada et al., 2006)
- 20S subunit of the proteasome *Saccharomyces cerevisiae* (Groll et al., 1997), *Thermoplasma acidophilum* (Löwe *et al*., 1995), *E. coli (*Bochtler et al., 1997).

Topological comparison of the -core structure of the members of Ntn-hydrolases revealed eight totally conserved secondary structural elements in three layers of a four-layered structure and all strands of -sheets are antiparallel to each other. The central -sheets in the core were found to have different packing angles, ranging from 5º to 35º. In all Ntn-hydrolases eight totally conserved secondary structure elements are present. The -sheets pack in such a way that there is a 30º rotation between the -sheets.

The catalytic action of Ntn hydrolases has similarity with the classical serine protease mechanism but with some catalytic economy. The side chain nucleophile atom of N-terminal residue attacks the scissile peptide's carbonyl group in the substrate to form a transition state complex known as tetrahedral intermediate. The amino acid residues in the vicinity of the nucleophile and interacting with it decide the reactivity of this group. During the reaction, the residues that constitute the oxyanion hole stabilize the covalently linked transition state intermediate (Perakyla & Rouvinen, 1996; Perakyla & Kollman, 1997). The conformational distortion that occurs with the formation of the tetrahedral intermediate causes the carbonyl oxygen of the scissile bond to move deeper into the active site and occupy the oxyanione hole.

The -amino nitrogen atom acts as base, which takes up the liberated proton (general base catalysis). The acylation step is complete when the -amino group of the nucleophile donate the proton to the nitrogen of the scissile peptide bond. The covalent bond between part of the substrate and the enzyme is formed (acyl-enzyme intermediate) and part of the substrate is released. The deacylation step begins when the basic -amino group of the nucleophile residue accepts a proton from the water molecule and the hydroxyl group of this water attacks the carbonyl carbon of the acyl-enzyme intermediate. The negatively charged intermediate is stabilized by oxyanion hole, as in the acylation step. The reaction is complete when the -amino group donates the proton

back to the nucleophile atom. This process is aided by the polarizing effect of the water molecule (Dodson, 2000). The variation found in the Ntn nucleophile group also indicate that they are more adaptable and perhaps more primitive.

All of the proteins belonging to Ntn hydrolase family catalyze the hydrolysis of amide bonds, but they differ in their substrate specificity. Accordingly, the shape and the size of the substrate-binding pocket differ, as do the spatial distribution of interacting residues. However, in addition to their similar types of oxyanion hole formation, proteaosome and PGA have substrate binding residues in the same places in the 5-strand (Ala20 in proteaosome and Phe24 in PGA), in the 11-strand (Met45 in proteaosome and Ser67 in PGA) and in almost the same places in the 12-strand (Arg33 and Ile35 in proteaosome and Phe57 in PGA). On the other hand, AGA and Grpp have substrate binding residues, Arg and Asn, which are located differently both in their structures and in their sequences.

The N-terminal nucleophile residues in the Ntn hydrolases could be substituted with one another, in some cases even without affecting the posttranslational processing or activity. For example, in PGA Cys can replace Ser without affecting processing but not without affecting activity (Choi et al., 1992). In Proteasome in which Thr is the nucleophile residue, Ser is an acceptable alternative for proteolytic activity only, whereas Cys is effective for processing only (Seemuller et al., 1996). In the case of AGA and GAT, Thr could be substituted by Ser but only for processing and not for catalysis (Fisher et al., 1993).

The removal of the signal peptide in cephalosporine acylase has little effect on precursor processing or enzymatic activity demonstrating that signal peptide has no influence on the folding and maturation of the enzyme. However, the substitution of the last residue Gly in the spacer peptide leads to an inactive and unprocessed precursor (Li et al., 1999).

The posttranslational autocatalytic processing can be either intermolecular between precursors or intramolecular within precursor. The precursors can be cleaved by a protease in bacteria that recognize the Gly-Ser bond in the case of *E.coli* or else by an intramolecular cleavage mechanism within the precursor. The former one is ruled out as no protease that recognize the Gly-Ser site could be found in bacteria. The co-expression

of the inactive precursor with the active form of enzyme resulted in no detectable increase in processing or activity (Li et al., 1999). All evidences suggest that the precursor might be processed through an intramolecular, but not intermolecular mechanism. The autoproteolytic potential of the precursor molecule might be originating partially from a highly strained, tight turn conformation at the scissile peptide bond. In addition, conserved residues are properly aligned as the general acid-base and oxyanion hole that are critical to facilitate the N› O or N› S acyl shift or to stabilize the reaction intermediate (Xu et al., 1999). There are also reports of self-cleavage being prevented by the conformational constraints (Ditzel et al., 1998)

The N-terminal methionine of a large number of proteins is cleaved by methionine aminopeptidase (MAP). The efficiency of removal of the initiator methionine is defined by a highly conserved local substrate specificity, which is determined by both methionine and its adjacent residues in the sequence. The seven aminoacids that have the smallest radii of gyration (Gly, Ala, Ser, Thr, Pro, Val and Cys) are substrates for MAP. Other residues downstream of these specificity-determining residues have little impact on the enzyme action (Sherman et al., 1985; Biossel et al., 1988).

1.7.4. Penicillin V acylases

PVA is industrially used in the hydrolysis of penicillin V (penV) to produce 6-aminopenicillanic acid (6-APA), which is the precursor molecule for semi-synthetic -lactam antibiotics (Shewale & Sudhakaran, 1997). PVA from *Bacillus sphaericus* is a homotetrameric protein of 37.5 kDa subunits. The crystal structure of *Bsp*PVA (Suresh et al., 1999) placed this protein in the N-terminal nucleophile (Ntn) hydrolase superfamily (Brannigan et al., 1995). It contains tripeptide before the active-site nucleophile and that must be unmasked by a post-translational processing event for the functional protein.

Among the penicillin V acylases the crystal structure of only *Bsp*PVA is reported. The functional molecule is a well-defined homotetramer of 222 organization formed by four monomers, which generates a flat disc-like assembly. The X-ray analysis revealed that the *Bsp*PVA monomer contains two layers of central anti-parallel β-sheets, above and below which is a pair of anti-parallel helices. There are two extensions, one from the upper pair of helices and the other from the C-terminal segment that interact with the neighbouring monomers of the tetramer and help to stabilize the quaternary structure

(Figure 1.7). The -sheets and the -helices and their connectivity and organization are characteristic of the members of Ntn-hydrolase family (Suresh *et al.*, 1999).
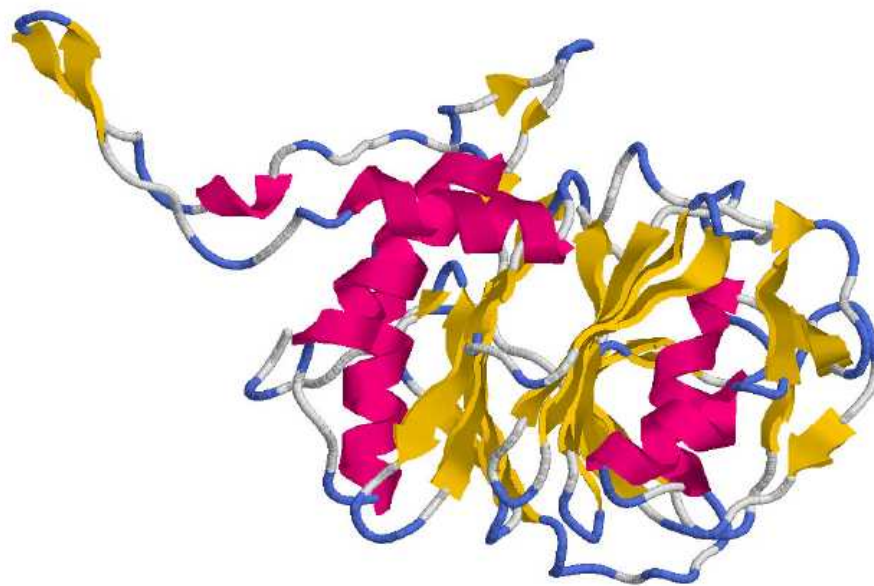
Figure 1.7: Structure of the Monomer of PVA from *B. sphaericus* (PDB ID: 3pva) showing the characteristic Ntn hydrolase fold.

1.7.5. Crystal structure of a conjugated bile salt hydrolase from *Clostridium perfringens*

The crystal structure reported is that of BSH from *Clostridium perfringens* (*Cp*BSH) in tetragonal space group $P4_122$ with two monomers in the asymmetric unit, and in complex with taurodeoxycholate in orthorhombic space group $F222$ with monomer in the asymmetric unit (Rossocha et al., 2005). The *Cp*BSH monomer is made up of a single globular domain with approximate dimensions of 40X50X55 Å which excludes an extended loop containing residues Gln188 to Pro225 which stretches out about 40 Å. Including the -sheet of this extended loop, the domain has a six-layered structure of composition (Figure 1.8). The core of the protein is composed of two antiparallel -sheets that contain N- and C-termini and are sandwiched by a layer of antiparallel -helices.
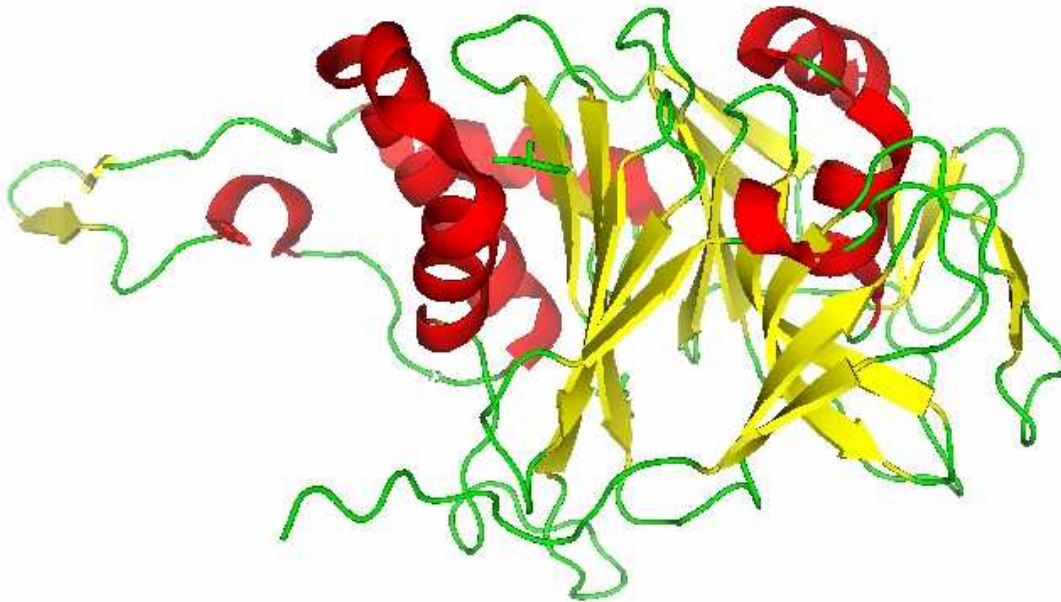
Figure 1.8: Structure of Monomer of *Cp*BSH (PDB ID: 2bjg) showing characteristic fold.

1.7.6. Penicillin G acylases

Penicillin acylases (penicillin amidohydrolases, PGAs, EC 3.5.1.11), a subclass of β-lactam antibiotic acylase superfamily, are enzymes that catalyze the selective hydrolysis of relatively stable amide bond in penicillins and some cephalosporins while leaving the labile β-lactam ring intact. Penicillin acylases are important in pharmaceutical industry for the production of semi-synthetic β-lactam antibiotics via the key intermediates 6-aminopenicillanic acid (6-APA) (Figure 1.12) and 7-amino-3-deacetoxycephalosporanic acid (7-ADCA) (Shewale et al., 1990; Shewale et al., 1989).
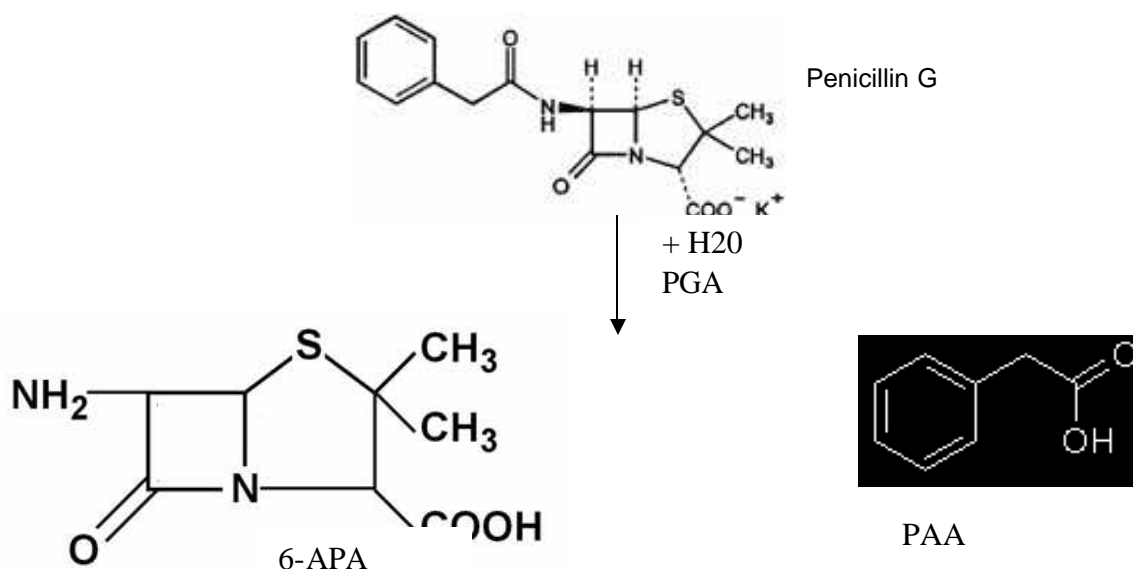
Figure 1.12: Reaction catalysed by PGA

Although PGAs are mainly used in the hydrolysis of natural β-lactams, they are capable of catalyzing the acylation of the amino group of key intermediates and thus can synthesize semi-synthetic β-lactam antibiotics (Valle et al., 1989, Bruggink et al., 1998, Savidge & Cole 1975). Enzymatic production of semi-synthetic β-lactam antibiotics has many environmental benefits compared to the classical chemical route. PGAs are also useful for other applications such as peptide synthesis (Fuganti & Grasselli, 1986), removal of protecting groups (Waldmann, 1988) and separation of racemic mixtures (Fuganti, 1986).

PGA is an intracellular enzyme produced by several microorganisms, including various bacteria, actinomycets, fungi, and yeast. Microorganisms have been extensively screened for penicillin acylase production. The cultures that are used most commonly in large scale production of PGA include *Escherichia coli*, *Bacillus megaterium*, *Streptomyces lavendulae*, *Achromobacter* sp., *Providintia rettgeri*, *Actinoplanes* sp., *Bovista plumbea*, *Kluyvera citrophila* (Barbero et al., 1986), *Pseudomonas melanogenum*, and *Pencillium chrysogenum*.

The crystal structures of PGA from *E. coli* and *Providintia rettgeri* are reported. The *Ec*PGA is kidney shaped in cross-section, with a deep cup-shaped depression in the center. Serine B1, its catalytic nucleophile is situated at the mouth of the binding pocket. A -sandwich comprising the N-terminal residues of A chain and much of B-chain is

flanked by -helices. The reported structure of the enzyme complexes with PAA and Phenylmethylsulphonoyl fluoride (PMSF) locates the binding site for the side chain of the substrate. The phenyl moiety of each of these compounds points towards the interior of the protein into a hydrophobic pocket. The complementary fit explains the specificity of the enzyme towards the phenyl moiety of a broad range of substrate. Structure of PGA shows a ring of six helices with a seventh protruding through a central hole and two further helices at the C-terminal end of the A chain wrap around the B-chain (Figure 1.9).

Figure 1.9: Crystal structure of PGA from *E.coli* (PDB ID: 1gk9) showing the heterodimeric association of unbunits.

The structure of PGA has been shown to be a heterodimer, composed of a small α and a large β subunit of molecular weights 24,000 and 62,000, respectively (Barbero et al., 1986; Alvaro et al., 1992). It is produced as a single-chain precursor consisting of 844 amino acid residues in the cytoplasm. Subsequently, the precursor undergoes processing through removal of a 26 residue signal or leader peptide meant for directing the protein to the periplasm and of a 54 residue spacer peptide, to produce mature enzyme in the periplasm in the form of a heterodimer consisting of an - and a -chain of 209 and 555

amino acid residues, respectively (Barbero et al., 1986). The serine residue with a newly generated free -amino group at the N-terminus of -subunit serves as nucleophile in the catalysis.

PGA from *Kluyvera citrophil*a (*Kc*PGA) compared to *Ec*PGA has attracted more interest due to its better and suitable features for industrial applications, such as ease of immobilization and greater stability with respect to temperature, pH, and presence of organic solvents (Fernandez-Lafuente et al., 1991 & 1996; Lie et al., 2006).

*Af*PGA also has a clear industrial advantage over other well-characterized penicillin acylases in -lactam conversions because of its higher thermostability and also for its high synthetic efficiency in enantioselective synthesis. This makes the *Af*PGA a more attractive biocatalyst both in hydrolysis and in synthetic conversions (Verhaert et al., 1997).

1.7.7. Aspartylglucosaminidase (AGU)

This enzyme cleaves the glycosylamide bond which links oligosaccharides to the peptide of asparagine-linked glycoproteins.

N^4-(-N-acetyl-D-glucosaminyl)-L-asparagine + H_2O $\longrightarrow$ N-acetyl-beta-glucosaminylamine + L-aspartate.
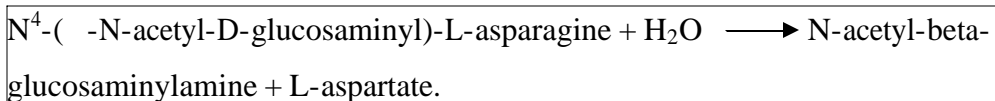
Aspartylglucosaminidase is a heterotetramer of two pairs of   and   chains. The catalytically essential residue, the N-terminal threonine of the -chain is situated in the deep pocket of the funnel-shaped active site (Figure 1.10). The sub-cellular location is lysosome. The functional defects in AGA cause aspartylglucosaminuria (AGU). AGU is an inborn lysosomal storage disease. Clinical features of AGU include mild to severe mental retardation manifesting from the age of 2, coarse facial features and mild connective tissue abnormalities.
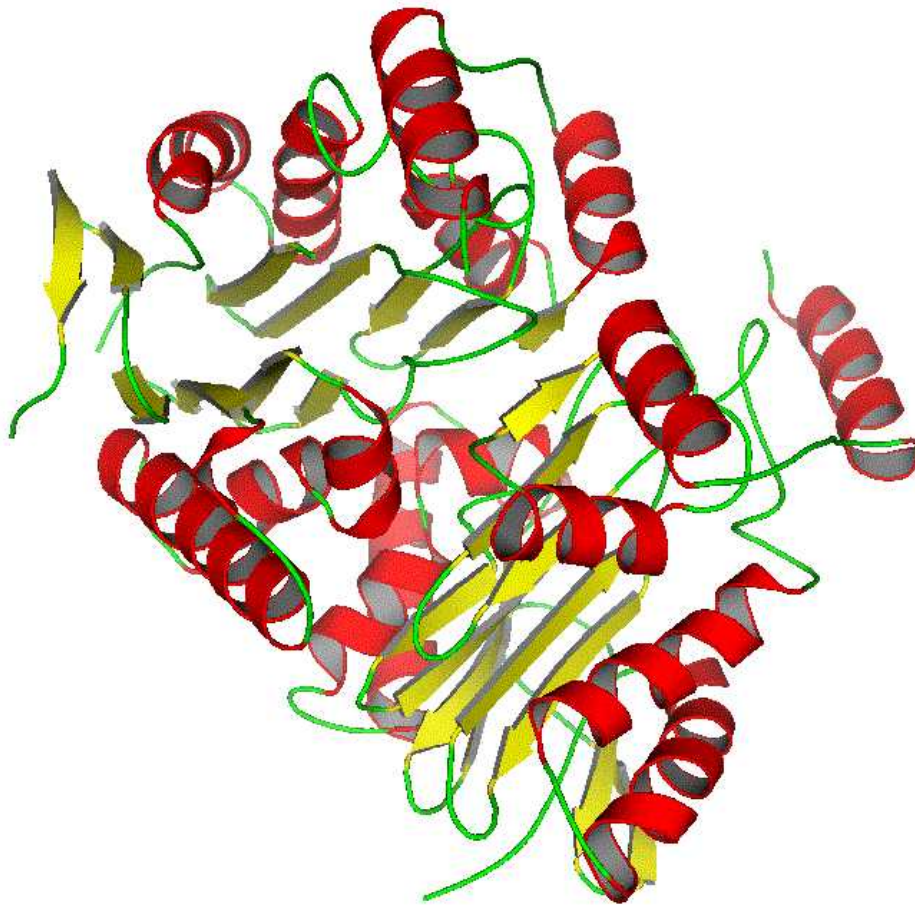
Figure 1.10: Crystal structure of Aspartylglucosaminidase from *F. Meningosepticum* (PDB ID: 1ayy) showing its characteristic   fold.

N-terminal cysteine nucleophile (Ntcn) hydrolases

1.7.8. Glutamine amidotransferase (GAT)

This protein transfers the glutamine amide nitrogen to a variety of substrates including amino acids, nucleotides and carbohydrates. Most GATs can use ammonia (NH_3) as an alternative nitrogen source. Sites for glutamine binding and for NH_3-dependent synthesis are localized in different domains (sometimes in different subunits), termed `glutamine' and `transferase' domains. Phosphoribosyltransferases (PRT) are involved in the biosynthesis and metabolism of nucleotides. They catalyse reactions at the position 1 of ribose with phosphoribosylpyrophosphate (PRPP)

> PRPP $+$ L-glutamine $+$ H_2O › 5-phosphoribosyl-1-amine $+$ L-glutamate $+$ PPi ;
>
> 5-phospho-beta-D-ribosylamine $+$ diphosphate $+$ L-glutamate › L-glutamine $+$

5-phospho-alpha-D-ribose 1-diphosphate $+$ H_2O.
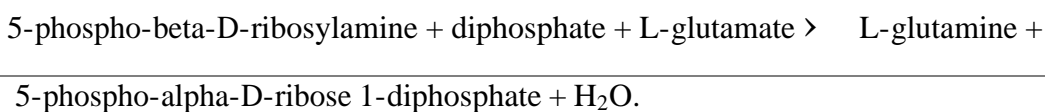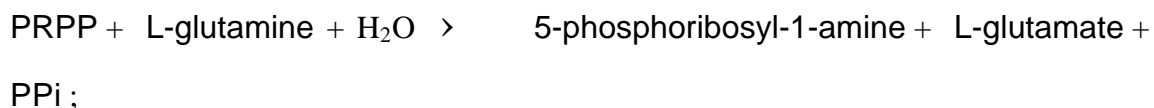
It consists of a homotetramer and contains one type-2 glutamine amidotransferase domain. Here the Cys1 is the catalytic nucleophile residue. This residue can adopt two conformations, one active and one inactive, glutamine binding locks the residues in a predetermined conformation. When a nitrogen acceptor is present Cys1 is kept in active conformation explaining the phenomenon of substrate induced activation of the enzyme, and that the residue Arg 26 has a central role in coupling (Smith *et al*., 1994).

1.7.9. Hexosephosphate aminotransferase, (L-glutamine-D-fructose-6-phosphate amidotransferase or Glucosamine-6-phosphate synthase) (GPS)

It catalyzes the first step in hexosamine metabolism, converting fructose-6P into glucosamine-6P using glutamine as a nitrogen source.

> L-glutamine $+$ D-fructose 6-phosphate $\longrightarrow$ L-glutamate $+$ D-glucosamine 6-phosphate.

It is a Homodimer and is commonly located in cytoplasm. It Contains one type-2 glutamine amidotransferase domain (Isupov *et al*., 1996).

1.7.10. L- Aminopepetidase-D-Ala-esterase/amidase (APA)

Aminopeptidases (alpha aminoacyl-peptide hydrolase) are exopeptidases that catalyze the hydrolysis of the amino-terminal residue from polypeptide substrates. Most of them recognize L-amino acids and are of critical biological and medical importance because of their key role in protein modification and degradation. They are mostly divalent cation-dependent or thiol enzymes (Taylor, 1993a & b). Only three aminopeptidases active on peptides containing N-terminal D-residues have been isolated so far, all of them are from the bacterium *Ochrobactrum anthropi*. The first two peptidases, DAP and DmpB, are two homologous strict D-stereospecific aminopeptidases, isolated from the strains SCRC C1-38 and LMG7991, respectively (Asano et al., 1989; Fanuel, 1999). The third one, DmpA, is a l-aminopeptidase showing the unique feature of hydrolyzing d-amides and d-esters also (isolated also from the strain LMG7991; Fanuel, 1999).

1.7.11. U34 peptidase family proteins

Peptidase U34 family consists of nearly 150 members and with unclear catalytic mechanism, from sequence comparison it is concluded that they might belong to the Ntn hydrolase superfamily. In this family the catalytic residue is cysteine and generated by the posttranslational modification. The strongest signal for all U34 family proteins resides in the motif containing the catalytic cysteine residue similar to the N-terminal -hairpin in the structure of PVA. Other common features in this motif include the hydrophobic pattern and position occupied mainly by small residues near catalytic cysteine (Pei & Grishin, 2003).

N-terminal serine nucleophile (Ntsn) hydrolases

1.7.12. Cephalosporine acylase (CPA) or Glutarylamidase (E.C. 3.5.1.11)

Like penicillin acylases are employed in the production of 6-APA, cephalosporin acylases are used in the production of 7-ACA, which in turn can be used to synthesize semi-synthetic cephalosporins. Irrespective of the low sequence similarity between PAs and CPAs the structural homology at their active sites are impressive. However, despite this structural conservation, they catalyse very different substrates (Kim *et al*., 2000). N-terminal Ser1 is involved in nucleophilic attack of the acyl group linkage of cephalosporine side chain. Based on similarity of their sequence four different classes of CPAs can be grouped into one family of Ntn hydrolase.

N-terminal threonine nucleophile (Nttn) hydrolase

1.7.13. Proteasomes

The proteasome, a multisubunit proteolytic complex, is the central enzyme of nonlysosomal protein degradation. In both the cytosol and nucleus. It is involved in various biological processes such as the degradation of misfolded protein, protein encoded by virus and other intercellular parasite and short-lived regulatory proteins. Proteins destined for destruction are conjugated to a molecule of ubiquitin that binds to the -NH2 group of lysine residue or any N-terminal amino group of a protein. The complex binds to ubiquitin-recognizing site(s) on the regulatory particle. The protein is unfolded by the ATPases using the energy of ATP. The unfolded protein is translocated

into the central cavity of the core particle and breaks down into short peptides (Figure 1.11).
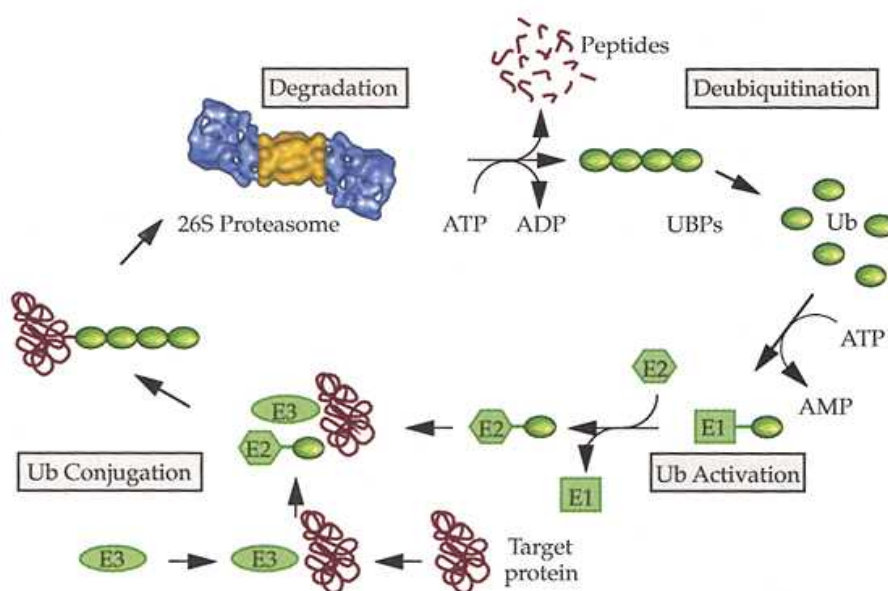
Figure 1.11: Mechanism of protein degradation by proteosome

1.7.14. Protein engineering that changed a PGA to a CPA

The production of semi-synthetic cephalosporin such as 7-aminocephalosporanic acid (7-ACA) is carried out in the pharmaceutical industry mostly by chemical methods, which are dependent on toxic compounds such as iminoether, nitrosyl chloride and methanol. The biggest problem of enzymatic industrial production is that cephalosporine acylase requires glutaryl-7-ACA (GL-7-ACA) as a primary substrate and its specificity for cephalosporin C (CpC) is too low to allow the industrial production of 7-ACA. The attempts to obtain a direct enzymatic transformation of CpC into 7-ACA by single enzyme were unsuccessful. So there is great need to develop an enzymatically efficient and economically viable enzyme to fulfill the pharmaceutical industry requirement (Kim and Hol, 2001; Fritz-Wolf et al., 2002).

A good candidate identified to improve the CpC specificity was PGA, because the residues forming the side-chain binding pocket in CPA matched well with the corresponding residues in the PGA active site (Gabor & Janssen 2004) (Table 1.3). So the protein engineering technique to alter the residues that form the side-chain binding pocket of PGA to appropriate residues was expected to produce substrate specificity similar to

that of CPA with improved CpC deacylation activity (Oh et al., 2003; Kim and Hol, 2001). This is one successful application of available structural information on two enzymes to change the substrate specificity of one to the other.

Table 1.3: Comparison between corresponding residues of PGA and CPA in catalytic and binding sites based on structural alignment

Enzyme›	PGA	CPA
Residues involved in catalysis	Ser B1	Ser B1
	Gln B23	His B23
	Ala B69	Val B70
	Asn B241	Asn B244
Residues as part of side chain binding pocket	Val B56	Arg B57
	Phe A146	Tyr A149
	Thr B32	Thy B33
	Ile B77	Phe B177
	Pro B49	Gln B50
	Trp B154	Tyr B153
	Phe B24	Leu B24

1.8. Outcome of this study

The detailed studies of bile salt hydrolase from bacteria have provided the following information. The study has essentially established the relationship between the PVA and BSH in terms of structure-function relationship and evolution. The penicillin acylases are used in the commercial production of 6-APA, the starting compound for the synthesis of semi-synthetic penicillins. The detailed comparison of the structure of a PVA with a related conjugated bile salt hydrolase enzyme with shorter C-terminal sequence, could help in reaching a better understanding of the mechanism of action and stability of the enzyme and could help in exploring the possibility of modifying the substrate specificity.

A better understanding of the role of BSH may be exploited in the selection and rational design of probiotic strains. Administration of bile salt-hydrolysing strains to control serum cholesterol levels shows much promise. This biological approach may

especially be appealing to an emerging health conscious society since the ingestion of prebiotics containing food may be considered more natural than other therapies using cholesterol-lowering drugs.

From recent research findings with respect to intestinal commensals there is clear picture emerging on the relationship of these hydrolases in microbial-host interaction process. The commensal genomes are adapted to survive in intestinal microenvironment on a mutual basis. So if there is any disturbance in the microenvironment some factor is getting dominated and the balance is upset. This study will provide the more comprehensible picture about the positive role of BSH in bacterial cell and in the host cell. There are also some reports on the negative effects of BSH activity. For example, in SSB patients the same beneficial role can just turn into a harmful one. In colorectal cancer it induces the cell proliferation. So the picture may look more complicated or even contradictory if we look at the single process in isolation. So a holistic scenario of these processes should be considered before trying any therapeutic intervention.

Even though there is no strong evidence to show an essential role for BSH in bacterial cell, many speculations are presented to assign a natural role for it. This starts from utilization of amino acid to tolerance of toxic environment. Newly emerging trends in research show that the enzyme can be accommodated as a good candidate in the path of parasitic defense mechanism. This is especially true in the role like sigma dependence of the BSH involvement in glutathione metabolism.

The structural and biochemical study of the BSH and PGA is expected to provide information on residues directly involved in catalysis of a set of stereochemically related substrates, which in turn can help in protein engineering towards modification of substrate specificity for novel applications. The comparative study of these enzymes are expected to throw light on their origin and evolution into three major branches and members of all the sub-families appearing in different forms from Archea bacteria to the most evolved humans.

CHAPTER 2
MATERIALS AND METHODS

2.1 Materials

Chemicals used in fermentation and protein expression were Yeast Extract, Tryptone, Sodium Chloride (NaCl), Ampicillin, Kanamycin, Isopropyl thiogalactopyranoside (IPTG) were purchased from Sigma chemical company, St. Louis, USA.

Chemicals used for the purification of BSH proteins from crude extract of cells such as Trizma, di-Potassium hydrogen phosphate, Potassium Hydrogen Phosphate, Streptomycin sulfate, Ammonium sulfate (AS), Sodium acetate, Nickel chloride, Glycerol, -mercapto-ethanol, Bromo-phenol-blue (BPB), Acrylamide, N,N'-methylene bisacrylamide, Sodium dodecyl sulfate (SDS), Acetic acid, Methanol, TEMED (N,N,N',N'-Tetramethylethylenediamine), Ammoniumpersulfate (APS), substrates and ligands such as penG, penV, 6-APA, PAA, POAA and compounds from Bile salt kit, Molecular weight marker kits for SDS-PAGE etc, were purchased from Sigma chemical company, St. Louis, USA.

Following chemicals were used extensively in the crystallization trials, obtained from different sources: Sodiumcacodylate buffer, Magnesium Chloride hexahydrate ($MgCl_2.6H2O$), Polyethylene glycols (PEG 400-20K), Lithium sulfate (Li_2SO_4). Crystallization additives tried for improvement of diffraction quality of crystals were Isopropanol, Dioxan, Dimethyl formamide (DMF), Dimethyl Sulfoxide (DMSO), Glycerol, Ethelene Glycol, NDSB-195, NDSB-201 (Non-Detergent Sulfobetaines), trimethylamine N-oxide (TMAO), 2-ethoxyethanol (ethyleneglycol monoethyl ether, Cellosolve), maltose, sucrose. All other chemicals used were of analytical grade and procured from local manufacturers. Multiwell trays used in crystallization were purchased from Becton Dickson and Company or Falcon. . The slab gel electrophoresis unit was purchased from Tarson, India. The other specialized instruments were used in the experiments are mentioned in the appropriate places.

Silicon graphics Octane workstations were used for graphics display, for running the crystallographic programs and for related calculations. Programs from 'CCP4 suite'

(Collaborative Computational Project, Number 4., 1994) were used extensively for calculations. The program QUANTA (Accelrys) was used for graphics display, visualization of protein structures and model building.

2.2. Cloning and Over Expression

2.2.1. Construction of the over-expression Plasmid pET26b/*bsh* and pET28b/*pga*

The *B. longum bsh* gene was fished out from the Plasmid pBH1351 (Tanaka et al., 2000) and *K. citrophila pga* gene from chromosomal DNA. The Vector pET-26b (Figure 2.1) and pET-28b (Figure 2.2) were used for over-expression of *Bl*BSH and *Kc*PGA, respectively.

The Promega miniprep kit was used for plasmid extraction purpose. Chromosomal DNA was isolated from *K. citrophila* by the bacterial DNA isolation protocol Sambrook et al. (1988). The quality of genomic DNA was checked through 0.8% agarose gel. The Genomic DNA was quantified by recording spectrophotometer readings to measure $A_{280/260}$. The Plasmid and chromosomal DNA were visualized by agarose gel containing ethidium bromide.

The primers overhanging restriction sites were designed to fish out the gene. The gene was amplified using PCR under standardized conditions. The PCR products were purified using Promega purification kit to remove the excess primers and other protein impurities.
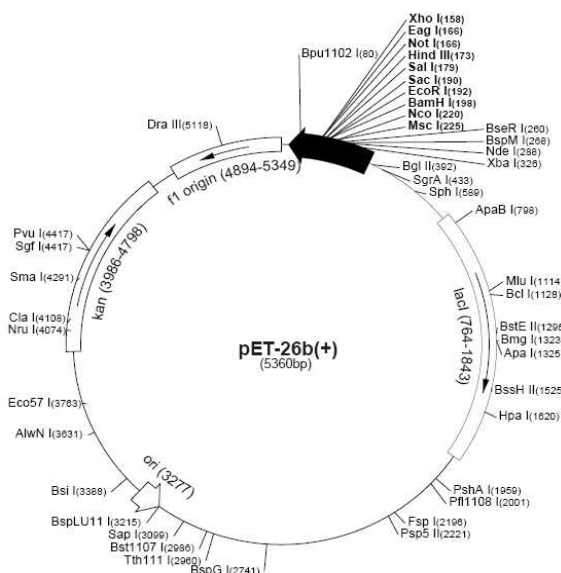
Figure 2.1: Vector map of pET-26b
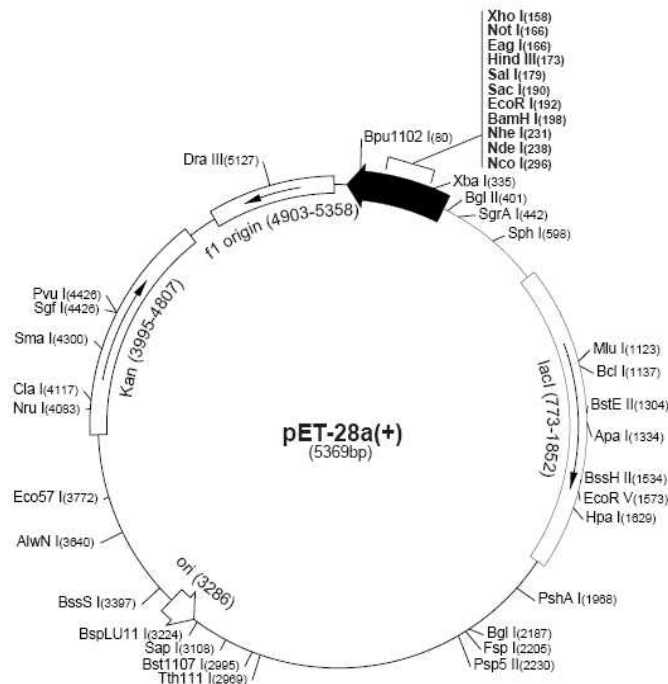
Figure 2.2: Vector map of pET-28a

2.2.2. Restriction digestion of vector and PCR product

The vector pET-26b and pET-28b, primers were digested with appropriate restriction enzymes. Both reactions were carried out at 37 °C with suitable buffers. After restriction digestion the products were loaded on to a 1.5 % agarose gels. The bands were carefully sliced from the gel. Then the purification of the vector and PCR products from the gel were carried out using gel extraction kit. The digested product was visualized on the gel for quantification.

2.2.3. Ligation of vector and inserts

Approximately 1:1 amount of vector and insert were incubated with T_4 DNA ligase with suitable buffer at 15 °C overnight.

2.2.4. Transformation into maintainer host

The ligation products were transformed into compatible cells of αDH5a. Competent cells were made using $CaCl_2$ method and the transformation of ligation product was done as described in Sambrook et al (1988). The transformed cells were plated out onto Luria-Bertain (LB) medium containing 35 µg ml^{-1} kanamycin and incubated overnight at 37 °C.

2.2.5. Checking for inserts from the colony

Following the overnight growth a number of colonies were picked up from the plate randomly and grown in LB liquid medium containing 35 µg ml^{-1} kanamycin and the plasmid DNA was isolated using Promega Wizard mini Prep kit. Restriction digestion with BglII and EcoRI (for pET-26b/*bsh*) or NdeI and EcoRI (for pET-28b/*pga*) were set up with each plasmid DNA prep in order to ascertain which colonies had the insert DNA (Figure 2.3)
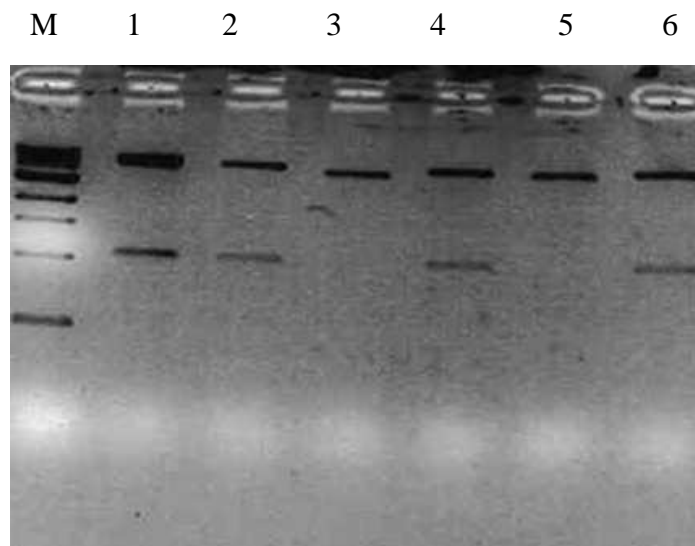
Figure 2.3: Checking for the insert (*bsh*) from the transformed colonies.

M-1 kb ladder Lane 1-6: digested plasmid from positive colonies.

The positive plasmids were transformed in to the expression host and screened for protein expression

2.3. Expression of cloned protein

Two liters of LB medium containing 35 µg ml^{-1} kanamycin was inoculated with 20 ml of stationary phase (BL21/pET-26b/*bsh* or BL21/pET-28b/*pga*) culture. Fermentation at 37 ℃ was continued till OD at 600 nm reached 0.6 which was fallowed by induction with 1 mM isopropyl β-D-1-thiogalctopyranoside (IPTG). After 5 h the cells were harvested by centrifugation and the cell paste was re-suspended in 50 mM potassium phosphate buffer pH 6.5 containing 10 mM DTT prior to storage at -20 ℃.

2.4. Protein purification

The cell suspension was thawed and lysed by sonication in an icebath, 2 min each five times on Biosonic III sonic oscillator (Bronwill Scientific Co., USA) at 20 KHz and 300 W. Cell debris were removed by centrifugation at 18,000xg for 20 min at 4 ºC.

2.4.1. Treatment with streptomycin sulfate

To remove residual DNA from cell free extract 1% (w/v) streptomycin sulphate was added and the suspension was stirred for 20 min. Then the mixture was centrifuged for 10 min at 18000xg at 4 ºC.

2.4.2. Ammonium sulfate fractionation

The cell suspension from the streptomycin sulfate fractionation was subjected to ammonium sulfate (AS) precipitation by slow addition of finely ground AS under stirring. The protein fraction that precipitated at 0.2 – 0.8 of AS saturation was collected by centrifugation at 10000xg for 20 min. The precipitant was re-suspended into the same buffer and dialysed overnight with two changes against 100 times volume of potassium phosphate buffer pH 6.5, containing 1 mM DTT.

2.4.3. Octyl- sepharose column chromatography

The clear dialyzate was treated with 24 % saturated AS and centrifuged to remove the undissolved solids. The supernatant was then applied on a 30 ml Octyl-sepharose column pre-equilibrated with 24% saturated AS in potassium phosphate buffer pH 6.5 containing 1 mM DTT. The column was washed with 5 times its volume of the buffer at a flow rate of 8 ml h^{-1} and eluted with reverse linear gradient of AS (24 to 0 %). Fractions (3 ml) were collected on an automatic fraction collector at flow rate of 10 ml h^{-1}. The fractions were analyzed by SDS-PAGE and those containing BSH were pooled and dialyzed overnight against Tris-HCl pH 8.0 containing 1 mM of DTT.

2.4.4. Q-sepharose column chromatography

The dialyzed fraction was applied to a Q-sepharose column pre-equilibrated with Tris-HCl pH 8.0 containing 1 mM DTT. The column was washed with 5 times its volume of buffer at a flow rate of 8 ml h^{-1}. The protein was eluted with a 0-1 M linear gradient of NaCl as described previously. The fractions containing BSH were pooled and concentrated with centriprep concentrator (Amicon, Bevesting, USA) and stored at –20 °C.

2.5. Isolation and purification of PGA enzyme directly from *Kluyvera citrophila*

2.5.1. Organism and growth media

Kluyvera citrophila (ATTC 21285) was grown on glutamate medium composed of yeast extract: 5, peptone: 10, sodium chloride: 2, sodium glutamate: 10 (gm/L); Phenyl-acetic acid (PAA) was added as an inducer to a final concentration of 15 mM and the pH set to 7.2. A 10 ml overnight starter culture was inoculated into 300 ml broth in 500 ml Erlenmeyer flask and incubated at 28 ºC and shaking at 150 rpm for 24 h.

2.5.2. Isolation of PGA

Cells were harvested, washed with 50 mM potassium phosphate buffer, pH 7.5 by centrifugation at 6000xg and 4 ºC for 15 min and the recovered cells were suspended in potassium phosphate buffer, pH 7.5. The intracellular PGA was extracted by cell disruption by sonication (20 KHz, 300 W) for 4 min (3-4 exposures of 1 min each) at 4 ºC. The cell debris was removed by centrifugation.

2.5.3. Treatment with streptomycin sulfate

To remove residual DNA from cell free extract 1% (w/v) streptomycin sulphate was added and the suspension was stirred for 20 min. Then the mixture was centrifuged for 10 min at 18000xg at 4 ºC.

2.5.4. Ammonium sulfate fractionation

The supernatant was subjected to AS precipitation with the precipitate being collected from the 40% to 80% AS saturated fraction. This was dissolved in minimum amount of 50 mM phosphate buffer, pH 7.5. The solution was dialyzed against the same buffer.

2.5.5. Octyl-sepharose column chromatography

The dialyzate was treated with 24 % ammonium sulphate and applied to an Octyl-sepharose column as described previously.

2.5.6. Hydroxy appatite coloum chromatography

Those containing PGA in Octyl-sepharose fraction were pooled and dialyzed against 10 mM phosphate buffer pH 7.5 and applied to a hydroxy appatite column (Amersham Pharmacia Biotech, Sweden) equilibrated with 10 mM phosphate buffer pH 7.5 and eluted with a 0.10 -1 M phosphate buffer linear gradient. The positive fractions

were pooled and concentrated and passed through Sephadex G-100 column and fractions were eluted with 10 mM potassium phosphate buffer pH 7.5

The purity of individual fractions and of the pooled fractions were checked using 12% (w/v) SDS polyacrylamide gels as described by Lamelli (1970). The gels were stained using coomassie blue and the protein concentration was determined in accordance with the method of Lowry et al., (1951) using BSA as standard. The pooled fractions were concentrated with Amicon membrane concentrator (Amicon, Bevesting,USA) and the protein was stored at –20 °C. Only about 0.5 mg of pure protein was obtained from 1 L of culture.

2.6. Biochemical assays

2.6.1. Bile Salt Hydrolase Assay

BSH enzyme activity was determined by estimating the amount of free amino acids released upon incubation of the enzyme sample with 1 mM sodium taurocholate or 1 mM sodium glycocholate at 40 ℃ in 10 mM sodium phosphate, pH 6.5 containing 10 mM DTT. After 10 and 30 min of incubation time, a 25 μL aliquot was withdrawn and the reaction arrested by mixing with 25 μL of 15% (w/v) trichloroacetic acid. The sample was spun at 10000xg for 1 min and the supernatant was mixed with an equal volume of 2% ninhydrin solution before boiling for 15 min. The absorption was recorded at 570 nm and the amount of product formed was estimated from a calibration curve (Lee & Takahashi, 1966). One unit of BSH activity is defined as the amount of enzyme that liberates 1 μmol of the amino acid from substrate per min. Specific activity was defined as the number of units of activity per milligram of the pure protein.

2.6.2. Penicillin G acylase assay

Enzyme sample contained in a final volume of 1 ml of 50 mM phosphate buffer and 1 mM penG was incubated at pH 7.5 and 40 ℃ for 10 min. 6-aminopenicillanic acid (6-APA) formed was estimated using p-dimethyl aminobenzaldehyde (Shewale *et al*. 1987). One unit of enzyme activity is defined as the amount of enzyme required to produce 1 μmol of 6-APA per min under assay conditions.

2.7. Crystallographic methods

X-ray crystallography is a powerful tool for three-dimensional structure determination of proteins. Although other methods like NMR and electron and atomic force microscopy are also useful tools to solve protein structures, they are still limited by the protein size and restricted structural information that can be obtained. The structure determined from X-ray crystallography provides information on structural chemistry of the interaction between protein and the cognate ligands. Thus, the knowledge of structural details allows to rationally design novel substrates or inhibitors in modern drug design for the pharmaceutical industry. In addition, functional details are derived from protein structures, such as their biological chemistry of enzyme activity.

The flowchart (Figure 2.4) describes a step-by-step protocol to proceed from a crystal to a protein structure, with the help of a known protein structure. From crystal to final model, the methods employed in the study will be briefly discussed, emphasizing those specifically used in the research reported here, including crystallization, cryo-crystallography, data collection, molecular replacement and refinement.

2.7.1. Crystallization

Crystallization of the molecule of interest is a limiting step in protein crystallography. Protein crystals are composed of approximately 50% solvent, though this may vary from 30 - 78% (Matthews, 1985). They are labile, fragile, and sensitive to external environments owing to their high solvent content, and the weak binding energies between protein molecules in the crystal (Littlechild, 1991). The only optimal conditions suitable for their growth are those that cause little or no perturbation in their molecular properties.

The crystallization of proteins from solution is a reversible equilibrium phenomenon. It can be considered as consisting of three stages: nucleation, growth and cessation of growth. Nucleation the necessary first step in the crystallization process, influence it decisively. Consequently, the ability to control it is of primary importance in crystallization experiments. Condition for nucleation, different from the supersaturation conditions, presents a free energy barrier which must be overcome in specific ways, and which subsequently makes crystal growth an energetically favorable process. The formation of crystals is due to decreasing free energy of the system by the formation of

Initial screening

↓

Crystal improvement

↓

Choosing correct cryoprotectant and method

↓

Data collection

↓

Data reduction

↓

Moleculaer replecement using 3 PVA as model
Rotation and translation function

↓

Calculation electron dencity map
Starting model building

↓

Refinement using Refmac
Structure factor calculation

↓

Manual rebuilding

↓

Calculation of new electron density map with new phases

↓

Final model

↓

Structure validation

Figure 2.4: Schematic representation of steps involved in protein X-ray crystallography.

many new interactions which outweighs the decreasing entropy of the system, allowing for a highly organized structure. In other words the reduction in free energy of the system is the thermodynamic driving force that causes ordering in crystals (McPherson, 1982).

The basic strategy of growing protein crystals is to generate certain degree of supersaturation in the solution. At the equilibrium point, the number of protein molecules entering the solution is same as the number leaving the solution. This is referred to as the solubility limit of a protein. When the amount of protein in the solution is below this limit, the solution is undersaturated. If the amount of protein is equal to this limit, the solution is in saturated state. Crystals can grow only when the solubility exceeds the solubility limit. Every protein has a unique solubility limit (Figure 2.5). Decreasing the solubility of the protein is the most effective way of creating supersaturation. Crystals will grow only from a non-equilibrium supersaturated solution. Supersaturation can be achieved by different approaches including altering the buffer pH, temperature, protein concentration, dielectric constant of the medium, and precipitant concentration.
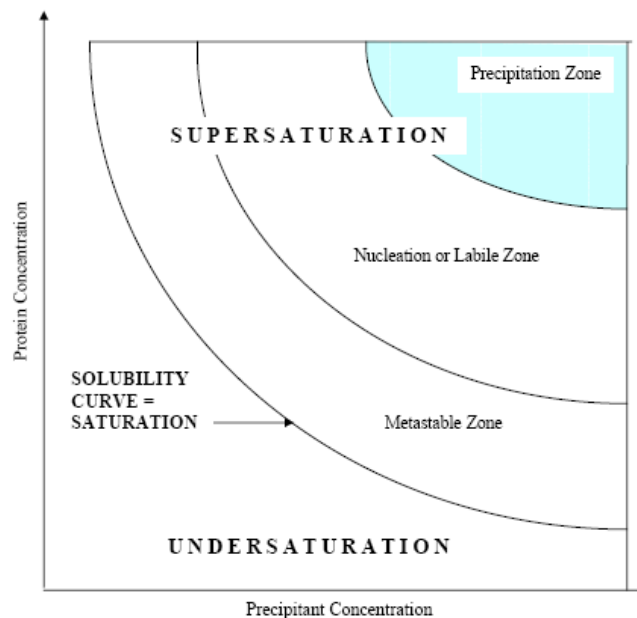
Figure 2.5: The solubility phase diagram of crystallization from solution (Blow, 2002)

The most popular experimental method used for crystallization is the vapor-diffusion method, both hanging-drop and sitting-drop method. The hanging-drop vapor-diffusion method is an efficient means of screening many crystallization parameters. The

advantage is that it requires only a small volume of droplet, which can be as low as 1 µL per experiment, so only a small amount of sample is consumed for screening and optimization of the crystallization conditions (McPherson, 1998). The reason for the popularity of the hanging-drop method is the ease of performing the experiment, only a 24 well-plate (such as Linbro plate), grease and cover slips are required (Figure 2.6).

In this method a drop composed of a mixture of macromolecule and precipitating solution is placed in vapor equilibration against a reservoir solution of precipitant and buffer. Hanging-drop vapor-diffusion method uses a VDX 24-well plate that had a fine bead of vacuum grease applied around the edge of the well. Precipitating solutions (500 µL to 1ml each) are composed of precipitant, buffer solution, additive, and distilled water pipetted into the 24 well reservoirs of the crystallization plate. Then 1 µL of protein solution was dispensed onto the center of a clean siliconized 22 mm circular glass cover slip and mixed with 1 µL of reservoir solution instantaneously. The cover slip was then inverted using tweezers without losing the drop and sealed over the reservoir with gentle pressure to ensure proper sealing. The plate is placed on crystallization rack and maintained at 20 °C or room temperature.
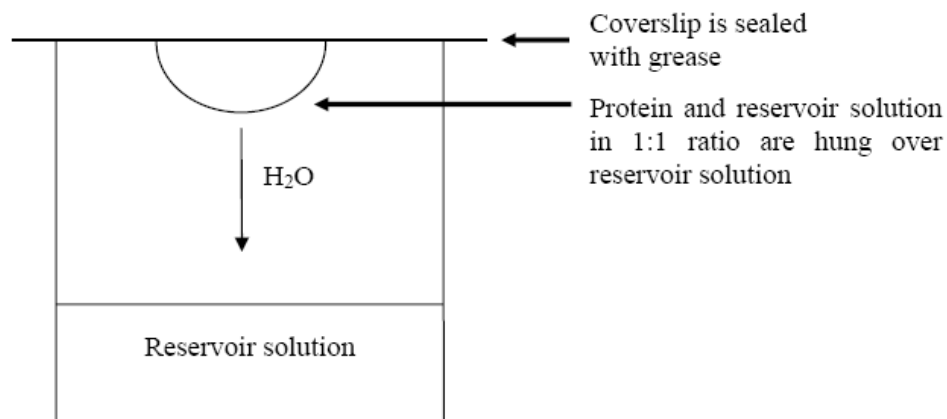
Figure 2.6: The hanging-drop vapor-diffusion method for protein crystallization.

The sitting-drop vapor-diffusion method was performed using 96-well Crystal Clear Strip plate. The method was similar to the hanging-drop method, but only 250 µL

reservoir solution was used instead of 500 μL. In this method, 1 μL of protein solution and 1 μL of reservoir solution were mixed together at the top of the ledge, and finally the plate was sealed with sealing tape, placed on crystallization rack and the temperature was maintained at 20 ºC or room temperature.

Seeding is a useful technique for growing crystals by producing a seed-stock solution in cases where spontaneous homogenous nucleation does not occur. It is generally believed that the nucleation will be initiated at a higher level of supersaturation, the labile zone (Stura and Wilson, 1991). To use seeding techniques, good quality crystal seeds have to be selected and introduced into the metastable labile zone. After seeding, the seeded crystals will continue to grow. The crystal growth conditions can also be optimized independently without the need to introduce nuclei of the protein itself (Luft & Detitta, 1999).

2.7.2. X-Ray diffraction study

X-rays are generated when electrons collide with atoms of a metal target, e.g. copper. The electrons liberated from a heated filament and accelerated by high voltage impinge the metal target to produce X-rays. X-rays are electromagnetic radiation of wavelength about 0.1 – 10 nm (1 -100Å) (Blundell & Johnson, 1976). The reason for selecting X-rays to study the three-dimensional structures of proteins is the convenient range of wavelengths of X-rays (0.5 Å – 1.6 Å) which is of the same order of magnitude as the bond length of the atoms within protein molecules. The bond length between atoms within a protein is about 0.15 nm or 1.5 Å, thus these wavelengths can be utilized to visualize the geometry and structure of protein molecules through X-ray diffraction (Blow, 2002). Individual atoms in a molecule can diffract X-rays; however, it is difficult to analyze an image from a single molecule. One can obtain analyzable diffraction pattern from a crystal rather than from single molecule. This is because a crystal is composed of a number of molecules arranged in a regular or ordered manner. Each molecule within the crystal contributes to the diffraction equally, and thus the diffracted X-ray beams are intense enough to be measured (Rhode, 2000). The X-rays from laboratory source used were copper radiation of wavelength 1.5418 Å.

Data were collected on frozen *Bl*BSH and PGA crystals on the in-house X-ray diffraction facility at NCL, Pune, India. Diffraction data were collected using the

oscillation method, in which a crystal is exposed to the X-ray beam and rotated about a spindle axis over a given oscillation range, while the resulting diffraction pattern is recorded on a image-plate detector. Crystals were cryoprotected using selected cryosolution. The procedure as described in the next section for cryoprotecting the crystals was followed. The crystal was mounted on the goniometer head using a cryo-loop. The liquid nitrogen jet produced by X-stream (Rigaku, Japan) was used to freeze the crystal. The temperature was set to – 160 ºC.

After mounting the crystal it was aligned to the center of the X-ray beam and the detector was placed at an appropriate distance. An initial image was collected to determine the cell dimensions of the crystal. The exposure time of initial image was generally 120 seconds. Using the initial exposure image the resolution limit of diffraction was determined and analysis of diffraction pattern was carried out using CRYSTALCLEAR software to determine the unit cell dimensions and tentative space group of the crystal.

2.7.3. Flash-cooling of protein crystals

When crystals are exposed to intense X-ray radiation, it can cause extensive damage to crystals. This is due to the weak lattice forces within the crystal structure (McPherson, 1982) and the X-rays can produce sufficient free radicals to cause specific chemical changes on the protein molecules such as to break the disulfide bonds (Ravelli & Garman, 2006; Ravelli et al., 2003). This problem can be overcome to some extent by collecting the data at lower temperatures. In cryocrystallography, the crystal is cooled to cryogenic temperatures. In an effort to minimize lattice damage during cryogenic cooling, cryoprotective reagents or cryoprotectants are added to the solution to prevent the formation of crystalline ice in the internal and external solution as well as at the interface between crystal and solution.

Organic solvents such as MPD, methanol, ethanol, isopropanol, ethylene glycol and glycerol are the commonly used cryoprotectants (Garman & Schneider, 1997). Glycerol is one of the most common cryoprotectants used in protein cryocrystallography. It has been reported that glycerol can be employed as a very useful cryoprotectant to provide cryoprotection for 50 typical protein crystallization solutions (Garman & Mitchell, 1996). Ethylene Glycol, MPD and lower molecular weight PEGs, glycerol,

70

glucose, sucrose and xylitol are also widely used. All of the above cryoprotectants have been tried to screen the cryo-conditions of *Bl*BSH and PGA crystals. The cryosolution was prepared matching the mother liquor. The composition of potential cryosolution included cryoprotectant, precipitant, buffer, salt, additive and MilliiQ water.

For instance, one of the best crystallization conditions of *Bl*BSH where crystals were consistently grown was chosen for this preliminary screening. Glycerol was the first cryoprotectant chosen for the screening at different concentrations ranging from 15% - 35%. Different concentrations of glycerol were used to prepare 500 μL of cryosolutions which were then tested in liquid nitrogen for 5 - 10 seconds (without any crystal in the loop) to see whether the cryosolution turned glassy upon freezing. Other cryoprotectants were also examined in the same way (Table 2.1).

There were two usual ways to cryoprotect crystals. The first method was to put 5 μL of cryosolution just beside the drop of crystal ready to be lifted up, with the use of CrystalCap and CryoLoop. A crystal was removed from the drop and transferred into the cryosolution for a given amount of time, after which the crystal was ready for X-ray exposure. The second method was to mix an equal volume of cryosolution directly onto the drop of the crystal (such as 5 μL + 5 μL), afterwards to transfer the crystal into the full cryosolution as quickly as possible. For PGA crystals that were grown in 45 % saturated AS conditions, a thin layer of denatured protein was formed on the drop; therefore, the second method was applied to dilute out the layer, and then the crystal could be transferred into the cryosolution without any difficulty. If the crystal is easily scooped out, the first method can be applied conveniently.

2.7.4. Handling protein crystals, mounting, cooling and storing

Historically the loops were made from copper or tungsten wires of 1-2 mm diameter and 25-75 μm thickness, but presently they are made from fine nylon fibers of diameter 0.05-1.0 mm and thickness 10 & 20 μm. These types of cryo-loops show minimal background diffraction due to the optically clear environment and the loops are thin enough to be conveniently freezed. Uniform results can be obtained as compared to the conventional mounting techniques such as capillary mounting (Teng, 1990).

Table 2.1: List of cryoprotectants used for flash-cooling the crystals

S.No.	Cryoprotectant	Concentration (%)
1	Glycerol	15-35
2	PEG 400	25-35
3	MPD	20-35
4	Ethylene glycol	15-25
5	2,3,5-hexane triol	15-30
6	2,4-dexanediol	15-30
7	Sucrose (1M)	20-30
8	Isopropanol	30-40
9	Mineral oil	90
10	Sodium malonate (1M)	40-50
11	Ammonium sulphate (saturated)	60-75
12	PEG 600	20-30
13	PEG 1000	30
14	PEG 2000	20-25

If the crystal needs a faster rate of cooling, or to be transported to synchrotron facility, the crystal can be rapidly plunged beneath the surface of liquid nitrogen for a few seconds to a few minutes. The crystal can then be mounted on a diffractometer for experiment or transferred to a storage tank for storage and transportation (Parkin & Hope, 1998).

2.7.5. Data collection and processing

In the case of most of the crystals discussed in this thesis the preliminary crystal characterization and the final data collection were carried out using Raxis IV^{++} area detector mounted on a Rigaku rotating anode X-ray generator operating at 50 kV, 100 mA with Osmic focusing mirrors to monochromate and focus X-rays. The determination of the orientation of the crystals and initial processing of the diffraction data were performed using the CRYSTALCLEAR software. Typically, data were collected in 0.5 oscillation frames, with exposure times ranging from 120 to 600 s. Exposure times were chosen to obtain sufficiently strong intensities and to make optimal use of the dynamic

range of the detector. The 2.25 Å resolution *Bl*BSH data were collected at the F-1 beam line at the European Synchrotron Radiation Facility (ESRF) Gronoble, France using an ADSC Quantum4 CCD detector.

Intensity data were processed using the programs DENZO and SCALEPACK of HKL suite of programs (Otwinowski & Minor, 1997).

The processing of the diffraction data involved following steps:

1. Indexing the diffraction pattern and determining the unit cell parameters

2. Refinement of the crystal and detector parameters

3. Integration of the diffraction intensities from contiguous frames

4. Refining the relative scale factors between equivalent measurements

5. Precision refinement of crystal parameters using whole data

6. Merging symmetry related reflections across frames and statistical analysis of the data

The first three steps could be carried out using DENZO, while steps 4 to 6 could be accomplished using SCALEPACK.

The two factors that influence the completeness of data are the geometric and the informative content. The geometric factor, arising from the symmetry of crystal lattice and the detector setup, is a quantitative factor related to a number of variables including orientation of reciprocal axes, the selection of rotation range appropriate for the crystal symmetry, crystal-to-detector distance, crystal mosaicity and beam divergence (Dauter, 1999). The informative factor includes the quality of the data, the dynamic range of detector and the R-factor. The longer the exposure time, the higher the intensities and better the signal-to-noise ratio (Dauter, 1997 & 1999). But this will also have the disadvantage of increased background intensity.

The X-ray data quality for macromolecular crystallography is assessed by a global indicator, the merging R-factor (Rmerge) or symmetry R-factor (Rsym). The merging Rfactor is defined by the following equation (Blundell and Johnson, 1976):

$$R_{merge} = \frac{\sum\limits_{hkl} \sum\limits_{i=1}^{N} |\langle I_{hkl} \rangle - I_{hkl}(i)|}{\sum\limits_{hkl} \sum\limits_{i=1}^{N} I_{hkl}(i)}$$

where $I_{hkl}(i)$ is the ith measurement of the reflection with indices hkl and I_{hkl} is the mean value of the N equivalent reflections.

R_{merge} is commonly used to guide decisions during data reduction, such as determining the resolution of the data. Diederichs & Karplus (1997) have argued that R_{merge} is seriously flawed, because it has an implicit dependence on the redundancy of the data. The R_{merge} is less appropriate than the Chi-square, because of its dependency on the multiplicity of the data and on the symmetry of the crystal. The desired crystal rotation per image is an important consideration during data collection. To prevent overlaps from successive zones, the crystal rotation per image must be selected depending on the resolution limit and the cell dimension along the beam direction. The oscillation range should be reduced to allow for the crystal mosaicity and beam divergence. The effect of cell dimension along the spindle axis will be uniform, since the crystal is rotated around that axis and will never lie along the beam direction assuming the crystal axes are perfectly orientated with respect to the beam.

2.7.6. Matthew's number

Once the space group and unit cell dimensions of the crystal are known, it is possible to estimate the number of molecules in the crystallographic asymmetric unit and the solvent content of the protein crystals based on a knowledge of the molecular weight of the protein. The following equation is used (Matthews, 1968).

$$Vm = \text{unit cell volume} * z / (MW * n)$$

Where Vm is the Matthew's number, MW is molecular weight of the protein, n is the number of molecules per asymmetric unit and z is the Avagadro's number. The Matthew's number for all the crystal forms of *Bl*BSH were calculated by assuming 35500 Dalton as the molecular weight of a monomer.

2.7.7. Structure solution

Multi-wavelength Anomalous Dispersion (MAD), Multiple Isomorphous Replacement (MIR) and Molecular Replacement (MR) are the three methods widely used in protein crystallography to solve protein structures. MAD technique is currently the most popular one owing to the availability of variable wavelength synchrotron radiation sources and sensitive detectors like Image plate and CCD to collect anomalous data on selenium derivative that can be prepared using molecular biology techniques for any protein (Beauchamp & Isaacs, 1999). Another advantage is MAD needs only one heavy atom derivative compared to MIR technique that requires more derivatives.

2.7.8. Molecular replacement method

The research reported in the thesis mainly used MR method in structure solution. Molecular replacement technique is the simplest of all and can be used when a homologous protein with structural similarity is available in the database. The success of the MR method depends on the structural similarity between the search model and the unknown structure. The pioneering studies of Rossman & Blow (1962) laid the foundation for the successful use of MR method. With the rapid expansion of protein data bank and with the increase in the number of structures available MR method is now used successively in many protein structure determinations.

The basic principle of molecular replacement is rooted in the Patterson function. A Patterson map contains information of two types of vectors:

Intramolecular or self-Patterson vectors of pairs of atoms in the same molecule. These vectors are relatively short and are thus clustered around the origin;

Intermolecular or cross-Patterson vectors, which are generally longer than self-vectors.

The self-vector cluster would be equal for non-crystallographically related molecules in the same unit cell but also very similar for similar molecules in different crystals, apart from a rotation difference. The cross vectors provide information on the required translation of molecules to their positions relative to symmetry elements. Thus the process of molecular replacement can be divided into two steps: orientation and positioning, obliterating the need for a direct six-dimensional search. If the structure of a molecule related to that under study is already known, it can be used to arrive at starting

phases through orientation and positioning of the known molecule into the unit cell of the unknown.

MR also exploits the presence of non-crystallographic symmetry to obtain phase information and reduce phase uncertainties. The ideas developed by Rossmann & Blow (1962) were extended by Bricogne (1976) to real-space symmetry averaging to improve the electron density.

The calculation of rotation functions used in the study of structures in this thesis was performed using the program package *AMoRe* (Navaza, 1994). The rotation function in AMoRe is based on the fast rotation function, using spherical harmonics and Bessel function expansions (Crowther, 1972). The adaptations made by Navaza permit more accurate computation of the rotation matrices (Navaza, 1994), and enhance the resolution of the rotation function peaks. Removal of the origin (usually performed by using an inner limit for the integration radius) is achieved by excluding the lower order spherically symmetrical Bessel functions.

AMoRe uses straightforward T0 translation functions (Navaza & Saludijian, 1997), with the option to incorporate information concerning already placed models (Navaza, 1994), enabling the positioning of additional molecules in the asymmetric unit. The program outputs the Eulerian angles , , and a translation vector (x,y,z). The calculation of an R-factor and a correlation coefficient are provided to help decide the correct solution.

The three-dimensional structure of PVA (pdb, 3PVA ; Suresh et al., 1999) has been used as model in molecular replacement method in the case of *B*lBSH and 1GK9 (Duggleby et al., 1995) for PGA. The position of the molecule in the unit cell of interest is defined in terms of three rotational and three translational parameters. A very powerful component of AMoRe is its fast rigid-body refinement procedure, which is very efficient in the evaluation of the correctness of the molecular replacement solution. It is based on ideas first developed by Huber & Schneider (1985). It is not a least-squares rigid-body refinement of coordinates, but rather a minimization of the misfit, where rotations and translations are calculated by interpolation and phase shifts.

2.7.9. Refinement

The next step after getting the coordinates using a preliminary model is its refinement. Refinement is the process of fitting the parameters of the model to achieve a closer agreement between the calculated and observed structure factors. The refinement of the model is by incrementing the positional parameters and the temperature factors of the atoms applying stereochemical restraints on the structure. At low resolution data the hydrogen atom is omitted since they have only one electron and their influence on X-ray scattering is low. The number of atoms and thus the number of parameters to be fitted are large in proteins.

Constraints are introduced into the refinement in the form of stereochemical criteria deduced from data of small molecular structures of amino acids and peptides in which the bond lengths and angles have been determined to high precision (Engh & Huber, 1991). The stereochemical information can be applied in two ways. For constrained refinement the geometry is considered rigid which reduces the number of parameters to be refined. In contrast, when the stereochemical parameters are allowed to vary around a standard value, the refinement is restrained by appropriate weights applied between geometric parameters and X-ray terms. Restraints on bond lengths, bond angles, torsion angles, and van der Waals contacts have the apparent effect of increasing the number of observations.

$$R = \frac{\sum_{hkl} ||F_{obs}| - k |F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

2.7.10. REFMAC5: program for maximum-likelihood refinement

Refinements of structures were carried out using restrained maximum-likelihood refinement implemented in the program REFMAC5 (Murshudov et al., 1996; Murshudov, 1997). Each cycle of the program can be considered to carry out grossly two steps:

• Estimates the overall parameters of likelihood using the free set of reflections

• Uses these parameters to build the likelihood function and refine the atomic parameters. To refine the atomic parameters only a working set of reflections were used. Maximum likelihood method of refinement performs the calculation of the first derivative and makes an approximation of the second derivative of the likelihood function with respect to refinement parameters and then estimates the shifts to be added to the parameters. The values of standard deviations are calculated for the geometric restraints used during the refinement. These values are also the estimated standard deviations that determine the relative weights of the corresponding restraints. The restraints used by the program include distances, bond angles, torsion angles, peptide planarity, chiral volumes, van der Waals radii and B values. Model building and manual fitting of the model were carried out using X-AUTOFIT module of QUANTA (Accelrys) and the addition of waters was with X-SOLVATE, again of QUANTA.

2.7.11. Structure validation and evaluating the quality of structure

The cycle of rebuilding the model in QUANTA and restrained refinement in REFMAC5 are followed by structure validation. The correctness and precision of the atomic parameters in a structure need to be assessed thoroughly, both during and at the end of the refinement. An initial indication of the reliability of a protein crystal structure depends on the extent of resolution of the diffraction data used in refinement. The conventional R factor (R_{cryst}) is an indicator of the general fit of the crystal structure with observed data. However, it is not a reliable independent validator, since it can be biased by the exclusion or overfitting of the data. A better assessment of the fit between observed and calculated structure factors is by calculating the free R factor (R-free) (Brunger, 1992)

Despite sterochemical restraints, it is possible to overfit or misfit the diffraction data: an incorrect model can be refined to fairly good R value. To overcome this problem, the R_{free} a statistical quantity was proposed.

$$R\text{-free} = \frac{\displaystyle\sum_{hkl \,\in\, T} ||F_{obs}| - k\,|F_{calc}||}{\displaystyle\sum_{hkl \,\in\, T} |F_{obs}|}$$

R-free refers to the 'test set T', a set of reflections (commonly 5-10% of the observed data) which are excluded from refinement. Refinement is carried out with the remaining reflections only, called 'working set W'.

In this scheme, the diffraction data are divided into two sets: a large 'working' set (usually comprising 90% of the data), and a complementary 'test' set (comprising remaining 10%). The data in the working set are used in the normal crystallographic refinement process, whereas the test data are not. R-free that measures the agreement between observed and computed structure factor for a 'test' set of reflections that is omitted in the modelling and refinement process. The R-free value is correlated with the accuracy of atomic models. In practice, this means that models with serious errors can be identified by a very high free R value (>0.40) irrespective of the value of the conventional R value, which may be very low (around 0.20). The advantage of the R-free value over knowledge based validation methods is that it can be applied to any type of model and does not depend on the availability of database-derived knowledge.

2.7.12. Estimate of coordinate precision

Isotropic temperature factors are related to the thermal vibrations of atoms and therefore to some extent a reflection of the precision with which these atomic positions can be measured. It has been demonstrated that the estimate of the precision of atomic positions depends on the magnitudes of B values (Daopin et al., 1994). The program PROCHECK (Laskowski et al., 1993) and SFCHECK (Vaguine et al., 1999) have been

used to examine the correct stereochemistry of the final refined models. PROCHECK calculates the phi-psi angles (Ramachandran plot). It estimates the percentages of residues within or outside the allowed or partially allowed regions in Ramachandran plot (Ramachandran & Sasisekharan, 1968). Unusual values of angles and eclipsed dihedral angles in side chains need attention. The exceptionally high B-values as well as unpaired charged residues in the interior of the molecule and abnormal van der Waals contacts are those that need checking. The program outputs a detailed list of deviations from the ideal geometry of residues along with the corresponding plots of the parameters.

In refinement cycle the function to be minimized follows a downhill path towards its minimum value and if the starting model is not too different from the real structure the refinement easily converges to the correct solution. However, if the difference between the model and the actual structure is rather large, the refinement may attain a local minimum instead of reaching the global minimum. To avoid this situation a refinement technique demands uphill as well as downhill search direction to overcome barriers in the Q-function. The method of simulated annealing is a superior refinement technique because it raises the temperature sufficiently high for the atoms to overcome local energy barriers and then cool slowly to approach the global energy minimum.

2.8. Biochemical and biophysical techniques for protein characterization

2.8.1. SDS - polyacrylamide gel electrophoresis (SDS-PAGE)

The purity of protein preparation was checked on a 12% sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE), according to Laemmli (1970). The gel was prepared using the Tarson SDS-PAGE apparatus with 1 mm spacers and the samples electrophoresed alongside molecular weight markers applying voltage 200 V for 1 h. Protein bands were visualized by staining the gel with 0.25% coomassie blue R-250 in 10% (v/v) glacial acetic acid and 25% (v/v) isopropanol. The destaining solution contained 7% (v/v) glacial acetic acid and 5% (v/v) isopropanol. Native PAGE was carried out on 9% (w/v) polyacrylamide slab gel, at pH 4.3, using methylene blue as marker.

2.8.2. Isoelectric focusing (IEF)

For IEF, 300 µL of the fractions (10 µg protein) were precipitated with 10% trichloroacetic acid, washed with ice-cold acetone and resuspended in 125 µL rehydration

80

solution (8 M urea, 2 M thio-urea, 2% Chaps, 2 mM dithiothreitol, 2 mM EDTA). Immobilized pH gradient (IPG) buffer (0.5%, pH 6–11, Amersham) was added, mixed and the sample was run on Immobiline DryStrip pH 3–10 (7 cm; Amersham) Focusing was performed for 6 h applying voltage of 500 to 8000 V. The strip was subsequently equilibrated in a solution containing 1% dithiothreitol, 50 mM Tris/HCl pH 8.8, 6 M Urea, 30% glycerol, 2% SDS, and stained with Coomassie brilliant blue.

2.8.3. Relative molecular mass (Mr) determination

Mr of the protein was determined using size exclusion chromatography performed at 4 ºC using a sephacryl 200 prepacked column (1.5X75 cm) from Pharmacia. The mobile phase consisted of 10 mM sodium phosphate buffer, pH 6.5 containing 1 mM DTT, freshly prepared and degassed before use. The flow rate was 0.30 ml/min. the apparent molecular mass of the protein was estimated from a calibration curve. For calibration wide range gel filtration molecular weight markers (Sigma) consisted of cyctochrome c (12.5 kDa), ovalbumin (43 kDa), BSA (68 kDa), apoferritin (440 kDa) were used.. Blue dextran was used to determine the void volume ($V_0 = 26.5$ ml) and the exclusion volume (V_T=117.6 ml) of the column.

2.8.4. Estimation of K_m and inhibition constant

The values of K_m and k_{cat} for different substrates (Glycocholic acid (GCA); glycodeoxycholic acid (GDCA), Taurucholic acid (TCA), Glycochenodeoxycholic Acid (GCDCA), Taurochenodeoxycholic acid (TCDCA)) were determined by incubating the enzyme sample with a range of concentrations of each substrate under standard assay conditions. Non-substrate ligands were compared by determining the K_i measured by incubating 1.12 mg/ml of *Bl*BSH with the respective inhibitors in the concentration range 0.5-5 mM and by increasing the concentration of glycocholic acid (GCA; 0.1 - 10 mM) under standard assay conditions. The constants were calculated by fitting the linear regression curve to the data on Lineweaver-Burk plots using Enzyme Kinetics!Pro (http://www.chemsw.com/16029.htm). The standard errors in the estimation of the values of K_m and k_{cat} were within limits of 10%.

2.8.5. Chemical modification

2.8.5.1. Arginine modification with phenylglyoxal and 2, 3-butanedione

The protein was incubated with different concentrations of phenylglyoxal (2-20 mM) or 2, 3-butanedione (10-50 mM) in 50 mM phosphate buffer at 25 ºC and pH 7.5. Enzyme incubated in the absence of modifying reagent served as control. Aliquots were withdrawn at different time intervals and the excess reagent was removed by passing through sephadex PD-10 (Amersham Pharmacia Biotech, Sweden) columns.

Phenylglyoxal can react with guanido groups of arginine residues with high specificity under mild conditions. The only comparably rapid reaction is with -amino groups. This reagent is superior to glyoxal, methylglyoxal, and 2,3-butanedione in terms of its speed and specificity of reaction with arginine residues. The reaction product with arginine contains 2 phenylglyoxal residues per guanido group and is quite stable under mildly acidic conditions (Takahashi, 1968; Yamasaki *et al.* 1980).

2.8.5.2. Estimation of Arginines

The *Bl*BSH in 50 mM phosphate buffer at pH 6.5 or PGA in 50 mM phosphate buffer at pH 7.5 was treated with 10 mM p-nitrophenyl glyoxal (pNPG) at 25 ºC. Aliquots were withdrawn at different time intervals and the residual activity determined as described before.

For determination of total number of arginine residues in PGA was treated with trypsin and subtilisin in 100 mM sodium pyrophosphate at pH 9.0 and 37 ºC for 3 h. The cooled mixture was treated with pNPG at 25 ºC for 30 min and the absorbance was determined at 475 nm. The arginine content was determined by comparison with a calibration curve prepared using a standard arginine solution. Undigested enzyme was used for pNPG treatment with protection by 50 mM phenylacetate solution and without it to determine the number of active site residues and surface/accessible arginine residues.

2.8.5.3. Estimation of the apparent inactivation rate constant

Appropriate enzyme amount was treated with different concentrations of (2-20 mM) phenylglyoxal or (10-50 mM) 2,3-butadione in 50 mM phosphate buffer at pH 7.5 and 25 °C. Aliquots were withdrawn at various time intervals (15-60 min) and the residual activity was determined as described previously.

The apparent inactivation rate constants at different reagent concentrations were calculated from the equation ln $(E_i/E_0) = -k_{app}t$, where E_0 is the activity of unmodified enzyme, E_i is the activity of chemically modified enzyme, t is time in minutes, k_{app}, the apparent inactivation rate constant (min^{-1}).

2.8.5.4. Modification of tryptophan with N-bromo succinimide (NBS)

To measure the effect of Trp modification on enzyme activity, enzyme in 10 mM sodium acetate buffer, pH 5.5 was incubated with different concentrations of NBS (10-100 µM). Ten µL aliquots were withdrawn every 2 min intervals till 15 min and the residual activity measured under standard assay conditions.

The number of tryptophan residues that reacted with NBS was calculated by measuring the decrease in absorbance at 280 nm as described by Spade & Witkop (1967). Enzyme in 10 mM sodium acetate buffer pH 5.5 was treated with increasing concentrations of NBS (5-200 µM). The reagent was added in aliquots of 10 µL in 10 instalments every 2 min intervals and absorbance recorded. After each addition, an aliquot of 10 µL was removed and the reaction arrested by the addition of 90 µL of 50 mM L-tryptophan. The residual activity was determined under standard assay conditions. NBS mediated inactivation was monitored by measuring the decrease in absorbance at 280 nm. The same procedure was followed in the presence of enzyme inhibitor phenylacetate. The number of tryptophan residues modified was determined using an estimated value of 5500 M^{-1}cm^{-1} for extinction coefficient. Enzyme samples incubated in the absence of NBS served as control.

N-bromo succinimide (NBS) is a potent oxidizing agent. Being a highly reactive source of electrophilic bromonium ions (Br$^+$), it is capable of adding bromine to the γ-δ carbon-carbon double bond of Trp which subsequently gets cleaved.

2.8.5.5. Modification of tryptophan with 2-hydroxy 5-nitrobenzylbromide (HNBB)

Appropriate amount of enzyme in 10 mM sodium acetate buffer, pH 5.5 was incubated with 10-40 mM HNBB. The solution of HNBB was freshly prepared in dry acetone. Aliquots of 10 µl were withdrawn every 2 min after a lapse of 10 min till the end of 20 min and the residual activity measured under standard assay conditions. The number of tryptophan residues modified was determined based on estimated OD$_{410}$ nm and assuming a molar absorption coefficient of 18,000 M^{-1}cm^{-1}.

83

2.8.5.6. Modification of Cysteine residues

*Bl*BSH in 100 mM potassium phosphate buffer, pH 8.0, was incubated with 1 mM 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) at 30 ℃ for 60 min. Aliquots were removed at defined time intervals and the residual activity was determined. The protective effect of substrate and inhibitor during cysteine modification was determined by incubating the enzyme with an excess of substrate or inhibitors prior to treatment with modifying reagent under similar experimental conditions. Excess reagents were removed by passing through a PD10 column. The residual activity was quantified under the standard assay conditions. The number of cysteine residues modified was followed by monitoring the increase in absorbance at 412 nm using an extinction co-efficient of 13,600 M^{-1} cm^{-1}.

2.8.5.7. Modification of tyrosine residue

The enzyme in 0.05 M sodium borate buffer, pH 7.5, was incubated with 1 mM N-acetylimidazole (NaI) for 30 min, at room temperature (25 ℃), followed by the estimation of the residual activity under standard assay conditions. The enzyme incubated in the absence of NaI was taken as control. The number of tyrosine residues modified was calculated by using a molar absorbance coefficient of 1160 M^{-1} cm^{-1} at 278 nm.

2.8.6. Substrate protection studies

In all chemical modification reactions, the effect of protective ligands was determined by incubating the PGA with penG (substrate) or phenylacetate (competitive inhibitor) and *Bl*BSH with GCA (substrate) or CA (competitive inhibitor) to a final concentration of 50 mM, prior to treatment with modifying reagents. Thereafter, the mixture was passed through sephadex PD 10 column and the residual activity determined under standard assay conditions.

2.8.7. Circular Dichroism measurement

The enzyme solution filtered through a 0.20 µm membrane and dialyzed against 10 mM potassium phosphate buffer at pH 7.5 was used for CD spectroscopy. The CD spectra were recorded on a J-715 spectro-polarimeter with a PTC343 Peltier unit (Jasco, Tokyo, Japan) at 25 ℃ in quartz cuvette of 0.1 cm path-length. All spectra were corrected for buffer contributions and converted to mean residue weight ellipticity. Each CD spectrum was accumulated from five scans at 50nm/min with a 1nm slit width and a time

constant of 0.5 s for nominal resolution of 1.7 nm. Data were collected in the range of wavelength 190 to 320nm.

2.8.8. Fluorometric studies

Fluorescence was measured using a PerkinElmer Life Sciences LS50 fluorescence spectrophotometer connected to a Julabo F20 water bath. To eliminate background emission, the signal produced by either buffer solution, or buffer containing the appropriate quantity of substrate or inhibitor was subtracted..

Appropriate amount of enzyme in 10 mM sodium acetate buffer, pH 5.5 was excited at 280 nm and the emission was recorded in the range 300-400 nm. Fluorescence quenching due to modification with NBS and HNBB was measured by adding 2 μl aliquots of the modifier from stocks, till the relative fluorescence intensity lowered substantially. Suitable controls were included to correct for changes in enzyme dilution. Fluorescence quenching due to substrate and competitive inhibitor was measured by progressive addition of 10 μl aliquots of either penG or phenylacetate from a 500 mM stock. Substrate protection against chemical modification was determined by incubating the enzyme sample with 150 μM penG prior to the addition of 200 μM NBS, the fluorescence quenching was then measured.

2.8.9. Fluorescence Measurement for ligand binding

The sample was excited at 295 nm and the emission recorded in the wavelength range 300-400 nm at 25ºC. The slit-width on both the excitation and emission were set at 5 nm, and the spectra were obtained at 100 nm/min. The binding constant (K_d) of the various substrates and inhibitors was determined by monitoring the effect of fluorescence emission upon titration with a given ligand. The binding of ligands to protein can be described as

$$K_d = \frac{\text{[Protein] [inhibitor]}}{\text{[Protein inhibitor]}} \qquad \text{(eq. 1)}$$

where K_d is the apparent dissociation constant, [Protein] is the concentration of the protein, [Protein inhibitor] is the concentration of complexed protein, and [inhibitor] is the concentration of unbound inhibitor. The proportion of inhibitor-bound protein as described by equation 1 is related to measured fluorescence emission intensity as

(Fo-F)/(F-F$_\infty$) = [Protein inhibitor] / [Protein]$_T$ (eq.2)

where Fo is fluorescence intensity of enzyme alone, F is observed fluorescence intensity at a given concentration of inhibitor, F$_\infty$ is intensity of Protein saturated with inhibitor, and [Protein]$_T$ is the total protein concentration. If the total inhibitor concentration, [inhibitor]$_T$, is in large molar excess relative to [Protein]$_T$, then it can be assumed that [inhibitor] is approximately equal to [inhibitor]$_T$. Equations 1 and 2 can then be combined as,

(Fo-F)(/F-F$_\infty$) = [inhibitor]$_T$ / (K_d + [inhibitor]$_T$)   (eq. 3)

The values of K_d were determined from a nonlinear least square regression analysis of titration-data by using equation 3. The stoichiometry of binding was established using a linear version of the Hill equation,

log {(Fo-F/(F$_\infty$-Fo)} = n log[inhibitor]- log K' (eq.4)

where n is the order of the binding reaction with respect to inhibitor concentration and K' is the concentration of the ion that yields 50% of F-F$_\infty$. The thermodynamic parameters G, H, and S were determined according to equation 5,

   $-RT$ ln $K =$ $G =$ $H - T$ S (eq.5)

where R is the gas constant, and T is in absolute temperature.

2.8.10. Tryptophan accessibility

To study the nature of tryptophan microenvironment 8 M acrylamide, 5 M KI and 5 M CsCl, each separately has been added to the enzyme sample and tested. 10 mM sodium thiosulfate was included in the case of KI to prevent the formation of ions of I$_3^-$. Corrections made for contribution from buffer and quenchers to fluorescence and to compensate for the dilution by the addition of quenchers. At the highest concentrations of

the quenchers, the volume changes accounted for only less than 10 % of the initial enzyme solution. The Stern-Volmer equation was used to analyze the quenching of tryptophan fluorescence

$$F_0/F \ = 1 + K_{sv}[Q] \ \ldots\ldots\ldots\ldots\ldots.(4)$$

where F_0 and F are the fluorescence intensities in the absence and in the presence of the quencher, [Q] is the concentration of the quencher, and K_{sv} is the effective quenching or the Stern-Volmer constant (Eftink & Ghiron, 1976). The fraction of the fluorophore accessible to the quencher was determined using modified Stern-Volmer plot (Lehrer, 1971)

$$F_0/(F_0 - F)) \ = \ [1/ (f_a \ K_{sv}) \ (1/[Q] \ ) + 1/f_a] \ \ldots\ldots. \ (5)$$

where f_a is the effective fractional accessible fluorescence. The parameters F, F_0, and [Q] are same as defined in equation 4.

2.8.11. Fluoremetric measurement for thermal denaturation studies

The protein solution of 0.024 mg/ml was excited at 280 nm and the emission recorded in the range of wavelengths 300-400 nm at 30 ºC. The slit-width on both the excitation and emission were set at 5 nm, and the spectra were obtained at 100 nm/min.

2.8.12. Thermal denaturation

The effect of temperature on the enzyme was studied using a thermostatic cuvette holder connected to an external constant temperature circulation water bath. The protein sample was incubated for 15 min at specified temperature before taking the scan. The samples were removed from specified condition for measuring the activity under standard assay conditions. For renaturation experiments the samples were cooled back to 30 ºC and left for one hour before recording the spectra recording. Fluorescence spectra were recorded as described above.

2.8.13. Effect of pH

Enzyme samples were incubated in appropriate buffer for 4 h at 30 ºC over a pH range of 1-12. The following buffers (100 mM) were used for these studies: Glycine-HCl

buffer for pH 1-3, glycine-NaOH for pH 10-12, acetate buffer for pH 4-5, phosphate buffer for pH 6-7 and Tris-HCl buffer for pH 8-9. For refolding experiments the pH of the samples were adjusted back to pH 7 and incubated at 39 ºC for 1 h before recording the spectral measurement. Fluorescence spectra were recorded as described above.

2.8.14. Guanidium hydrochloide (Gdn-HCl) mediated unfolding

Protein samples were incubated in 0.25-7 M denaturant solution at pH 7 for 4 h to attain equilibrium. After recording the scans, aliquots were removed from the sample to check for activity. Refolding experiments were conducted on samples diluted 10 times to dilute the Gdn-HCl and incubated at 30 ºC for 1 h before recording the spectra.

2.8.15. ANS-binding assay

The intermediate state of denatured and native BSH at different denaturation conditions was analysed by hydrophobic dye (ANS) binding. A final ANS concentration used was 40 µM, excitation wavelength, 375 nm and total fluorescence emission was monitored between 400-550 nm. Reference spectrum with ANS at each buffer of respective pH and Gdn-HCl was subtracted from the spectrum of the sample.

2.8.16. Light scattering

In order to estimate the protein aggregation upon thermal, pH and Gdn-HCl denaturation conditions, the fluorimeter was employed to measure scattering at 400 nm. The same experimental setup used for fluorescence studies was used for this purpose. Both excitation and emission wavelengths were set at 400 nm. The excitation and emission slit width were set at 5 nm and 2.5 nm, respectively. Scattering was observed for 120 sec.

2.8.17. Isothermal titration calorimetry (ITC)

Measurements using isothermal titration calorimetry (ITC) were carried out using a VP-ITC microcalorimeter (Microcal, Inc., Northhampton, MA) following standard instrumental procedures. The enzyme was dialyzed overnight at 4 ºC in 10 mM sodium phosphate buffer pH 6.5. Phosphate buffer was chosen by virtue of its small ionization enthalpy change; hence, the binding enthalpies reported for the protein-ligand interaction do not reflect a contribution due to buffer protonation. The enzyme solution was in the calorimetric cell and titrated with ligand dissolved in the dialysate buffer at a concentration of 100 mM. The actual protein concentrations for the replicate titrations

were determined from absorbance measurements at 280 nm using 37 220 cm^{-1} M^{-1} as the molar extinction coefficient for the dimer, which was estimated from the protein primary sequence by program Sednterp (V1.02) for Windows 95 and Windows NT (Created by David Hayes, Tom Laue, and John Philo). This free-ware program was obtained from the Reversible Associations in Structural and Molecular Biology (RASMB) software archive at http://www.bbri.org/rasmb/rasmb.html. Ligand samples were prepared from powder stocks by adding an appropriate aliquot of the material into the dialysis buffer and shaking for 2 h. The undissolved ligand was pelleted by centrifugation, and the supernatant was collected for the titrations. All experiments were performed at 25 ºC, with a 250 μL injection syringe. A typical binding experiment involved 25-30 injections varying between 5 and 15 μL aliquots of ligand solution into the titration cell with stirring at 400 rpm. Heats of dilution were determined by titrating ligand against buffer and buffer against protein. However, the heats of protein dilution were considered insignificant, so only the ligand dilutions were used to correct the total heats of binding prior to data analysis.

2.8.18. Processing of PGA precursor

The purified precursor was incubated at varying pHs, (5-8), temperature (20-35 ºC) and ionic strength (1-100 mM). Samples were withdrawn at periodical intervals for the SDS-PAGE analysis and activity was determined under the standard assay conditions.

2.8.19. Dynamic light scattering (DLS)

A dynamic light scattering experiment was performed on a DynaPro instrument using protein solutions (Lakewood, USA). The measurements were made at 18 ºC on the purified protein at 0.1-0.5 mg/ml in buffer solution containing 10 mM potassium phosphate pH 7.5. Protein samples were either centrifuged at 10,000xg or sterile filtered to remove dust particles before analysis. The samples were then illuminated at 780 nm using a solidstate laser. Each sample was evaluated at least 20 times, with the quoted values being the mean of these independent evaluations. DLS experiments provide a direct determination of the diffusion coefficient (D) of the particles in solution (www.proteinsolutions.com). Using the diffusion coefficient, the hydrodynamic radius (RH) of the particles can be calculated from the Stokes-Einstein equation

$$D = kT / 6\eta RH$$

where k is the Boltzmann constant, T is the absolute temperature, and η is the solvent viscosity. Approximate molecular weights for each of the species were determined from RH calculated using the software package DYNAMICS supplied with the instrument.

2. 9. Homology modeling of PGA and docking studies

The sequence alignments derived from multiple sequence alignment were modeled based on a homologous PDB structure (code, 1GK9) to generate 10 models in the MOE (Chemical Computing Group, Inc http://www.chemcomp.com/) program. These models were subjected to a coarse energy minimization procedure in order to remove any possible van der Waals' short-contacts between atoms. The best model predicted by MOE had a score of ±2.9. The predicted models were further evaluated for correct geometry, stereochemistry, and minimum energy distributions. The programs SFCHECK and PROCHECK (Vaguine et al., 1999) were used to establish reliability of the comparative structures and to assess the predicted models of PGA. In addition, to compare the variability of models, C atom traces and backbone atoms were superposed onto the template crystal structure and RMSD for the positional differences between equivalent atoms calculated. The protein structures were visualized and analyzed by using PyMOL- 0.97 (http://www.pymol.org) and QUANTA (Accelrys).

2.9.1. Preparation of protein for docking

The coordinates of the refined structure (2.5 Å resolution) of *Bl*BSH after the addition of hydrogen atoms were used. After adding hydrogen atoms at appropriate positions using standard geometries, the structure was minimized, keeping the heavy atoms fixed at their crystallographically determined positions. The resultant structure was energy minimized using 1000 steps of the steepest descents minimization followed by conjugate gradient minimization until convergence of 0.01 kcal/mol/Å was achieved with a distance dependent dielectric constant of 80, to mimic implicit water. This resulting molecule served as starting structure for the docking experiments.

2.9.2. Binding Site Identification

The Alpha-site finder module of MOE is used for identifying and characterizing protein binding sites and functional residues in proteins. This program searches the BSH structure for binding sites in a selected chain by locating cavities on the surface, which in turn can be used to guide the protein-ligand docking experiments. By comparing the

conserved residues in protein sequences and by combining this information with search results, we can predict the binding site of protein. After choosing an appropriate site dummy atoms are included which serve as anchor for the ligand during its docking process.

2.9.3. Ligand modeling

Ligands were created using MOE Molecule builder panel and the geometries of these molecules optimized. All the ligands designed were energy minimized using above mentioned parameters.

2.9.4. Docking of Ligands

Pharmacophore Dock (Ph4Dock, 21) of MOE (ver. 2004.02, Chemical Computing Group: Montreal, Canada) was performed for ligands such as glycine-, Taurine- conjugated bile salts and inhibitors like penV, penG, PAA, POAA, 6-APA. The structures of the ligands binding with highest affinity to the receptor, based on the criteria of total energy of the ligand/receptor complex, were automatically detected in this procedure, which employs a combination of Monte Carlo type and stimulated annealing procedures to dock a ligand molecule on a protein. Exclusive volumes were automatically generated with Ph4Dock with reference to the atomic positions of protein. The potential charge of the complex was assigned considering the force field. Interaction energy ($U_{\text{ele}}\_U_{\text{vdw}}$; *i.e.*, summation of intermolecular electrostatic and van der Waals energy) was calculated after energy minimization, which was carried out by the MMFF94s force field (Halgren, 1999a & b) for substrate molecules and the side chains of the enzyme within 4.5 Å distance in the case of each substrate. Default values were used for all the other parameters. Ten possible conformations were generated for the substrate inside the protein, and ranked according to the sum of the ligand's internal energy, van der Waals and electrostatic energy terms of the potential energy function, while keeping the side chains flexible.

The first binding mode has the lowest van der Waals term and is compatible with an orientation allowing deconjugation to take place. Two best-ranked binding modes for ligands were selected and included in the protein and the protein+substrate complexes and energy minimized using the MMFF94s force field in MOE assuring that the final energy gradient was 0.01 kcal/mol/Å. In the 'protein + ligand' minimizations, all protein

side chains were left fully flexible. Interaction energies between the energy minimized protein and the ligand were calculated as the difference between the total potential energy of the minimized complex, and the sum of the individual energies of protein and ligand components of the complex. The van der Waals and electrostatic contributions to this interaction energy were recorded and reported.

2.10. Sequence alignment and phylogenetic tree construction

Protein sequence is downloaded from RCSB protein data bank (www.rcsb.org) for PGA from *E.coli* (1AI4), PVA sequence of *Bacillus sphaericus* (3PVA) and (-N-acetylglucosaminyl)-L-asparaginase from *Flavobacterium meningosepticum* (2GAW) were also downloaded. Initially protein sequences similar to these sequences were identified from protein-protein BLAST using individually the -chain of *Ec*PGA, L-asperginase and monomer of *Bsp*PVA. Each set of sequences obtained were aligned using Clustal X & W. A Phylogenetic tree is also constructed using NJplot and Unrooted in ClustalX package. More homologous sequence across organisms were identified by selecting closer sequences and using them in BLAST search.

CHAPTER 3

CRYSTAL STRUCTURE ANALYSIS OF *Bl*BSH

3.1. Introduction

This chapter describes the results of crystallization experiments, especially details about the conditions exploited to improve the diffraction quality of *Bl*BSH crystals. Three-dimensional structures of proteins provide information about the over all architecture and also the orientation of catalytic and binding site residues. Three-dimensional structure of bile salt hydrolase was determined using X-ray crystallographic technique. The details of cryoprotectant screening and strategies tried for obtaining crystals of enzyme-substrate complex are discussed. In this process, several sets of X-ray data of wild type and mutant *Bl*BSH crystals in various space groups were collected. The structure of wild type *Bl*BSH, determined using molecular replacement method using PVA co-ordinates (PDB: 3pva), was refined at resolutions 2.30 Å (space group $P6_122$) and 2.5 Å (space group $P3_22_1$). These structures were compared with the reported structures of PVA from *Bacillus sphaericus* (*Bsp*PVA) and BSH from *Clostridium perfringens* (*Cp*BSH).

3.2 Results

3.2.1. Crystallization of bile salt hydrolase from *B. longum*

The technical details of cloning and overproduction of *Bl*BSH is already described in chapter 2. Cloning and overexpression assured a constant and reliable supply of protein for crystallization experiments. The purity of the final preparation of protein was confirmed by SDS-PAGE. Analysis by dynamic light scattering confirmed the presence of tetramers in solution.

Different methods for screening crystallization conditions were tried. The standard screening kits contained pre-prepared solutions, while the solutions for the grid screens were prepared from mixing of stock solutions of precipitant, salts and buffers. In grid-screens the precipitant concentration and pH were varied. Sparse matrix screens (Crystal Screen I and II of Hampton Research, CSS-I and II of Molecular Dimension) showed that the protein crystallized in a variety of conditions but the crystals were of
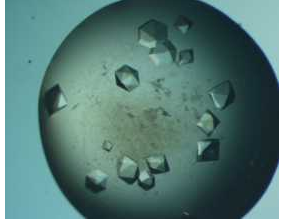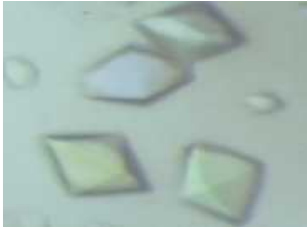
small size and possessed different morphologies (Table 3.1). These tiny crystals were not suitable for X-ray measurement owing to their small size and growth defects. Several other crystallization conditions were also tested for growing better crystals. Morphologically the best crystals obtained were from PEG 400 (30% w/v), 100 mM HEPES, pH 7.2 and 0.2 M Mgcl$_2$. Equally good crystals were also grown from 30% saturated ammonium sulphate, 0.5 M sodium formate, 100 μL of sucrose (1M) and 40 μL of β-octyl-glucosidase (BOG) (0.50% w/v) at pH 6.7. The crystals generally appeared after four days.

3.2.2. Improving crystal quality

After optimizing the initial conditions various additives were added to crystallization solutions to improve the crystal size and resolution. Additives including sugars, alcohols, cations, polyamines, detergents or surfactants and dioxane are reported to improve crystal quality (Sauter et al., 1999). Detergents help to alter polycrystalline state, promote growth of single crystals instead of clusters, improve diffraction quality and achieve better reproducibility of results (Guan et al., 2001). BOG has been successfully used in the crystallization of a number of proteins, RNA and protein-nucleic acid complexes (McPherson et al., 1986a; McPherson et al., 1986b). Compounds such as polyols like glycerol are known to increase structure stability (Sauter et al., 1999; Sousa, 1995). Reducing agents such as DTT, cysteine, -mercaptoethanol, glutathione etc. maintain the reduced state of reactive sulfyhydryl groups important for enzyme activity and structural integrity (McPherson, 1998). The concentration of BOG was optimized and crystals of max dimensions 0.4 x 0.2 x 0.2 mm of both crystals forms could be grown (Figure 3.1). Growing at temperature 303 K instead of 293 K also improved the quality of the crystals. For collection of X-ray diffraction data at room temperature, crystals were wet mounted into quartz glass capillaries of 1.0 mm diameter. Typically, one single crystal was drawn along with mother liquor into the capillary. The crystal was taken out of the mother liquor and made to stick to the walls of the capillary by the surface tension of mother liquor surrounding the crystal, or otherwise the mother liquor on the crystal was dried out using thin strips of filter paper. A small amount of mother liquor was left in the capillary. The ends of the capillary were sealed with bee wax. The capillary was then mounted on a goniometer head with a small amount of plasticine.

Table 3.1: Initial screening conditions in which crystals appeared

S. No	Condition	No. days for crystal growth	Crystal figure	Diffraction quality[#]
1	25% 2KMME, 0.2 M KSCN, 100mM Na-Cacodylate pH 6.5.	5		No diffraction
2	25% PEG 4K, 100mM Na acetate pH 4.6, 0.2 M AS	4		4-5
3	30% PEG 400, 100 mM HEPES pH 7.5, 200 mM MgCl2	2		6-8
4	5% Isopropanol, 2M ammonium Suplhate	7		No diffraction

5	30% saturated AS, 100 μL sucrose (1M) and 40 μL of β-octyl glucopyranoside (50%) at pH 6.7	4		3.5-4
6	500 mM Na formate, 10% PEG 8K, 10% PEG 1K	3-4		No diffraction
7	8% PEG 20K, 8% PEG 550 MME, 200 mM Ca Acetate	7-8		4-5

For initial diffraction experiments either glycerol 25% or PEG 400 25% was used as cryoprotectant for crystals from all conditions.

To achieve higher resolution and to obtain whole data from one crystal the data collection at low temperature was tried. In this technique the main challenge is the cooling of the crystal to cryogenic temperatures and at the same time avoiding the formation of crystalline ice within the sample. So, a suitable cryoprotectant is added to the crystal that can slow down the nucleation of ice crystals so that water forms glass instead of ice crystals in the lattice. The simplest approach to introduce a cryoprotectant is to include the cryoprotectant in the mother liquor by trial and error. The optimal concentrations of cryoprotectant was chosen by an initial examination of effective concentration by adding small amounts of cryo-protectant as mentioned in Table 2.1 to the mother liquor and testing.

A
B
C
D
E
F
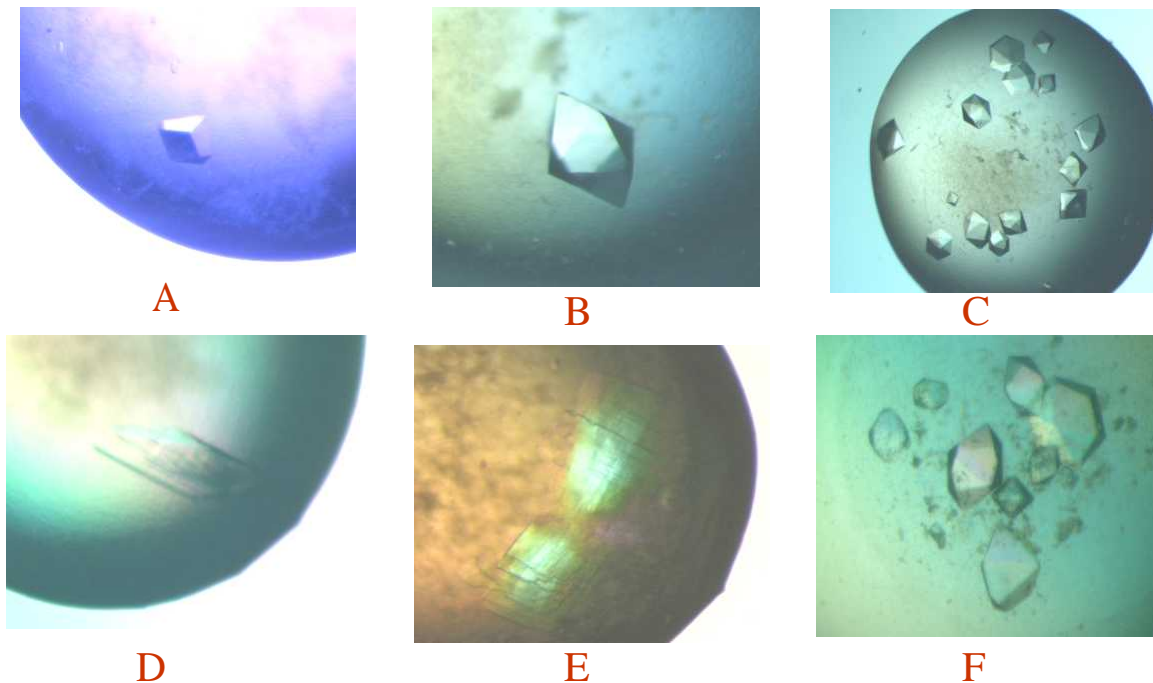
Figure 3.1: Crystals of native and mutant *Bl*BSH grown from optimized crystallization conditions. A) Trigonal crystals of *Bl*BSH in space group P3$_2$2$_1$. B) Hexagonal crystals of *Bl*BSH in space group P6$_1$22. C) *Bl*BSH-T2A mutant crystallized in the presence of penV. D) *Bl*BSH-C1A mutant crystallized in the presence of GCA. E) *Bl*BSH-C1A mutant crystallized in the presence of TDCA. F) *Bl*BSH-T2A mutant crystallized in the presence of GCA

3.2.3. Improving *Bl*BSH mutant crystals

Mutant proteins (C1A and T2A) were crystallized with bile salts present in the mother liquor from the same condition as native protein but with slight difference in the concentration of additive. The crystals obtained from these conditions were thin plates and were sticking together as clusters. Attempts to grow separated good quality crystals in these conditions failed. So, seeding technique was tried to grow better crystals. In microseeding technique submicroscopic seeds are introduced in the crystallization solution. The serial dilution method was followed for microseeding method, where a few small crystals were selected and washed, and then the crystals were crushed so that a seed stock solution was prepared. Normally, the seed stock solution was serially diluted to 10^{-7} to avoid too many nuclei inside the solution. Each diluted seed solution was used to examine the best seed concentration for growing large single crystals by the addition of a small aliquot of seed solution (Stura & Wilson, 1990 & 1991). Single crystals of mutants were obtained successfully using seeding technique. However, the diffraction data collected from these crystals turned out to be of poor quality due to very high mosaicity displayed by the diffraction spots (Figure 3.2).
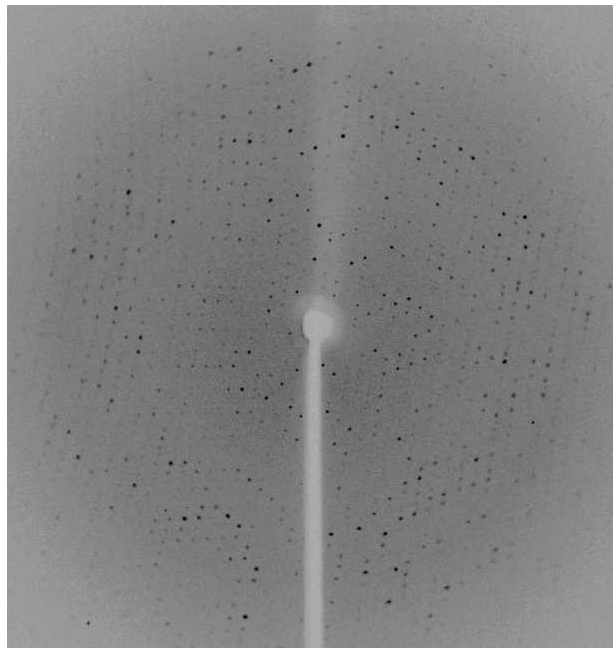
Figure 3.2: Diffraction image obtained from C1A crystals of *Bl*BSH with high mosaicity. Space group is C2 and oscillation angle 0.5° diffracted upto 3

3.2.4. Diffraction data collected on *Bl*BSH crystals

Many data sets were collected from different crystal forms obtained (Table 3.2). The best crystals diffracted better than 2.5 Å resolution and belonged to either space group $P6_122$ with unit-cell parameters $a=b= 123.98$, $c= 219.56$ or to space group $P3_22_1$, with unit cell parameters $a=b=125.24$ $c=117.02$ Å. Detailed data collection statistics are given in Table 3.3.

The crystal structure of wild type protein has been determined at resolutions 2.5 for trigonal and 2.3 for hexagonal crystals (Table 3.3). The resolution of data from crystals of mutants C1A and T2A were 3.2 and 3.0 , respectively. One of the mutants, C1A, crystallized in monoclinic space group C2 with unit cell parameters $a= 186.36$, $b=71.10$, $c=133.65$ and $=109.10º$ while the other, T2A, crystallized in $P3_22_1$ with same wild type unit cell parameters $a= b= 125.24$, $c= 117.03$.

Table 3.2: List of crystal types of *Bl*BSH and mutants characterized.

S. No	Crystal form & space group	Unit cell (, °) parameters	Mathews number	Solvent content (%)	Number of monomers
1	Trigonal (Wild) $P3_22_1$	$a=b=125.24,$ $c=17.03$	3.6	65.9	2
2	Hexagonal (Wild) $P6_122$	$a=b=123.98,$ $c=219.56$	3.52	65.1	2
3	Hexagonal (T2A) $P6_122$	$a=b=126.04,$ $c=116.27$	3.6	65.9	2
4	Monoclinic (C1A) P2	$a=86.36$ $b=71.101$ $c=133.65$ $=109.10$	2.7	61.6	4
5	Monoclinic (C1A) P2	$a=112.03$ $b=88.305$ $c=165.24$ $=109.10$	2.72	54.77	4
6	Hexagonal (wild) $P6_122$	$a=b=163.75$ $c=116.43$	3.66	66.4	2
7	Monoclinic (C1A) C2	$a=165.67$ $b=113.49$ $c=269.57$ $=89.48$	2.23	44.86	8

From among above crystal forms we could collect complete data from trigonal, hexagonal and monoclinic (C2) crystals only. The trigonal and hexagonal data were used for structure solution and refinement. Although we could solve the structure using data from monoclinic crystals (C2) no further refinement could be realized, presumably due to poor quality of data.

Table 3.3: Data collection statistics of *Bl*BSH cystals

	P3$_2$21	C2	P6$_1$22
Detector and wavelength (Å)	Raxis-IV, 1.541	Raxis-IV, 1.541	ESRF
Unit-cell parameters			
(Å)	*a=b* 125.24	a= 186.36 b= 71.10	a=b=124.94
	c=117.02	c= 133.65 ß = 109.10°	c=218.95
Matthews coefficient[#] (Å^3 Da^{-1})	3.6	2.7	3.52
Molecules per AU	2	4	2
Solvent content (%)	65.9	61.6	65.09
Resolution range (Å)	20-2.5 (2.28-2.2)	20-3.2 (3.20-3.11)	20-2.25 (2.25-2.29)
Total observations	154427	578088	1648459
Unique reflections	37170	78599	36287
Average $I/\sigma(I)$	10.6 (2.1)	7.2 (5.2)	_
R_{merge} (%)	6.0 (39.0)	13.6 (24.6)	3.1
Completeness (%)	99.5 (100)	98.5 (89.9)	99.7 (97.9)

3.3. Structure solution

Reflections were phased using molecular replacement (MR) method. The processed form of *Bsp*PVA monomer (3pva) was used as search model. The data in the resolution range 20.0-3.0 Å for trigonal form were used for MR calculations using

AMoRe program (Navaza, 1994) of CCP4 suite (Collaborative Computational Project, 1994). The solutions with the highest correlation coefficients (Cc) obtained from rotation function calculation (Table. 3.4a) were used in the translation function calculation (Table. 3.4b) to place the molecule in correct positions in their respective unit cells. The application of rigid body fit in AMoRe program improved the solutions in trigonal form (Table. 3.4c). The initial electron density map allowed assignment of all 316 residues in each of the subunits A and B of the asymmetric unit. The data in space group $P3_22_1$ gave better Cc and R-factor.

The refined structure of *Bl*BSH monomer was used as search model using recently developed PHASER program for other crystal forms. The data used for MR calculations were in the resolution range 20.0-2.30 Å for hexagonal form and 20.0-3.5 Å for monoclinic form. In this case acceptable values of Z-score and R-factor were obtained only for data in the space group $P6_122$.

Table 3.4a : Molecular replacement solution obtained using AMoRe program. Model coordinates were from PDB file 3pva

Crystal form	Solution	( )	( )	( )	Cc(%)
Trigonal	Sol_1	109.20	25.25	10.80	22.8
	Sol_2	56.80	14.98	2.62	22.1
	Sol_3	23.42	41.28	15.3	18.9
	Sol_4	74.8	87.67	76.88	12.2

Table 3.4b

		alpha	beta	gamma	Xfrac	Yfrac	Zfrac	R-fac	Corr
Sol_TF_1	1	37.81	12.28	142.74	0.707	0.551	0.130	0.448	0.575
Sol_TF_2	_1	29.35	12.46	90.07	0.625	0.842	0.170	0.452	0.569

Table 3.4c

Sol_ RF		TF	theta	phi	chi	tx	ty	tz		Rfac	Corr
Sol___6___2	1		91.79	119.54	170.43	0.177	0.794	0.554	.	0.427	0.602
Sol___9___1	2		170.04	-126.88	61.00	0.381	0.263	0.812	.	0.428	0.600

3.4. Refinement of the structure

The structures in trigonal and hexagonal crystal forms (wild type) only could be refined. The initial phases obtained from molecular replacement using the search model were improved by subsequent rigid body refinement keeping the sequence of the input model. This was followed by several cycles of positional refinement using the data in the resolution shell of 20.0-2.5 (trigonal), 20.0 -2.3 (hexagonal) resulting in the improvement of both R-factor and R-free. Individual B-factors were refined in the final cycles. Electron density maps were calculated at this stage. Side chains were mutated according to the sequence of *Bl*BSH and their conformations were selected based on the observed density and their new positions were refined. The difference map (Fo-Fc) showed three distinct blobs of density close to S atom of the N-terminal residue Cys1. These were interpreted as oxygen atoms due to formation of sulfonic acid. Extra density in the vicinity of the N-terminal cysteine was left unidentified and may represent an oxidized DTT molecule. Also sulphate molecules could be fitted into some of the electron densities. In the last four refinement cycles, solvent molecules were placed at peaks of (Fo-Fc) density above 4 (3 in the final cycle) providing the sites could participate in hydrogen bonds with protein atoms. 229 water molecules were fitted into the asymmetric unit. The final refined model obtained from trigonal was used as search model for hexagonal form. The other refinement procedure were followed as same as mentioned above. 255 water molecules were fitted into the asymmetric unit.

3.5. Structure validation

During final cycles of refinement the stereochemistry and the geometry of the models were checked using PROCHECK. The overall G factor considered to be a measure of stereochemical quality of the model output by PROCHECK was -0.034 (troginal) –0.032 (hexagonal). This is within the limits expected for a correct structure that is refined at 2.5 and 2.3 resolution. The Ramachandran (,) plot (Ramachandran et al., 1968) showed that the residues were placed in the most favored region or partially allowed region of the map (Figure. 3.3. A,B). The refinement statistics and the values of refined parameters are presented in Table 3.5.

Table 3.5: Summary of data collection and refinement statistics.

Crystal system / Space group	Trigonal / $P3_22_1$	Hexagonal / $P6_122$
Resolution range (Å)	20.0 - 2.49 (2.55 – 2.50)	20.0 - 2.30 (2.35 - 2.30)
Completeness (%)	99.5 (100)	99.7 (97.21)
Number of unique reflections	37135	36287
R_{sym} (%) [a]	6.0 (39.0)	3.1
Refinement statistics		
R_{cryst} [b]	17.4	19.2
R_{free} [c]	21.1	22.8
Number of protein atoms	5154	5152
Number of water molecules	229	254
Overall B value ($Å^2$)	38.9	25.4
Root mean square deviation		
Bond length (Å)	0.024	0.029
Bond angle (°)	2.175	2.269
Dihedral angles	0.142	0.207
Ramachandran plot (% residues)		
Most favoured region	87.1	87.5
Additionally allowed region	11.8	10.7
Generously allowed region	1.1	1.8
Disallowed region	0	0

[a] $R_{sym} = \ |I – (I)|/\ I$

[b] $R_{cryst} = \ |Fo\text{-}Fc|/\ Fo$

[c] $R_{free} = \ |Fo\text{-}Fc|/\ Fo$ where the F values are test set amplitudes (5 %) not used in refinement

Data in brackets correspond to the outer shell
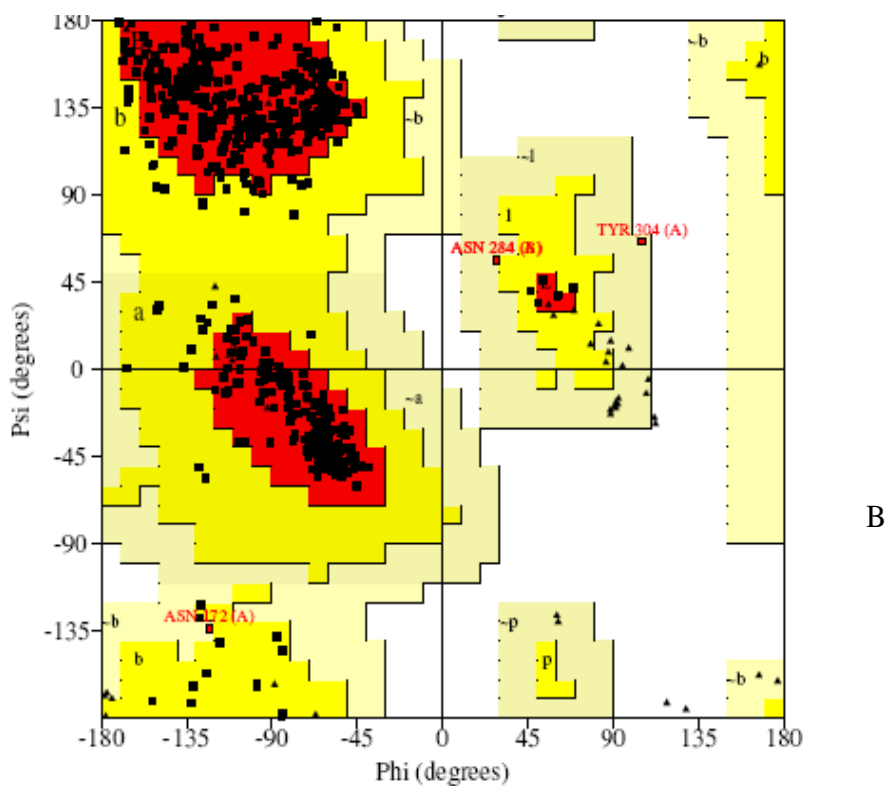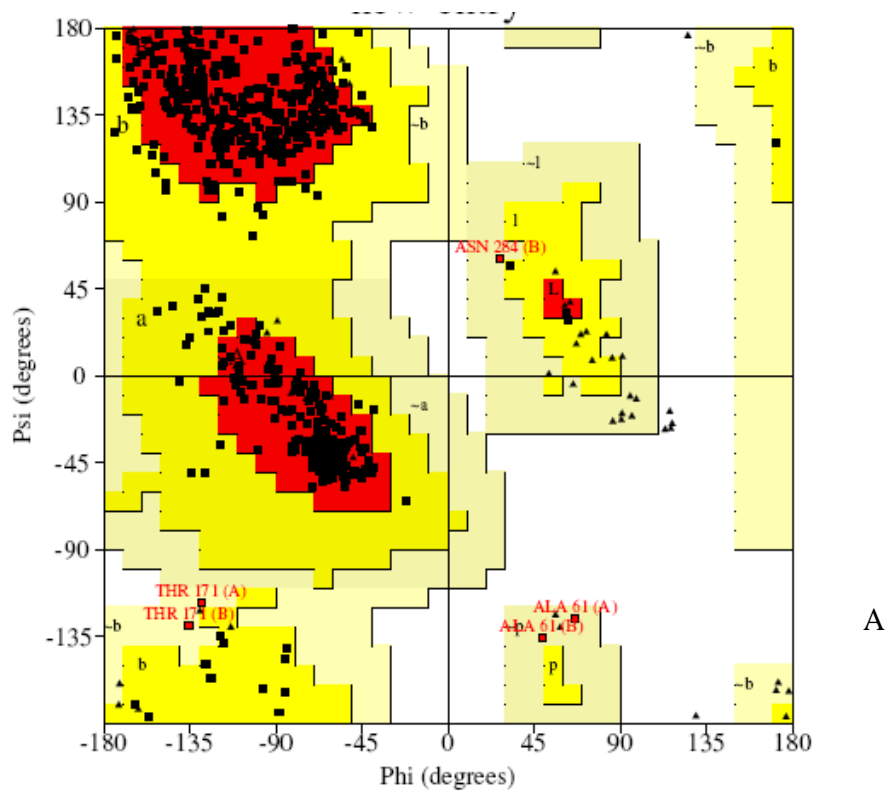
Figure 3. 3: Ramachandran plot corresponding to A) trigonal form B)hexagonal form generated using PROCHECK. Triangles represent glycines**.**

3.6. Description of the structure

The overall structure of BSH from *B. longum* (*Bl*BSH) confirms the characteristic Ntn hydrolase fold comprising of a four-layered β-α-β-α core structure that is formed by two antiparallel β-sheets packed against each other, and these β-sheets are sandwiched between the layers of α-helices on either side. The approximate dimensions of the monomer are 75 x 38 x 44 Å. The secondary structure of BSH contains two six stranded anti-parallel β-sheets (Figure 3.4 and Table 3.6). There are many main chain-main chain hydrogen bonds in each and mostly involving β-sheet formation. Listed in table. 3.7 are the number of various types of observed hydrogen bonds in the two crystal structures. The β-sheet I includes the N- and C- termini and is composed of five strands (Figure 3.5). The β- sheet II comprises of seven strands. The topology of the strands in the first β-sheet is NH_2- β1, β2, β16, β17 and β18-COOH and the second β-sheet is β13, β12, β11, β8, β7, β6 and β3 (Figure 3.6). The β-sheet I is more flattened and the angle between the strands of the two sheets is ~30º.

Few residues are involved in the loops connecting the β-strands and α-helices. However, a large loop of about 26 Å length comprising residues 188 to 220 contained in β-sheet II, is a prominent feature in the structure, which extends into the neighbouring molecule of the tetramer. Two other major loops (58-65 and 129-139) enclose the active site. The distances from the active site residue Cys1 to the termini of the above two loops are about 15.5 and 13.4 Å, respectively. Out of a total of 12 loops, nine contain glycine residues that are more flexible in their possible conformation than other amino acids and influence the loop structure. These glycine residues appear to be conserved in all the Ntn-hydrolases. There is a preponderance of hydrophobic residues between the β-sheets and many hydrophobic side chains are also present in the interface between β-sheets and α-helices and between α-helices. The β-sheet II and α-helix I layer together form a substructure reminiscent of a T-fold (Colloc'h et al., 2000).
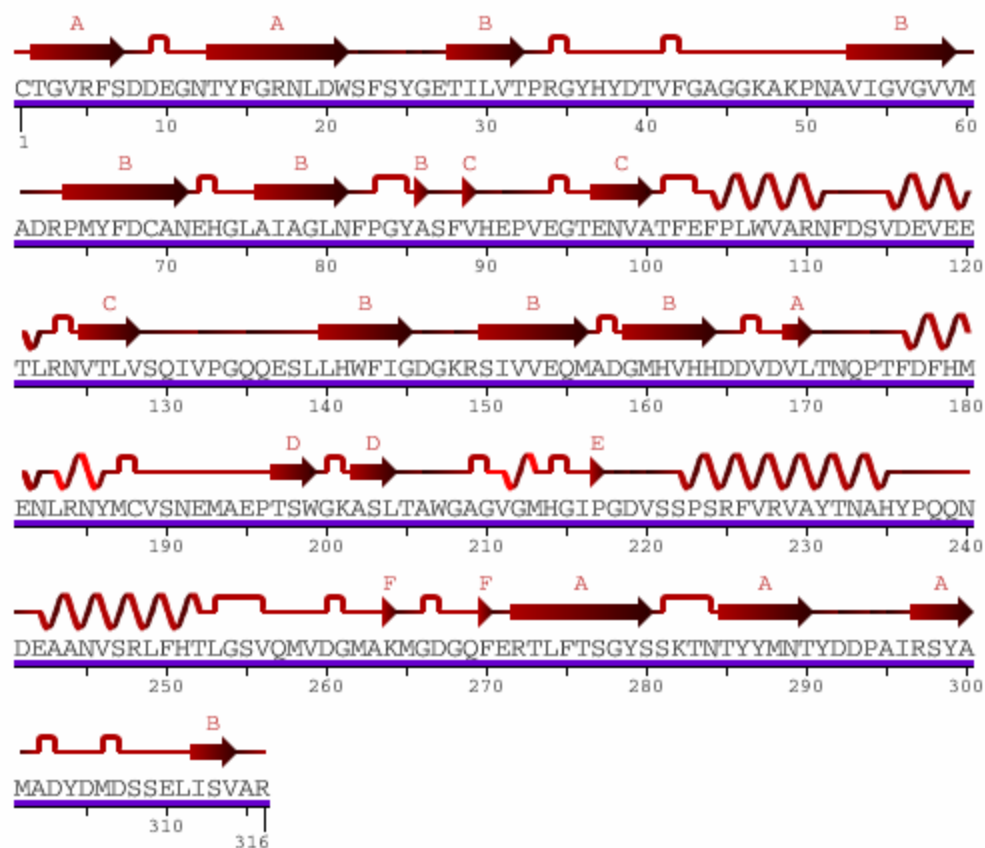
Figure 3.4: Arrangement of secondary structure elements in *Bl*BSH

Table 3.6: List of helix and sheets in BlBSH structure

Helix number	Helix ID	Initial residue	Terminal residue	Helix class
Helix 1	1	103	112	Right-handed
Helix 2	2	114	123	Right-handed
Helix 3	3	175	183	Right-handed
Helix 4	4	184	187	Right-handed 3_{10}
Helix 5	5	210	214	Right-handed 3_{10}
Helix 6	6	221	236	Right-handed
Helix 7	7	241	253	Right-handed

Strand number	Initial residue	Terminal residue	Strand sense
A sheet			
Strand 1	169	170	first strand
Strand 2	1	7	anti-parallel
Strand 3	13	21	anti-parallel
Strand 4	272	280	anti-parallel
Strand 5	285	290	anti-parallel
Strand 6	297	300	anti-parallel
B sheet			
Strand 1	159	164	first strand
Strand 2	150	156	anti-parallel
Strand 3	140	145	anti-parallel
Strand 4	76	81	anti-parallel
Strand 5	64	71	anti-parallel
Strand 6	53	59	anti-parallel
Strand 7	28	32	anti-parallel
Strand 8	312	314	anti-parallel
C sheet			
Strand 1	97	100	first strand
Strand 2	125	128	parallel
D sheet			
Strand 1	197	199	first strand
Strand 2	202	204	anti-parallel

Table. 3.7. The number of intra-subunit hydrogen bonds classified according to their types: main chain-main chain (Mc-Mc), main chain- side chain (Mc-Sc), side chain-side chain (Sc-Sc) and protein-water (Prot-Wat), observed for each subunit in the two crystal forms of B*l*BSH.

Interaction Subunits	Mc-Mc	Mc-Sc	Sc-Sc	Prot-Wat
Trigonal				
A	499	63	48	317
B	497	68	49	319
AS	499	63	44	309
BS	495	63	47	320
Hexagonal				
A	492	53	51	334
B	495	50	51	366
AS	493	51	52	326
BS	495	54	50	359
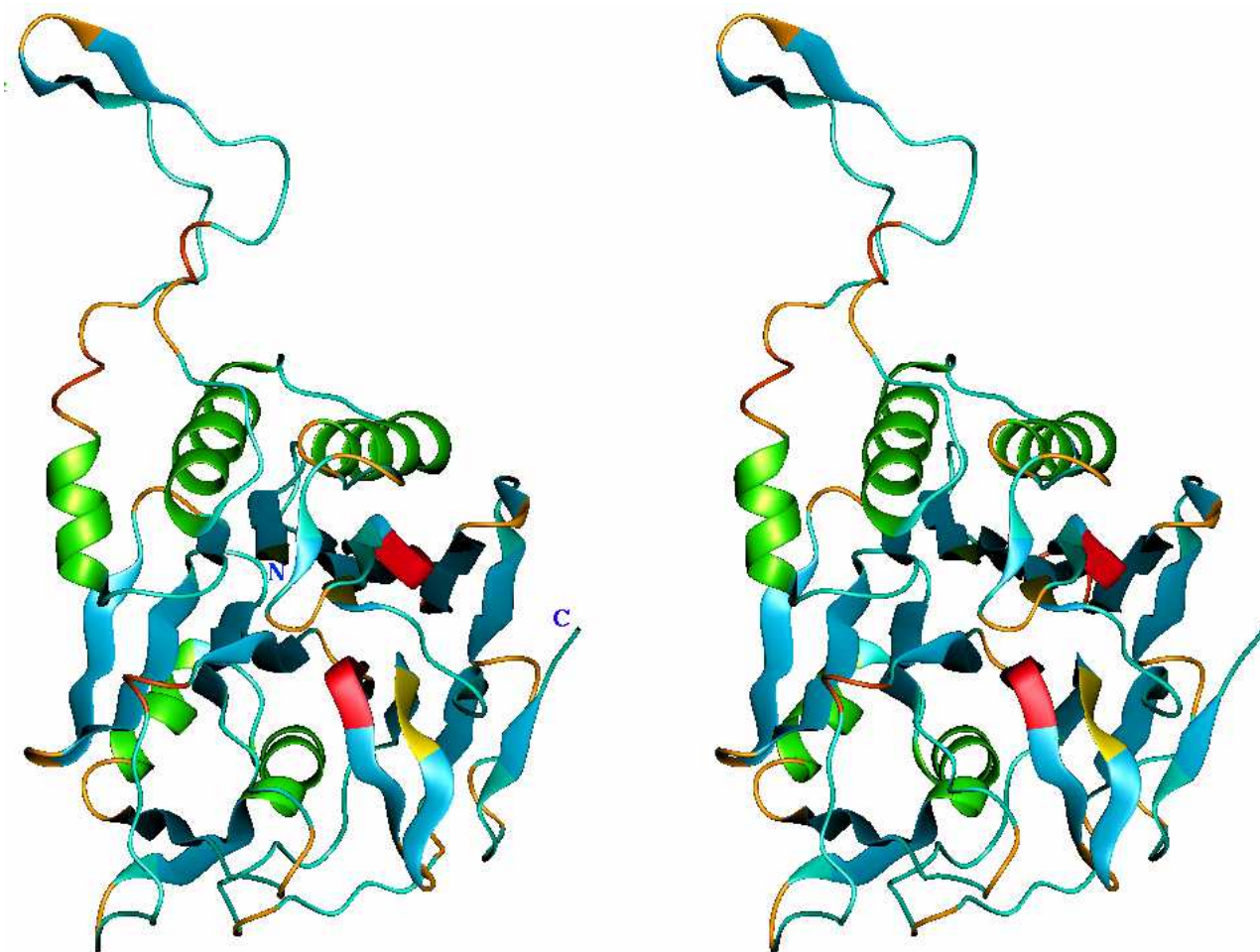
Figure 3.5: Stereo representation of the three-dimensional structure of *Bl*BSH monomer.
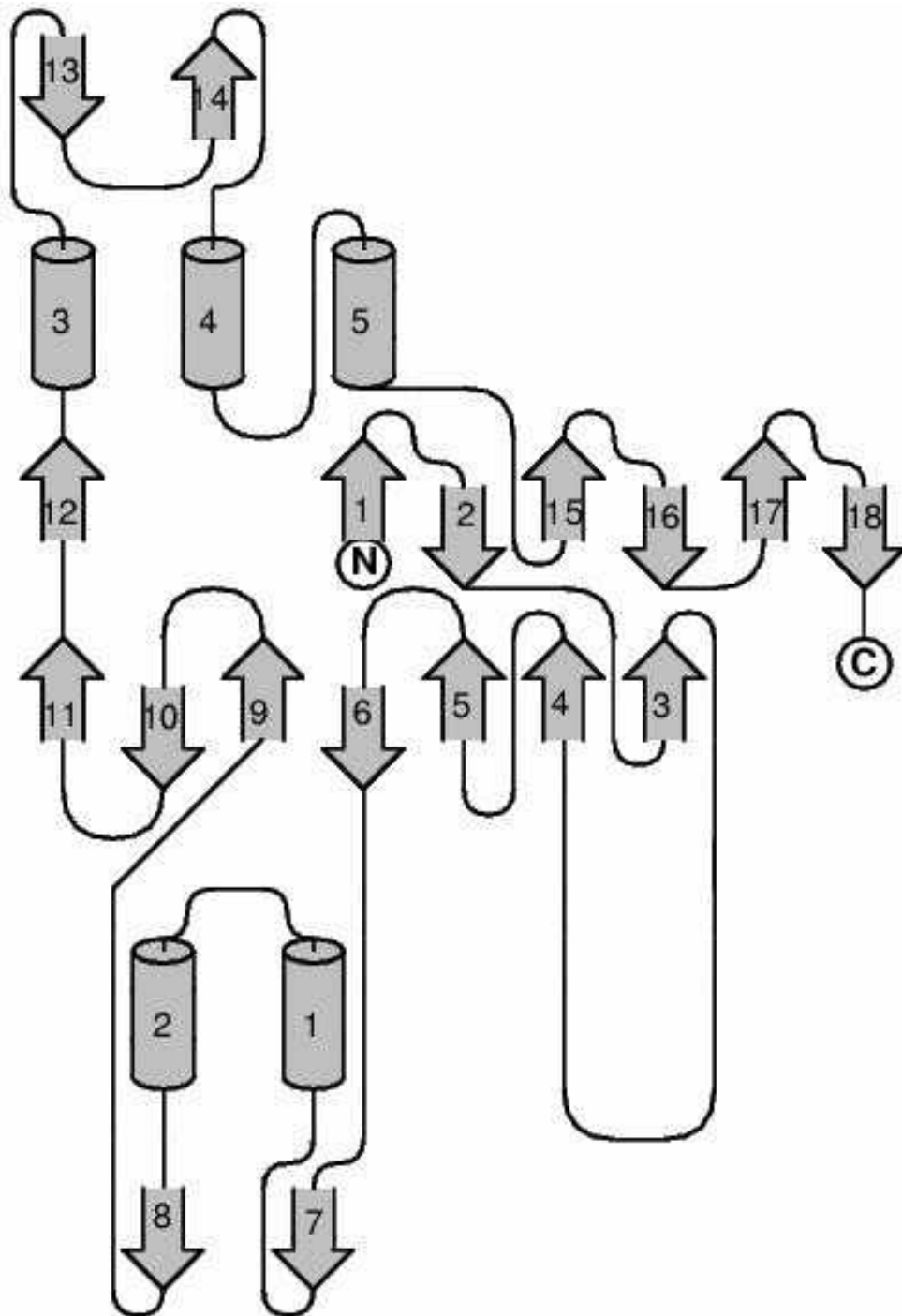
Figure 3.6: Topology diagram of the secondary structure content of *Bl*BSH. The α-helices are numbered and represented by cylinders. The β-strands are labelled alphabetically and represented by flat arrows.

3.6.1. Quaternary structure of *Bl*BSH

The tetrameric association of BSH subunits in solution was suggested by gel-filtration and dynamic light scattering studies. The tetramer may be considered a dimer of dimers. Two dimers related by crystallographic diad axis (monomer A related to AS and B to BS) form the tetramer in the crystal (Figure 3.7). A major contribution to the formation of the dimer in the asymmetric unit comes from interaction of the loop consisting of residues 188 to 220, which extends from one monomer across to two of the others in a manner somewhat reminiscent of domain swapping (Figure 3.7). Extensive interactions between monomers A and AS (12 hydrogen bonds) A and BS (14 hydrogen bonds) and the equivalent B with AS, and B with BS subunits exist. All these associations also include hydrophobic and electrostatic interactions. The total solvent accessible area of the whole tetramer is about 43,500 ^{2}.

The differences between the sum of the accessible surface areas of the individual subunits and the accessible surface area of the oligomer gives the surface area buried as a result of oligomerization. The surface area buried upon association of any two subunits is computed in order to understand the nature of interactions among the subunits. The computation was performed using the method of Lee & Richards (1971) by employing a probe radius of 1.4 Å. Table 3.8 lists the total surface area buried and the non-polar surface area buried with the percentage of non-polar surface area estimated for a pair of subunits. In the tetramer, each monomer interacts with the remaining three monomers. The buried surface area between monomers A and AS is 13% of the surface of each monomer. The main interactions between these monomers are through the long loop (residues 188-200). The buried surface between subunits A and B is only 3% of the monomer surface area while that between A and BS is about 10%. About 26% of the total tetramer surface is involved in subunit interaction. Thus the maximum interactions within the dimers are between subunits A and AS or B and BS. The dimers are held together by the interaction of A with B and AS with BS. Each of the monomers has approximately one-fourth of the surface contributed in tetramer formation (Table 3.8a).
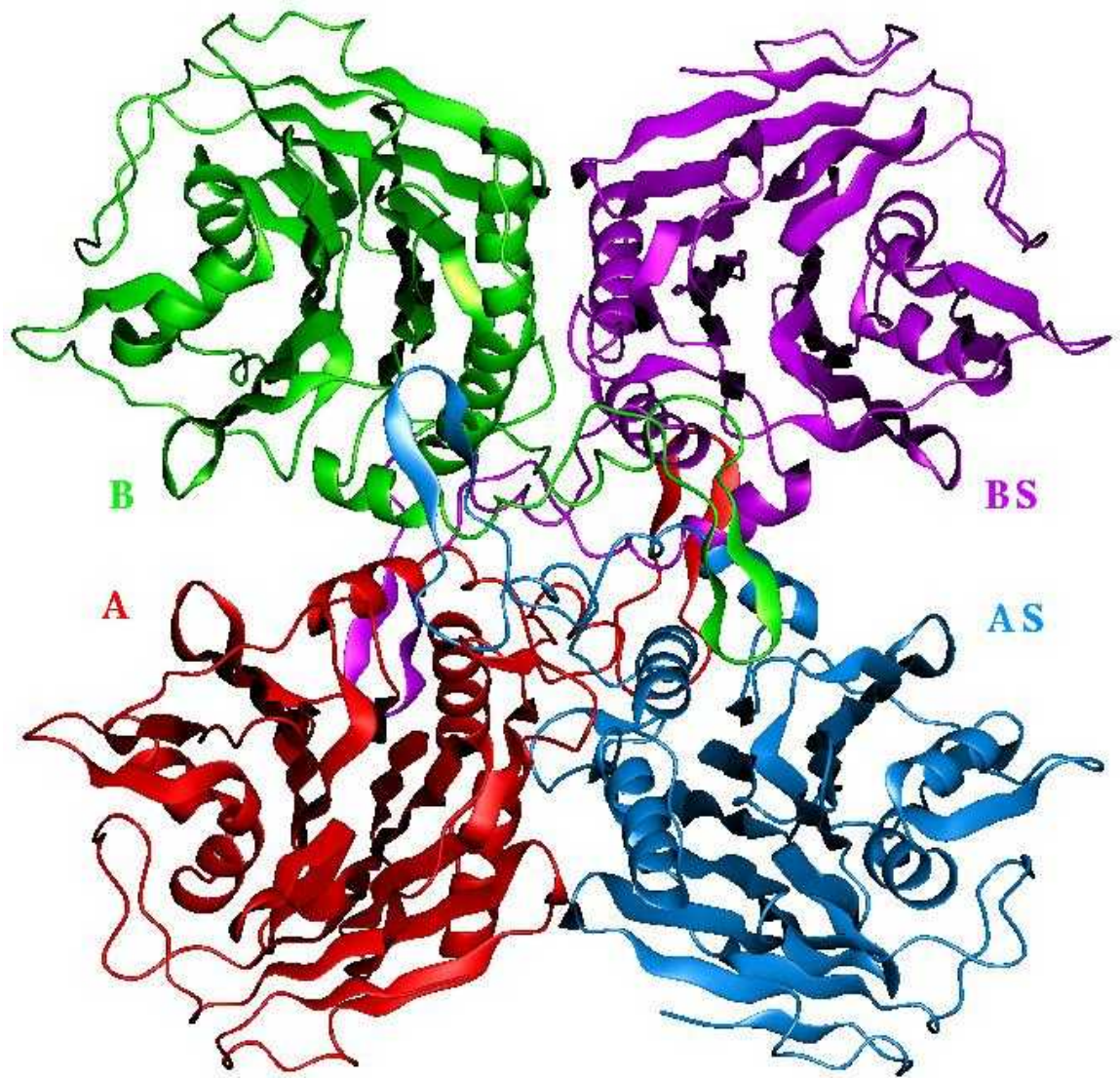
Figure 3.7: The quaternary structure of *Bl*BSH is displayed. The symmetry related monomers generated from A and B are labelled as AS and BS, respectively. The four subunits of the tetramer are shown in different colours.

Table 3.8a: Comparison of accessible surface area (ASA) (Å^2) during complex formation in trigonal and hexagonal form.

	Trigonal form	**Hexagonal form**
Absolute sum ASA for complex	54239.6	53721.1
Absolute sum ASA for isolated chain A	18019.3	17956.7
Total ASA per chain in complex	13559.9	13430.3
Difference per chain from isolated chain A to complex formation	4459.4	4526.4

Table 3.8b: Results of buried surface area calculations for the *Bl*BSH structure

Interacting subunits	Total surface area (Å^2)	Total surface area buried (Å^2)	Percentage of surface area buried
A-B	28870.4	6843.3	23.7
A-AS	25891.9	12961	50.4
A-BS	22408.3	4484.3	20.0
Tetramer (total)	44439.50	22212.6	49.9

The inter-subunit hydrogen bonds observed between subunits A and B are listed in Tables 3.9a and b classified according to the type.

Table 3.9a: Hydrogen bonds observed between the subunits A and B in the trigonal crystal form.

The type of bond	Subunit 1	Subunit 2	Distance
	Residue, chain number & atom	Residue, chain number & atom	(Å)
Main chain-Main chain	Asn A 191 O	Lys C 264 N	2.83
	Ile A 216 O	Gly C 218 N	2.77
	Gly A 218 N	Gly C 218 N	3.25
	Gly A 218 N	Ile C 216 O	2.76
	Lys A 264 N	Asn C 191 O	2.81
	Lys A 201 N	Asp D 166 O	3.34
Main chain-Side chain	Asn A 191 ND2	Gly C 261 O	2.87
	Tyr A 231 OH	Met C 262 O	2.88
	His A 251 NE2	Tyr C 291 O	2.72
	Gly A 261 O	Asn C 191 ND2	2.86
	Met A 262 O	Tyr C 231 OH	2.86
	Tyr A 291 O	His C 251 NE2	2.70
	Lys A 148 O	Lys D 201 NZ	3.43
	Arg A 184 NE	Thr D 205 O	3.23
	Lys A 201 NZ	Lys D 148 O	3.45
	Thr A 205 O	Arg D 184 NH2	3.46
	Thr A 205 O	Arg D 184 NE	3.00
Side chain-Side chain	Tyr A 85 OH	Glu B 181 OE1	3.02
	Tyr A 85 OH	Arg B 184 NH1	2.64
	Tyr A 85 OH	Glu B 181 OE2	2.45
	Glu A 181 OE1	Tyr B 85 OH	3.05
	Glu A 181 OE2	Tyr B 85 OH	2.36
	Arg A 184 NH1	Tyr B 85 OH	2.68
	Glu A 192 OE2	Lys C 264 NZ	2.54
	Tyr A 236 OH	Tyr C 291 OH	2.92
	His A 251 ND1	Gln C 257 NE2	3.29

The type of bond	Subunit 1	Subunit 2	Distance
	Residue, chain number & atom	Residue, chain number & atom	(Å)
	Gln A 257 NE2	His C 251 ND1	3.31
	Lys A 264 NZ	Glu C 192 OE2	2.52
	Tyr A 291 OH	Tyr C 236 OH	2.91
	Arg A 184NH1	Trp D 207 NE1	2.99
	Glu A 181 OE1	Trp D 207 NE1	3.23
	Arg A 184 NH2	Thr D 205 OG1	3.43
	Thr A 205 OG1	Arg D 184 NH2	3.12
	Trp A 207 NE1	Glu D 181 OE1	2.92
	Trp A 207 NE1	Arg D 184 NH1	3.44

Table 3. 9b: Hydrogen bonds between the subunits A and B as observed in the hexagonal form.

Nature of the bond	Subunit 1	Subunit 2	Distance
	Residue, chain number & atom	Residue, chain number & atom	(Å)
Main chain-Main chain	Asn A 191 O	Lys B 264 N	2.74
	Ile A 216 O	Gly B 218 N	2.76
	Gly A 218 N	Ile B 216 O	2.81
	Gly A 218 N	Gly B 218 N	3.32
	Lys A 264 N	Asn B 191 O	2.83
	Asp A 166 O	Lys D 201 N	3.31
	Lys A 201 N	Asp D 166 O	3.32
Main chain-Side chain	Asn A 191 ND2	Gly B 261 O	2.68
	Gly A 210 O	Asp B 219 OD1	3.48
	Asp A 219 OD1	Gly B 210 O	3.41
	Tyr A 231 OH	Met B 262 O	2.70
	Ser A 247 OG	Tyr B 291 O	3.12
	Gly A 261 O	Asn B 191 ND2	2.75
	Met A 262 O	Tyr B 231 OH	2.66

Nature of the bond	Subunit 1	Subunit 2	Distance
	Residue, chain number & atom	Residue, chain number & atom	(Å)
	Tyr A 291 O	His B 251 NE2	2.71
	Tyr A 291 O	Ser B 247OG	3.01
	Lys A 148 O	Lys D 201 NZ	2.68
	Arg A 184 NE	Thr D 205 O	2.81
	Arg A 184 NH2	Thr D 205 O	3.21
	Lys A 201 NZ	Lys D 148 O	2.59
	Thr A 205 O	Arg D 184 NH2	3.13
	Thr A 205 O	Arg D 184 NE	2.69
Side chain-Side chain	Glu A 192 OE1	Lys B 264 NZ	3.40
	Glu A 192 OE2	Lys B 264 NZ	2.66
	His A 214 ND1	Asp B 219 OD2	3.48
	Gln A 257 NE2	His B 251 ND1	2.58
	Lys A 264 NZ	Glu B 192 OE1	3.46
	Lys A 264 NZ	Glu B 192 OE2	2.72
	Tyr A 291 OH	Tyr B 236 OH	2.67
	Tyr A 236 OH	Tyr B 291 OH	2.63
	Tyr A 85 OH	Glu C 181 OE2	2.65
	Tyr A 85 OH	Glu C 181OE1	3.26
	Tyr A 85 OH	Arg C 184 NH1	2.99
	Glu A 181 OE1	Tyr C 85 OH	3.26
	Glu A 181 OE2	Tyr C 85 OH	2.62
	Arg A 184 NH1	Tyr C 85OH	2.97
	Glu A 181 OE1	Trp D 207 NE1	2.90
	Arg A 184 NH1	Trp D 207 NE1	3.38
	Arg A 184 NH2	Thr D 205 OG1	3.15
	Thr A 205 OG1	Arg D 184 NH2	2.94
	Trp A 207 NE1	Arg D 184 NH1	3.32
	Trp A 207 NE1	Glu D 181 OE1	2.87

In this table C is same as AS and D is same as BS of figure 3.7

3.6.2. Crystal Packing

The asymmetric unit of both the crystal forms contained a dimer. Since the unit cell dimensions of the two crystal forms were different, the molecular arrangement in their crystals is expected to differ. The molecular packing diagrams viewed along the three unit cell axes of both the forms are shown in Figure 3.8. In trigonal form totally 322 contacts were observed between symmetry related molecules A and BS. The interacting amino acid residues at the intermolecular contact region that stabilize the crystal packing in trigonal form are Asp A267 with Glu BS91, Gln A269 with Glu BS 130 and Tyr A 25 with Gln BS135 of a subunit from a symmetry related neighboring tetramer.

The trigonal form contained 266 water molecules in the primary solvent shell and established a total of 317 hydrogen bonds with the protein atoms. The numbers in the hexagonal form were 248 waters and 334 hydrogen bonds, respectively. In total, there were 220 water-water hydrogen bonds in the trigonal form and 150 in hexagonal form. The r.m.s. deviations between the two forms were shown in Table 3.10.
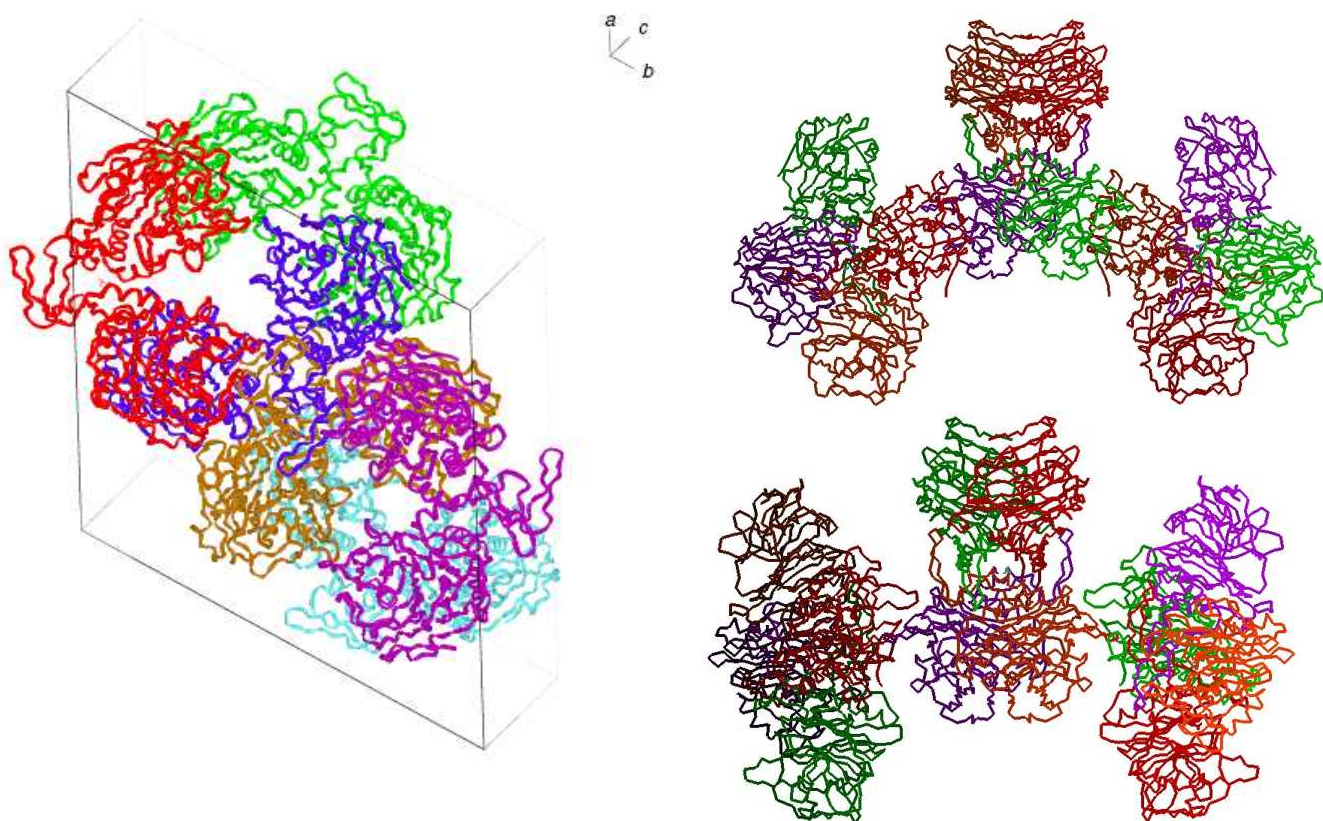
Figure 3.8a: Crystal packing in trigonal form
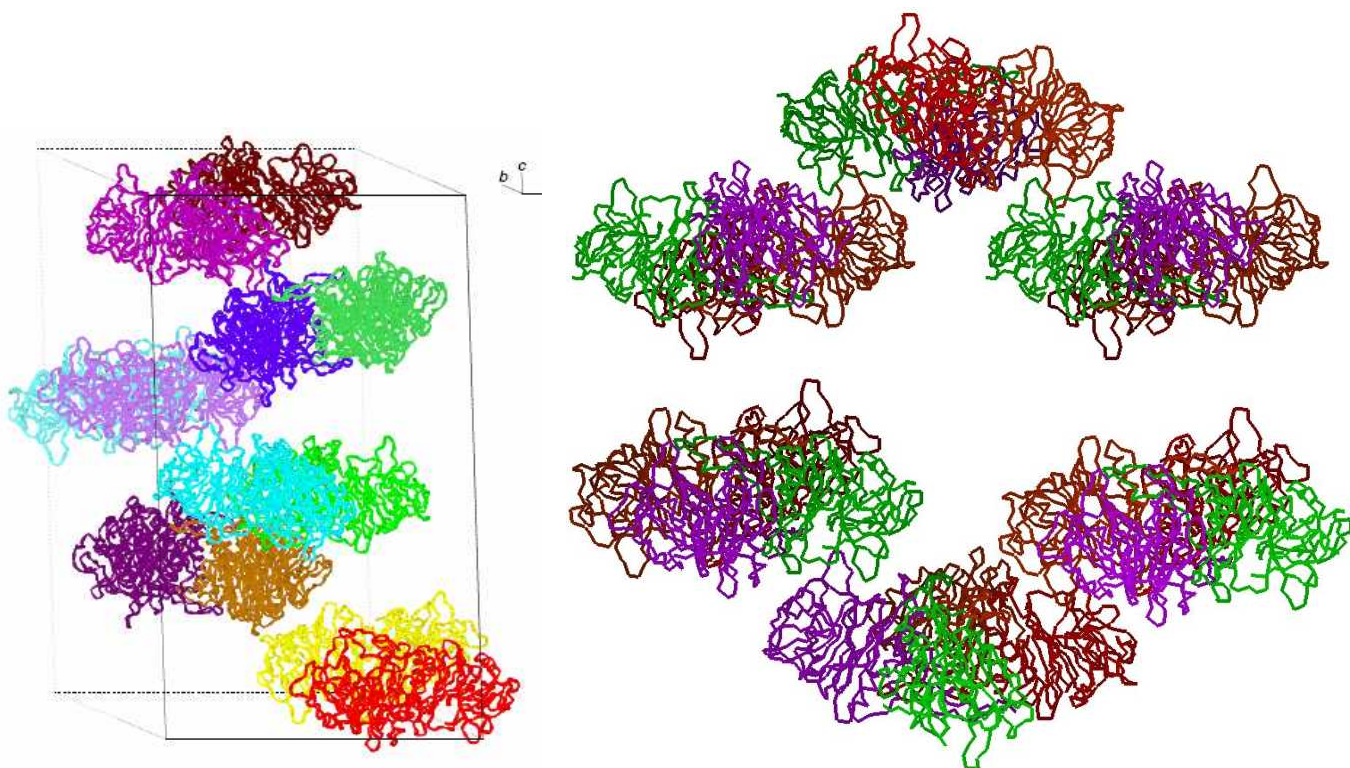
Figure 3.8b: Crystal packing in hexagonal form

Table 3.10: The r.m.s. deviation in C positions upon pair wise superposition of A and B subunits of one crystal form on to subunits of other.

Crystal Form	Trigonal A	B
Hexagonal A	0.212	0.296
B	0.286	0.207

3.6.3. Active site and activity studies

Chemical modification of BSH with 2 mM cysteine-specific DTNB lowered the activity by 78%. In the presence of 20 mM GCA, the drop in activity was only 12% indicating that the modified residue or residues are at, or near the active site. The conformity between the CD spectra of the native and the chemically modified BSH indicates that the loss of activity is due to specific residue modification and not due to structural perturbations. The putative role of Cys1 in catalysis was further clarified by the replacement of Cys1 by Ala through site directed mutagenesis. The C1A mutant turned out to be completely inactive. The residue Cys1 is situated at the tip of strand 1. The proximity of this residue to other putatively important residues such as Trp21, Thr171, Asn172 and Arg225 is evident from the structure (Figure 3.9).

The binding site was also confirmed by superposing the structure of the enzyme-substrate complex of *Cp*BSH. The residues that are identified as responsible for catalysis (Cys1, Arg17, Asp20, Asn174 and Arg227) by comparison with PVA (Suresh et al., 1999) are conserved in both the structures. Residues involved in substrate binding are conservatively replaced by Leu19, Tyr25, Met60, Phe67, Gln136 and Phe42, respectively The substrate-binding site consists of residues of -sheet II located in strands 6, 7 and 11 that stabilize the geometry of the active site. The higher B-factors of the loop comprising residues 262-273 which is proximal to the active site reveal its enhanced conformational flexibility (Figure 3.10). Based on sequence alignment (Figure 3.11) and substrate binding studies we suggest that Trp21 could be playing an important role in the binding of bile salt and at the same time may be suppressing productive binding of penV. This residue is located about 6.3 away from Cys1.
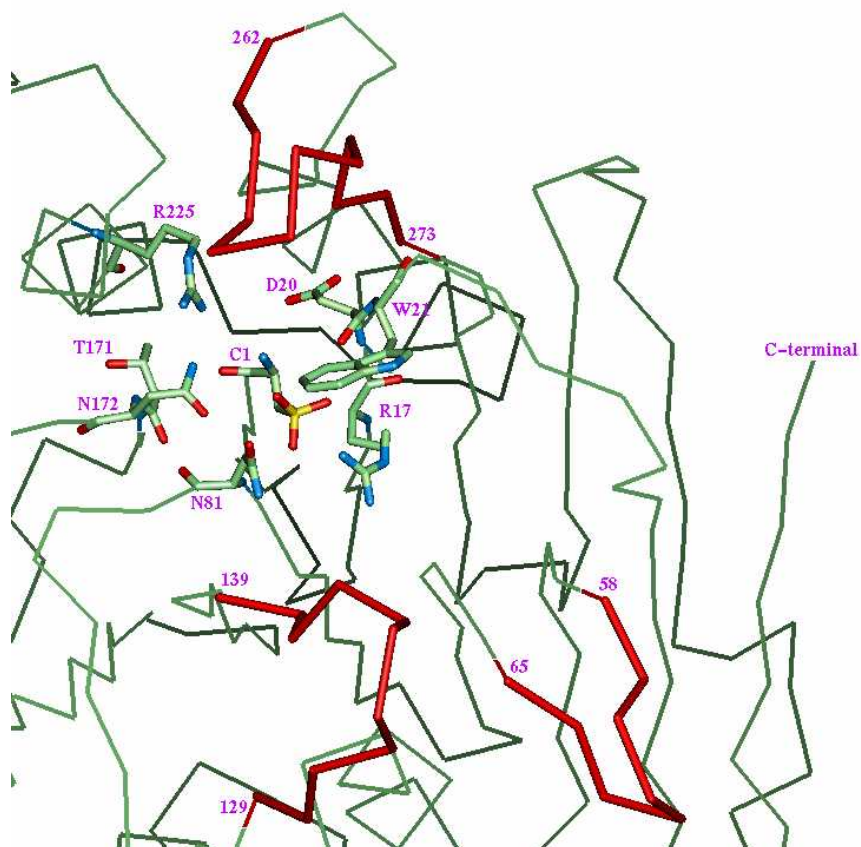
Figure 3.9: Active site and the substrate-binding site of *Bl*BSH are displayed.

120

3.7. Crystal structures of mutants C1A and T2A

The superposition of the mutant and native *Bl*BSH structures showed that there was little effect of the mutations on the geometry of the catalytic site. Residue Thr2 is conserved in all BSHs but not in PVA where serine takes its place (Figure 3.11). This mutation was tested for possible effects on processing or activity. Both mutant proteins processed normally, by simple removal of the initiator methionine (Sherman et al., 1985). The C1A mutant is completely inactive while T2A mutant is partially active, with a reduced k_{cat}.

3.8. Discussion

3.8.1. Structural comparison of BSH with PVA

The structure of *Bl*BSH is highly similar to those of *Cp*BSH (Rossocha et al., 2005) and *Bsp*PVA (Suresh et al., 1999). On superposition of 294 and 291 equivalent C atoms of *Bl*BSH on to other two structures, the r.m.s. deviation with *Cp*BSH is 1.5 Å and with *Bsp*PVA is 1.6 , respectively (Figure 12 and 13). The pattern of secondary structures is also similar in all three proteins. The -sheet arrangements are shown in Figure 3.6, *Bl*BSH contains an additional -strand in the first sheet. The prominent differences observed between these three structures is the reduced length of the C-terminal tail in the case of *Bl*BSH. This extra strand, present in *Cp*BSH and *Bsp*PVA provides additional subunit interaction by forming an intermolecular -sheet. So the association of the subunits in *Bl*BSH might not be as strong as in *Cp*BSH and *Bsp*PVA due to the absence of these C-terminal interactions. Further investigations are required to understand the significance of this terminal strand in subunit interactions and stabilization of tetramer association.

Detailed examination of the superposed structures of these three enzymes has shown significant difference in the orientation of the loops near the active site. Only in *Cp*BSH structure these loops are seen with ligand bound enzyme and free enzyme. Comparison of ligand bound and free enzyme structures show considerable movement of loops comprising of residues 58-65 and 129-139 and minor changes in loops of residues 38-47 and 21-27. The largest displacement of 10 is observed between the loops 60-65 of *Bsp*PVA and the corresponding loop of *Bl*BSH. All the above loops might be playing a role in substrate specificity. These loops are more open and placed further away from the

binding pocket as we go from the structure of a 'good' PVA to a 'good' BSH (Figure 3.10). In *Bsp*PVA it is nearer to a closed conformation while it has a more open conformation in *Bl*BSH. In general the surface loops show considerable deviation between the three structures.
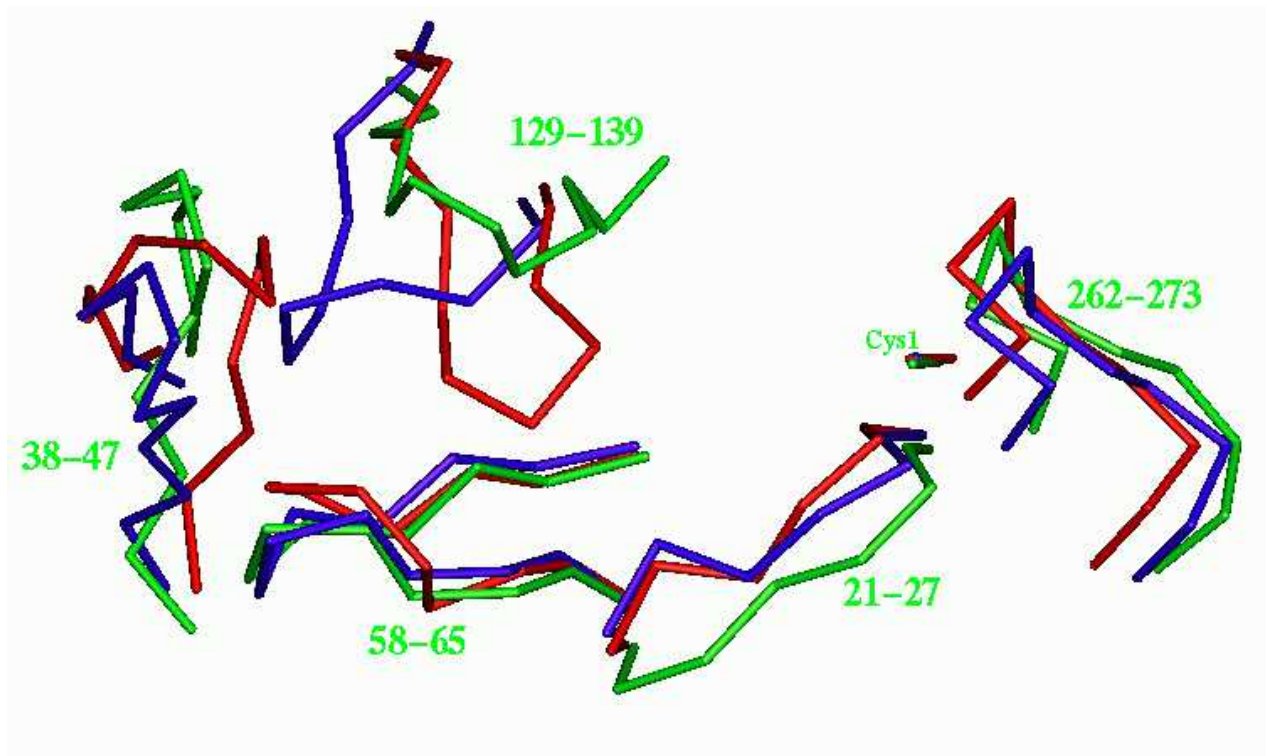
Figure 3.10*:* Superposition of active site loops represented by C trace of the three
structures: BlBSH in green, CpBSH (PDB code 1BJG) in blue and BspPVA
(PDB code 3PVA) in red. Residue numbers shown are with respect to
BlBSH structure. Superposition was performed using the program
SUPERPOSE and the figure is produced in QUANTA (Accelrys).

β1   β2   β3   β4   β5

```
            1         10        20        30        40        50        60
BlBSH  CTGVRFSDDEGNTYFGRNLDWSFSYGETILVTPRGYHYDTVFGAGGKAKPN.AVIGVGVVMADRPMYFDC
CpBSH  CTGLALETKDGLHLFGRNMDIEYSFNQSIIFIPRNFKC.VNKSNKKELTTKYAVLGMGTIFDDYPTFADG
BsPVA  CSSLSIRTTDDKSLFARTMDFTMEPDSKVIIVPRNYGIRLLEKENVVINNSYAFVGMGSTDITSPVLYDG
```

β1   β2   β3 β4 β5 β6   β7

β6   β7 α1   α2   β8

```
            70        80        90        100       110       120       130
BlBSH  ANEHGLAIAGLNFPGYASFVHEPVEGTENVATFEFPLWVARNFDSVDEVEETLRNVTLVSQ.IVPGQQES
CpBSH  MNEKGLGCAGLNFPVYVSYSKEDIEGKTNIPVYNFLLWVLANFSSVEEVKEALKNANIVDIPISENIPNT
BsPVA  VNEKGLMGAMLYYATFATYADEPKKGTTGINPVYVISQVLGNCVTVDDVIEKLTSYTLLNEANIILGFAP
```

β8   β9 η1 α1   α2 η2 β10

β9   β10 β11 β12   α3 η1 β13 β14

```
            140       150       160       170       180       190       200
BlBSH  LLHWFIGDGK.RSIVVEQMADGMHVHHDDVDVLTNQPTFDFHMENLRNYMCVSNEMAEPTSWGKASLTAW
CpBSH  TLHWMISDITGKSIVVEQTKEKLNVFDNNIGVLTNSPTFDWHVANLNQYVGLRYNQVPEFKLGDQSLTAL
BsPVA  PLHYTFTDASGESIVIEPDKTGITIHRKTIGVMTNSPGYEWHQTNLRAYIGVTPNPPQDIMMGDLDLTPF
```

β11 β12 β13 β14   α3 η3 β15 β16

η2   α4   α5   β15

```
            210       220       230       240       250       260       270
BlBSH  GAGVGMHGIPGDVSSPSRFVRVAYTNAHYPQQNDEAANVSRLFHTLGSVQMVDGMAKMGDGQFERTLFTS
CpBSH  GQGTGLVGLPGDFTPASRFIRVAFLRDAMIKNDKDSIDLIEFFHILNNVAMVRGSTRTVEEKSDLTQYTS
BsPVA  GQGAGGLGLPGDFTPSARFLRVAYWKKYTEKAKNETEGVTNLFHILSSVNIPKGVVLTNEGKTDYTIYTS
```

η4   α4   α5 η5   β17

β16   β17   β18

```
            280       290       300       310
BlBSH  GYSSKTNTYYMNTYDDPAIRSYAMADYDMDSSELISVAR
CpBSH  CMCLEKGIYYYNTYENNQINAIDMNKENLDGNEIKTYKYNKTLSINHVN
BsPVA  AMCAQSKNYYFKLYDNSRISAVSLMAENLNSQDLITFEWDRKQDIKQLNQVNVMS
```

β18   β19   β20   β21

Figure 3.11: Sequence alignment of *Bl*BSH, *Cp*BSH and *Bsp*PVA. Identical residues have a black background. Secondary structure elements of *Bl*BSH (above) and *Bsp*PVA (below) are shown.
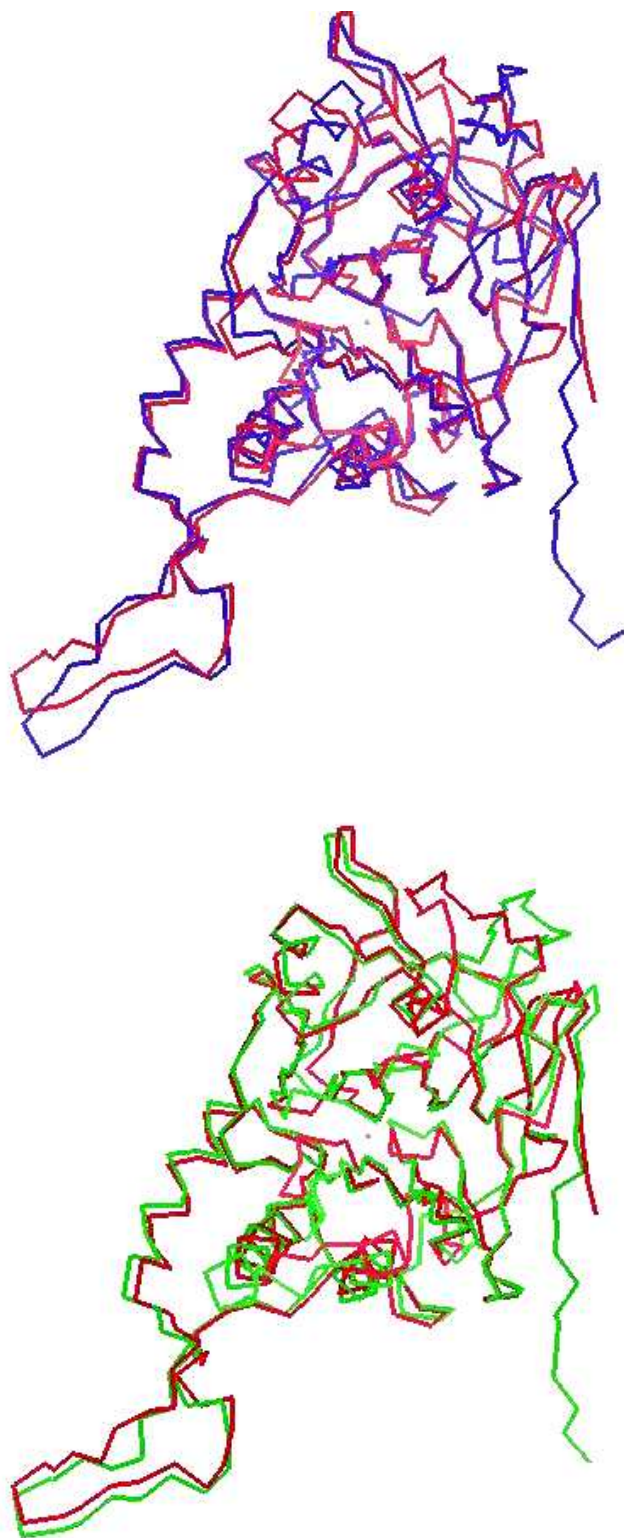
Figure 12: The superposition of C positions between structures of A) *Bl*BSH subunit (blue) with *Bsp*PVA (blue) and B) *Bl*BSH (red) with *Cp*BSH (green).
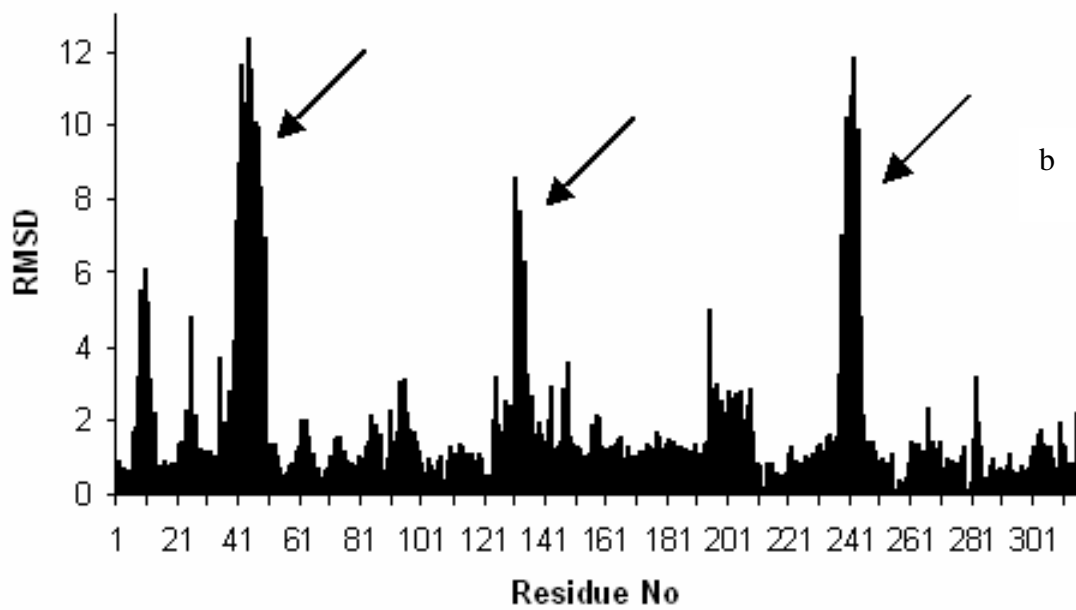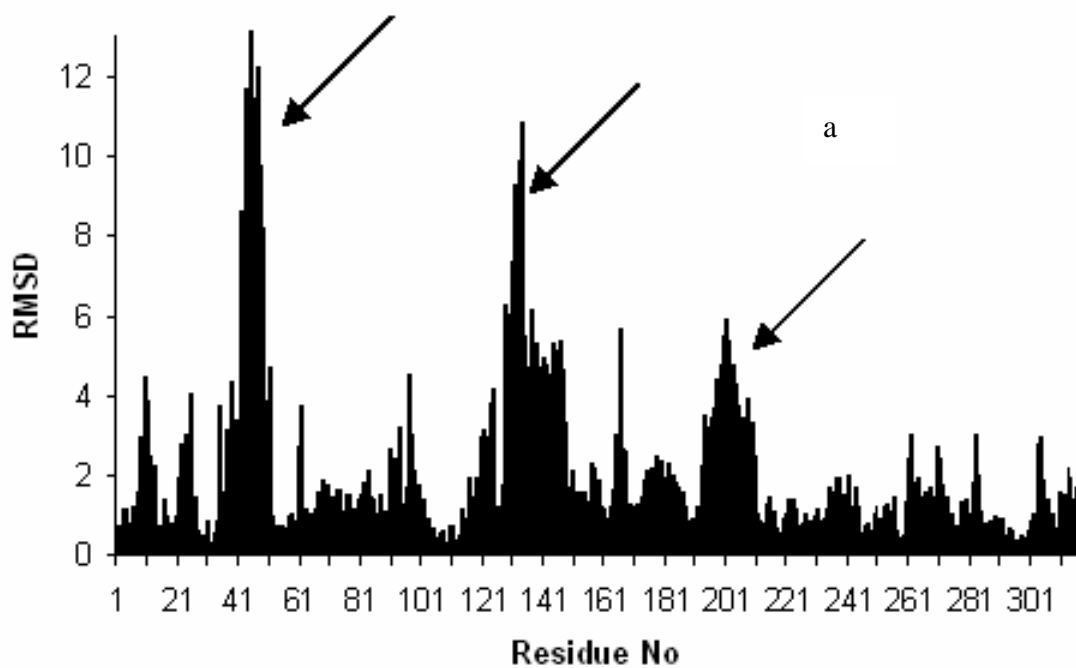
Figure 13: The r.m.s.d. estimated between the positions of main chain atoms of
(a) *Bl*BSH *vs Bsp*PVA (b) *Bl*BSH *vs Cp*BSH.

3.8.2. Comparison with other Ntn hydrolases

Three-dimensional structures of a number of proteins belonging to N-terminal nucleophile hydrolase family have been solved. They have been listed in chapter 1. Table 3.11 lists the protein structures in PDB identified through DALI alignment using *Bl*BSH. Even though their primary sequences share limited homology and a wide range of variation exists in their quaternary association, the members of the Ntn hydrolase family have a well-conserved arrangement of critical residues and an equivalent stereochemistry is preserved at the active site, even when the nucleophilic residue varies (Cys/Thr/Ser). A structural alignment of *Bl*BSH with other Ntn hydrolases using the conserved -strands as template, reveals spatial overlap of all residues identified important for catalysis. The enzyme possesses a catalytic cysteine residue (Cys1), which corresponds to the catalytic nucleophile identified in other enzymes such as PVA Cys1; PGA SerB1; cephalosporin acylase (CPA), Ser170; glutamine amidotransferases (GAT), Cys1; glycosyl asparaginase (AGA), Thr206 and proteaosome (PRO), ThrB1. The -amino group of this residue is the only candidate close to the nucleophilic sulphur atom that can act as base. The backbone >NH group of Tyr82 (in PVA) is in ideal conformation to play the role of the first element of the oxyanion hole (correspondingly PVA: Tyr82; PGA: AlaB69; CPA: Val239: GAT: Gly103; AGA: Gly235; PRO: Gly47B). Similarly, in PVA Asn175 is most likely playing the role of the second contributor to the oxyanion hole (correspondingly PVA: Asn 175; PGA: AsnB241; CPA: Asn413; GAT: Asn102; AGA: Thr234; PRO: Ser129).

Table 3.11: Structural alignment of *Bl*BSH with related structures (DALI alignment)

PDB ID	Z	RMSD	LALI	LSEQ2	%IDE	PROTEIN
3pva	39.3	2.2	309	334	28	PVA
1fm2	14.6	3.6	220	520	10	GACA
1ajq	13.9	3.5	211	557	7	PGA
1pma	9.9	3.4	158	203	6	proteasome
1kuu	8.7	3.7	161	202	8	Hypothetical protein
1pma	8.6	3.4	155	221	8	proteasome
1ryp	8.2	3.6	157	233	5	proteinase
1ned	7.8	3.1	140	180	9	hydrolase
1ryp	7.1	3.5	145	222	4	proteasome
1q5q	6.8	3.7	146	219	10	proteasome
1txo	2.7	4.5	120	230	11	putative bacterial enzyme
1apy	2.3	3.8	94	141	5	aspartylglucosaminidase

GACA: glutaryl 7-aminocephalosporanic acid acylase

3.8.3. Post-translational processing in bile salt hydrolases

All Ntn hydrolases are synthesized as inactive precursors. In the case of PGA and AGA, a single chain precursor is converted into an active heterodimeric form after the precursor undergoes autocatalytic processing. In *Bsp*PVA, the precursor contains a tripeptide at the N-terminus that blocks the active site and prevents the formation of a free amino group for Cys residue that acts as nucleophile. The auto-processing is so fast that in the case of native enzyme no processing intermediate can be trapped. However, it is reported that in *Bsp*PVA the Cys to Ala/Gly/Ser mutants lack auto-processing capability with the result that the first methionine residue is retained in the mutant protein (Chandra et al., 2005). This observation raised the question whether the same phenomenon is occurring in the case of *Bl*BSH also. From the crystal structures of both the mutants C1A and T2A of *Bl*BSH it is inferred that the processing is not hampered by

these mutations. Thus, in the case of BSH the active form is produced by the removal of the N-terminal formylmethionyl residue. As no further post-translational processing is involved, BSH may prove to be a good model system for the further study of Ntn-hydrolases. Especially this molecule is simple among Ntn hydrolases as far as post-translational processing is concerned.

3.8.4. Function-sequence comparison and phylogenetic analysis

The sequences of an exhaustive list of related proteins presently available in databases were aligned (Figure 3.14). The phylogenetic analysis suggests a diversion of PVA and BSH during evolution despite their apparently conserved structures. From Table 4.3 we can clearly observe a correlation between sequence homology and the enzyme activity. *Bsp*PVA shows good activity towards both penV and taurocholic acid substrates. As the sequence departs from PVA towards the *Bl*BSH there is a diminishing PVA activity and a correspondingly increasing in BSH activity, whereas *Cp*BSH sequence represents an intermediate between these two extreme cases.

Sequence alignment figure.

```
              β1        TT      β2           β3       TT        TT                      ____  ..
Q9KK62_BIFLO  1         10        20          30        40         50
```

Block 1 (positions 1–55):

```
Q9KK62_BIFLO  CTGVRFSDDEG..NTYFGRNLDWSFSYG.ETILVTPRGYHYDTVFG......AGGKAKPNAVIGVGVV..
Q6J263_9BIFI  CTGVRFSDDEG..NMYFGRNLDWSFSYG.ETILVTPRSYQYDYVYG......AEGKSEPNAVIGVGVV..
Q74LX7_LACJO  CTGLRPTDDQG..NLYFGRNLDVGQDYG.EGVIITPRNYPLPYKF.......LDNTTTKKAVIGMGIV..
Q9AHJ7_9LACO  CTSILYSPK....DHYFGRNLDYEIAYG.QKVVITPKNYEFEFTD.......LPAEKSHYAMIGVAAV..
Q74IV4_LACJO  CTSIVYSSNN...HHYFGRNLDLEISPG.EHPVITPRNYEFQYRK.......LPSKKAKYAMVGMAIV..
Q75N78_ENTFC  CTAITYVTS....DHYFGRNFDYEISYN.EVVTITPRNYKLNFRK.......VNDLDTHYAMIGIAAG..
Q8KUA7_ENTFA  CTAITYVSK....DHYFGRNFDYEISYN.EVVTITPRNYKFSFRE.......VGNLDHHPAIIGIAAG..
Q8Y5J3_LISMO  CTSITYTTK....DHYPGRNFDYELSYK.EVVVVTPKNYPFHPRK.......VEDIEKHYALIGIAAV..
CBH_LACPL     CTAITYQSY....NNYFGRNFDYEISYN.EMVTITPRKYPLVFRK.......VENLDHHYAIIGITAD..
CBH_CLOPE     CTGLALETKDG..LHLFGRNMDIEYSPN.QSIIFIPRNFKCVNKSN......KKELTTKYAVLGMGTI..
Q8Y9S7_LISMO  CTSFVLETLDG..KKLLSRTMDFAPILE.ANPTISPRNYGWKSSTDG.....ANYIN.RYAFVGAGR...
Q65D34_BACLD  CTSITLTTAGR..EHLLARTMDFDPELN.GEVLLHPRYKWKSEADG.....SEHAG.RYAFIGMGR...
Q732V9_BACC1  CTSLTLQTKND..QHLPARTMDFTLDMN.QEVIIIPRHYQWNNITG......EIINTKHATVGMGIN..
Q81WS0_BACAN  CTSLTLETKNG..QHLPARTMDFTLDMN.QEVIIIPRHYQWNNITG......EIINTKHATVGMGIN..
Q6HF48_BACHK  CTSLTLETKNG..QHLPARTMDFTLDMN.QEVIIIPRHYQWNNITG......EIINTKHATVGMGIN..
Q81H11_BACCR  CSSFTLESQDN..KHFLSRTMDFQIEMA.EQILFIPRNKEIGFAHSN.....ETIRTKKAVVGMGIN..
Q9CEK2_LACLA  CSSLSIRTTDD..KSLPARTMDFTMEPD.SKVIIVPRNYGIRLLEKE.....NVVINNSYAFVGMGST..
P12256_BACSH  CSSLSIRTTDD..KSLPARTMDFTMEPD.SKVIIVPRNYGIRLLEKE.....NVVINNSYAFVGMGST..
Q890F5_LACPL  CSLTYTNSHG..GHFLARTMDFNVDPE.TRIMFMPRHYRVTGDLG.....DFTTTYGPIGAGR...
Q6AP13_DESPS  CTSLRLVAGDG..GVVVGRTMEFEVDMK.SNVLVIPAGTEMTSSLAD.....KSQGMKYKTKYGMIG...
Q8YIA2_BRUME  CTSFVLPTSDG..GMVYGRTMDFGPNLK.SDMIAIPRNYTITASGPD.....GAAGKKWKGKYATIG...
Q8ZPG8_SALTY  CTTLAIQDKQG..DIFHGRTLEYMQDLP.SWLTYYPAGTQFDKKTPD.....GSQGISYQAKYPILAITS
Q8Z714_SALTI  CTTLAIQDKQG..DIFHGRTLEYMQDLP.SWLTYYPAGTQFDKKTPD.....GSQGISYQAKYPILAITS
Y284L_CHVP1   CSGLRIIADDG..TVVVGRTLEFGENIL.KFKKFVNG.NIRGISTPD.....G........ ..
Q6ARU2_DESPS  CTRSLYKAGDG..RFLTGRSMDW.DTNLHSNLWVFPRGMERNGG......IDNKSIKWTSKYGSLTASA
Q9KL70_VIBCH  CTRILYETGNQ..SYISGRSMDWADPSAATALWVFPKGMKRDGT.....VGDNPIKWSAKYGSIIVSF
Q6E7I9_9CYAN  CSRLVYDGSSN..GPITARSMDWLNDPG.TELWVFPKGMERDGT......AGCNPITWTSKYGSIVTSS
Q983W5_RHILO  CTRAVYLGAKG..EVITARSMDWKTDVG.TNLWVFPRGMKRSGE.....AGLNSIRWTSKYGSVIASG
Q8U768_AGRT5  CTRFVYIGENN..QVMTARSMDWKTDVG.TNLWVFPRGMERSGE.....AGPNSVKWTSKYGSVIASG
Q6FAR4_ACIAD  CTRVVYKGSNQ..NIMTARSMDWKSDVG.TNLWILPSGVKRSGN.....AGKDSIHWTSKYGSVVATG
Q6D291_ERWCT  CTRFVYLDPHNPDYPITARSMDWKTDVG.TNLWIFPQELKRSGG.....AGQYSLEWTSKYGSVIASA
Q7NNP2_GLOVI  CTRAVYIGPQD..TIITVRSMDWFADIG.TNLWAFPRGLQRDGA......AGAKTIRWTSKYGSVVAAG
Q5N0N8_SYNP6  CTRAVYQGPEG..TIITGRSMDWKDEIP.ANLWLFPRGLQDSGA.....VGPNSVTWVSKYGSVVASA
Q8A8A8_BACTN  CTRVVYSGSNG..MVATGRSMDWKTDMH.SNLWVFPRGMKRNGE......TGENSLEWTSRYGSVVTSA
Q8A600_BACTN  CTRAVYLGPDR..MVVTGRTMDWKEDIM.SNIYVFPRGMQRAGH.....NKEKTVNWTSKYGSVIATG
Q64PP5_BACFR  CTRAVYIGPDN..MVVTGRSMDWKEDIQ.SNLWLFPRGIKRAGY.....NKGNTVEWISKYGSVIATG
Q6MQ22_BDEBA  CTRILWDGKNQ..DVIVGRNMDWSEDTG.SNMWLFPRGMERDGM.....ASKNSAKWTSKYGSVILSM
Q8TSP0_METAC  CTRILYETGTG..TYVTGRSMDWNDLQDMNTTFPRVMKRDGG......VGADSITWTSKYGSLIFSA
P72765_SYNY3  CTRILWNNNKL..AVVVGRTMDWPESTE.PELMVFPQGIKRNGGMLGDMLVVKDNPAQWTSKYGSVVTTV
Q7TU08_RHOBA  CTRILWNTNGV..LTVAGRTMDWAESTE.PRLHVFPRGMHRNGGRVGPLDLGDENPLTWMSRYGSVVVSA
Q5X6U9_LEGPA  CSALVFNKNKP..ATVAVNLDWKYREG..AFVIHPRNAMMVSS....VDYHSMHPLIWKSQYGSVIFHG
Q5ZXC8_LEGPH  CSALVFNKNKP..ATVAVNLDWKYREG..AFVIHPRNAMMVSS....VDYHSMHPLIWKSQYGSVIFHG
Q5WYA2_LEGPL  CSALVFNKNKP..ATVAVNLDWKYREG..AFVIHPRNAMMVSS....VDYHSMHPLIWKSQYGSVIFHG
Q5NFV9_FRATT  CTWFSPENDKG..HYFIGRTMEWPGDLN.TKITLVPRGYKFGNFT........THYGFVGMSHN..
Q82ZP5_ENTFA  CTGIKIISKTN..DIFYGRTMDFTPDFFGNEDPIAPKIPTLIAQPPKG.TVLNSQLNPWTAKYAFMGLAM
```

Block 2 (positions 60–110+):

```
              β4          TT    β5       TT    β6       TT    β7       α1
Q9KK62_BIFLO  ........  60      70      80       90      100      110

Q9KK62_BIFLO  .................MADRPMYFDCANEHGLAIAGLNFPGYASFV.HEPVEG.TENVATFEFPLWVARNPDSVDE
Q6J263_9BIFI  .................MVDRPMYFDCANEHGLAIAGLNFPGYASFA.HEPVEG.TENVATRNPDTVSE
Q74LX7_LACJO  .................VDGYPSYFDCYNEDGLGIAGLNFPHFAKFS.DGPIDG.KINLASYEIMLWVTQNPDTHVSE
Q9AHJ7_9LACO  .................ADNTPLYCDAINEKGLGVAGLSFAGQGKYP.PNAAD..KKNIASPEFISYLLATYETVDQ
Q74IV4_LACJO  .................ENNYLAYYDAANEEGLGIAGLNFDGPCHYF.PENAE..KNNVTPPELIPYLLSQCTTVAE
Q75N78_ENTFC  .................IADYPLYYDATNEKGLSMAGLNFSGYADYK.EIQEG..KDNVSPPEFIPWILGQCSTVGE
Q8KUA7_ENTFA  .................IADYPLYYDAINEKGLSMAGLNFSGYADYK.KIEEG..KENVSPPEFIPWILGQCSTVDE
Q8Y5J3_LISMO  .................MENYPLYYDATNEKGLSMAGLNFSGNADYK.DFAEG..KDNVTPPEFIPWILGQCATVKE
CBH_LACPL     .................VESYPLYYDAMNEKGLCIAGLNFAGYADYK.KYDAD..KVNITPPELIPWLLGQFSSVRE
CBH_CLOPE     .................PDDYPTADGMNEEGLCAGLNFPVYVSYS.KEDIEG.KTNIPVYNFLLWVLANFSSVEE
Q8Y9S7_LISMO  .................ELDKYIPADGLNEEGLSCASLYLPGEAVYA.PAPVEG.KINLAPQEFLLWLLGTCATIKD
Q65D34_BACLD  .................RLKNALPADGLNEEGLSCAALYFPGYAYE.NEAKEQ.SRNLAPHEFVTWVLSECGDLED
Q732V9_BACC1  .............HQGRIIMADGVNEAGMTCATLYFPGFATYS.QSIDDN.TTNLAPPDFVTWSLTQPNSVKE
Q81WS0_BACAN  .............HQGRIIMADGVNEAGMTCATLYFPGFATYS.QSIDDN.TTNLAPPDFVTWSLTQPNSVKE
Q6HF48_BACHK  .............HQGRIIMADGVNEAGMTCATLYFPGFATYS.QSIDDN.KMNLAPPDFVAWSLTQPNSVEE
Q81H11_BACCR  .............FGGRVMMADGVNEAGMTCATLYFPGFATYS.SHVDSN.KTNVAPPDFVTWSLTQPNSVEE
Q9CEK2_LACLA  .............EEGHPVLFDGVNEKGLMGATLYYATFATYA.DEPKKG.TTGINPVYVISQVLGNCVTVDD
P12256_BACSH  .............DITSPVLYDGVNEKGLMGAMLYYATFATYA.DEPKKG.TTGINPVYVISQVLGNCVTVDD
Q890F5_LACPL  .............QLNHEIPTDGVNEGGLSVIAALYFPNHAIYQ.PHSNQD.KIDLAPHDFVAWVLGKITSVAD
Q6AP13_DESPS  ........ANGPDLPIIADGINDQGLYVGTLLLPGYASYPDVDSKNA.SRALASEWPDVSKNAPANGPDPST
Q8YIA2_BRUME  ..........MNAFGIVALTDGMNEKGLAGGLLYFPEYAKYQDPSTAKP.EDSLAPWDFPLTWALANFSTVAE
Q8ZPG8_SALTY  .......ITDGDSRDILEGMNSAGLSFSENMIMNAQLPP.LPASEY.KQAIPVTSLGEWALARPATVGE
Q8Z714_SALTI  T......ITDGDSRDILEGMNSAGLSFSENMIMNAQLPP.LPASEY.KQAIPVTSLGEWALARPATVGE
Y284L_CHVP1   .................KLLDGMNEHGLVIPFKNYAKYG.CPS.QT.KLNIKPTEVALPLLQKANVKD
Q6ARU2_DESPS  F........G.....AATADGINEKGLVTNLLYLVE.SDYG....KSN.KPTLSVGAWAQYLDNYATVAE
Q9KL70_VIBCH  Y........D.....AGTADGMNEKGLVANMLYLKE.AKWGDAT.KAG.KPTLTAGAWAQYFLDNYATVEE
Q6E7I9_9CYAN  Y........N.....MGTVDGLNEKGLVANLLPLTA.TDYG.KI.EPG.DQTLSIGGWGQYVLDNYATVNE
Q983W5_RHILO  Y........D.....ISTTDGMNEAGLNANVLWLVE.SSYPKYD..GK.TPGLSLAAWAQYVLDNYATVSE
Q8U768_AGRT5  Y........D.....VSTTDGMNEAGLNANVLWLVE.SSYPDYD..GK.SPGLSLAAWAQYVLDNPATVEE
Q6FAR4_ACIAD  Y........D.....ISTTDGVNEVGLNANLLWLVE.SQYPDVSQ.SK.KPKLTIAAWAQYVLDNPATVDE
Q6D291_ERWCT  F........DGRKGMASTTDGVNEKGLAANVLWLAE.SEYPKTKPTAK.KPGLSVAAWAQYVLDNPATVDE
Q7NNP2_GLOVI  Y........D.....IASADGMNEKGLVANLLYLTE.SEYVKATNADK.RLPLSISAWAQYVLDNYATVAE
Q5N0N8_SYNP6  F........D.....IASADGMNEKGLVANLLWLAE.SKYPEYK..PS.QKGLSVAAWVQYVLDNPATVEE
Q8A8A8_BACTN  F........E.....IASTDGMNEKGLVANLLWLPE.TEYPVRD..KN.KPGLAITAWVQYMLDNPATVRE
Q8A600_BACTN  Y........D.....IGTCDGMNEKGLVASLLFLPE.SVYSLPG..DT.RPAMGISIWTQYVLDNPATVRE
Q64PP5_BACFR  Y........D.....IGTCDGMNEKGLVASLLFLPE.SIYVRRN..DT.RPVMGISIWTQYVLDNPATVSE
Q6MQ22_BDEBA  Y........D.....VGTADGVNEKGLVANLLYLSE.SHFGTRD..EK.VPGMSVSMWAQYFLDNPATVEE
Q8TSP0_METAC  Y........N.....LVTSDGVNEAEMAGNLLYLAE.ADYGDAK.TSG.KPTISVGALLQYLLDNYGSVAE
P72765_SYNY3  Y........G.....MGTVDGLNEKGLMANLLYLTA.TDFGPRN..PK.KAGVQAGLWGQYLLDNAATVEE
Q7TU08_RHOBA  Y........G.....LASVDGLNEKGLAMHLLFLTA.TDYGPRN..PK.KAGLSAGLWGQYLLDNAATVEE
Q5X6U9_LEGPA  G........N.RPKAGPSADGMNEKGLTASILMLDS.SIYPSNS...E.LPALKTTEWVQYILDNFQTVQE
Q5ZXC8_LEGPH  G........N.RPKAGPSADGMNEKGLTASILMLDS.SIYPSNS...E.LPALKTTEWVQYILDNFQTVQE
Q5WYA2_LEGPL  G........N.RPKAGPSADGMNEKGLTASILMLDS.SSYPSNS...E.LPALKTTEWVQYILDNFQTVQE
Q5NFV9_FRATT  ...............GNFSDGMNEYSLAVSALWLSG.SKYP....DKK.DADYNITMLLPYVLGNTTTVAQ
Q82ZP5_ENTFA  SGTDQPANDGKTVSLAITDGINEAGLSGDIQYLMESSTAPAESLADRGLTPHIAEEVLAYILSNPESVDE
```

129

```
                 α2            β8                                   β9            β10            β11
                 ΩΩΩΩ         ───►                                 ───►          ───►   T.T    ───►   TT
                 120           130                    T.T          140           150            160
                 .             .                                   .             .              .
Q9KK62_BIFLO  VEETLRN..VTLVSQ..............IVP.GQQESLLHWFIGDGKRSIVVEQMA.D.GMHVHHDDVD
Q6J263_9BIFI  VEBALKN..VTLVSQ..............VVP.GQQESLLHWFIGDGTRSIVVEQMA.D.GMHVHHDDVD
Q74LX7_LACJO  VKEALKN..VNLVNE..............AINTSPAVAPLHWIISDSDEAIIVEVSK.QYGMKVPDDKVG
Q9AHJ7_9LACO  VKESLTN..ANISNV..............SFAKNTPASELHWLVGDKTGKSIVVESD.EKGLHVYNNPVN
Q74IV4_LACJO  VKDALKD..VSLVNI..............NFSEKLPLSPLHWLMADKTGESIVVEST.LSGLHVYDNPVH
Q75N78_ENTFC  AKKLLKN..INLANI..............NYSDELPLSPLHWLLADKE.KSIVIESM.KDGLHIYDNPVG
Q8KUA7_ENTFA  AKKLLKN..LNLVNI..............NFSDELPLSPLHWLLADKE.QSIVVEST.KEGLRVPDNPVG
Q8Y5J3_LISMO  ARRLLQR..INLVNI..............SFSENLPLSPLHWLMADQT.ESIVVECV.KDGLHIYDNPVG
CBH_LACPL     VKKNIQK..LNLVNI..............NFSEQLPLSPLHWLVADKQ.ESIVIESV.KEGLKIYDNPVG
CBH_CLOPE     VKEALKN..ANIVDI..............PISENIPNTTLHWMISDITGKSIVVEQT.KEKLNVFDDNNIG
Q8Y9S7_LISMO  VEAKLSV..INLVDQ..............PVPLLGITTPLHWIFTDKSGRCVVIEPT.ETSLRIKENPVG
Q65D34_BACLD  VKKAAVS..LNIVER..............EVSLLSTVTPLHWLLTDRSGASVVIEPT.ADGIQIHDNPVG
Q732V9_BACC1  LKKSVDS..ITFLDI..............PLPDLGLTPPLHWILADKWGDCIVLEPT.SEGLKLYDNPLG
Q81WS0_BACAN  LKKSVDS..ITFLDI..............PLPDLGLTPPLHWILADKWGDCIVLDPT.SEGLKLYDNPLG
Q6HF48_BACHK  LKKSVDS..ITFLDI..............PLPDLGLTPPLHWILADKWGDCIVLEPT.SEGLKLYDNPLG
Q81H11_BACCR  LRKSIDS..IAFIDV..............PLPILGVTPPLHWILADKSGECIVLEPT.ADGIKVYDNPIG
Q9CEK2_LACLA  IIDLFDKK.PVIIND..............TNPTLGLTPPLHPIFSDSSGQSLIIEPR.QGGLSIIKDSIG
P12256_BACSH  VIEKLTS..YTLLNE..............ANIILGFAPPLHYTFTDASGESIVIEPD.KTGITIHRKTIG
Q890F5_LACPL  LRERVKD..VQLISS..............TAELINEIPPLHPIISDQTGETAVLEPT.SGELRLLNNPVG
Q6AP13_DESPS  VRANFNKA.VIVSHP..............PISKPLAGQSPFHYVVHDSTGAAVVIEPV.NKEIKIYEDPIG
Q8YIA2_BRUME  VKDALSTI.SIVDVK..............QKDLG.FTPPAHYTLHDATGASIVIEPI.DGKLKVYDNKLG
Q8ZPG8_SALTY  VKQAIKEG.KFWSPE..............LHRFGDLKSPFHYAFYDKKGGSIVVEVE.NGKFHVYDNPTR
Q8Z714_SALTI  VKQAIKEG.KFWSPE..............LHRFGDLKSPFHYAFYDKKGGSIVVEVE.NGKFHVYDNPTR
Y284L_CHVP1   VKAIAKTL.NVIHES..............YPPFTETP.PMHWLVTDASGKSIVLEPLGNGELTVPDNPMG
Q6ARU2_DESPS  AVDGLRKAPPQIIAP.............LLP..GKVAAGVHLAISDSQGDSAIAEYLNGTLTIHHGKGYL
Q9KL70_VIBCH  AVTAMANPPPTIIAP.............SLP...NGDAAALHLSISDVNGDSAIFEYIDGKLVIHHGSQYK
Q6E7I9_9CYAN  AVTELQQBPPRIIPATAKDILSRKLGLDIPENRSQLEFTLHLSLSDPSSGDSAIFEYINGELKIHHGKEYN
Q983W5_RHILO  AVDALAKQPPTVVTA.............NVPG..EDRLATLHLSLSDASGDSAIIEYIDGKQVIHHSRDYQ
Q8U768_AGRT5  AVRVLEKNPPIIVTD.............SVPG..EERLATLHLSLSDASGDSAIVEYIDGKQVIHHGRQYQ
Q6FAR4_ACIAD  AVKALQKBPPIIVSD.............AVPG..EQRLTTLHLSISDATGDSAIIEYINGKQVIHHNPDYQ
Q6D291_ERWCT  AVKSLQQBKPILVTK.............QVEG..QKRLATLHLSLSDSSGDSAIIEYIDGKQVIHHSKNYQ
Q7NNP2_GLOVI  AVEDLRKBPPYVVPT.............MTP...DGKPGQLHLSISDSTGDSAIFQYIGGKLNIHHGRQFQ
Q5N0N9_SYNP6  AVNALRSBPPVVVSA.............NIPG..TDRFATLHLSVSDAKGDNAIFEYIDGKLVIHHSMDYT
Q8A8A8_BACTN  AVAYIDEDTPQVVSD.............MMPD..GSRLATLHLSISDATGDCAIFEYIGGKLTVYHSKEYK
Q8A600_BACTN  AVDEMKKETPRIDAP.............RMPN..GGPESTLHMAITDETGNTAVIEYLDGKLSIHHGKEYQ
Q64PP5_BACFR  AVEELKKQTPRIDAP.............DMPN..G.SASTLHMAITDETGNSAVLEYLDGNLIIHEGKEYQ
Q6MQ22_BDEBA  AVSSMEKSPIQVLMA.............SVPTAHGVRQGTVHLALSDKTGDSVVMEYIDGKAKFYHGKEYR
Q8TSP0_METAC  AVEGIRSDPVTVIAA.............NAPN...GRPVSVHFSLSDATGDSAVLEYIGGKLQIHHDREYR
P72765_SYNY3  ALALMNEIQPVMVAL.............DDMKATVHLAIEDASGDSAILEYVEGKLVHHGPEYR
Q7TU08_RHOBA  ALERMKDVQPIMVEH.............AGFKSSLHLAIEDASGDSAIIEYVDGQPKIYHGSEHQ
Q5X6U9_LEGPA  VIDDSSNYQLVADTY.............RGITMNVHLMVSDAEGKSAVFEYLIHTQDNLT
Q5ZXC8_LEGPH  VIDDSSNYQLVADTY.............RGITMNVHLMVSDAEGKSAVFEYLDGKLVIHTQDNLT
Q5WYA2_LEGPL  VIDDSSNYQLVADTY.............RGITMNVHLMVSDAEGKSAVFEYLDGKLVIHTQDNLT
Q5NFV9_FRATT  AVEFIKKHKPYTSVL.............PEISDKVEITVHFAISDAQGNSVVVEFING.KTKIYDNKIK
Q82ZP5_ENTFA  VKVAFEKIGLLDQKFQ............LDSLGEVHFTLHWTINDKNNNSIVLQPTDNGAFVIYDS.IG
```

```
                 β12           α3       η1          β13   β14         η2              α4
                 ─►          ΩΩΩΩΩΩΩΩΩ  ΩΩΩ         ───► TT....───►   ΩΩΩ            ΩΩΩΩΩΩΩΩΩΩ
                 170           180       190         200         210           220           230
                 .             .         .           .           .             .             .
Q9KK62_BIFLO  VLTNQ.PTFDFHMENLRNYMCVSNEMAEPTSWGKA....SLTAWGAGVGMHGIPGDVSSPSRFVRVAYTN
Q6J263_9BIFI  VLTNQ.PTFDFHMENLRNYMCVSNEMAEPTSWGKA....ELSAWGAGVSMHGIPGDVSSPSRFVRVAYTN
Q74LX7_LACJO  VLTNS.PDFNWHLTNLGNYTGLNPHDATAQSWNGQ....KVAPWGVGTGSLGLPGDSIPADRFVKAAYLN
Q9AHJ7_9LACO  ALTNA.PLFPEQLTNLANYASVVPGEPDNNFLPGV....NLKLYSRSLGTHHLPGGMDSESRFVKVCFAL
Q74IV4_LACJO  VLTNN.PEFPGQLRNLANYSNIAPAQPKNTLVPGV....DLNLYSRGLGTHPLPGGMDSASRFVKIAFVR
Q75N78_ENTFC  VLTNN.PSFDYQLFNLNNYRVLSSETPKNNFSNQI....SLNAYSRGMGGIGLPGDLSSVSRFVKATFTK
Q8KUA7_ENTFA  VLTNN.PTFDYQLFNLNNYRVLSTRTPKNNFSDQI....ELDIYSRGMGGIGLPGDLSSVSRFVKATFTK
Q8Y5J3_LISMO  VLTNN.PTFDYQLFNLNNYRVLSSETPENNFSKEI....DLDAYSRGMGGIGLPGDLSSMSRFVKATFTK
CBH_LACPL     VLTNN.PNFDYQLFNLNNYRALSNSTPQNSFSEKV....DLDSYSRGMGGIGLPGDLSSVSRFVRAAFTK
CBH_CLOPE     VLTNS.PTFDWHVANLNQYVGLRYNQVPEPKLGDQ....SLTALGQGTGLVGLPGDFTPASRFIRVAFLR
Q8Y9S7_LISMO  VMTNT.PRIEWHIENLRNYTGLQATQLAPVKFGEY....MAKPFSQGTGTSKLPGGYTPPERFVRAAYLK
Q65D34_BACLD  VLTNS.PDFPWHLTNLRNFIGLQPGQFAAKKMGGL....TLSAFGQGSGLSGLPGDFTPPSRFVRAAFLK
Q732V9_BACC1  VMTNS.PEFNWHLQNLRQYIGLKSQPFAPTEWSNL....PLSAFGQGSGSMGLPGDFTPPSRFVRAAYGK
Q81WS0_BACAN  VMTNS.PEFNWHLQNLRQYIGLKSQPFAPTEWSNL....PLSAFGQGSGSMGLPGDFTPPSRFVRAAYGK
Q6HF48_BACHK  VMTNS.PEFNWHLQNLRQYIGLKSQPFAPTEWSNL....PLSAFGQGSGSMGLPGDFTPPSRFVRAAYGK
Q81H11_BACCR  VMTNS.PEFSWHLQNLRQYIGLKSQPFAPTEWGDV....PLSAFGQGSGSMGLPGDFTPPSRFVRAAYGK
Q9CEK2_LACLA  VMTNS.PDYQWHLTNLRNYLSFTPNQKESIELLGK....TLKPFSQGSGTPGFPGDFTPPSRFVRTAYLK
P12256_BACSH  VMTNS.PGYEWHQTNLRAYIGVTPNPPQDIMMGDL....DLTPFGQGAGGLGLPGDFTPSARFLRVAYWK
Q890F5_LACPL  VLTNS.PNLKWQLQNLSKYGTLTNTERPLNKFHY....QPGSQGPGTGALGLPGDYTSMSRFARTVFLK
Q6AP13_DESPS  VLTNS.PPFDWQVTNLSNYVNLNVNNVPPVDLSGL....KITNYGQGSGMHGLPGDFTPPSRFVRAVVFS
Q8YIA2_BRUME  VMTNS.PSFDWHMTNLRNYVYLSRENPKPLQILGE....TIQSFGQGAGMHGIPGDTTPPSRFVRASAYV
Q8ZPG8_SALTY  VMTNG.PAFPWHLTNLNNYTQLTNVDRSSGTLGGI....KVMQPDSGIAIADLPSSDTSVSRFIRGVYYT
Q8Z714_SALTI  VMTNG.PAFPWHLTNLNNYTQLTNVDRSSGTLGGI....KVMQPDSGIAIADLPSSDTSVSRFIRGVYYT
Y284L_CHVP1   IFTNA.PTFKWHMESAKKALEHLSPISDPNAAS.........QGTGALGLPGDFSSASRFIRLAFFS
Q6ARU2_DESPS  VMTNS.PSFDKQLALNAYWQELG..................GITMLPGTNRASDRFARASFYT
Q9KL70_VIBCH  VMTNS.PTFDQQLALNTYWQQIG.................GSKPLPGTINAADRFVRASYAL
Q6E7I9_9CYAN  VLTND.PTFDEQLAINNNWNQLND..............VVGGSFPVPGSNHPSSRFIRASYYL
Q983W5_RHILO  VLTNS.PIFEKQLALNEYWKQIG...............GTIMLPGTNRASDRFARASFYI
Q8U768_AGRT5  VMTNS.PTFDEQLALNAYWTQIG...............GTVMLPGTNRASDRFVRASFYA
Q6FAR4_ACIAD  VMTNS.PIFEEQLALNEYWKKIG...............GTTFLPGTNRAADRFARASFYI
Q6D291_ERWCT  VMTNS.PTFDQQLTNLAYWDQIG...............GNVMLPGTNRAADRFVRASFYV
Q7NNP2_GLOVI  VMTNS.PIYDQQLALNEYWKEIG...............GTTMLPGTNRASDRFVRASFYI
Q5N0N9_SYNP6  VMTNS.PTFEKQLALDGYWQQVG...............GLNMLPGTNRAADRFVRASFYI
Q8A8A8_BACTN  VMTNS.PTYNKQLALSEYWKSIG...............GLSFLPGTNRAADRFARASFYI
Q8A600_BACTN  VMTNS.PRYELQLAVNDYWKEVG...............GLQMLPGTNRSSDRFVRASFYI
Q64PP5_BACFR  VMTNS.PRYDLQLAVNDYWKEVG...............GLNMLPGTNRSSDRFVRASFYI
Q6MQ22_BDEBA  FMTNS.PPFDQQLKSMKQYQGFG...............GSKKLPGTTBAADRFVRAAYYT
Q8TSP0_METAC  VMTNS.PIYDQQLAINAYWDLIG...............GKNPLPGTISAADRFVRASYAL
P72765_SYNY3  VMTND.PTYDQQLKFLETWDFTN................ATRQTPLPGNVDPKDRFVRASYYQ
Q7TU08_RHOBA  VMTND.PPYDEQLIELEKWDFAN................ATRQTPLPGNVNPIDRFVRANYPL
Q5X6U9_LEGPA  TPVLTNTDYKLSLALLEEYQDHG...............GSKALPGGYDSKSRFVRAAHYL
Q5ZXC8_LEGPH  TPVLTNTDYKLSLALLEEYQDHG...............GSKALPGGYDSKSRFVRAAHYL
Q5WYA2_LEGPL  MPVLTNTDYKSSLALLEEYKDHG...............GSKALPGGYDSNSRFVRAAHYL
Q5NFV9_FRATT  VLTND.PAYDEQLKLLKKYQDREIK..............EDKFIAFDYSPQGRFAKMVAFN
Q82ZP5_ENTFA  VVTNS.PEYNYHLTNARNYIGMRNYAIKEPYTLKSGATLDPIEGGTSYGLLGIPGDPTSPSRFIRALYYS
```

```
                      ℓ......ℓ          αααααααααααα            TT    TT                              β15
Q9KK62_BIFLO                            240       250              260             270               280

Q9KK62_BIFLO    A......HYP......QQNDEAANVSRLFHTLGSVQMVDGMAKMGDGQ...........PERTLPTSGYS
Q6J263_9BIFI    T......HYP......QQNNEAANVSRLFHTLVSVQMVDGMSKMGNGQ...........PERTLPTSGYS
Q74LX7_LACJO    V......NYP......TAKGEKANVAKFPHILKSVAMIKGSVVNDQGK...........DEYTVYTACYS
Q9AHJ7_9LACO    N......HAP.....KDSDEVENVTNFPHILESVEQAKGMDQIGPNS...........FEYTMYTSCMN
Q74IV4_LACJO    A......HSP.....QGNNELSSVTNYPHILHSVEQPKGTDBVGPNS...........YEYTIYSDGTN
Q75N78_ENTFC    L......NSV.....SGDSESESISQFFHILGSVEQAKGLCDVGDGK...........YEYTIYSSCCN
Q8KUA7_ENTFA    L......NSV.....SRSSEYESISQFFHILSSVEQQKGLCDVGDEK...........YEYTIYSSCCN
Q8Y5J3_LISMO    L......NSV.....SGDSESESISQFFHILGSVEQQKGLCDVGGGK...........YEHTIYSSCCN
CBH_LACPL       L......NSL.....SMQTESGSVSQFFHILGSVEQQKGLCEVTDGK...........YEYTIYSSCCD
CBH_CLOPE       D......AMI.....KNDKDSIDLIEFFHILNNVAMVRGSTRTVEEK...........SDLTQYTSCMC
Q8Y9S7_LISMO    E......NII....KAKNEEBAITNVWYVLNSVRIPNGAVIKESGD...........PDPTQYVASMC
Q65D34_BACLD    E......HMK.....PVSDETBGVSAAFHVLSNMNIPKGVVMTDQGE...........DDYTQYTAVMC
Q732V9_BACC1    Q......NIQ.....GIDSEEBGVSALPHILSNCEVPKGGVITEEGA...........LDNTIYTSVMC
Q81WS0_BACAN    Q......NIQ.....GIDSEEBGVSALPHILSNCEVPKGGVITEEGA...........LDNTIYTSVMC
Q6HF48_BACHK    Q......NIQ.....GIDNEEBGVSALPHILSNCEIPKGGVITEEGT...........LDNTIYTSVMC
Q81H11_BACCR    Q......NIQ.....GIENEEBGISAIPHILSNCEVPKGGAVITEDGI..........LDNTIYTSAMC
Q9CEK2_LACLA    N......YAE.....KPSNELAAITLCHHILESVIIPKGIVITEHGA...........SDFTCYSAYMC
P12256_BACSH    K......YTE.....KAKNETBGVTNLPHILSSVNIPKGVVLTNEGK...........TDYTIYTSAMC
Q890F5_LACPL    H......YAQ.....VPATTTDTVNLLQHILNAVTIPKGAKVAANGQ...........ATYTEYRSYMD
Q6AP13_DESPS    E......TAV.....KLPTTEKTVPQVFHIMNMPDIPVGAMREVLGKKV.....VHDYTNWTSVSD
Q8YIA2_BRUME    L......SAK.....KVPSGLESVRLAEHIANNPDIPKGWSEEQN.........MPEYTQWTAFAD
Q8ZPG8_SALTY    T......YAP.....QATSAHDAMNTLAHIMSRPDRPKNITVDYMGSEGEGNATRKPVSEYTVWTTLSD
Q8Z714_SALTI    T......YAP.....QATSAHDAMNTLAHIMSRPDRPKNITVDYMGSEGEGNATRKPVSEYTVWTTLSD
Y294L_CHVP1     Q......TIE.....IPRTSAGAVNTLFHVLNNPDIPKGVVASINMNTG......KHVYEKTIYTVIYN
Q6ARU2_DESPS    K......STP.....AFKDERMAVGAIFSIIRNASVPLGFTDPKQPN...........IATTLWRSVSD
Q9KL70_VIBCH    S......ASP.....KYRDNDLALASVFSQIRAVSVPLGITDPDRPN...........LAMTLWRTFAD
Q6E7I9_9CYAN    G......QMNNACLSKLLKNRRNELAAALGIIRNVSDPLGLSG...LA.........ISTTQWRTVSD
Q983W5_RHILO    N......AIP.....KSEKPDEAVASVFSVIRNTSVPYGIATPDQPN...........ISSTRWRTVSD
Q8U768_AGRT5    N......AIP.....KSENPVEAIASVFSVIRNVSVPYGITTPDQPN...........ISSTRWRTVID
Q6FAR4_ACIAD    N......AIP.....KDADSKATLASVFSVIRNVSVPYGLNTEEEPN...........ISSTRWRAVVD
Q6D291_ERWCT    KNVNPNKLIPGVAEKGKIEKDKADLATAFSIIRNASVPYGYSLPDMPN.........IASTRWRTVVD
Q7NNP2_GLOVI    K......AIP.....QTANNTEAVASAFSVIRNVSVPLGFSTPGQPN...........ISSTIWRTVAD
Q5N0N8_SYNP6    D......AIP.....KTTDTRVAVASVFSVIRNASVPFGISSEEEPN...........ISSTRWRSLSD
Q8A8A8_BACTN    N......ALP.....ETDDERIAVASVFSVVRNASVPYGISTPESPE...........ISTTQWRTVSE
Q8A600_BACTN    H......AIP.....QTADAKIAVPSVLSVMRNVSVPFGINTPEKPH...........ISSTRWRSVSD
Q64PP5_BACFR    H......AIP.....QTSDAKIAVPSVLSVMRNVSVPFGITTPDKPY...........ISSTRWRSVSD
Q6MQ22_BDEBA    Q......HLP.....DPKDYREAVASVLRNVSQPFGTPDPARPY...........ISTTRWRTVAD
Q8TSP0_METAC    K......SSP.....RFKDRRAAVAAVLSQMRSIGVPPGRAIP...........TSLTSRRRCGE
P72765_SYNY3    M......MLP.....EPKTEQBAIARNTSVPFGAPNNIPGS...........LYNTEYRTAID
Q7TU08_RHOBA    K......TLR.....KPEGERBAIATTLSLARNASVPFNAPYKTPGT...........IYDTEYRTVVN
Q5X6U9_LEGPA    KKLP..........SFIAKEBHIAYAFNGLNVVTQAPGTSSP...........TQLSMVFD
Q5ZXC8_LEGPH    KKLP..........SFIAKEBHIAYAFNGLNVVTQAPGTSSP...........TQLSMVFD
Q5WYA2_LEGPL    KKLP..........SFIAKEBHIAYAFNGLNVVTQAPGTSSP...........TQLSMVFD
Q5NFV9_FRATT    TTKAN.........TPDDLTAVNHAWSMINTVDIPQGSLYWRWVSSK.........PQFTSYSVVDD
Q82ZP5_ENTFA    DN...........LQEFDSSEGIMQLYRAFQTVMIPRGIGHLGQSNS.........LSDPTHYWSGYD


                      β16             β17  η3     TT        β18
Q9KK62_BIFLO          290           300         310

Q9KK62_BIFLO    SKTNTYYMNTYDDP.AIRSYAMADYDMD.SSELISVAR.................................
Q6J263_9BIFI    GKTNTYYMNTYEDP.AIRSFAMSDFDMD.SSELITAD.................................
Q74LX7_LACJO    SGSKTYYCNFEDDF.ELKTYKLDDHTMN.STSLVTY.................................
Q9AHJ7_9LACO    LEKGILYFNCYDDS.RISAVDMNKEDLD.SSDLVVYDLFKKQDISFIN.....................
Q74IV4_LACJO    LETGTFYYTNYENN.QINAIELNKENLN.GDELTDYKLIEKQTINYQN.....................
Q75N78_ENTFC    VDKGIYYYRTYEDS.QITAIDMNKEDLD.SHKLISYPIYRKTTN........................
Q8KUA7_ENTFA    LEKGIYYYRTYDNS.QITAVDMNKENLE.KDSLIVYPMVETQQINYAN.....................
Q8Y5J3_LISMO    IDKGIYYYRTYGNS.QITGVDMHQEDLE.SKELAIYPLVNEQRLNIVNK....................
CBH_LACPL       MDKGVYYYNTYENN.QINSVSLNHEHLD.TTELISYPLRSEAQYYAVN.....................
CBH_CLOPE       LEKGIYYYNTYENN.QINAIDMNKENLD.GNEIKTYKYNKTLSINHVN.....................
Q8Y9S7_LISMO    SESKTYYFTSYANS.QINSVTLSKEVIENMKEPTTYQVNKLV...........................
Q65D34_BACLD    SETCSYYFHHYQNR.QIQKVCLLQEDLD.SNELKAFQPKQEETVHSLQS....................
Q732V9_BACC1    MESGTYYYHTYDCR.QIIAIHLFHENLD.TDEIKAYPFQRKQKIFYEN.....................
Q81WS0_BACAN    MESGTYYYHTYDCR.QIIAVHLFHENLD.TDEIKAYPFQRKQKIFYEN.....................
Q6HF48_BACHK    MESGTYYYHTYDCR.QIIAIHLFHENLD.TDEIKAYPFQRKQKIFYEN.....................
Q81H11_BACCR    MESGTYYYHTYDCR.QIIAIHLFNEDLD.TAEIKTYPFQRKQKIFYEN.....................
Q9CEK2_LACLA    SETLSYYFSTYGNQ.RIRKISLSESLKN.EKEPKNFPIVNEEDILELN.....................
P12256_BACSH    AQSKNYYFKLYDNS.RISAVSLMAENLN.SQDLITFEWDRKQDIKQLNQVNVMS...............
Q890F5_LACPL    LNHQTYALELYENPGVIQQVNLTDHLLEKQTVPLEYVLSRTPHVQLLTPDVATLPAAH..........
Q6AP13_DESPS    LKHLTWAYKTYDDQ.SIRSVDVRKACGAAGNEIKTIKMHSVQPIEDVSTKFN................
Q8YIA2_BRUME    MKNDVYYIKTYDDQ.VLRSFSF.KDFDVDSKDILTIKFEPKLDAPSLKK....................
Q8ZPG8_SALTY    LMHGEMMVRGYNDI.NYKTWSLSQFKNATAPVFEKINVKG.............................
Q8Z714_SALTI    LTHGEMMVRGYNDI.NYKTWSLSQFKNATAPVFEKINVKG.............................
Y294L_CHVP1     IKSKEIVFKHYNDQ.NIQKL.................................................
Q6ARU2_DESPS    QKSLTYYFESTISP.NIFWVDLHKLDPN.STAQVKELNLNAQ..PIFAGEVSAKFVPAKPFKWLAPGK..
Q9KL70_VIBCH    HTAKIYYFESAVFP.AVSWLDMSKVDLT.EGAAPKVVRVERG..QPLAGELSAALKPAEPFKWLGAE...
Q6E7I9_9CYAN    QTNLTYFFEPTKSP.NVFWVEMSKLNLE.EGAPVMKLENVDS..LSLVGEVSGEFQKAEPFSWDLSTVCP
Q983W5_RHILO    HKRKLYFFESALTP.NTFWMDLNKFDFSEKTGKVLMLDLGPEQTHTYSGLANDKFEKTKPFKFLGL....
Q8U768_AGRT5    HKRKLYFFESALTP.NIFWIDMTKLDLSKETGAVKKLDLGANQIHIYSGMANESLKDFKPFKFLGL....
Q6FAR4_ACIAD    HKRKLYFFESALSP.NIFWVDLKQINF..KDGITRKLDLGKNQENIYSGDATAKFKPSTPFHFLGIDD..
Q6D291_ERWCT    HKSLQYFFESAVSP.NIFWVDLKKINPAPRGGSAAKLDLGPNQSTIYSGQASGHPKPAQPFEFAGL....
Q7NNP2_GLOVI    QKNKRYFFESARSP.NVFWVNLTDLDFA.AGKPTKKLTLTNG..TTFAGNTAGQFQPAEPFKFLPAEVK.
Q5N0N8_SYNP6    QKNLVYYFETTLTP.NTFWVDLKKVDFS.ENRPVKRLLASNN..EVYAGDALNSFQNAKPFEFAGL....
Q8A8A8_BACTN    SKNLRYYFESTLTP.NTFWVNLKDFDLS.EGAPVFKLSIANG..EMYHGNTAKNFKTALPFKFMGVKG..
Q8A600_BACTN    QKNKVYYFESTLTP.NLFWLDLKKIDFS.PKAGVKKLSLTKG..EIYAGDAVKDLKDSQSFTFLPETPVM
Q64PP5_BACFR    QKNRVYYFESTLTP.NLFWIDLHKVDFS.PNASIKKLSLTHG..EVYAGDVVKDPKDSRSFTFMPELPK.
Q6MQ22_BDEBA    LTRGLYFYESTLSP.YLVWVELPKLDFK.KGTAVKKINLVKN..YELIGDVTGKFKTTPVFQFLKPDDQG
Q8TSP0_METAC    PLSTTMQNVTTSTP.......................................................
P72765_SYNY3    LTNRRYFFELTTSP.NVIWVNLDQLNLA.PGAPVLSLDPDNL...DLSGNVTDKFTKVLKSPF......
Q7TU08_RHOBA    LDSRRFFFELTTSP.NVIWVNLEDLALS.PGSPVLTLNPDNI...HLSGNVNKDPTPVAEPPF......
Q5X6U9_LEGPA    IPTKTIYFRSINEA.ALRVIPMDSIDFD.....YLYQPIALNVYQHLSGDVTDKFQPMIG.........
Q5ZXC8_LEGPH    IPTKTIYFRSINEA.ALRVIPMDSIDFD.....YLYQPIALNVYQHLSGDVTDKFQPMIG.........
Q5WYA2_LEGPL    IPTKTIYFRSINEA.ALRVISMDGIDFD.....YLYQPIALNVYQHLSGDVTDKFQPMIG.........
Q5NFV9_FRATT    LSNRIYYFRTWDNY.DIRKVDLSKIDPAKAKYQADTIFGQADYQEFKFK....................
Q82ZP5_ENTFA    VTNLTMYVQPESTT.SFTKYTLDPALTEVTTPAVSNELLLTDLNQ........................
```
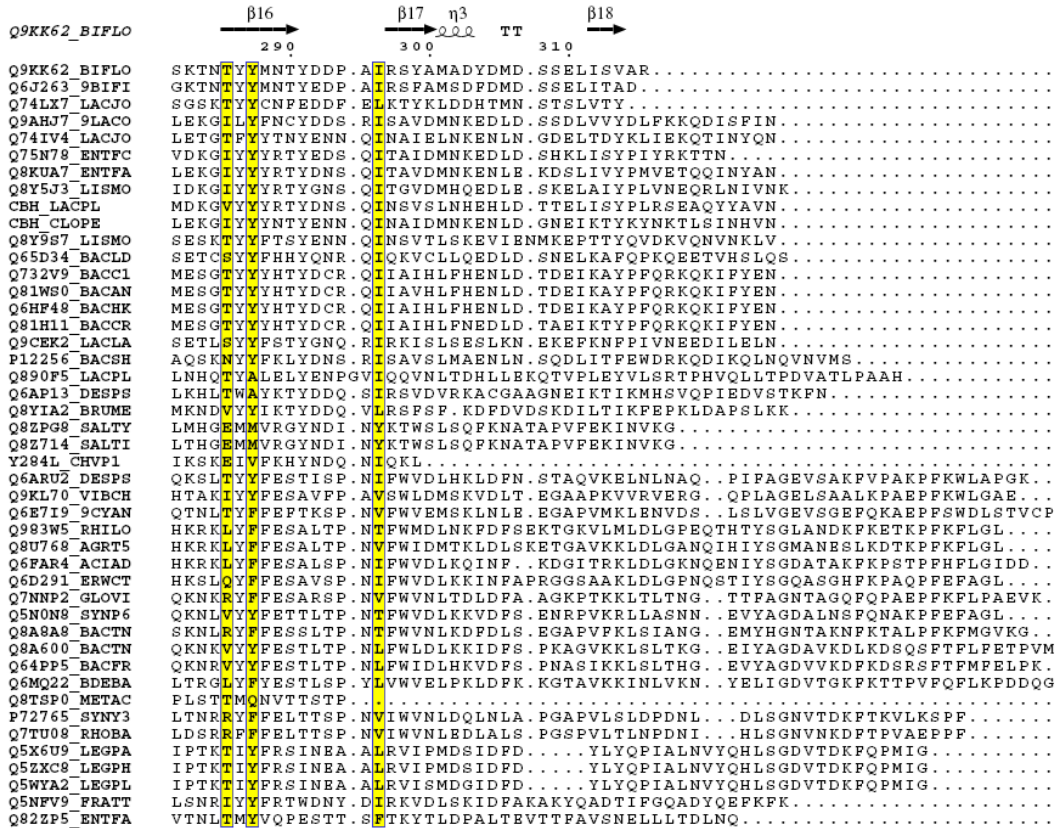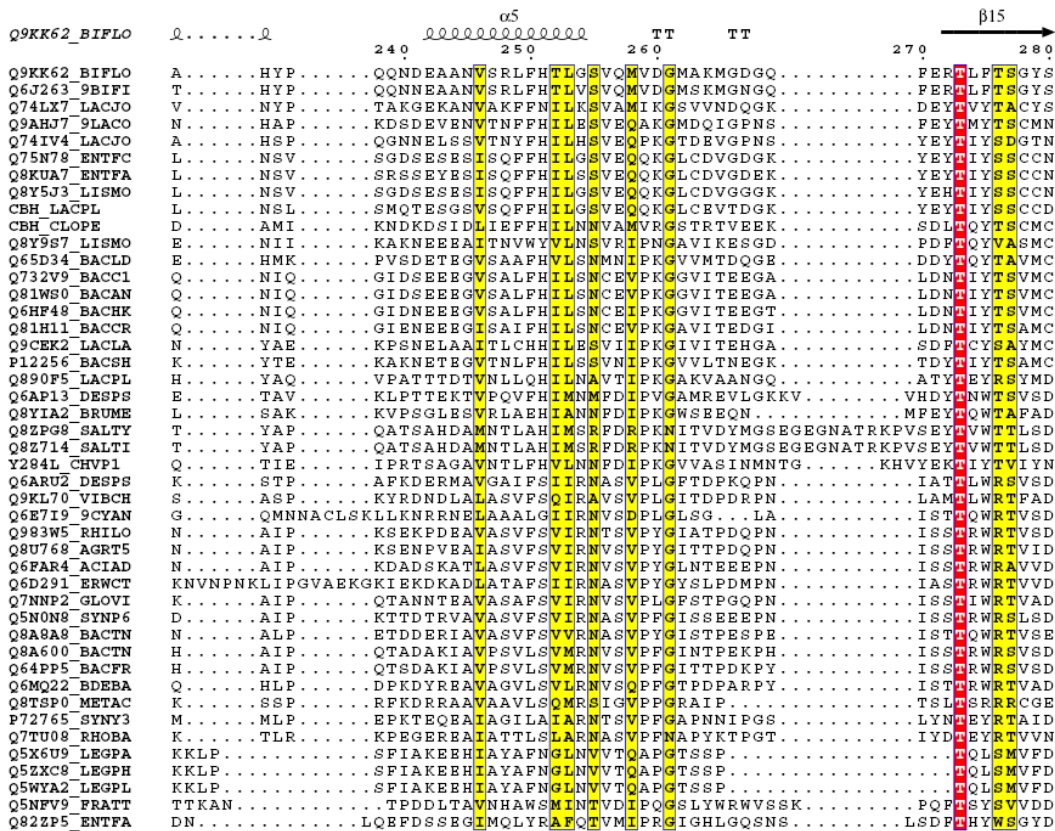
Figure 3.13: Sequence alignment of available BSH and PVA

131

3.8.5. Comparison of Active site geometry

The residues reported in the binding pocket of *Cp*BSH are Cys1, Arg17, Met19, Phe25, Phe60, Ala67, Ile132, Ile136 and Lys41 (Rossocha et al., 2005). Among these the residues Cys1, Arg17 and IIe132 are strictly conserved in *Bl*BSH. The other residues are replaced by residues Leu19, Tyr25, Met60, Phe67, Gln136 and Phe42, respectively. In *Cp*BSH there are movements of two side chains, those of Trp143 and Phe60, upon substrate binding. Residue Trp143 is conserved (Trp142 in *Bl*BSH), while Met60 replaces Phe60 in *Bl*BSH. Two residues (Arg 17 and Asp 20) are also highly conserved in most of the PVA/U34 proteins.

In the structure of 3pva, Arg 17 makes hydrogen bonds with atoms of the opposite -sheet, two with the main chain carbonyl groups (Tyr 68 and Met 80) and one with the side chain Asp 69 (Suresh et al. 1999). Thus Arg 17 could be important for maintaining the overall stability of the structure. Moreover, one side-chain nitrogen atom of Arg 17 is only 3.8 Å away from the catalytic sulfhydryl group, suggesting that Arg 17 could also be involved in catalysis. The position corresponding to Arg 17 is usually occupied by a positively charged residue in PVA/U34 homologs (Pei and Grishin, 2003). The side chain of Asp 20 makes a hydrogen bond with the free backbone amino group of the catalytic cysteine. This interaction is critical for maintaining the orientation of the cysteine residue for catalysis. Conservation of Arg 17 and Asp 20 is unique in the PVA/U34 family, compared with the other Ntn-hydrolases.

Another important component of the catalytic machinery is the oxyanion hole, which stabilizes the negative charges developing on the substrate.The oxyanion hole is normally formed from two residues. The distances remain similar between the nucleophile and the residues that form the oxyanion hole in most Ntn hydrolase structures, approximately 4.5–5.5 Å. In PVA it is the main chain amide nitrogen of Tyr82 and side chain atom Asn125N [2]. Only in proteasome this distance is larger, around 6.1 Å (Lowe et al., 1995). One of the two oxyanion hole formers is located in a structurally and topologically equivalent place, near the end of the 11-strand and in most of the Ntn-hydolase this oxyanion hole is formed from main chain N of gly41. PGA is the exception to this rule since it contains the main chain N of Ala69 in the same location in the residues of the oxyanion hole (Done et al., 1998). The position of the second residue of

oxyanion hole is different. In AGA and Grpp this residue is the glycine residue preceding the β11-strand. In PGA it is formed by Asn241 which is located in the β3-β3-loop. In proteasome the possible candidate for the second oxyanion hole residue is 131SerOγ, originating from the same the β3-β3-loop.

A complex of *Bl*BSH with bile salt would have revealed the interaction between the enzyme and the substrate. Unfortunately, no mutant yielded good diffraction quality crystals even after repeated attempts. Attempts to obtain complexes in crystals of *Bl*BSH with potential inhibitors or reaction products, by soaking crystals in solutions containing substrates or co-crystallization with inhibitors or substrate, none were successful.

The absence of a Met residue at the N-terminal end could be explained by the fact that the Cys1 residue becomes a catalytic center after removal of the initiation formyl methionine by an autoproteolytic process, which is one of the common features of the Ntn hydrolase superfamily (Brannigan et al., 1995). The importance of SH group of cysteine for catalysis in BSH and PVA is confirmed by the fact that replacing Cys with other two potential nucleophilic residues Ser or Thr possessing OH groups only inactivates the enzyme.

CHAPTER 4

ACTIVE SITE CHARACTERIZATION AND
SUBSTRATE SPECIFICITY OF *Bl*BSH

4.1. Introduction

Chemical modification experiments were conducted on BSH enzyme by chemically modifying selected type of amino acid residues and measuring the change in enzyme activity to assess the specific role of these residues in catalysis and substrate binding. Two sets of molecules, the true substrates and substrate analogs, were tested for activity. The substrates, the bile salts, showed variations in activity depending on the type of conjugated amino acid moiety (glycine or taurine) present and also depended on the position of the –OH group in the cholesterol ring. The second set of compounds was basically the inhibitors such as penV. The inhibition of bile salt hydrolysis by these compounds was confirmed by inhibition kinetics. The effective inhibition was measured using inhibition constant K_i. Most of the inhibitors showed competitive inhibition pointing to the possibility that they bind at the same site as substrates. Since cysteine was suspected involved in activity cysteine-specific chemical modifiers were used and found complete loss of activity. The cysteine modification was hampered in the presence of substrates and inhibitors that again indicated that inhibitors, like substrates, were binding at the active site and blocking it. Further studies were conducted on a representative compound of each type and thermodynamic quantities and energetics of binding were estimated using fluorescence studies and isothermal calorimetry. To explore the molecular interactions involved in binding *insilico* experiments such as docking and molecular dynamics simulations were attempted. Thus, this chapter provides an account of the determination of biochemical and biophysical properties of bile salt hydrolase. The optimum values of parameters such as pH and temperature, the value of kinetic constant and substrate specificity provide insights into the catalytic efficiency of the enzyme.

4.1. 2. Enzymatic properties of BSH

Many BSH producing bacteria are identified and BSH from those bacteria were characterized. Table 4.1 shows that the BSH are active mostly in acidic environment and they themselves are acidic in nature. Four isozymes of BSH have been purified from the cytosol of cells of *Lactobacillus* sp. strain 100-100. Agents that oxidize thiol groups (e.g., *p*-mercuribenzoate, iodoacetamide, Hg^{2+}, Cu^{2+}, and Cd^{2+}) have been shown to strongly inhibit BSH activity in *C. perfringens* (Gopal-Srivastava et al., 1988; Hirano et al., 1981), *L. reuteri* (Elkins & Savage, 1998), and *B. longum* (Grill et al., 1997). The most of the reports are only about the basic characterization of BSH from many sources. There is no comprehensive investigation available till now on the substrate specificity and its evolutionary association with PVA enzyme.

Batta et al., (1984 & 1990) reported the substrate specificity of cholylglycine hydrolase using bile acid conjugates with modifications in the steroid ring system, the side chain, or the amino acid moiety. Epimerization at C-3 and C-7 did not affect the activity of the enzyme while oxidation of the three nuclear hydroxyl groups reduced the affinity of the enzyme toward the substrate. Elongation of the side chain by one or three carbons inhibited enzyme activity. Conjugates prepared from C_{24} bile acids and analogs of taurine and glycine with one or two methylene groups were effectively hydrolyzed, whereas conjugates with a tertiary amide group completely resisted hydrolysis. An investigation by Moser and Savage (2001) revealed that *L. buchneri* JCM1069 expressed taurodeoxycholic acid-hydrolase activity but not taurocholic acid-hydrolase activity. However, the majority of kinetic data available in the literature indicates that substrates are predominantly recognized by the amino acid moieties, and most BSHs are more efficient at hydrolyzing glyco-conjugated bile salts than tauro-conjugated bile salts (Tanaka et al., 2000; Coleman & Hundson et al., 1995)

Sequence and structural comparison of BSH and PVA were reported by Tanaka et al. (2000) and Rossocha et al., (2005). This comparison helped to identify inhibitors and classical enzyme kinetics proved that their binding was in the same site. The binding of the inhibitor is further confirmed through the active site protection studies, isothermal calorimetry and fluorescence spectroscopy.

CA

DCA
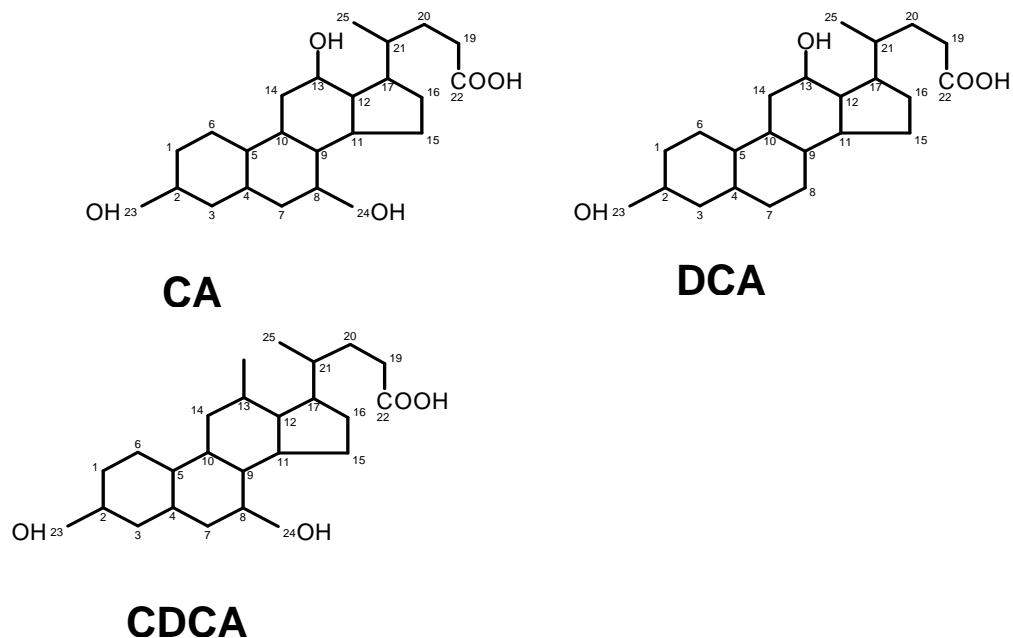
CDCA

Figure 4.1: Major forms of bile acid: CA- Cholic acid, DCA- Deoxycholic acid,
CDCA- Chenodeoxycholic acid

Table 4.1: Properties of BSH enzyme from various sources

S. No	Source	Molecular weight (kDa)	iso-electric point (pI)	Optimum pH	Binding constant (K_m)	Reference
1	*Bacteroides fragilis* subsp. *fragilis*	250	-	5.0-6.0	-	Stellwang et al., 1976
2	*Clostridium perfringens*	140	5.26	5.26	0.5	Gopal-Srivastava et al., 1988
3	*Enterococcus faecium*	-	4.97	4.8	-	Wijaya et al., 2004
4	*Xanthomonas maltophilia*	100	4.1	7.9-8.5	-	Dean et al., 2002
5	*Listeria monocytogenes*	-	4.7	5.02	-	Sue et al., 2004
6	*Lactobacillus*	126	5.29	3.5-4.5	1.1	Corzo &

S. No	Source	Molecular weight (kDa)	iso-electric point (pI)	Optimum pH	Binding constant (K_m)	Reference
	acidophilus					Gilliland, 1999
7	*Lactobacillus plantarum 80*	126	5.12	-	2.4	Christiaens et al., 1992,
8	Bacteroides	140	-	-	-	Kawamoto et al., 1989
9	**Lactobacillus casei**	126	-	6.0	2	Pereira et al., 2003
10	Enterococcus	-	-	4.8	-	Wijaya et al., 2004
11	*Bifidobacterium adolescentis*	126	4.3	6.0	-	Kim et al., 2005
12	*B. animalis*	126	4.1	5.5	0.43	Lepercq et al., 2005
13	*B. bifidum*	140	4.7	6.5	0.25	Kim et al., 2004
14	*B. longum*	140	4.71	6.5	0.16	Tanaka et al., 2000
15	*B. infantis*	140	4.2	6.0	-	Liong et al., 2004
16	*B. breve*	126	4.5	6.0	-	Liong et al., 2004
17	*Lactobacillus reuteri*	80	-	4.5-5.5	29	Taranto et al., 1999; Kanasaki et al., 1975

4.2. Construction of the over-expression Plasmid pET26b/*bsh*

The *bsh* gene was fished out from the Plasmid pBH1351 (Tanaka et al., 2000). This plasmid was originally constructed from the *Bifidobacterium* chromosomal DNA. Plasmid was extracted from the *E. coli* containing pBH1351. The primers overhanging restriction sites were designed to fish out the gene from pBH1351.

Primer A - GGAGTCATTAATGTGCACTGGTGTCCGTTTC

Primer B – GGAAGAATTCATCGGGCGACGCTGATGAG

The gene was amplified using PCR under the following conditions. Denaturation step at 92 ºC for 3 min, annealing at 55 ºC for 30 sec and extension at 72 ºC for 1 min. These steps were repeated for 20 cycles using VENT DNA polymerase. The vector pET-26b (Figure 2.1) was digested with NdeI and EcoRI. The PCR products were digested with AspI and EcoRI. Both reactions were carried under 37 ºC with suitable buffers. The other details are already described in material and methods.

The plasmid DNA from one of the construct that had successfully taken up the insert was used to transform $CaCl_2$ competent BL21 expression host cells. The overnight culture grown from the transformants were checked for the protein expression in SDS-PAGE. (Figure 4.2)
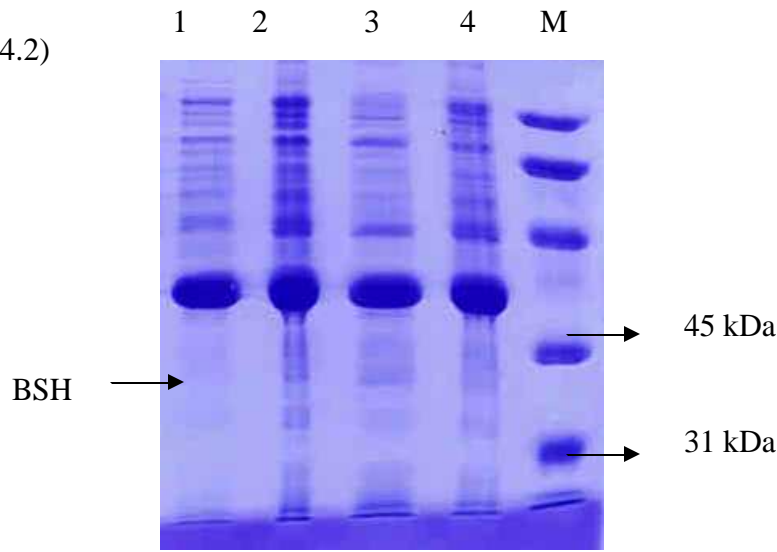
Figure 4.2: Protein expression profile of *E. coli* over expressing *Bl*BSH. M- Markers, Lanes 1-4: total soluble proteins from positive colonies.

4.3. Purification of *Bifidobacterium longum* BSH (*Bl*BSH)

The steps followed for purifying wild type and mutant forms of *Bl*BSH were described under materials methods in chapter 2. The elution profile of the last step using Q-sepharose column is shown in Figure 4.3. In every step 1mM DTT was added to the buffer to prevent oxidation of the enzyme active site cysteine. However, when the concentrated protein was taken through 2 to 3 freeze and thaw cycles the sample showed

protein aggregation and precipitation. Any effort to dissolve back the protein in to the solution was unsuccessful. So to avoid precipitation, the purified enzyme was kept frozen in dilute condition and concentrated only before experimental use. The purity of *Bl*BSH preparation was shown using SDS-PAGE in Figure 4.4. About 15 mg of the purified protein was obtained from 1 liter of the culture.

The mutant protein C1A and T2A were purified using the same procedure and the yield of pure protein was 8 and 12 mg/L, respectively.

<u>Purification chart of *Bl*BSH</u>

Sonication of the Cells

↓

Streptomycin sulphate precipitation

↓

Ammonium Sulphate precipitation

↓

Hydrophobic chromatography (Octyl sepharose)

↓

Anion exchange chromatography (Q- sepharose)

↓

Purified protein

BSH
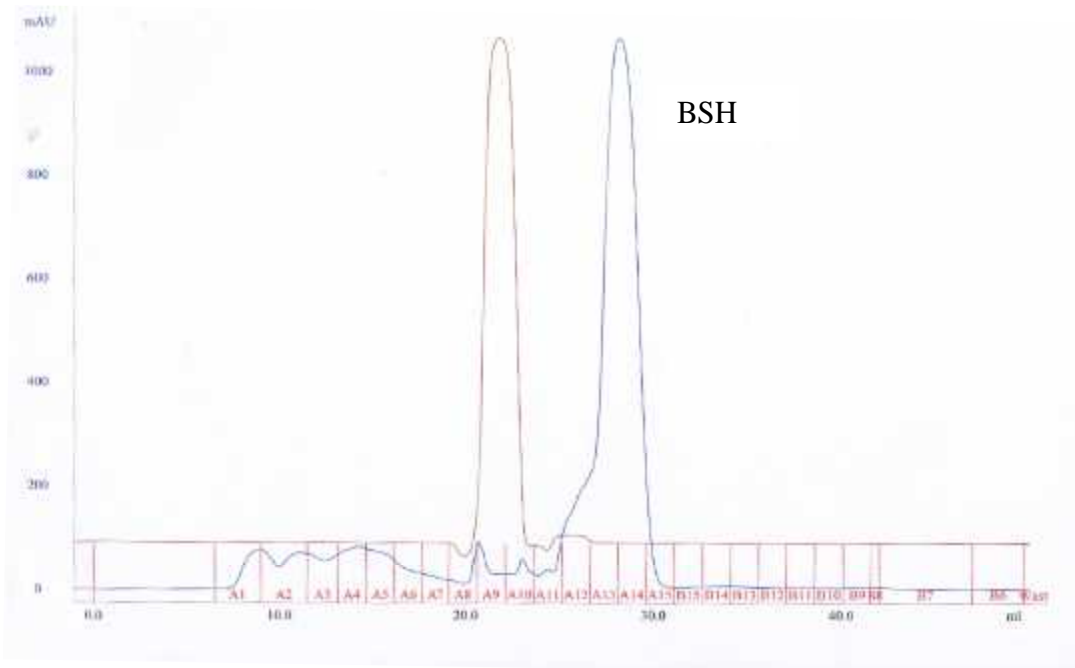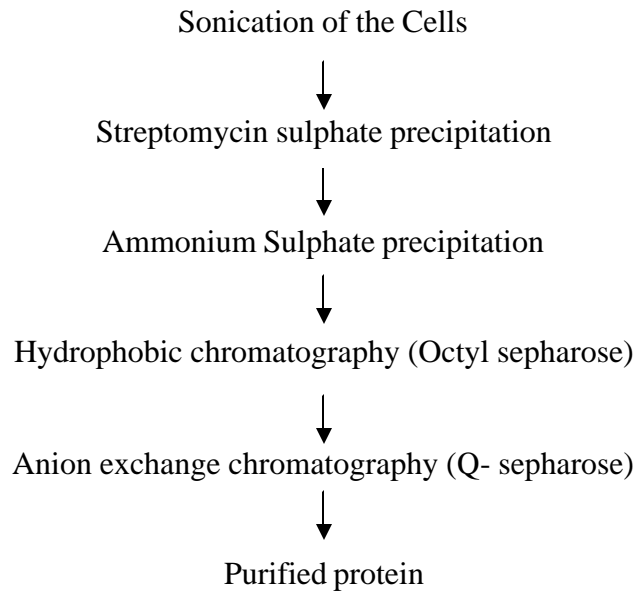
Figure 4.3: FPLC elution profile of *Bl*BSH from Q-sepharose column OD-218 (Blue) protein content (Brown)

Figure 4.4: SDS-PAGE showing purified *Bl*BSH, lane A: molecular weight markers, lane B: purified *Bl*BSH

4.4. Preliminary Biochemical characterization of *Bl*BSH

4.4.1. Molecular weight determination

The molecular weight of native enzyme was determined using gel chromatography with FPLC; the subunit molecular weight was determined by performing four independent SDS-PAGE analyses in which standard proteins of medium molecular weights (Sigma, USA) were used. The molecular weight of native *Bl*BSH was found to be 130,000. The subunit molecular weight determined using SDS-PAGE was 36,500 which indicated that the native enzyme was a tetramer (Figure 4.5).
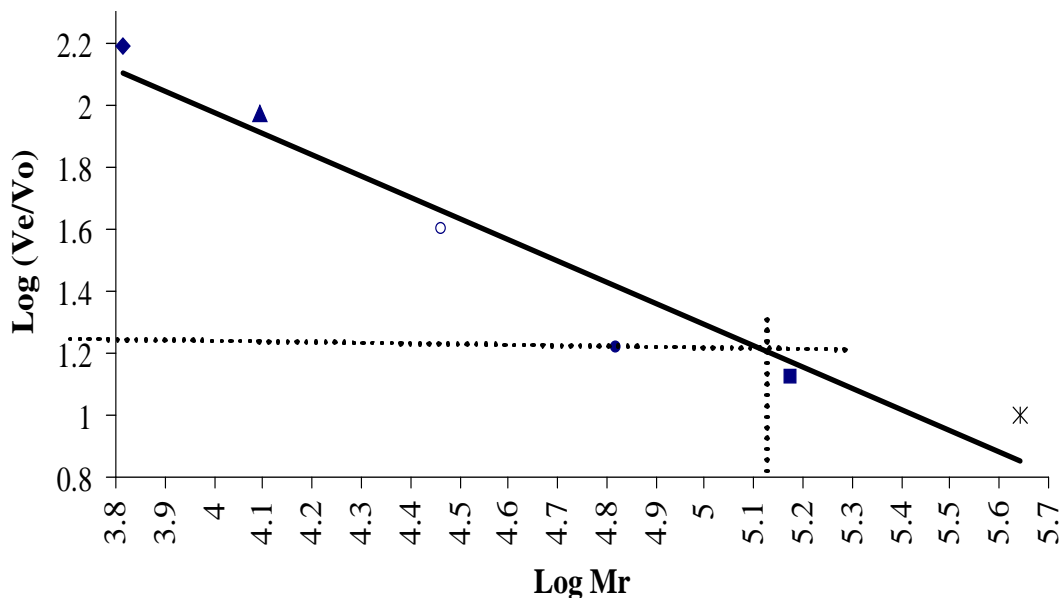
Figure 4.5: Molecular weight determination of *Bl*BSH using sephadex G-200 Gel chromatography. [◆ Aprotinin (6.5 kDa) ^cyctochrome c (12.5 kDa), 0 Carbonic anhydrase (29 kD), BSA (66 kDa), | Alcohol dehydrogenase (150 kDa), Apoferritin (440 kDa)].

4.4.2. pH and temperature corresponding to optimum activity and stability

The activity of purified *Bl*BSH was determined in the pH range 3 to 9 under standard assay conditions (Figure 4.6a). The buffers used were: 100 mM sodium citrate phosphate buffer (pH 3 to 4), the sodium acetate buffer (pH 5 to 6), 100 mM phosphate buffer (pH 7 to 9). 10 mM DTT was present in all the buffers. The enzyme has an optimum pH of 6. The maximum activity was observed at temperature between 40 and 50 ºC (Figure 4.6b). The enzyme was stable in the pH range 4 to 8 and up to 40 ºC for 30 min; keeping at 50 ºC caused 70% loss in activity.
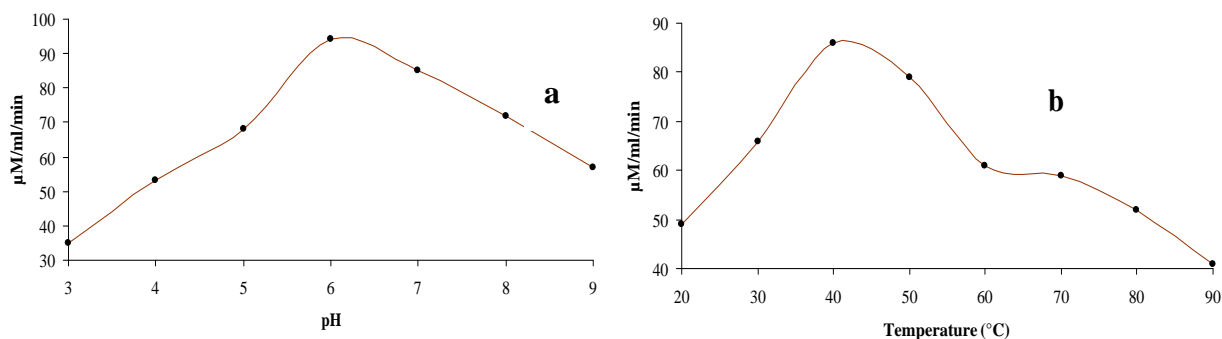
Figure 4.6: Determination of optimum pH (a) and temperature (b) for activity of *Bl*BSH

141

4.4.3. Iso-electric point

In iso-electric focusing experiment the purified protein showed a single distinct band in the gel and no other isoforms were observed. The iso-electric point estimated was 4.1 (Figure 4.7). The theoretical value of pI calculated using amino acid composition by ProtParam tool (http// www.expasy.ch) was 4.71
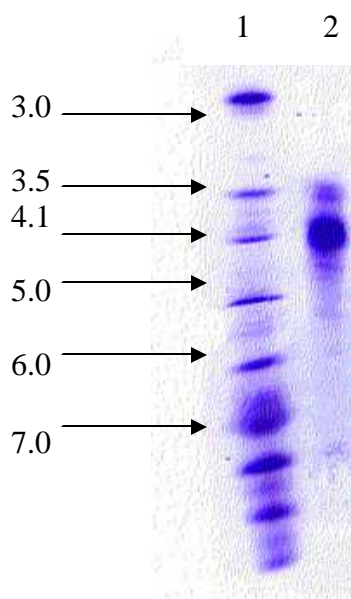
Figure 4.7: Iso-electric point determination of the *Bl*BSH. Lanes: 1, IEF markers;
2, BSH protein

4.4.4. Theoretically calculated parameters of *Bl*BSH

The protein contains 30 valines, about 9.5% of the total amino acid composition. The total number of negatively charged amino acids (Asp + Glu) is 41 which is double the number of positively charged amino acids (Arg + Lys = 20). The Extinction coefficient of *Bl*BSH at 280 is 47770 $M^{-1}cm^{-1}$. (i.e Absorption for 0.1 [1g/L] sample is 1.359) as calculated by Protparam. The instability index (II) computed is 41.11 that classify the enzyme as unstable. The grand average hydropathy index is -0.286.

4.4.5. Michaelis-Menten kinetics of *Bl*BSH

The kinetic studies using different bile salts as substrates of native enzyme and of both the mutants were carried out. The following bile acid salts were used for estimating K_m and V_{max}: Glycocholic acid (GCA); glycodeoxycholic acid (GDCA), Taurucholic acid

(TDCA),Glycochenodeoxycholic Acid (GCDCA), Taurochenodeoxycholic a c i d (TCDCA). The highest levels of activity were observed with GDCA. In general higher affinities were observed for glycine-conjugated bile acids than for taurine-conjugated bile acids (Table 4.2 and Figure 4.8). The kinetic data on native and T2A mutant are shown in Table 4.3.

Table 4.2. Kinetic constant of BlBSH estimated against various bile salts.

Compounds	K_m (mM)
Glyco cholic acid (GCA)	0.22 ± 0.07
Glyco deoxy cholic acid (GDCA)	0.18 ± 0.05
Glyco chenodeoxy cholic acid (GCDCA)	0.28 ± 0.11
Tauro cholic acid (TCA)	0.32 ± 0.13
Tauro deoxy cholic acid (TDCA)	0.49 ± 0.04
Tauro chenodeoxy cholic acid (TCDCA)	0.42 ± 0.15

Table 4.3: Comparison of kinetic constants of native and mutant (T2A) BlBSH

	K_m (μM)	k_{cat} (s^{-1})	k_{cat} / K_m (μM^{-1} s^{-1})
Wild type BSH			
GCA	22 ± 7	85 ± 9	3.8
TCA	32 ± 13	76 ± 9	2.4
T2A mutant			
GCA	24 ± 5	44 ± 3	1.8
TCA	30 ± 0	40 ± 1	1.3
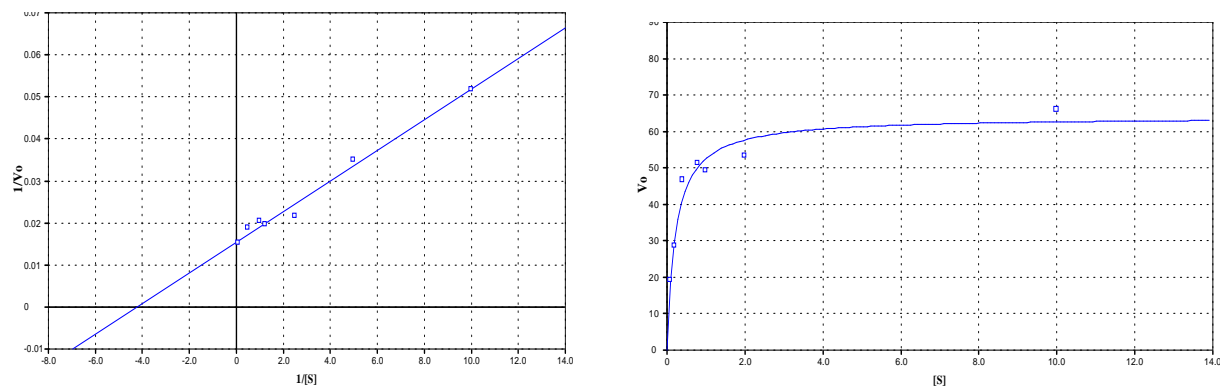
Figure 4.8: Determination of kinetic constant K_m and V_{max} of *Bl*BSH in the presence of GCA. a) LB plot b) simple curve plot.

4.5. Active site characterization

4.5. Chemical modification

The effect of amino acid specific modifying reagents is shown in Table 4.4. The modification of amino acid such as cysteine, arginine, tryptophan and tyrosine showed considerable inactivation of the enzyme indicating the participation of these residues in catalysis or binding and also on their presence at active site or near the active site. These modifications were further characterized in detail.

4.5.1. Modification of Cysteine residue

Reaction with 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB): the *Bl*BSH (4 μM, 1ml) in 100 mM, potassium phosphate buffer, pH 8.0, was incubated with 1mM DTNB at 30ºC for 60 min. Aliquots were removed at definite time intervals and the residual activity was determined. The 80 % of inactivation was achieved in 2 mM DTNB, and the number of residues modified was calculated to be 1. Site directed mutant C1A confirmed the participation of Cys1 in the catalysis. The total inactivation was achieved in the DTNB treated sample only after partial denaturation of the protein.

144

Table 4.4: Effect of amino acid specific modifying reagent on *Bl*BSH

Reagent	Concentration	Possible residue modification	Residual Activity %	Buffer
None	-	-	100	Sodium Phosphate, 50mM, pH 6.5
TNBS	1 mM	Lys	98	Sodium Bicarbonate , 100 mM, pH 8.5
Citronic anhydride	5 mM	Lys	94	Sodium Phosphate 50 mM pH 7.8
DEPC	1 mM	His	94	Sodium Phosphate 50 mM pH 7.0
PMSF	1 mM	Ser	100	Sodium Phosphate 50 mM pH 7.5
DTNB	1 mM	Cys	20	Sodium Phosphate 50 mM pH 8.0
PCMB	10 μM	Cyc	27	Sodium acetate, 50 mM, pH 5.8
NEM	10 mM	Cys	10	Sodium Phosphate 50 mM pH 7.5
Butanedione	10 mM	Arg	18	Sodium Borate, 50 mM, pH 8.0
Phenylglyoxal	3 mM	Arg	29	Sodium Bicarbonate , 50 mM, pH 8.0
PNPG	2 mM	Arg	22	Sodium Bicarbonate , 50 mM, pH 8.0
NBS	50 μM	Trp	60	Sodium acetate, 50 mM, pH 5.0
HNBB	20 mM	Trp	72	Sodium acetate, 50 mM, pH 5.0
NAI	10 mM	Tyr	35	Sodium Borate, 50 mM, pH 7.5
TNM	50 μM	Tyr	28	Sodium Borate, 50 mM, pH 7.5
EDAC	40 mM	Asx/Glx	97	MES, 25 mM, pH 5.0

4.5.2. Modification of Arginine residue

The enzyme (10mg/ml) in 100 mM, sodium borate buffer, pH 8.0 was incubated with phenylglyoxal (PG) (1.0 – 3.0 mM effective concentration) for 30 min at 30 ºC and

the activity got reduced by 93% as compared to unmodified enzyme. The inactivation was dependent on the time of treatment and concentration of the modifying reagent. Blocking the active site with 1 mM GCA o **a** protected the enzyme against inactivation, indicating that the modified residue was in the active site. Semi-logarithmic plots of residual activity as a function of time of inactivation at various concentrations of phenylglyoxal (Figure 4.9) was linear up to 18% of the initial activity indicating that the inactivation follows pseudo-first order kinetics. Arginine may possibly be positioning the substrate optimally to enhance the rate of formation of tetrahedral complex. Evidences point to involvement of arginine residues in catalysis of mechanistically related enzyme *Bsp*PVA (Manish et al., 2005)

4.5.3. Modification of Tryptophan residue

Enzyme activity was inhibited by NBS at an enzyme to NBS ratio of 1:20 which resulted in 65% inhibition. Inactivation of the enzyme was dependent on the concentration of the modifier, and the residual activity plot as a function of time at different NBS concentration was linear up to 10% residual activity. NBS mediated inactivation followed pseudo-first order kinetics at any fixed concentration of the reagent and analysis of the order of reaction yielded a slope 1 and a second-order rate constant of 2.23 µM of NBS. Titration with NBS showed a decrease in activity by 80% when one residue was modified. The apparent K_m of the partially modified enzyme increased from 0.18 to 0.34 and the k_{cat} remained constant indicating decreased binding of GCA to *Bl*BSH on enzyme modification, thus suggesting that tryptophan is involved in substrate binding. Incubation of enzyme with excess GCA resulted in almost 80% protection against inactivation.
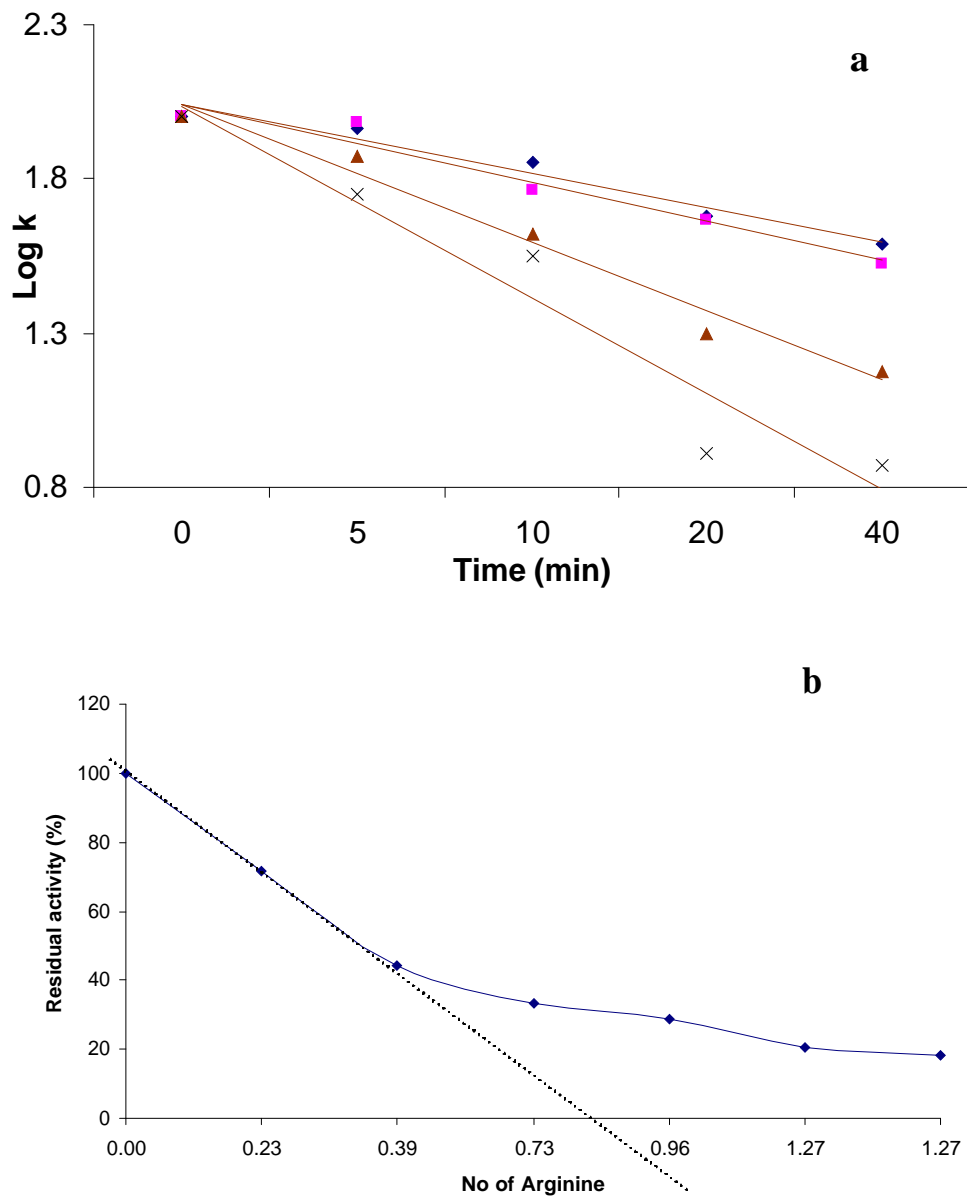
Figure 4.9: a) Determination of the order of modification reaction of *Bl*BSH with respect to arginine specific modification reagent phenylglyoxal (PG) at pH 8 and 25 °C. PG concentrations were 10 µM (♦), 20 µM (|), 40 µM (^) and 100 µM (**X**). b) The titration of *Bl*BSH against PG is plotted.

4.5.4. Substrate protection offered in chemical modification studies

       In all chemical modification reactions, the effect of substrate protection was studied by incubating the enzyme with excess amount of bile salt mixture followed by treatment with modifying reagent under similar experimental conditions. Then excess reagents were removed by PD10 (Amersham Bioscience) column. The residual activity was quantified using the standard assay (Table 4.5). The results suggest that the modification of the amino acid by specific modifying agents get subdued in the presence of substrate or inhibitor in the reaction mixture. This is indication that the modified amino acid may be in the active site and gets protected by the inhibitor or substrate during the modification reaction.

Table 4.5: Active site protection assay in the presence of various amino acid-specific modifiers.

Treatment	Enzyme activity (% of the initial activity)
None	100
PG (20 mM)	22 ± 1
GCA (1 mM) + PG (20 mM)	88 ± 5
CA (1 mM) + PG (20 mM)	95 ± 4
DTNB (2 mM)	22 ± 2
GCA (1 mM) + DTNB (2 mM)	85 ± 6
NBS (20 μM)	40 ± 3
GCA (1 mM) + NBS (20 μM)	78 ± 4
CA (1 mM) + NBS (20 μM)	72 ± 6

4.5.5. Modification of tyrosine residue

       The titration of *Bl*BSH with N-acetylimidazole demonstrated a concentration dependent and time dependent inactivation of the enzyme. Spectrophotometric analysis of N-acetylimidazole mediated inactivation has showed an increase in absorbance at 278 nm rather than decrease, suggesting possible cleavage of the protein.

4.5.6. Circular Dichroism experiments

The CD spectra of native and chemically modified enzyme samples were recorded in a J-715 spectro-polarimeter using filtered protein solution in 10 mM potassium phosphate buffer, (pH 7.5), at 25 ºC in quartz cuvettes of 0.1 cm path-length. The conformity of CD spectra of the modified enzyme with that of unmodified enzyme rules out any possibility of modifier mediated structural changes as cause of inactivation (Figure 4.10).
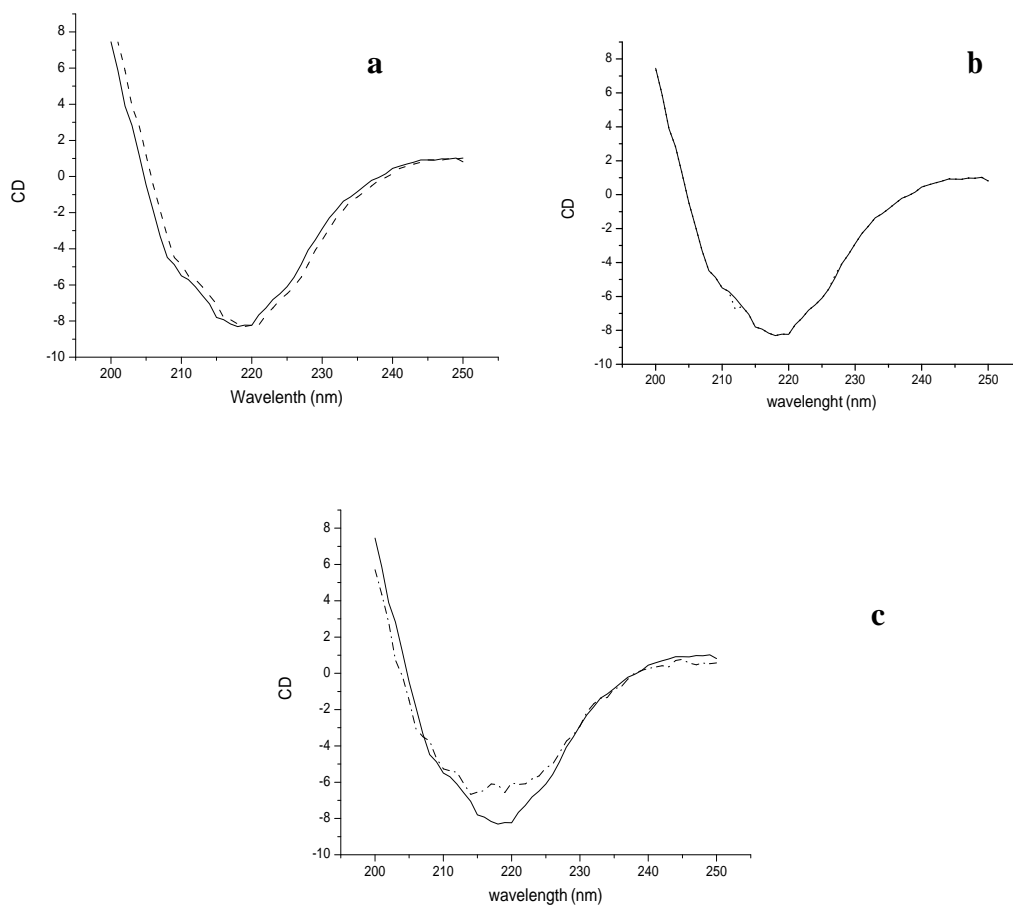
Figure 4.10: Circular Dichorism spectra of Native ( ) and Chemically Modified Enzyme (--) of *Bl*BSH. Targeted amino acids, a) Cystein b) Tryptophan c) Arginine.

4.6. Characterization of tryptophan microenvironment

The quenching of fluorescence using acrylamide or KI provided linear Stern-Volmer plots, suggesting that the Trp residue responsible for emission is accessible to soluble quenchers. No shift in enzyme's observed emission maximum at 333.5 nm and no decrease in activity in the presence of quenchers indicates that the quenching is due to some physical process. From Table 4.6 it is clear that the Trp residues causing emission are in a relatively polar environment. A downward curvature of KI plot implies that Trp residues are in two environments, one that is readily accessible to KI and another which is not freely accessible (Figure 4.11). The K_{sv} value for CsCl is lower (0.325) than that for KI (0.660). This reveals that the Trp environment is electropositive. The f_a value indicates that a fraction of the fluorescence is accessible to quenching but it need not match with the fraction of Trp residues accessible to quenching.
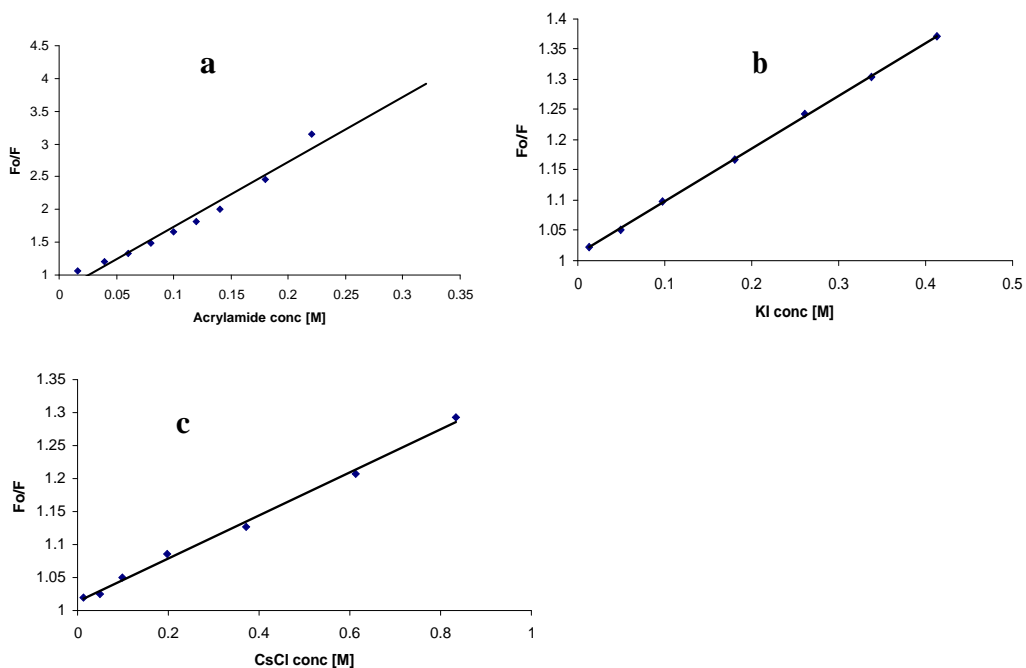
Figure 4.11: Modified Stern-Volmer plots. (a) Acrylamide, (b) KI, (c) CsCl.

Table 4.6: The effects of neutral and ionic quenchers on the fluorescence of *Bl*BSH

Quencher	15 % quenching	K_{sv} (eff)	f_a
Acrylamide (0-0.2 M)	0.016	4.75	1
KI (0 – 0.3 M)	0.012	0.66	0.33
CsCl (0-1.0 M)	0.012	0.325	0.1

4.7. Determination of substrate specificity

The preliminary studies identified many substrate analogs and substrate mimics which inhibit BSH activity. Mainly some substrate analogs and reaction products were tested for their inhibitory effects. The inhibition of *Bl*BSH activity was characterized using different biochemical and biophysical methods. The variation of binding constant reflecting systematic changes in a series of substrates or substrate analogues can provide considerable information about the interaction of substrate molecule with the active site of the enzyme. Thermodynamics studies are particularly valuable in determining the driving forces behind the binding process, including contribution from hydrogen bonding, hydrophobic and van der Waals interactions.
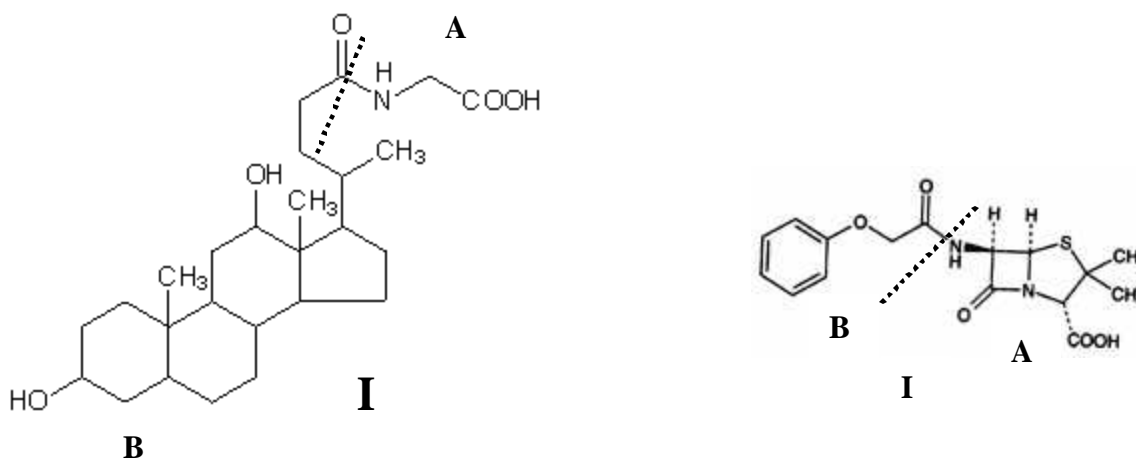
Figure 4.12: Schematic drawings of Glycocholic Acid (**I**) and penV (**II**) are shown. Label **A** indicates the putative leaving group and **B** indicates the group participating in tetrahedral intermediate.

4.7.1. Substrate specificity of *Bl*BSH

*Bl*BSH exhibited preference for glycine-conjugated bile salts over taurine-conjugated ones, but did not appear to discriminate between deoxy- and dihydroxy-bile salts. The BSH activity of the enzymes *Cp*BSH and *Bsp*PVA are 73% and 20%, respectively, of the *Bl*BSH activity. The latter is totally inactive towards penV, while *Bsp*PVA enzyme has the highest PVA activity. The PVA activity of *Cp*BSH is only 11% of the activity of *Bsp*PVA. There appears therefore to be an inversion of substrate preference within this protein sub-family. For comparison, the values of their activities at their respective pH optima are tabulated (Table 4.7).

Table 4.7: Sequence similarity between *Bsp*PVA, *Cp*BSH and *Bl*BSH and their activity against penV and bile salt are shown.

Source of the enzyme	Sequence Alignment similarity (%)		PVA activity (U)	BSH activity (U)
	B. sphaericus	*C. perfringens*		
B. sphaericus			20.2 ± 2.3	6.2 ± 1.4
C. perfringens	34		1.6 ± 0.3	18 ± 2.1
B. longum	28	35	Nil	30 ± 2.6

4.7.2. Inhibition kinetics

Initial kinetic evaluation revealed that CA and DCA competitively inhibited bile salt hydrolase. The other related compounds like POAA, PAA and 4-amino-PAA also showed competitive inhibition, while penV and penG showed mixed type inhibition. The K_i values of these compounds are listed in Table 4.8 and Figures 4.13a and b. In an independent method to determine the value of K (the rate constant for the conversion of conjugated bile salt to deconjugated bile salt) high concentrations of enzyme and inhibitor were pre-incubated for sufficient time to allow the system to reach equilibrium. The enzyme-inhibitor mixture was then diluted 100-fold and assayed at saturated

substrate concentration (100 x K_m), which resulted in the dissociation of inhibitor and regeneration of enzyme activity. The inhibition of enzymatic activity by penV was investigated, given that this was a substrate mimic of conjugated bile acid. From Lineweaver-Burk plots, penV shows mixed inhibition against the hydrolysis of GCA. In order to characterise the system, both Dixon (i.e. $1/V$ *vs* I) and Cornish-Bowden (i.e. S/V *vs* I) plots were drawn. On a Dixon plot the five regression lines corresponding to various GCA concentrations intersected at a point that was clearly negative for both $1/V$ and I.

As already stated penV and penG showed mixed inhibition while it is strictly competitive in the case of CA and DCA. This implies that the conformation of the enzyme's active site in complex with inhibitor may be substantially different from that formed with CA and DCA. The pattern shown by kinetic results can be understood in terms of complementarity between the structure of inhibitor and the binding site. Evidently, a more elaborate study on enzyme-substrate complex will be required to understand the diverse interactions between *Bl*BSH and its substrates or inhibitors such as penV.
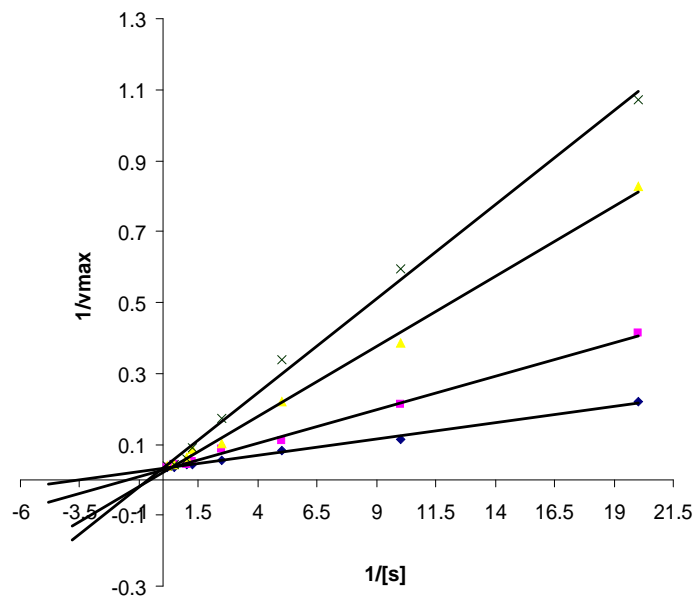
Figure 4.13a: LB plot for the determination of inhibition constant of CA

153
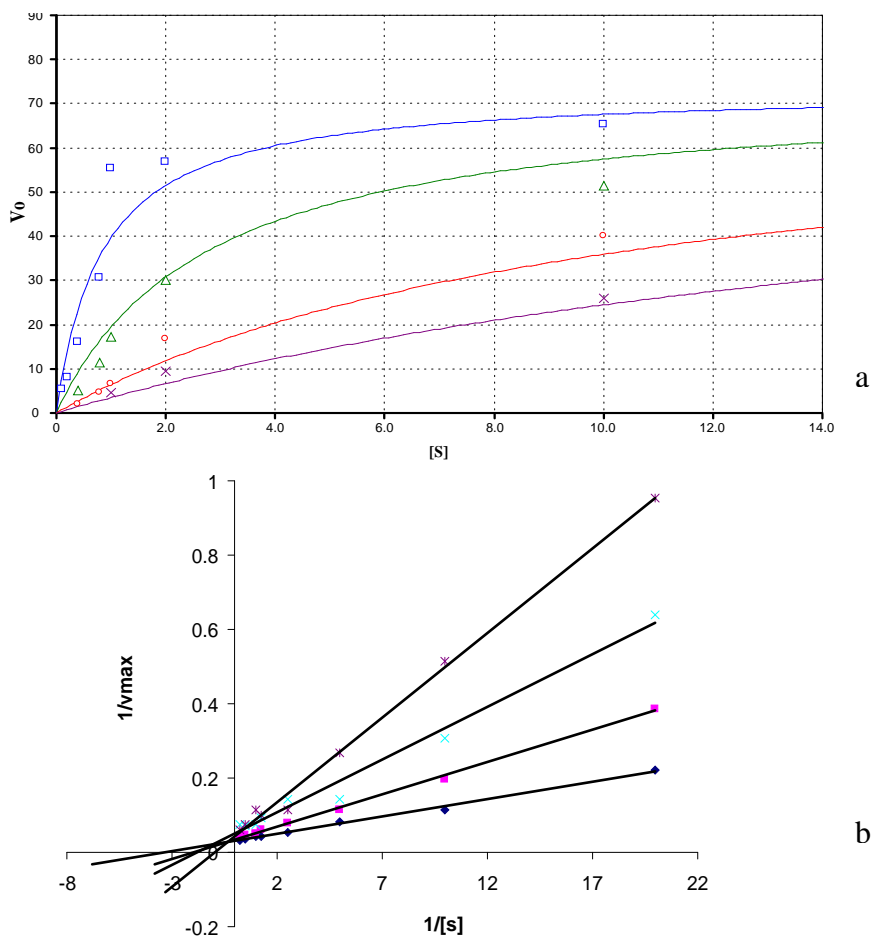
Figure 4.13b: Determination of inhibition constant of penV a) Simple curve fitting
b) LB plot

Table 4.8: Inhibition constants (mean ± standard deviation) for bile salt analogs binding
to *Bl*BSH

Compounds	Inhibition Type	K_i (mM)
Cholic acid (CA)	Competitive	0.32 ± 0.07
Deoxycholic acid (DCA)	Competitive	0.46 ± 0.09
Phenyl acetic acid (PAA)	Competitive	0.94 ± 0.17
Phenoxy Acetic acid (POAA)	Competitive	0.82 ± 0.21
PenV	Mixed	1.2 ± 0.32
PenG	Mixed	2.0 ± 0.48
4-amino phenyl acetic acid	Competitive	0.86 ± 0.26

4.7.3. Active site protection studies to determine ligand affinity

We have already discussed about substrate protection offered to chemical modification. In a similar type of experiment the affinity of various ligands for *Bl*BSH enzyme was determined by their protective effect during modification of the active site cysteine by DTNB. The degree of protection offered by various ligands is tabulated in Table 4.9 as percentage of residual activity with reference to the activity of unmodified sample.

Table 4.9: Active site protection provided by various ligands during cysteine modification by DTNB (20 mM)

Treatment	Activity (%)
Enzyme alone	100
DTNB	25 ± 1
CA + DTNB	75 ± 4
DCA + DTNB	72 ± 2
PAA + DTNB	84 ± 6
POAA + DTNB	50 ± 4
4-PAA + DTNB	62 ± 2
PenV + DTNB	75 ± 6
PenG + DTNB	74 ± 4

4.7.4. Fluorescence spectroscopy

The fluorescence spectra of *Bl*BSH showed emission maxima at 333.5 nm (λ_{max}). The 315 residue length sequence of the protein has five Trp and twelve Tyr residues. To excite Trp selectively, a wavelength of 295 nm was chosen. The inhibitors tested have shown quenching of fluorescence. The absorption spectra of ligands do not overlap with the tryptophan fluorescence spectrum (Figure 4.14); hence, an energy transfer from tryptophan to ligands may be ineffectual. To rule out the possibility of collision-quenching of Trp fluorescence, we measured the fluorescence parameters of free Trp in

the presence of all ligands. Fluorescence of free Trp is not quenched by the ligands; thus, the collision mechanism of *Bl*BSH fluorescence quenching is ruled out. Figure 4.15 shows the quenching of Trp as a result of penV binding to *Bl*BSH. When denatured *Bl*BSH was used, no quenching of the fluorescence intensity occurred in the presence of penV. Therefore it can be concluded that the quenching of Trp fluorescence of the active form of *Bl*BSH is a direct consequence of enzyme–substrate/inhibitor interaction. From the changes in fluorescence spectra, the dissociation constant and enthalpy change due to ligand binding were calculated (Table 4.10). The slope of the plot of log {(Fo-F)/(F- F)} *vs* log [S] was unity for all the inhibitors used, indicating the formation of a 1:1 complex between enzyme and inhibitor (inset of Figure 4.15c). The titration against penG resulted in a blue shift of the fluorescence maxima from 333.5 to 338 nm. The titration with substrates (GCA & GDCA) as well as with the end products (CA & DCA) both resulted in enhancement of fluorescence (Figure 4.16).
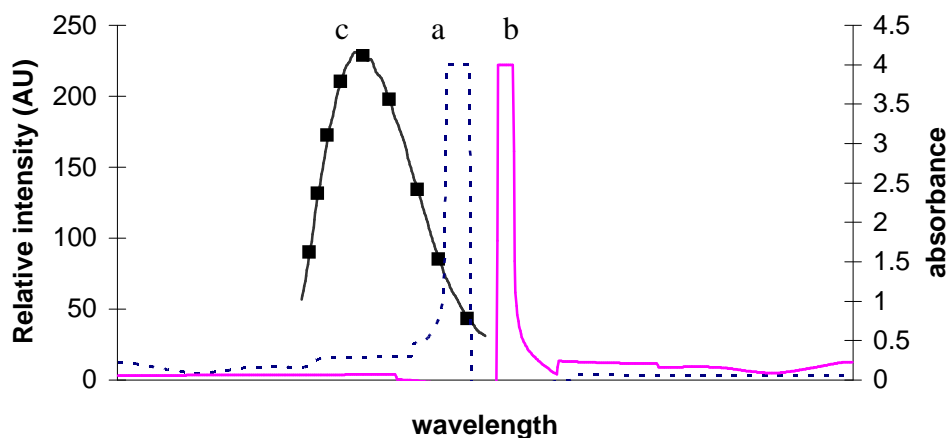
Figure 4.14: Absorption spectra of (a) penV (b) penG and (c) fluorescence spectra of *Bl*BSH at pH 6.5.
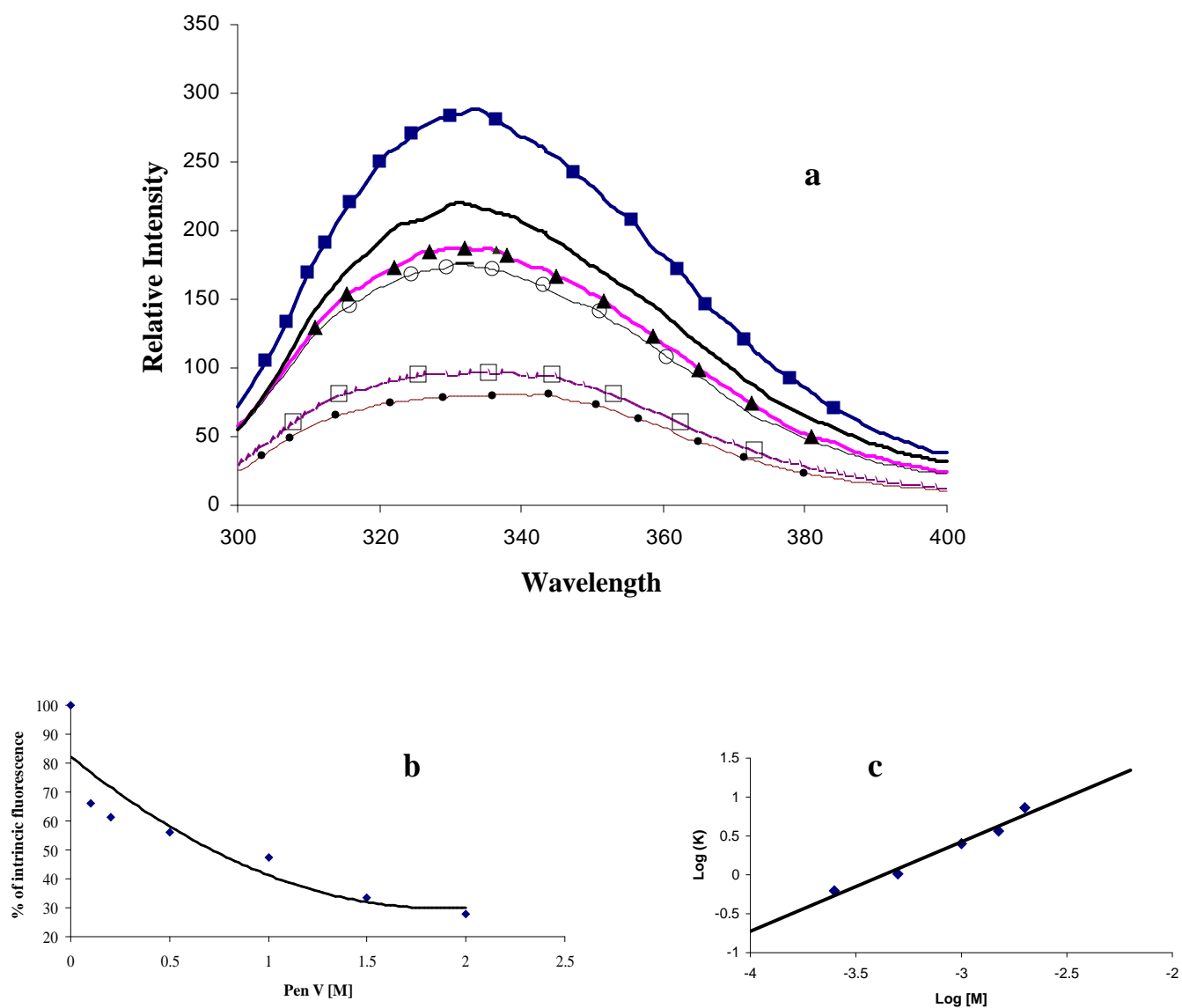
Figure 4.15: a) Fluorescence quenching observed due to penV binding obtained at different concentrations: no added substrate (|), 0.5 (-), 0.75 (^), 1 (0), 1.5 ( ) and 2 mM ( ). b) Percentage of quenching by the addition of increased molar concentrations of penV. c)Log (K = (Fo-F)/(F-F$_\infty$)) was plotted against Log [M = molarity] of penV.

Table 4.10: Thermodynamic parameters and binding constants estimated for *Bl*BSH-ligand interaction estimated from fluorescence titration study.

Ligands	K_d (x 10^{-4})	- G^a (kJ mol^{-1})	- H (kJ mol^{-1})	S (kJ mol^{-1} K^{-1})
PenV	4.6 ± 0.8	18.57 ± 1.3	61.05 ± 4.6	0.14 ± 0.02
PenG	5.2 ± 1.4	18.31 ± 1.1	62.65 ± 5.5	0.14 ± 0.05

G, free energy change;   H, enthalpy change;   S, entropy change of binding.
[a] values are determined at 298 K.
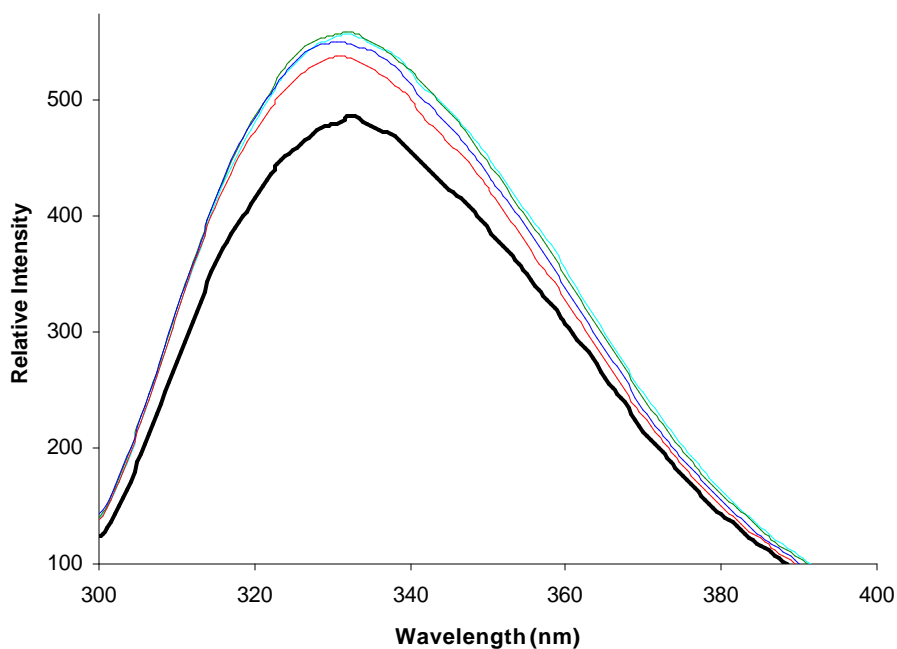
Figure 4.16: Fluorescence enhancement of *Bl*BSH in the presence of different concentration of GDCA: no GDCA (-) 0.1 ( ) 0.25 ( ) 0.5 mM ( )

4.7.5. Isothermal calorimetry

The thermodynamics quantities that are most appropriate for the characterization of ligand binding equilibria are the Gibb's binding energy change (ΔG), the binding enthalpy change (ΔS) and the binding heat capacity change (ΔC). A typical calorimetric titration curve on the addition of GCA to 0.8 ml of penV at pH 6.5 and 298 K is shown in Figure 4.17. The titration curve exhibits a monotonic decrease in exothermic heat of binding in each successive injection. The binding enthalpies range from -5.4 to -2.0 kJ mol^{-1}. The ITC titration in the presence of GCA could not be measured even after repeated efforts because of froth formation by stirring on a concentrated solution of both protein and bile salts in the calorimetric cell.

The calculated van't Hoff enthalpies and entropies of binding of penV to *Bl*BSH differed substantially from the value obtained from ITC experiments. Such differences between the ITC binding enthalpies and van't Hoff enthalpies are not uncommon. Moreover, the intrinsic fluorescence changes of the enzyme upon binding of penV or GCA are rather small and fine interpretation of the values would be misleading. But the pattern of change between the entropy and enthalpy remains same in the both experiments.

4.8. Substrate and inhibitor docking studies
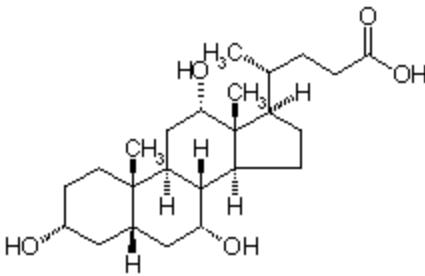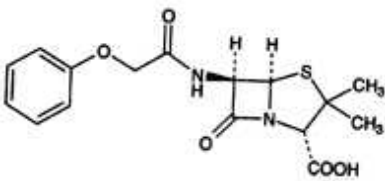
The biochemical knowledge obtained on *Bl*BSH also suggests on the competitive nature of inhibitors (Kumar et al., 2006). Thus the primary assumption in docking study is that the substrate and inhibitors bind at the same or closely disposed pockets in the enzyme. The properties of the chosen molecules such as conjugated bile salts and penV are listed in the Table 4.11. The pharmacophore identified here for Ph4Dock consists of 5 hydrophobic patches and 3 hydrogen bond donor and acceptor sites. In the docking procedure, the flexibility of side chains surrounding the ligand was parameterised for energy minimization of the complex.
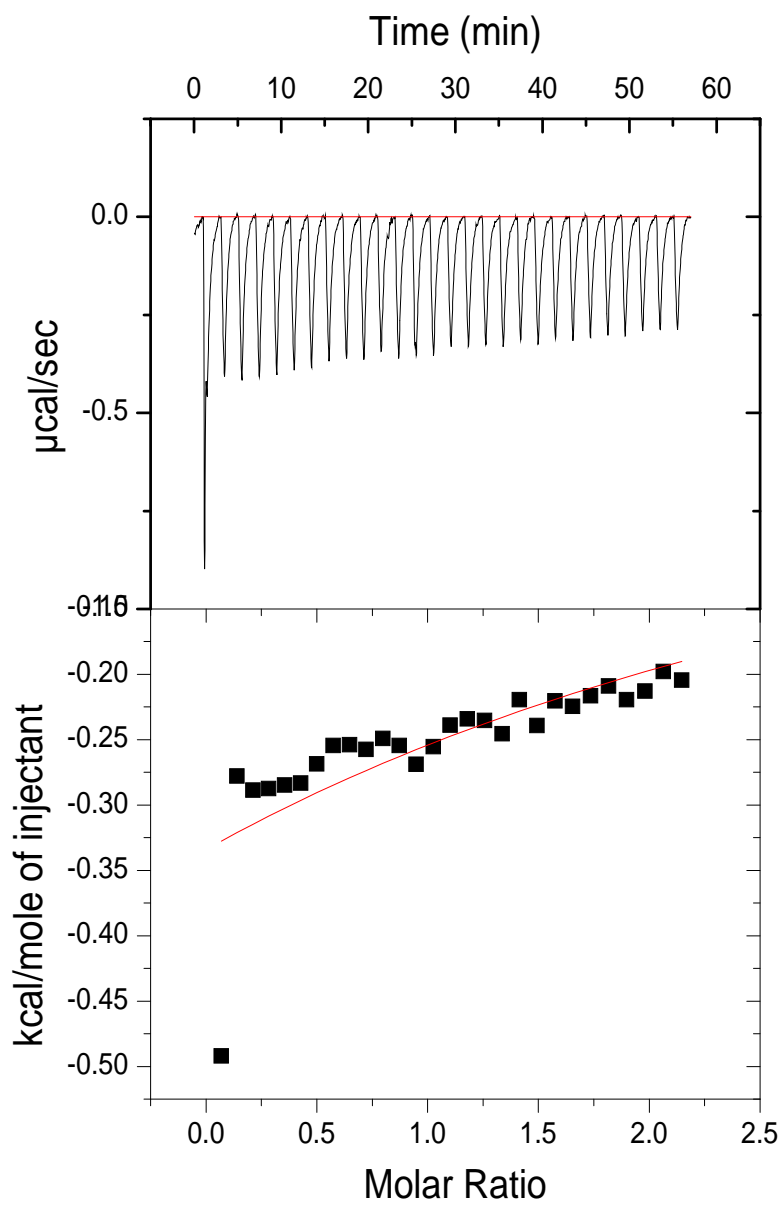
Figure 4.17: Isothermal calorimetric measurement of *Bl*BSH in presence of penV in 10
mM phosphate buffer pH 6.5 at 25 ℃

Table 4.11: The properties of GCA and penV

S. No.	Property	Conjugated Bile Salt	PenV
1	Structure		
2	Molecular weight	487.60	388.48
3	No. of rotatable bonds	10	6
4	Binding molecule	*Bl*BSH	*Bsp*PVA
		*Bsp*PVA	*Bl*BSH
5	Positive enzymatic action	*Bl*BSH	*Bsp*PVA
		*Bsp*PVA	
6	Residues in the binding pocket: conserved & non-conserved	Cys1, Arg17, Ile132, Leu19, Tyr25, Met61, Phe68, Asp137, Phe42	Cys1, Arg17, Asn132, Met19, Pro25, Asp61, Tyr68, Gly137, Gln42

4.8.1. Identification of binding site

The Alpha site finder was used to correctly identify the binding pocket of the enzyme. The Alpha site finder methodology is based upon Alpha shapes which are a generalization of convex hulls developed by Edelsbrunner et al., (1995). The Alpha site finder identified 19 possible binding sites, among those only two had included in them the nucleophile residue Cys1 (Site No. 1 and 7). The binding-site residues of consequence in the enzyme have been identified by calculating total combined interaction energy between the substrate and each of the amino acid residues in the enzyme. This method of identification, combined with another definition based on the distance from the substrate, can clearly compile the relative significance of every residue.

The site number 1 has thus been selected as the most conducive binding site consisting of 185 hydrophilic contact atoms (140 of them are side chain atoms) and another 33 are hydrophobic contact atoms.

4.8.2. Ligand docking
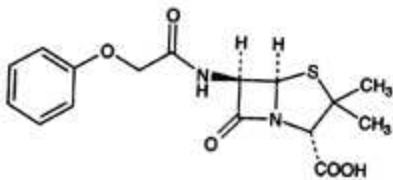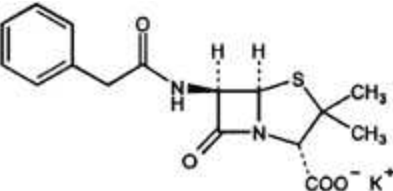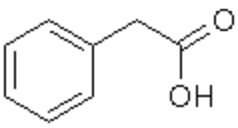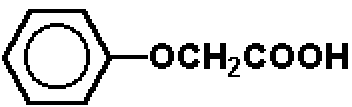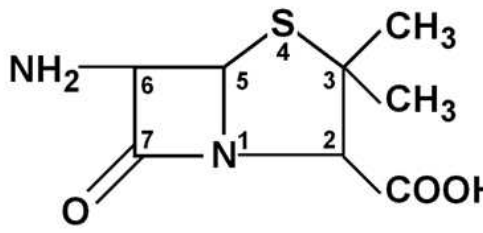
The energy-minimized ligands were docked against the selected binding site. The energy calculated for the complex as well as ligand alone and other relevant parameters are listed in table 4.12. For each of the ligands ten favourable conformations were ranked based on their individual total energies (U_{total}). Among them the best conformation was selected based on the range of energy levels (Figure 4.18). The overall free energy of interaction is the main determinant of favourable binding of a ligand with protein and it indirectly indicates the degree of affinity of the protein for the ligand. Since hydrogen bonds play an important role in structure and function of biological molecules, especially in enzyme catalysis, those identified in the complexes modeled are (listed in Table 4.12) shown in figures of protein docked with ligand.
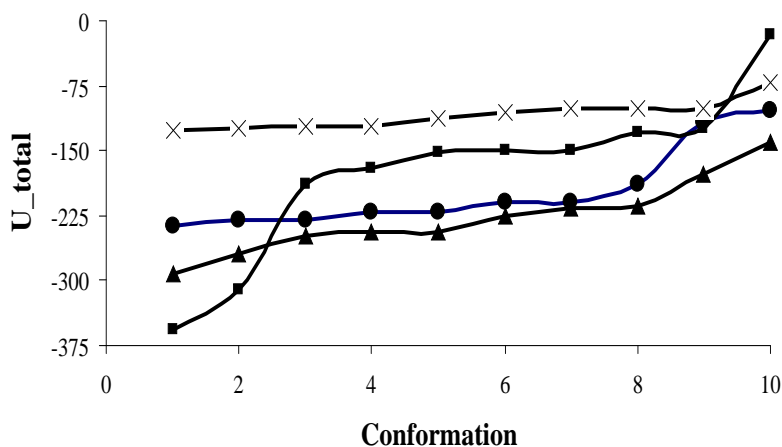
Figure 4.18: The Total free energy of different docked conformation of GCA (|)

TCA (^) penV ( ) 6-APA (×)

Table 4.12: Minimum energy conformations of ligands generated by Ph4Dock, affinity for best and the total free energy corresponding to that complex are listed.

S. no	Ligand	K_m or K_i	No of best conformations	RMSD	Total energy
1	GCA	0.22 ± 0.07	1	1.06	-356.31
2	GDCA	0.18 ± 0.05	2	1.14	-293.17
3	GCDCA	0.28 ± 0.11	1	0.99	-355.42
4	TCA	0.32 ± 0.13	2	1.08	-291.36
5	TCDCA	0.42 ± 0.15	1	0.99	-378.62
6	Pen V	1.2 ± 0.32	5	0.88	-236.62
7	Pen G	2.0 ± 0.48	2	0.38	-206.14
8	PAA	0.94 ± 0.17	3	0.75	-260.58
9	POAA	0.82 ± 0.21	3	0.8	-148.54
10	6-APA	-	4	1.02	-126.31

4.8.3. Mode of binding of Bile salts

There are differences in the orientation of binding between the bile salts on one side and inhibitors on the other. The differences are primarily due to the differences in chemical nature of the side chain groups of the ligands. The nuclophilic residue (Cys1) has no accessibility for amide bond in the case of inhibitors as a result of modified binding. The binding pocket in the enzyme for bile salts consists of 26 residues including Arg17, Asp20, Trp21, Tyr25, Asn81 and Arg225. Among these residues Arg17 form most frequently hydrogen bond with the carboxylic group of the conjugated bile acids. Interestingly, in the X-ray determined structure of the complex of *Cp*BSH this hydrogen bond is present even with the product formed (Rossocha et al., 2005). In this case the carboxylic group of the cholic acid generated by substrate-cleavage remains hydrogen bonded to Arg17 (Figure 4.19). This particular residue can be thought to play a very crucial role in substrate binding and may also serve as the anchor to bring correct orientation of substrate. Similarly, the residue Trp21 is only 4.5 – 7 Å away from the substrate in the modeled energy minimum complexes. This residue also may be interacting with the incoming substrate through some stacking interaction and helping to select an initial favourable orientation of the substrate.

In the crystal structure of the complex of *Cp*BSH with reaction products, residues Pro134, Arg17 and Tyr25 also take part in enzyme-substrate interactions. Another hydrogen-bonded interaction is through a water molecule. These hydrogen bonds are observed in the docked structures of various ligands as well. In addition to this, some of the modeled complexes also showed hydrogen bonds between ligands and residues Asn81 and Gly79.

Generally the protein structures may not undergo any major changes due to ligand binding, exceptions if present are usually of functional significance. The most noticeable changes that can be found are the movement of certain flexible loops on the surface of the protein. In the structure of *Cp*BSH such a displacement is observed in the loop containing residues 60-65. No great conformational changes were observed between the structures of the native and the modeled complexes of *Bl*BSH.

No apparent differences are observed in the binding pattern of substrates irrespective of the type of scaffolds (cholic acid or deoxycholic acid) or side chains

(glycine and taurine) they possess. The amide bond is positioned just inside the cleft of the enzyme, very close to the N-terminal cystein, in a favourable position to undergo nucleophilic attack. An orientational and directional preference of the substrate with respect to the nucleophile and the base of the enzyme are observed in the case of Ntn hydrolases and serine proteases (Isupov et al., 1996). Even in these modeled complexes the same archetype of disposition is observed between the enzyme and the true substrate.

4.8.4. Mode of binding of inhibitors

The inhibitors of *Bl*BSH such as penV, penG, PAA, POAA and Cholic acid have been used for docking studies. Between penV and penG the former one is more preferred at least in terms of energy estimation (Table 4.12). From the figure 4.20 it is clear that the orientation in this case is not favorable for the enzymatic action on the amide bond. The amide bond is placed away from the nucleophile. Other binding conformations and binding modes were also observed in the case of penV, but could be ruled out as energetically unfavorable. PenV forms hydrogen bond with N of Asn81. The side chain of Trp21 undergoes considerable change in conformation and this residue is possibly be interacting directly with the -lactam ring of penV. It is worthwhile to note that this particular Trp is absent in *Bsp*PVA and strictly conserved in all BSH type enzymes.
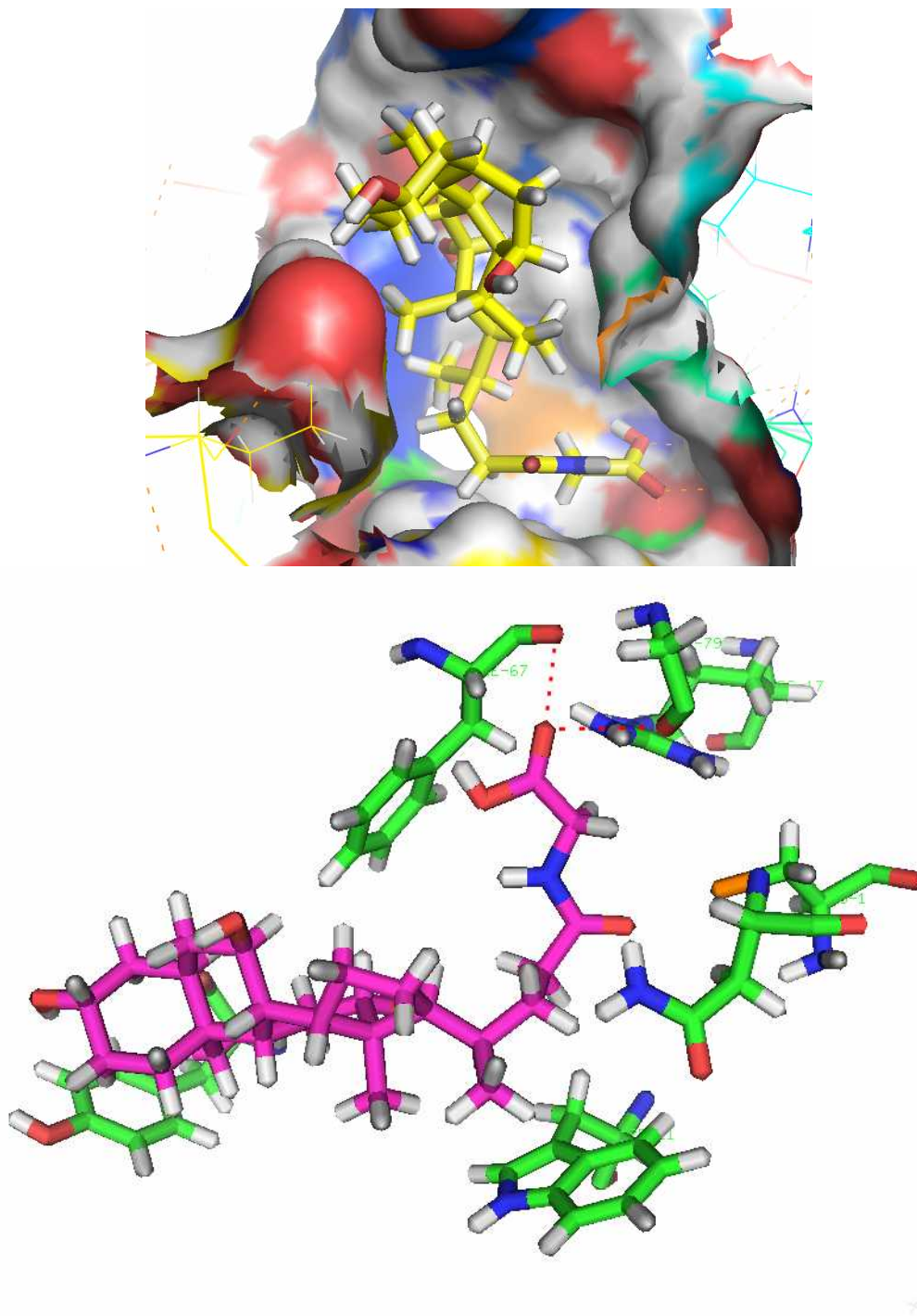
Figure 4.19. Docking of GCA on *Bl*BSH a) solid surface view

b) showing those residues interacting with GCA
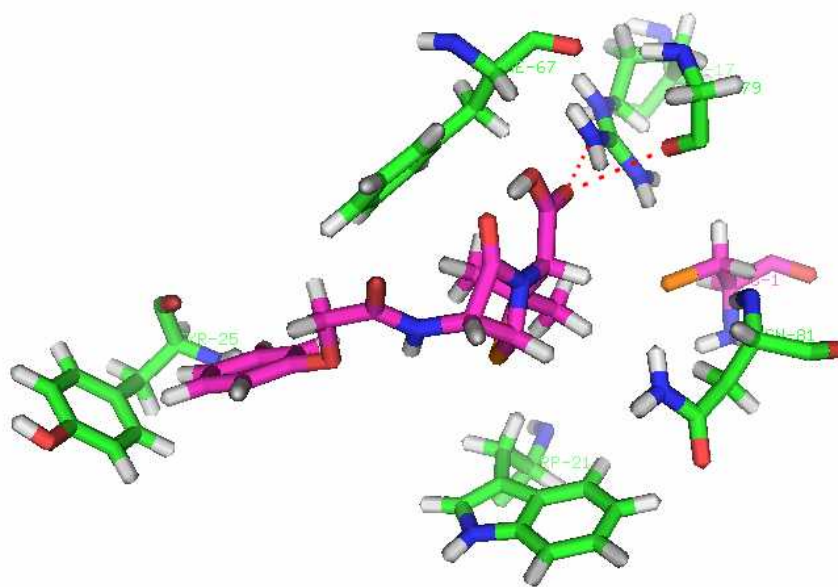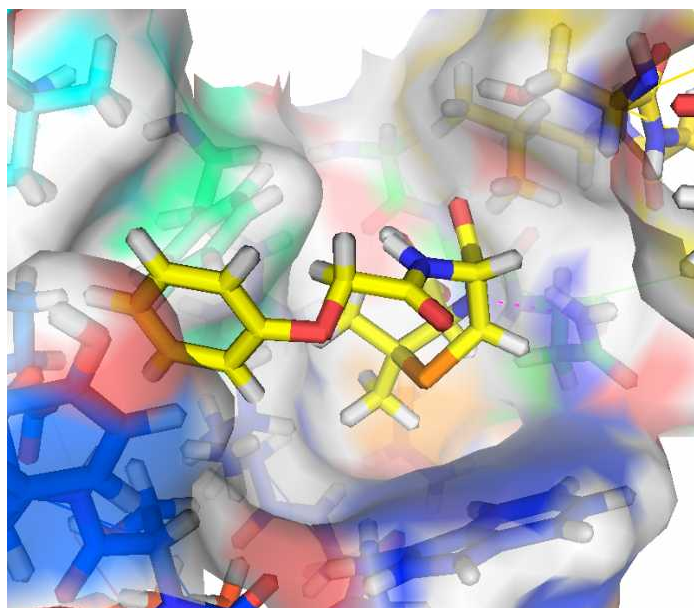
Figure 4.20. Docking of inhibitor penV on *Bl*BSH a) solid surface view

b) showing those residues interacting with penV

4.9. Discussion

4.9.1. Biochemical Characteristics of *Bl*BSH

On prolonged storage at 4 °C in phosphate buffers, the enzyme progressively lost activity that could be recovered only partially by treatment with reducing agents such as -mercaptoethanol or DTT. The loss of activity coincided with the appearance of insoluble inactive aggregates in the enzyme preparation. This may be due to some firm intermolecular interactions that can lead to inseparable association taking place under the storage conditions used. Nevertheless the purified enzyme was found more stable in the absence of rather than in the presence of reducing agent like -mercaptoethanol. The loss of enzyme activity can also be explained by the mechanism that in the presence of air thiol groups have a propensity to generate hydrogen peroxide and superoxide anions, which in turn can lead to irreversible enzyme damage (Misra., 1974; Costa et al., 1977).

4.9.2. Substrate specificity

Albeit a higher degree of sequence identity and three-dimensional structural similarity between BSH and PVA, the difference in their substrate specificity and activity was evident through data obtained from biochemical experiments. The enzyme catalytic studies collectively suggest that catalysis could be understood in terms of physical organic principle. One prominent feature, seen repeatedly, is that the catalytic elements in an active site are precisely positioned for their function. This precision facilitates passage through the transition state intermediate and thus provides a more satisfying picture of what might be understood as transition-state stabilization. This alignment is achieved by the pre-organisation of an enzyme's active site and allows the selection of subpopulation of the substrate ensembles that approach the configuration of the relevant transition state and bound with high affinity. This is achieved through the energy of binding and the imposition of steric considerations and non-bonded interactions within the active site cavity.

There are numerous reports of conformational changes taking place wherein loop elements or entire domains move several angstroms to sequester the substrate or to create a tunnel for a product to transit or for the precise alignment in the transition

reorganization of the active site. The symbiotic combination of experiment and theory is vital for the elucidation of complete picture of enzyme catalysis.

The nature of the active site was revealed from substrate specificity analysis. The inhibitors showed competitive and mixed type inhibition. The protection provided by these ligands to chemical modification of catalytic residues to various degrees showed that these inhibitors could be binding at the active site, but in an orientation not conducive for activity. Among various inhibitors tested, CA, DCA and POAA showed good protection towards chemical modification.

The inhibition depended to various extents on the side chain of the inhibitor. The K_i is very much lower for cholic than deoxycholic acid, which is an indication of the strong affinity of BlBSH towards the oxy substrates. Substitution in the cholic acid ring does not alter the K_i much, which is in good agreement with the previous study (Batta et al., 1984). There is some evidence to show that the variation in cholic acid side chain affects the catalytic efficiency (Batta et al., 1984). It has been noted that substrates lacking a carboxyl group in the side chain are not hydrolysed. The other reports showed that there was no change in catalytic efficiency on substitution of amino acid conjugates, glycine by other molecules such as taurine, alanine, 4-ANBA (Maeda et al., 1989; Maeda and Takahashi, 1989).These results indicate that the active site of BlBSH recognizes only the cholic acid side chain and the -carboxyl group, while allowing flexibility in the side chain amino acids. During hydrolysis by BlBSH the leaving group is glycine or taurine whereas in hydrolysis by PVA the leaving group is 6-APA. Both moieties show little effect when substituted with a wide variety of compounds, indicating that there is no stringent binding condition for these leaving groups. In BSH the cholate and in PVA the PAA groups must be accommodated as tetrahedral intermediate within the enzyme's catalytic site, requiring these groups to possess shapes specific for and complementary to the active site binding pockets.

The fluorimetric titration with different substrate and inhibitors provided some insight into their binding. The interaction of the enzyme with inhibitor or substrate led to a perturbation in the microenvironment of the Trp residue as demonstrated by the quenching or enhancement of Trp fluorescence of the enzyme upon binding. The fluorescence enhancement with the addition of substrate like GCA, GDCA and the

quenching of fluorescence with the addition of penV, penG, and POAA was observed. It may be due to the difference in interaction of the inhibitor (PAA group) compared to substrate (CA/DCA group) with Trp21. Interestingly this Trp residue is conserved in true BSHs but not in known PVAs as well as enzymes which have both PVA and BSH activities. We believe that, in addition to other unknown factors, Trp21 also plays a role in the selectivity between bile salts and penV as substrates for these hydrolases.

The variation of binding constant with systematic changes within a series of substrates and substrate analogues may provide considerable information about the interaction of substrate molecule with the active site of the enzyme. Thermodynamical studies are particularly valuable in determining the driving forces for the binding process, including contribution from hydrogen bonding, hydrophobic and van der Waals interactions. In general, the geometric complementarities between protein and ligands are crucial for their binding to each other. It is equally important that they match in terms of electron distribution, hydrogen bond donor-acceptor properties, hydrophobicity and polarizability. Hydrogen bonds are prominent interactions between pairing polar groups, and water molecules frequently act as intermediaries in such interactions. Rigidity of the docking procedure perhaps was the major reason for false positive or false negative prediction, as the geometry of the complex and docking score are sensitive to the orientation of individual side chains.

The most reasonable binding mode of ligands was explored using the pharmacophore docking program Ph4Dock in the MOE suite. Pharmacophore dock is a new docking methodology, which involves a ligand-based pharmacophore being introduced in an automated docking procedure. The merit of Ph4Dock is that the binding mode, consistent with a predicted ligand-based pharmacophore, can be obtained by considering structural information of target protein.

Computational and modeling studies suggest that steric effects are not the main component of enzyme catalysis, but rather the electrostatic effects are responsible for most of the catalysis (the enzyme reduces the electrostatic reorganization free energy). A more general perspective is that a combination of steric and electrostatic interaction work together to align the substrate and catalytic elements of the active site. The results from modeling studies suggest that although the inhibitors discussed here bind at the same site

as substrates, their orientation at the binding site of *Bl*BSH does not provide nucleophilic atom with free access to the amide bond to trigger a nucleophilic attack, This also explains the high K_m values observed for these enzymes (in millimolar units) towards their identified substrates as a result of their extended nature of specificity.

CHAPTER 5

STABILITY AND DENATURATION STUDIES ON *Bl*BSH

5.1. Introduction

Conformational transitions and functional stability of the bile salt hydrolase (BSH; cholylglycine EC: 3.5.1.24) from *Bifidobacterium longum* (*Bl*BSH) cloned and expressed in *E. coli* were studied in thermal, chemical and pH mediated denaturation conditions using biophysical techniques such as fluorescence and CD spectroscopy. The major findings are i) Thermal and Gdn-HCl mediated denaturation of *Bl*BSH is a multistep process of inactivation and unfolding. Enzyme activity seems sensitive to even minor conformational changes at the active site. ii) Thermal denaturation as such did not result in any insoluble protein aggregate. However, on treating with 0.25 to 1M Gdn-HCl the enzyme showed increasing aggregation at temperatures 40-55°C indicating more complex structural changes taking place in the presence of chemical denaturants. iii) The enzyme secondary structure was turning more compact at extremely low pHs (1-3 units). The perturbation in the tertiary structure at the above pH was detected through freshly formed solvent exposed hydrophobic patches on the enzyme. It shows that these changes can be interpreted as due to formation of an acid induced molten globule-like state.

The three-dimensional structure assumed by a protein, unique and suitable for its function, can in general be considered to be thermodynamically the most stable conformation adopted by the polypeptide chain. A polypeptide can also adopt a less rigid or more flexible conformation, different from its functional native form, responding to changes in environment. Thus, proteins are only marginally stable because of the functional requirement for their inherent dynamic state and due to delicate balancing of interactions involved in stabilizing or destabilizing particular structure. In this context, detailed denaturation studies of partially folded conformation of a protein or a family of proteins is important for understanding the general principles that govern protein-folding pathways. By recording changes in intrinsic tryptophan fluorescence of protein in response to tailored changes in surroundings, one can establish presence of interesting structural intermediates relevant to structure-function relationship of the protein. An

assessment of changes in the secondary and tertiary structural features will complement these observations.

The aggregation of protein is a phenomenon associated with many neurodegenarative diseases, such as amyloidoses, prion disease and cataracts, that are caused by non-specific protein interactions (Dobson, 2006). The conformational stability of proteins is studied by denaturation of protein using different agents such as temperature, pH, chaotropic agents and pressure, either individually or in combination of the denaturants. Denaturation is followed by monitoring the changes in any measurable physical property of the protein such as intrinsic fluorescence, CD spectra and corresponding change in the protein function are checked by measuring its catalytic activity.

The aim of this study is to understand the importance of a correctly folded structure and active site geometry with respect to its functional stability. To our knowledge, this is the first time any such studies on the conformational changes of a member of Ntn-hydrolase family are conducted. Since all the members of these family share common mechanistic and structural features, the study reported here has wider implications to structure, function and stability of this class of enzymes in general.

5.2. RESULTS

5.2.1. Thermal denaturation of *Bl*BSH

The monomer subunit of *Bl*BSH contains 5 Trp residues and one Trp at the active site of the enzyme has been suggested to be crucial for substrate binding (Kumar et al., 2006). The intrinsic fluorescence of the protein due to these Trp residues was monitored at different temperatures in the present study. The fluorescence emission spectrum of the native enzyme shows a maximum (λ_{max}) at 330 nm, which characterizes the non-polar environment of Trp residues (Figure 5.1a & b). The fluorescence intensity of the enzyme gradually decreases with increasing temperature. This must be due to the deactivation of the singlet-excited state by non-radiative processes that are competitive with fluorescence. The observed λ_{max} remained same (330 nm) till 55 °C, while above 60 °C showed a red shift to 350 nm presumably due to Trp exposure to solvent (Figure 5.1c). The activity was also measured simultaneously. A loss in enzyme activity of 20% at 50 °C and 90% at 55 °C were recorded (Figure 5.1c). The ratio of fluorescence intensity

(F330/ 350 nm) 1.25 at 30 °C reduced to 1.15 at 55 °C indicating that slight changes occurred in the Trp environment with increasing temperature. The above observations suggest that even minor rearrangement of Trp environment can lead to complete loss of enzyme activity.
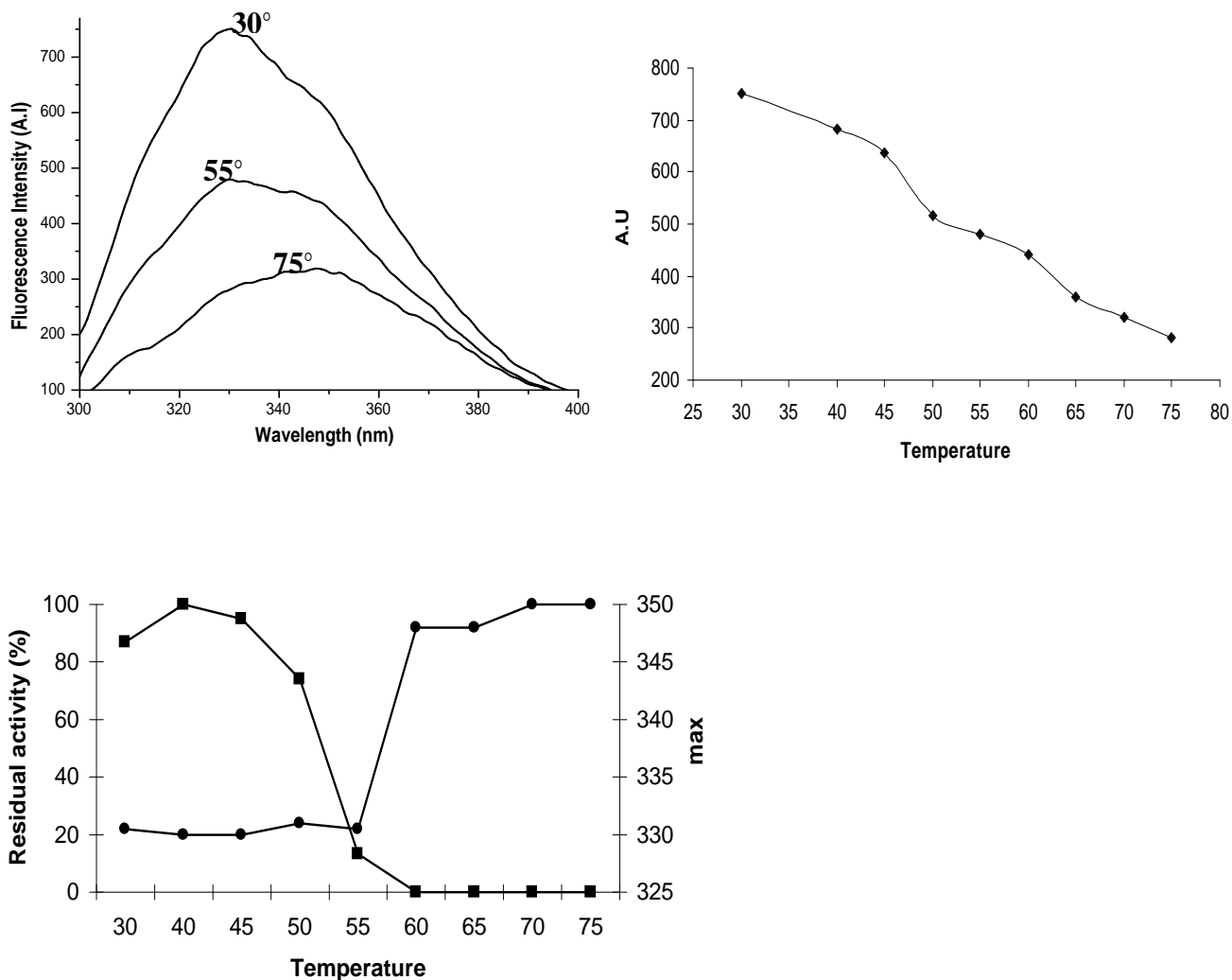
Figure 5.1: Thermal denaturation of *Bl*BSH. a) Fluorescence scans of native enzyme incubated at 30, 55 and 75 °C for 15 min. b) Fluorescence intensity at 330 nm plotted against different temperatures. c) Effect of temperature on *Bl*BSH: $_{max}$ of the enzyme (^) at different temperatures and residual activity (|) against different temperatures

174
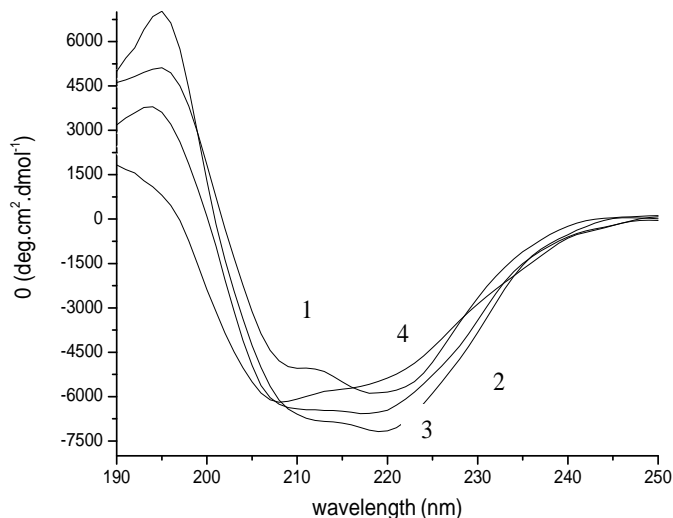
Figure 5.2: Far UV CD spectra of *Bl*BSH at 30 (1), 50 (2) 55 (3) 60 °C (4)

The far UV CD spectra of the enzyme recorded at different temperatures are shown in figure 5.2. When temperature was raised until 50 °C the secondary structure of the protein was becoming more and more compact without major loss in activity. At 55°C, there is no major change in the structure but there is a substantial loss of activity as already shown. Thus, thermal inactivation of *Bl*BSH takes place even before any major changes in the conformation of the protein happen. At 60 °C, minima at 218 nm in the CD spectrum are substantially affected showing loss of -sheet structure. The near UV spectrum has showed the unfolding of protein after 55 °C (Figure 5.3)
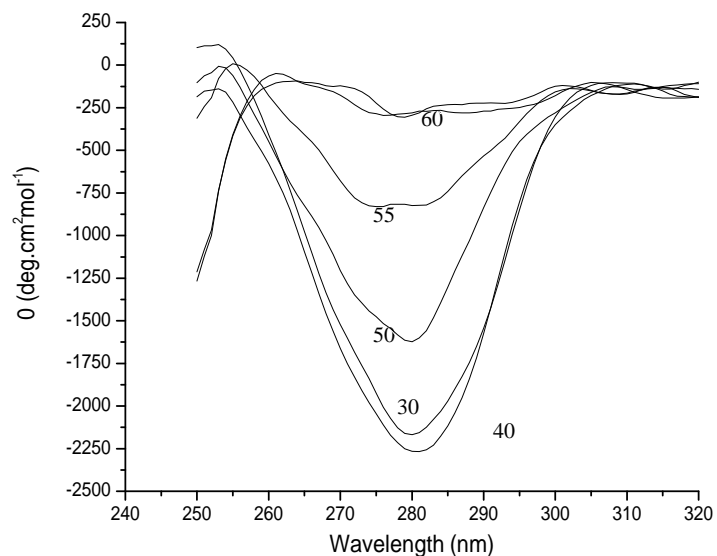
Figure 5.3: Near UV CD spectra of *Bl*BSH at 30, 40, 50, 55, 60 and 65 °C

175

At none of the above temperatures (till 80 °C) the protein showed ANS (hydrophobic dye) binding indicating that exposure of hydrophobic patches did not take place during unfolding (Figure 5.4). The unfolding observed in the above experiment seem to be partial as the red shift observed in λ_{max} can be due to rearrangement in the Trp environment.
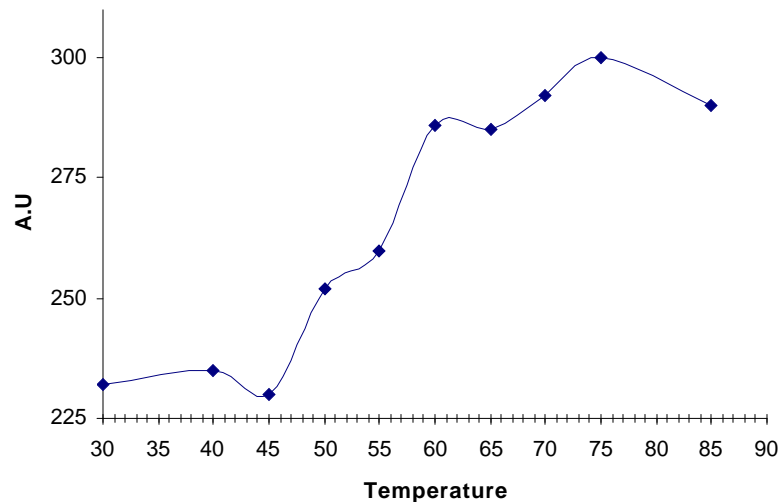
Figure 5.4: Light scattering of *Bl*BSH at different temperatures

Renaturation or cooling of the heated protein samples back to 30 °C , the enzyme incubated beyond 45 °C there was no reactivation of the enzyme was observed indicating irreversibility of the inactivation. The enzyme incubated in 40 °C there was regain of 78% of its original activity and at 45 °C only 33% of its original activity was retained. The fluorescence scans of the renatured protein samples, however, showed blue shift in λ_{max} from 350 to 342 nm. Since activity was not restored this blue shift could be due to solvent shielding of Trp through a misfolded structure. Thus the thermal denaturation of *Bl*BSH seems irreversible. Very little change in the light scattering intensity of protein was observed till 85 °C (Figure 5.4), ruling out any aggregation due to thermal denaturation. Thus, thermal inactivation of *Bl*BSH takes place without complete unfolding, or major loss of secondary structure, with no exposure of hydrophobic amino acids and no formation of insoluble aggregate.

5.2.2. Effect of Guanidine hydrochloride

Under denaturing conditions when Trp environment inside the protein changes the λ_{max} shifts with decrease in fluorescence intensity at 330 nm and increase at 350 nm. The fluorescence spectra, the enzyme inactivation and the decrease in ratio (F330/350) of intrinsic fluorescence of *Bl*BSH in the presence of Gdn-HCl are shown in figure 5.5. The red shift of λ_{max} was observed with increasing concentration of Gdn-HCl indicating the exposure of Trp to the solvent. The negative ellipticity at 218 nm due to change in secondary structure caused by Gdn-HCl is presented in Figure 5.6. In the presence of 0.5 M Gdn-HCl the enzyme loses 60% activity with no indication of any change in the Trp environment, whereas there is a 10 % loss of the secondary structure (Figure 5.6). At 1 M Gdn-HCl 90 % of the enzyme activity is lost with only a slight change in the Trp environment and there is 30 % loss in secondary structure. At 2 M concentration of Gdn-HCl there is total loss of activity accompanied by major changes in the Trp environment indicating unfolding of the protein. The loss of secondary structure is same till 4M Gdn-HCl whereas 70% loss is recorded at 5 M Gdn-HCl. Thus denaturation of *Bl*BSH by Gdn-HCl seems to be a multistep process.
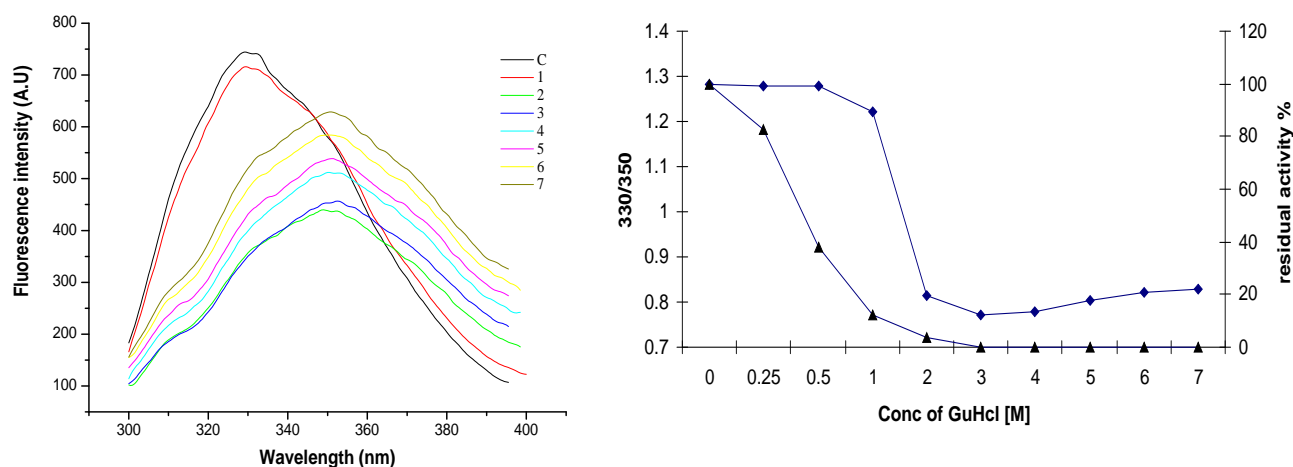
Figure 5.5: Effect of Gdn-HCl on BSH: a) fluorescence spectra of the enzyme in the presence of different concentrations of Gdn-HCl   (1 - 6M Gdn-HCl)
b) Plot of ratio 330/350 (^) and activity (|) against Gdn-HCl
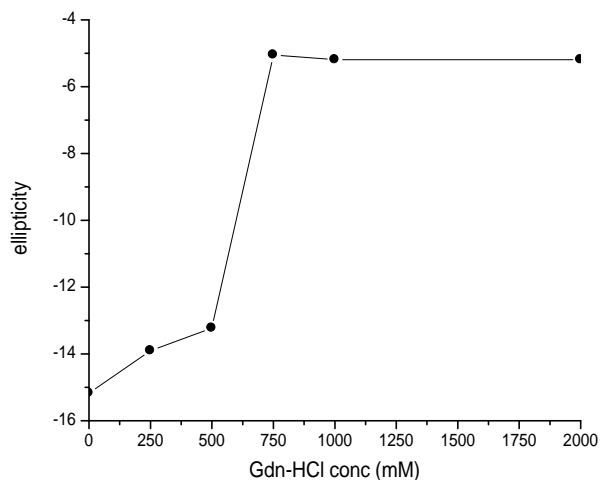
Figure 5.6: The ellipticity at 218 nm of Far UV CD of Gdn-HCl treated samples.

The drastic inactivation of the enzyme in presence of 0.5-0.75 M Gdn-HCl can be due to modification of carboxylate group at the active site of *Bl*BSH. The structure has shown one Asp residue in the active site region of protein (Kumar et al., 2006). The positively charged guanidium group of Gdn-HCl can interact electrostatically with the carboxylate group of Asp causing inactivation.

From figure 5.5b and Figure 5.6 the midpoint value of Gdn-HCl measured by fluorescence, inactivation and negative ellipticity are 1.7, 0.5 and 0.7 M, respectively. The inactivation occurs before any detectable conformational changes taking place in the enzyme molecule. Gel filtration data of the enzyme in presence of 1M Gdn-HCl showed two peaks, one that of tetramer and the other that of dimer (Figure 5.7).
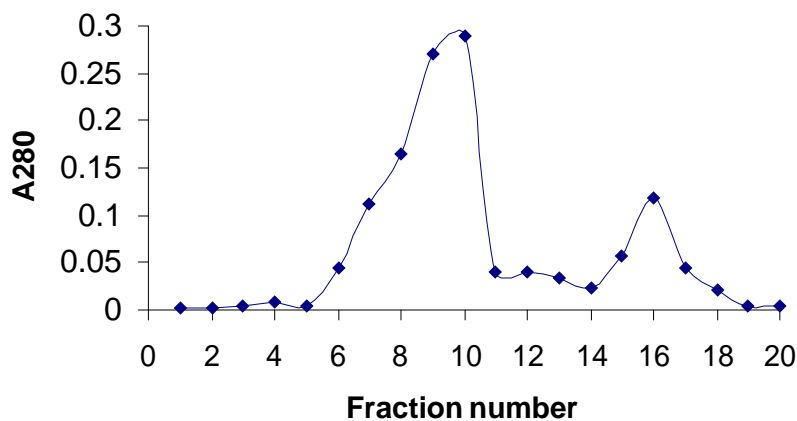
Figure 5.7 : Gel filtration profile of *Bl*BSH in presence of 1M Gdn-HCl.

However, dissociation need not be the only cause of inactivation. The loss of enzyme activity can also occur because of minor structural changes if the active center is too sensitive to conformational variations of the enzyme.

To reactivate the enzyme the denaturant in the reaction mixture was diluted and incubated for 1 h. The activity and the spectroscopic features of the enzyme were subsequently monitored. The enzyme incubated in 250 mM of Gdn-HCl there was gained 72% of its original activity and 500mM of Gdn-HCl showed 37% of its original activity. Beyond this concentration no reactivation of the enzyme was observed indicating irreversibility of the inactivation. However, the spectrum of Gdn-HCl denatured protein with $_{max}$ at 350 nm changed into a plateau in the region 340 to 350 nm, after renaturation indicating that partial shielding of Trp residue from the solvent can be due to a misfolded structure. Thus, both inactivation and unfolding of the protein are irreversible.

5.2.3. Effect of Gdn-HCl at different temperatures

Rayleigh light scattering studies are used to monitor the formation of aggregates in the protein solution under different denaturing conditions. Exposure of the enzyme at 55 °C in presence of different concentrations of Gdn-HCl showed high light scattering intensity when Gdn-HCl concentration was increased to 1 M indicating the formation of insoluble aggregates (Figure 5.8). In the presence of 0.25 M and 0.5 M Gdn-HCl, the protein showed high light scattering intensity when temperature was raised to 50 °C while in the presence of 0.75 M and 1M Gdn-HCl fast aggregation took place at a lower temperature of 40 °C. No aggregation was observed at higher temperature and at concentrations 2 M or above. At 0.25-0.75 M Gdn-HCl, progressive inactivation and distortion of secondary structure of the enzyme were observed (Figure 5.9). No aggregation of the protein was observed due to thermal denaturation in the absence of Gdn-HCl. Thus, due to distortion in the secondary structure, protein turns into insoluble aggregate on heating. The Trp environment of the *Bl*BSH intermediate structure formed in the presence of 0.25-1 M Gdn-HCl has not changed. But the enzyme progressively lost catalytic activity and secondary structure. The protein under these conditions was susceptible to forming insoluble aggregates.
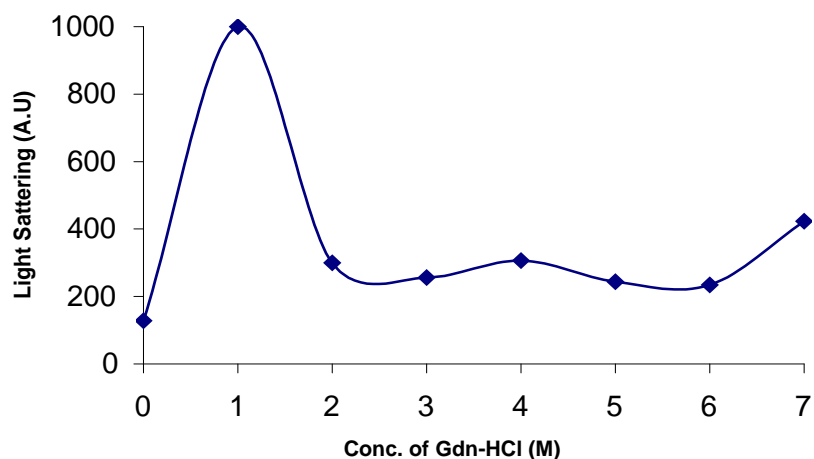
Figure 5.8: Light scattering of BSH in presence of Gdn-HCL at 55 °C.
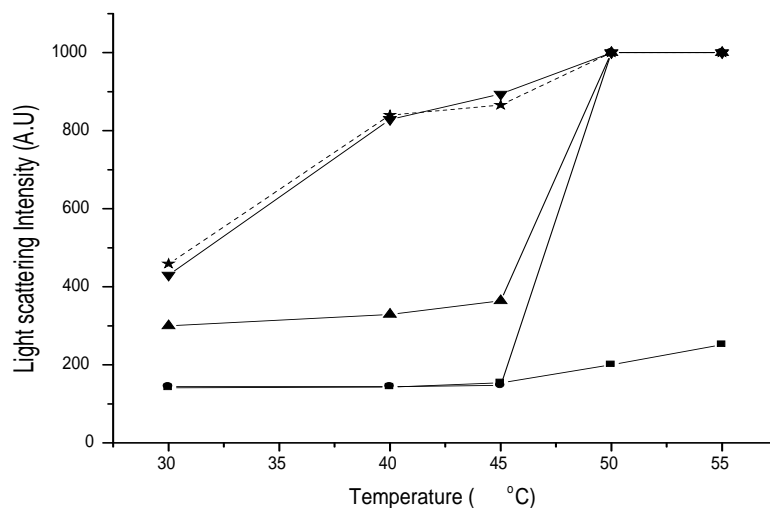
Figure 5.9: Light scattering of BSH in presence of Gdn-HCl at different temperatures. Protein in the absence of (|) and in presence of 0.25 ( ), 0.5 (^), 0.75 (˘) and 1 M (*) Gdn-HCl was incubated at different temperature for 10 min.

The aggregation process usually involves conformational change of the whole protein or of a specific domain and it is often attributed to the association of partially unfolded/folded molecules (Vetri & Militello, 2005). Aggregation process appears to

strictly depend on temperature and starts only after a rearrangement of the protein structure (Stirpe et al., 2005). Multiple unfolded states have been reported in the case of glutathione transferase (Sacchetta et al., 1999). Reversible equilibrium of unfolding of triophosphate isomerase from *Trypanosoma cruzi* in Gdn-HCl has been reported (Chanez-Cardenas et al., 2005). Irreversible unfolding of -amylase was analysed by unfolding kinetics. For all -amylases the irreversible process was found to be fast and the preceding unfolding transition was the rate-limiting step (Duy & Fitter, 2005).

5.2.4. Effect of pH on the structure and function of *Bl*BSH
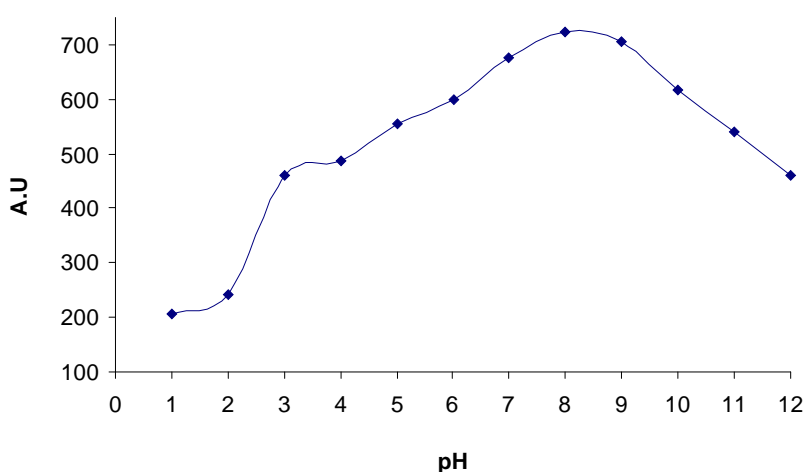
Figure 5.10: Fluorescence intensity at 330 nm with respect to different pH

The fluorescence intensity in different pH at 330 nm is shown in figure 5.10. The *Bl*BSH was irreversibly inactivated under extreme pH conditions. At extreme alkaline pH the protein did not bind ANS, while at extreme acidic pH there was strong ANS binding (Figure. 5.11). The binding of hydrophobic dye (ANS) to *Bl*BSH showed an approximate increase in fluorescence intensity 4, 3 and 2 times the intensity of protein in that buffer at pH 1, 2 and 3, respectively, with simultaneous blue shift in λ_{max} from 520 to 490 nm. This indicated the exposure of hydrophobic patches in protein at the extreme acidic pH. There was major retention of the secondary structure in the pH range 1.0 – 3.0 as compared to pH 7.0, as observed in the far UV CD spectra (Figure 5.12a). CD spectrum of the native *Bl*BSH in far UV region shows minimum at 218 nm and a shoulder at 208 nm indicating presence of prominent β and a small fraction of α structure in the enzyme.

The three-dimensional structure of *Bl*BSH is reported to have αββα folding pattern (Kumar et al., 2006).
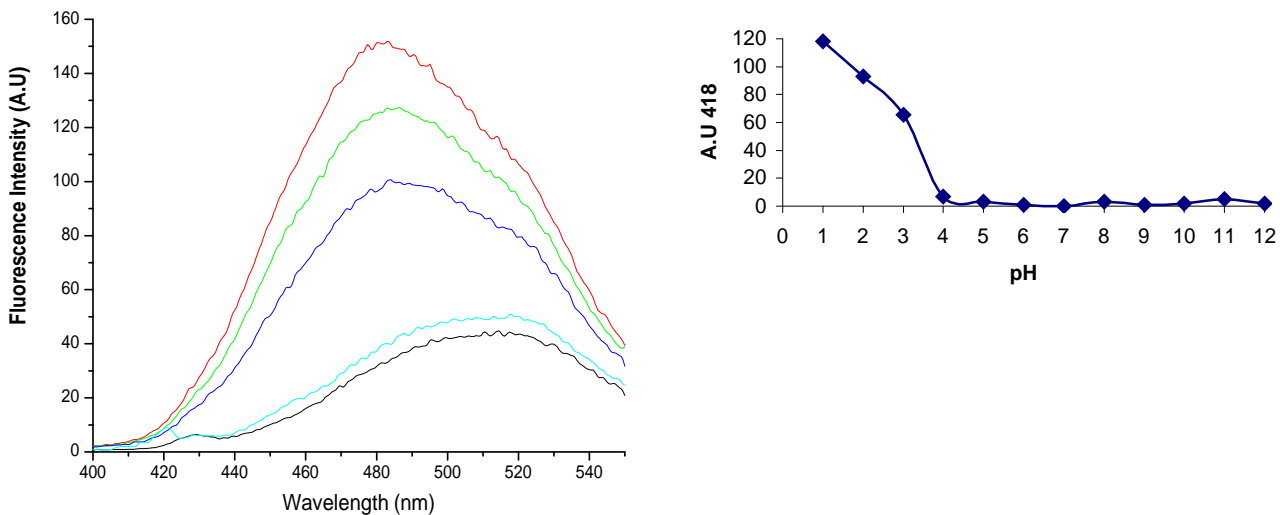
Figure 5.11: a) ANS binding of BSH at different pH. Fluorescence spectra of buffer containing ANS (|), protein samples at pH1 (|), pH2 (|), pH3 (|).and pH4 (|) excited at 375 nm. b) AU 418 at different pH

Figure 5.12b shows the near UV CD spectra of *Bl*BSH at different pHs. Spectrum of the native enzyme shows that it contains ordered tertiary structure, which is disturbed at extreme acidic pH. During this change, the hydrophobic patches must be getting exposed on the surface of protein, thereby binding the hydrophobic dye. Thus a compact secondary structure, major distortion in the tertiary structure, exposure of hydrophobic patches and loss of activity, together show as indications of the existence of molten-globule like structure in the extreme acidic pH. Since some tertiary structure was still present at pH 1.0 the conformation can be called as an incomplete molten globule. The structural changes at pH 1.0 and 2.0 seem to be reversible since readjusting the pH from 1.0 to 7.0 reduced the ANS binding capacity of the protein almost 4 fold.
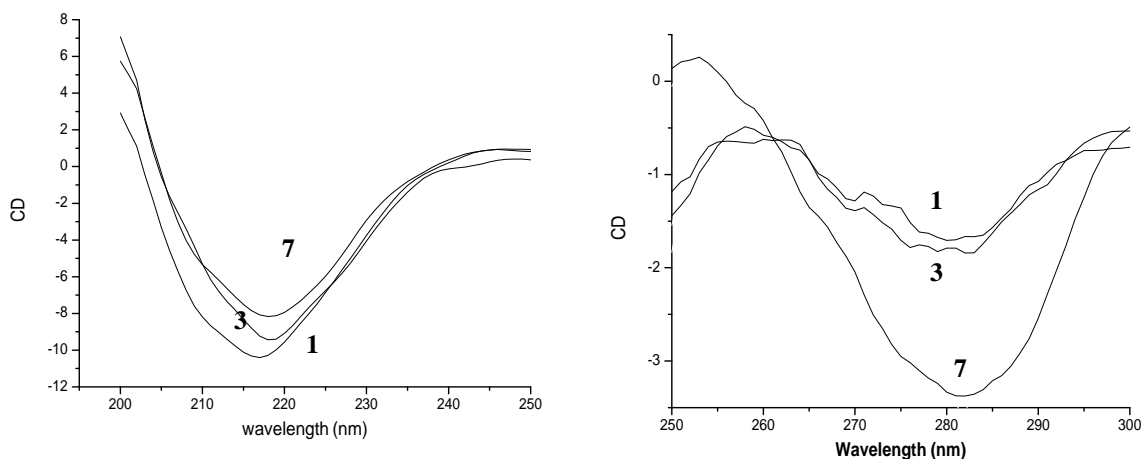
Figure 5.12: a) Far UV CD spectrum of BSH at pH 1.0 , 3.0 and 7.0

b) Near UV CD spectrum of BSH at pH 1.0, 3.0 and 7.0

The fluorescence spectrum of the renatured protein (Figure 5.13) also showed increase in the fluorescence intensity as well as blue shift in the λ_{max} from 350 to 342 nm. Thus structural changes at pH 1.0 and 2.0 are partially reversible as the exposed hydrophobic patches are buried again after adjusting the pH to 7.0. However, the refolding may be incorrect as partial shielding of Trp from solvent was observed in the renatured sample. The partially renatured structures in case of thermal and pH denatured enzyme are different from the renatured structures of Gdn-HCl treated enzyme as they have the peaks with λ_{max} at 342 nm.
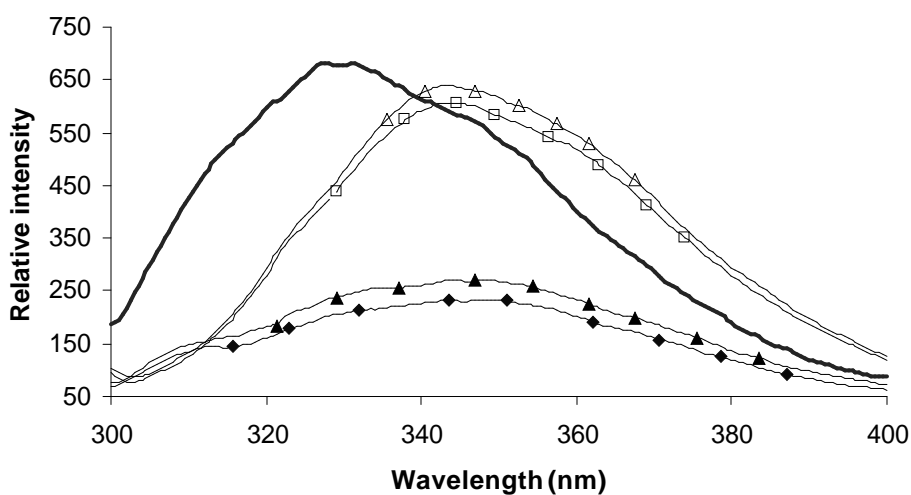
Figure 5.13: Fluorescence spectra of renatured *Bl*BSH with respect to pH. pH 7 (-)

pH 1 (|) pH 2 (^) readjusted sample of pH1 to 7 ( ) pH2 to 7 ( ).

When light scattering studies were carried out on protein samples incubated at different pH, *Bl*BSH showed intense light scattering at 45 °C in the pH range 4-5 indicating the presence of insoluble aggregates (Figure 5.14). The estimated pI of the *Bl*BSH is 4.3. Thus with net charge close to zero on the surface, the protein molecule aggregates immediately on heating.
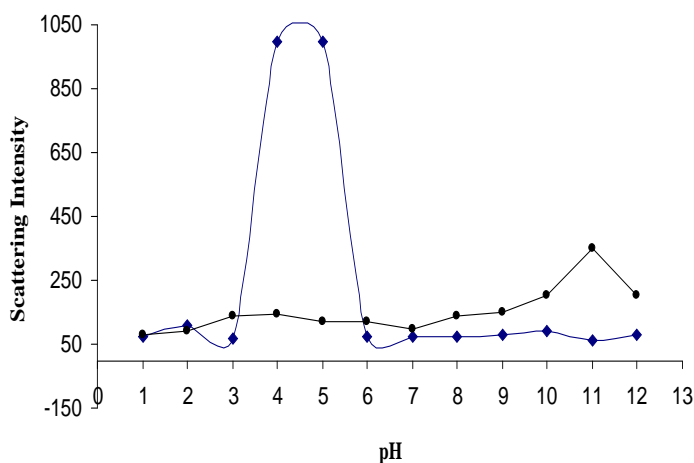
Figure 5.14: Rayleigh light scattering of BSH at different pH at 30 °C ( ), 45 °C (|)

5.3. Discussion

The conformational stability of proteins is studied by denaturation of protein using different agents such as temperature, pH, chaotropic agents and pressure, either individually or in combination. Proteins, due to delicate balance of stabilizing and destabilizing interactions, are marginally stable. Denaturation is followed by monitoring the changes in any measurable physical property of the protein such as intrinsic fluorescence, CD spectra. With every change the corresponding change in protein function is checked by measuring the enzyme activity. The summery of the result for denaturation of *Bl*BSH has been given in Table 5.1

Table 5.1: Denaturation of *Bl*BSH under various conditions.

Conditions	Activity	Fluorescence	Secondary structure	Tertiary structure	ANS Binding	Aggregation
Thermal (55°C,15min)	No	Decreased Intensity	Distorted	Distorted	No	No
Acid induced (pH 1-3)	No	Unfold	Intact	Distorted	Strong binding	No
pH 4-5 at 55°C	No	Decreased intensity	Distorted	-	No	+
Gdn-HCl (1M)	No	Folded	Distorted	Intact	No	+ (at 55°C)

Acids and bases are known to denature proteins by disrupting their electrostatic interactions. Protonation of all ionisable side chains below pH 3.0 leads to charge-charge repulsion and consequently protein unfolding. This induces loosening of side chain packing and the exposure of hydrophobic groups to solvent. Several proteins have been shown to exist in molten globule form at extremely acidic pH. glucose/xylose isomerase (Pawar & Deshpande, 2000), -crystallin (Rajaraman et al., 1999), xylanase (Nath & Rao, 2001), glutathion transferase (Luo et al., 2002), and glucose oxidase (Khatun Haq et al., 2003) are some of the examples. Molten globule may be one of the first conformations embraced by the polypeptide chain in folding from the unfolded state (Vamvaca et al., 2004). It has been suggested that the molten globule could be the state in which non native proteins exist in living organisms and thus, is important for physiologically important processes such as translocation and interaction with chaperones and release of ligands.

Thermal and chemical denaturation inactivate the enzyme due to minor changes in the conformation. The irreversibility of the structural changes becomes obvious when renaturation of protein does not yield the correctly folded original structure Combination of two denaturants lead to protein aggregation, which is not observed when treated with each denaturant separately to the same extent. The changes in the conformation of

enzyme with increasing temperature in the presence of low concentrations of Gdn-HCl lead to formation of insoluble aggregates. The biological significance of the labile nature of the active center of *Bl*BSH might be decided by the nature of its functional environment and its involvement in the reduction of serum cholesterol level.

CHAPTER 6

BIOCHEMICAL CHARACTERIZATION OF *Kc*PGA AND STUDY OF ITS POST-TRANSLATIONAL PROCESSING

6.1 Introduction

This chapter discusses the preliminary biochemical characterization and active site residue modification studies of PGA from *Kluyvera citrophila* (*Kc*PGA). This hydrolase is produced as an inactive precursor that undergoes post-translational modification to turn to active form. During purification of *Kc*PGA along with the processed form of the enzyme the 96 kDa precursor was also found present in the preparation. The presence of this precursor was confirmed by chromatographic and light scattering experiments. The preparation was maintained in different pH and temperature for various time durations to achieve complete *in vitro* processing.

6.1.1. Post-translational processing of *Kc*PGA

The important rate-limiting step in the production of active enzyme is the efficiency of proteolytic processing of the precursor. The temperature dependent proteolytic processing has been already reported (Oh et al., 1987, Lindsay & Pain, 1991). Surprisingly, this processing has a lot of resemblance with the processing of some higher eukaryotic peptide preprohormones. This processing phenomenon, very rare in prokaryotic systems, is also observed in *Bradyrhzobium japonicum* cytochrome *bc,* (Trumpower, 1990) *Bacillus subtilis* spore coat proteins (Aronson et al., 1989) and *Bacillus polynyxa* amylase (Uozumi et al., 1989).

Considerable progress has been made in the understanding of post-translational modification process in prokaryotic systems (Thony-Meyer et al., 1992). This process constitutes two essential steps, the translocation of precursor to the periplasmic membrane and the processing of precursor by proteolytic cleavage (Figure 6.1)

A) For translocation, especially in *E. coli,* there is Sec translocase that selectively export the protein which contains some signature in the signal sequence (Economou, 2000). Sec system required the sec chaperone for the translocation

process (Chou et al., 1999). Another major pathway identified in the protein translocation is the twin arginine translocation (Tat) pathway, completely Sec-independent and is involved in the translocation of proteins that are incompatible with the Sec translocase (Ignatova et al., 2003).

B) For the proteolytic cleavage an enzyme DegP was identified. It is found in *E. coli* periplasm and function simultaneously as the protease and chaperone (Lin et al., 2001).

Still a lot of speculation exists regarding the exact mechanism of post-translational processing in the prokaryotic system. Varying the temperature, media composition, and point of induction are found to affect the processing towards mature enzyme. Dynamic light scattering (DLS) is used to measure and monitor the intensity fluctuations in scattered light as a function of time. This technique is one of the most popular methods used to determine the hydrodynamic radius (Rh) of sub-micron sized molecules such as proteins. The hydrodynamic radius is the effective size of the molecule as they undergo constant diffusion or so-called Brownian motion in the solution. The measured intensities caused by the movement of protein particles in the solution are used to determine the diffusion coefficient and subsequently the hydrodynamic radius of the protein molecules. Thus, the polydispersity and molecular weight of the protein sample can also be estimated (Ferre-D'Amare & Burley, 1997).

This study also intend to define the operational condition for the phase of inoculum preparation (germination phase), aims at reducing fermentation time and achieve optimization for the production of enzyme by choosing quantity of inducer like PAA. The production of precursor protein in wild type organism and the standardized method for removing the precursor protein are described. The possible condition for the *in vitro* processing of the precursor to active enzyme is also defined.
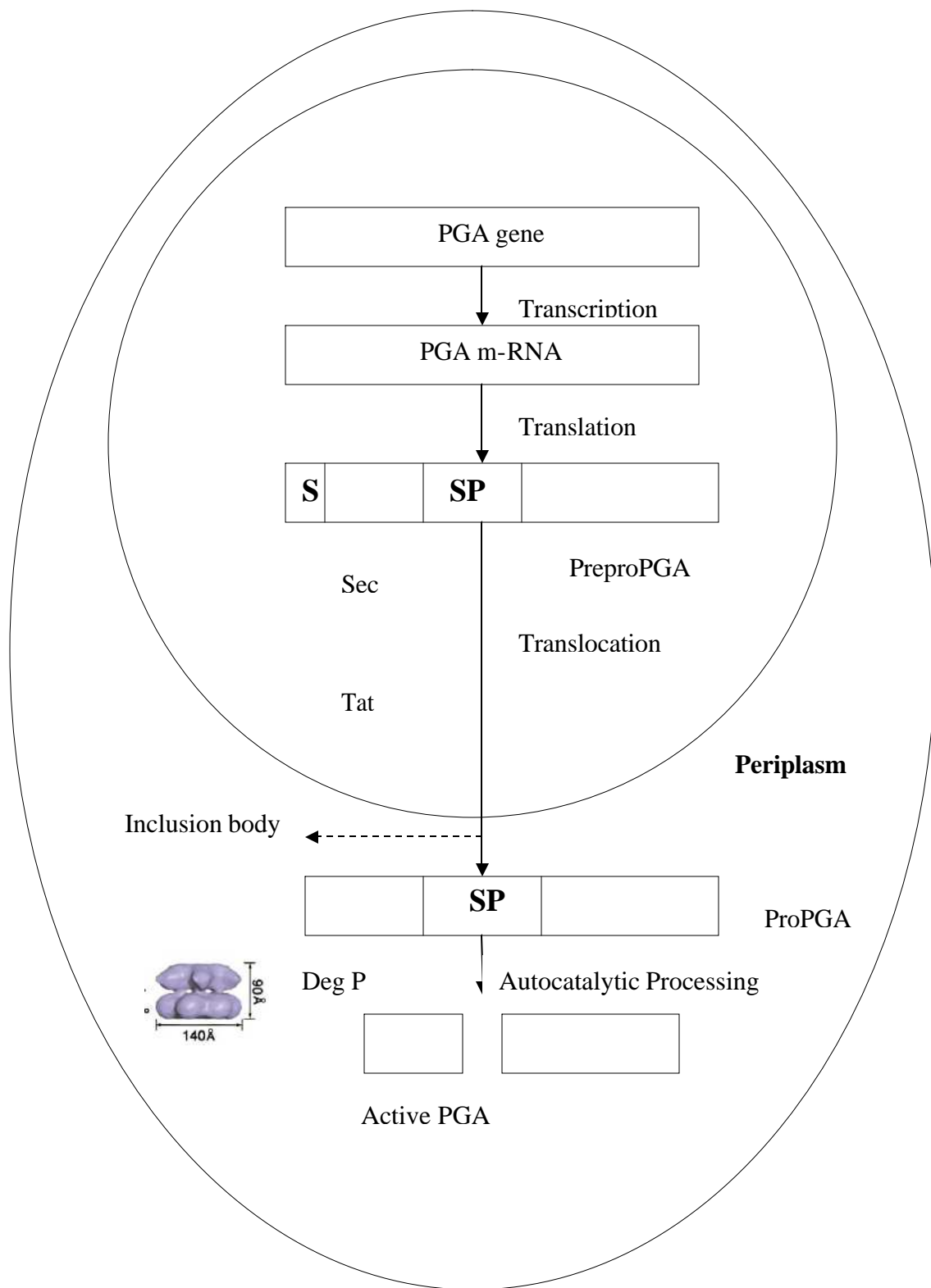
Figure 6.1 : Present holistic view of PGA processing in prokaryotes

6.2. Cultural conditions for the production of PGA from *K. citrophila*

After testing different strategies and varying the composition of the inoculum for growing *K. citrophila* to produce PGA, the studies showed that 24 h is enough for the microorganism to germinate adequately. The primary results, shake flank conditions, indicate that the production phase should go for 48 h, in order to reach the highest overall productivity of the cultivation in a bioreactor (Table 6.1). From our investigations it was found that a concentration of 25 mM PAA is optimum for the induction of maximum amount of PGA without being detrimental to cell growth. Above this concentration the increase in enzyme activity was accompanied with considerable decrease in cell growth (Table 6.2). The expression of the *pga* gene in *E. coli* wild-type strains is induced by PAA and repressed by glucose. These two compounds seem to affect gene expression at the transcriptional level.

Following theory can explain the principle behind PGA induction by PAA. Two major processes are involved in the synthesis and maturation of PGA, a post-translational intracellular proteolysis of the precursor of PGA to remove signal peptide, which is followed by processing of spacer peptide required for the formation of a mature enzyme. More than 90% of the precursor PGA produced is lost during the proteolysis in the cytoplasm. A higher concentration of glucose in the culture medium or an increase in cultivation temperature results in increased proteolytic loss of precursor PGA whereas the addition of PAA has the opposite effect of increasing the amount of correctly processed PGA in the cell.

Table 6.1: Inoculum and production phase for the optimum *Kc*PGA production.

Germination phase		Production Phase		
Period H	Growth g/L	PGA total Activity	Productivity	Product/biomass Yield (IU/g cell)
24 h				
24	2.87	36.07	2.87	12.57
48	3.24	45.50	3.24	14.04
72	3.45	32.80	3.45	9.51
48 h				
24	3.42	38.60	3.42	11.29
48	3.62	48.90	3.62	13.51
72	3.54	31.20	3.54	8.81
72 h				
24	2.96	34.5	2.96	11.66
48	2.83	32.1	2.83	11.34
72	3.14	32.3	3.14	10.29

Table 6.2: Effect of PAA on the *K.citrophila* cell growth and PGA production

Concentration of PAA (mM)	Growth (g/L)	Activity (μM/h/mL)
Control	3.14	28.34
10	3.24	41.40
25	3.93	48.20
50	3.62	43.45
75	3.60	42.14
100	3.43	43.76

6.3. Purification of *Kc*PGA

The enzyme was purified to homogeneity by the procedure summarized in Table 6.3. The purification of *Kc*PGA is complicated by the presence of precursor protein in the purified fraction. It is very difficult to remove the precursors from mature enzyme by methods such as gel filtration and isoelectric focusing owing to their close similarity. The protein purification and separation needs special type of ionic separation technique using a hydroxy apatite column. The principle behind separation in hydroxy apatite column is still unknown and needs further investigation. Hydroxy apatite is a form of calcium phosphate that often resolves complex substances when all other separation methods fail because its separation capability does not primarily dependent on the usual characteristics of protein (molecular weight, size and isoelectric point etc.). Non-specific adsorption of hydrophobic substances is minimized by the inorganic crystalline hydroxy apatite. It has negligible adsorptive capacity for low molecular weight substance such as mononucleotides, salts and amino acids.

Experimental evidence indicates that interaction between negatively charged moieties on the protein surface and Ca^{++} groups on the hydroxy apatite plays an important role in protein adsorption. The relative binding affinities of different proteins are evidently a function of the ratio of change of their surface charge density to mass and other factors which account for the unique fractionation properties of hydroxy apatite. Proteins are usually adsorbed on hydroxy apatite in a low ionic strength phosphate buffer at neutral pH and eluted often in a narrow phosphate range by increasing ionic strength. This chromatography technique was used to yield a monomer of bovine serum albumin, which contain no detectable heavier aggregates.

A major portion of the precursor eluted before the elution of mature active fractions of PGA (Figure 6.2). The fractions were checked using SDS-PAGE and those fractions containing pure processed protein were pooled and concentrated. As expected, there was no activity in the precursor protein fraction. The yield and amount of processed and precursor proteins varied from batch to batch. The factor influencing the processing of the protein may be attributed to the culture condition and the time taken for protein purification. The processed and precursor PGA are in equal amounts. It is already reported that the improper folding of protein leads to their aggregation in the cell which

192

ultimately causes intracellular proteolysis in the cytoplasm and the resultant degradation of more than 90% of the synthesized preproPGA (Ignatova et al., 2000). Because of all these reasons about 0.5 mg of the active *Kc*PGA only could be purified from 1 L of the culture.
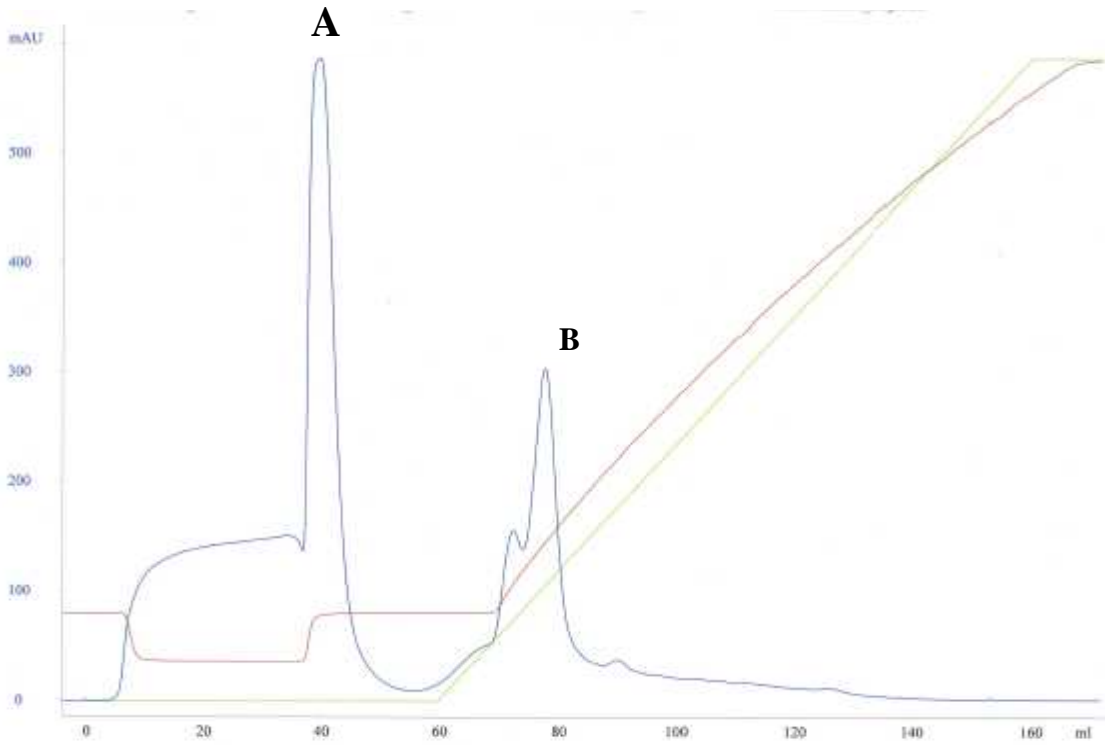
Figure 6.2 : Elution profile of hydroxy apatite chromatography of *Kc*PGA

The peaks are A: processed *Kc*PGA and B: precursor

Table 6.3: Purification steps and yields of *Kc*PGA

Fraction	Vol (mL)	Total Protein (mg)	Specific Activity (Units/mg)	Total Activity (Units)	Recovery %
Sonication	100	1172	0.12	140	100
AS precipitation	20	928	0.14	130	92
Octyl Sepharose	24	40	2	80	57
Hydroxy apatite	18	0.5	18	9	10

6.4. *In vitro* Processing

The 96 kDa precursor was always present in almost equal amount with processed form through all the purification steps. The presence of precursor was identified and characterized using light scattering experiments. Initial DLS experiments were performed using the precursor and matured protein together in the same buffer. The intensity-weighted fitting showed the presence of some aggregates of precursor (Figure 6.3A), which were shown by vanishing proportion of the aggregate population (Figure 6.3B). The presence of precursor somehow stimulates microaggregation of the protein. So a range of peaks corresponding to entities of different molecular weights was observed in the DLS experiment (Figure 6.3A). Many biocompatible detergents such as TMAO, NDSB etc. were tried to reduce aggregation and increase solubility, but without success. The precursor after separating from the mature enzyme was incubated under various pH, temperatures and buffer concentrations. Irrespective of the condition tried the sample incubated upto 24 h showed no increase in activity. The best results were obtained when the protein was incubated in 50 mM Tris-HCl pH 8.0 at 30 ºC for 48 h (Table 6.4). SDS-

PAGE showed that the heterodimer obtained by the processing of the precursor had the subunits of size 24 kDa and 62 kDa (Figure 6.4). This also rules out involvement of any external factor such as a protease or a chaperone in the maturation of PGA. Thus, the result provides evidence for the autocatalytic processing of the precursor.

Table 6.4:. Effect of various conditions of buffers (Phosphate pH 6 & 7 and Tris pH 8), temperatures and reaction times in the processing of *Kc*PGA is listed.

PH	Time (h)	Temperature (°C)	Activity (μM/h/mL)
Control (pH 7.5)	-	4	4.78
6	24	30	8.67
6	48	30	8.43
7	24	30	8.84
7	48	30	14.98
8	24	30	20.38
8	**48**	**30**	**34.98**
8	120	30	31.73
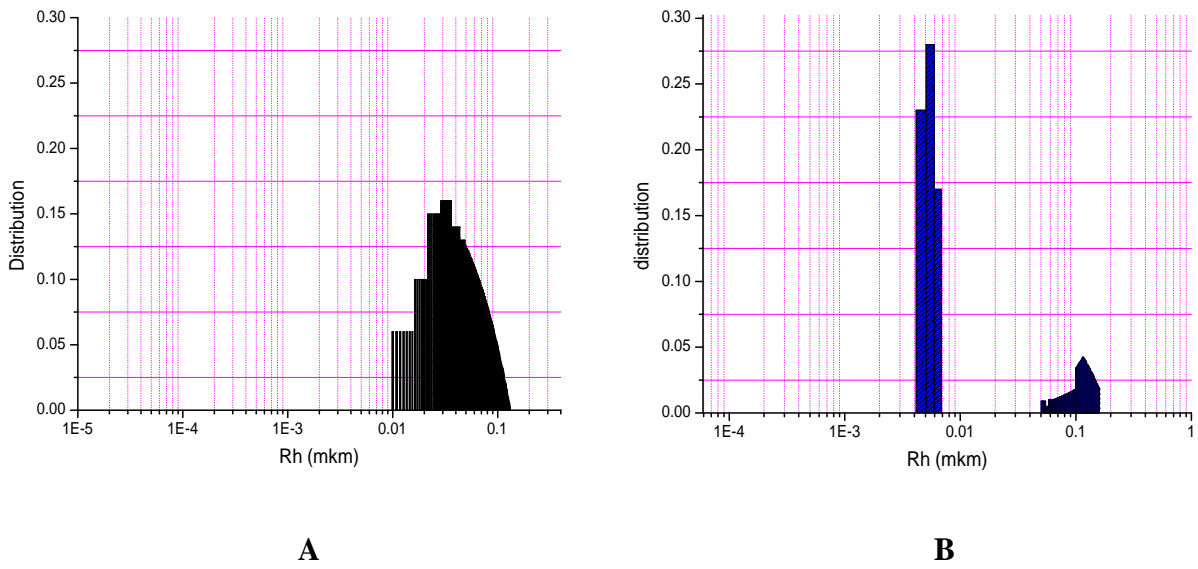
Figure 6.3: Dynamic light scattering (DLS) experimental plots. A) Precursor of *Kc*PGA showing the state of micro-aggregation, B) Processed *Kc*PGA
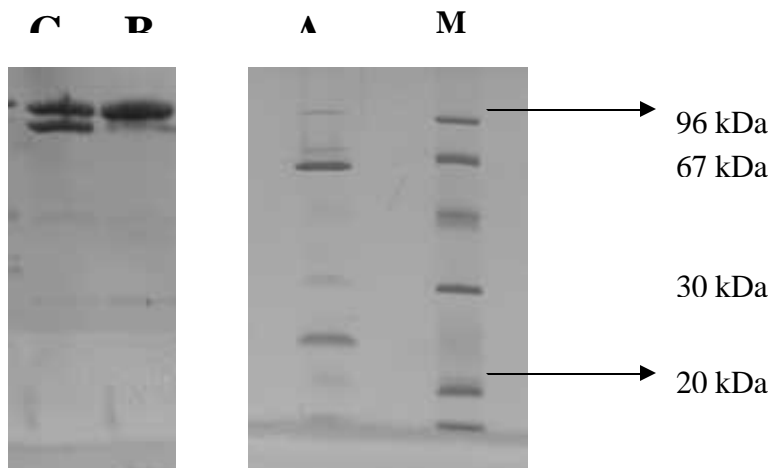
Figure 6.4: SDS – PAGE gel, the lanes contain *Kc*PGA in different forms. A: processed PGA, B: prepro-PGA, C: prepro-PGA and precursor, M: markers

6.5. Physicochemical properties of *Kc*PGA

*Kc*PGA is a heterodimer of two subunits α and β of molecular weights 24,000 Da and 62,000 Da, respectively (Alvaro et al., 1992). The active site topography has been investigated mainly using specific chemical modifiers since no three-dimensional structure of this enzyme is reported yet.

The active site of the enzyme has drawn special attention due to its importance in understanding the catalytic mechanism and substrate specificity for application in protein engineering (Roa et al., 1995; Martin et al., 1993). Since the three-dimensional structure of the *Kc*PGA is not available, investigation of the active site topography has been performed mainly using specific chemical modifiers. It is important to identify the residues that influence the specificity, activity and stability of the enzyme. The residues Arg 145, Phe 146 and Tyr 31 have been previously identified to participate in substrate binding (Done et al., 1998; Kumar et al., 2004; Kazan et al., 2001). Most of the enzymes possess certain amino acid residues in their active sites that are directly involved in catalysis. It is reported that generally residues such as tryptophan, cysteine, lysine, histidine, serine, arginine, aspartate and glutamate residues are found in the vicinity of active sites of enzymes (Roig & Kennedy, 1992). Several methods have been described in the literature for the identification of the catalytically essential amino acid residues of a particular enzyme by measuring the kinetic parameters of enzymes at different pH values; X-ray structural analysis and substrate specificity studies are examples. In cases where the enzyme is available in limited quantities; the chemical modification using amino acid specific reagents seems to be a convenient approach to identify amino acids directly or indirectly involved in catalysis (Roig & Kennedy, 1992).

Active site chemical modification studies of arginine and tryptophan have been carried out along with kinetic analysis, fluorescence spectroscopy, homology modeling, molecular dynamics and bioinformatics techniques that reveals the importance of particular Arg and Trp residues in substrate binding of PGA enzymes (Kumar et al., 2004; 2006b).

6.5.1. Optimum pH for activity and pH stability

The enzyme had optimum activity at pH 6.0 and the relative activities at pH 5.0 and 7.0 were 39% and 42%, respectively (Figure 6.5). KcPGA showed comparatively high pH stability and retained activity significantly ($< 70\%$) up to pH 5.5 and 9.0
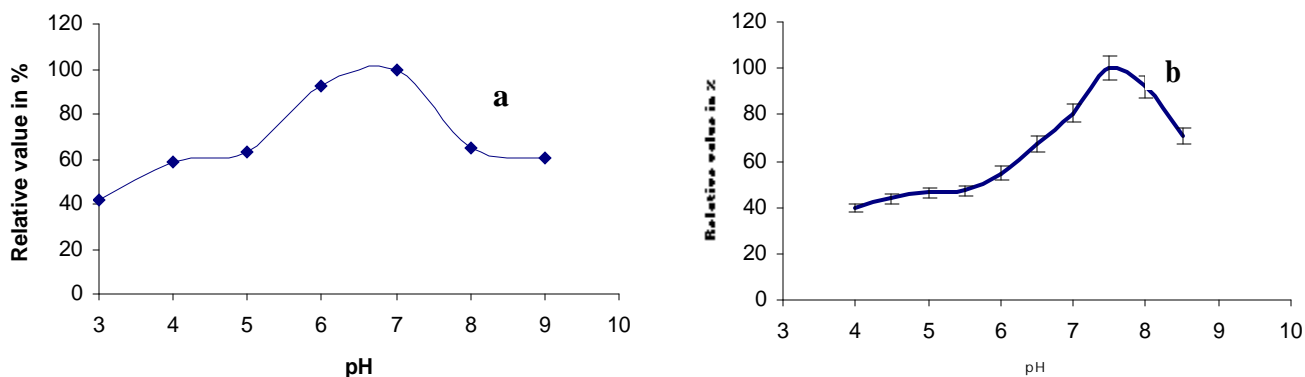
Figure 6.5: Profile of, optimum pH: plot-a and pH stability: plot-b

6.5.2. Effect of metal ions and EDTA

Among the divalent cations tested for their effects on PGA activity, Hg^{++} completely inactivated the enzyme whereas Cu^{++} inhibited the enzyme activity by 17%. None of the other metals and ions and EDTA (10 mM) had any significant effect on the enzyme activity.

6.5.3. Optimum temperature for activity and temperature stability

The optimum temperature for enzyme activity was 42 ºC. The enzyme showed high thermo stability and retained its full activity when stored at 60 ºC for 2 h, above that temperature there was considerable drop in activity (Figure 6.6).
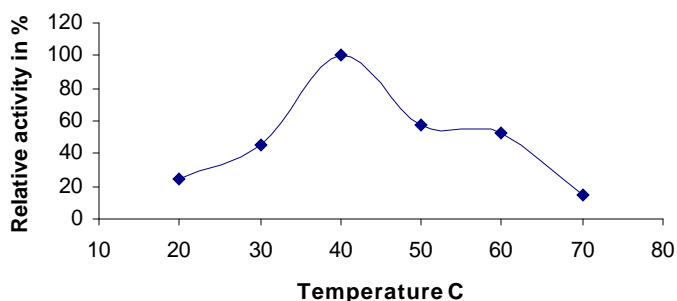
Figure 6.6: Activity of KcPGA plotted against temperature

6.6. Arginine modification

The treatment of *Kc*PGA with 20 mM phenylglyoxal in 50 mM phosphate buffer at pH 7.5 and 25 °C reduced the activity by 82% as compared to unmodified enzyme. 78% inactivation was observed when treated with 2,3-butadione under similar conditions. The inactivation was dependent on the time of treatment and concentration of the modifying agent. By blocking the active site with 50 mM penG or 50 mM phenylacetate the enzyme is protected against inactivation, indicating that the modified residue is present in the active site (Table 6.5). Phenylacetate is known to act as a competitive inhibitor of PGA and evidence is reported for its binding at the active site of the protein (Prabhune & SivaRaman, 1990).

Table 6.5: Protection of *Kc*PGA against inactivation by arginine-specific reagent

Treatment [a]	Enzyme activity[b] (% initial activity)
None	100
phenylglyoxal (20 mM)	22
penG (50 mM) + phenylglyoxal (20 mM)	88
phenylacetate (50 mM) + phenylglyoxal (20 mM)	95

The inactivation followed a biphasic pattern with an initial rapid phase and a subsequent slow phase. This pattern was evident in all enzyme and modifier concentrations tested. The rapid-phase was linearly dependent on phenylglyoxal concentration. Semi-logarithmic plots of residual activity as a function of time of inactivation at various concentrations of phenylglyoxal (Figure 6.7) was linear up to 18% of the initial activity indicating that the inactivation follows pseudo-first order kinetics.
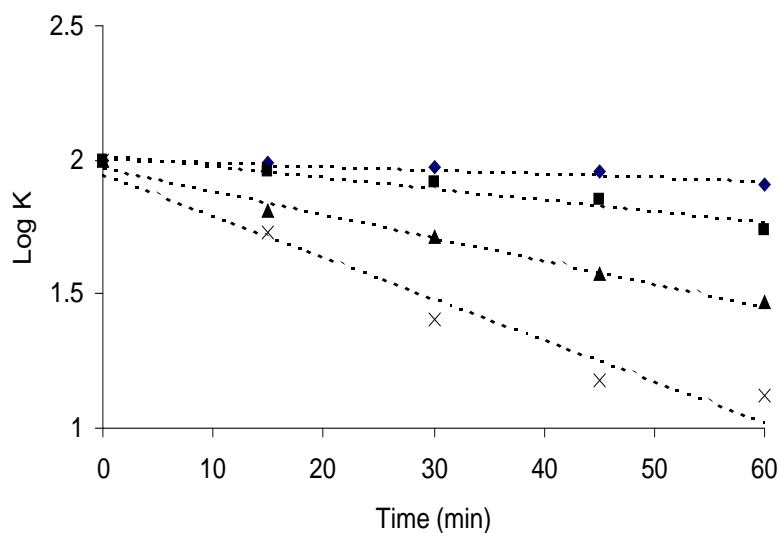
Figure 6.7: Inactivation of KcPGA using arginine-specific modifying reagent phenylglyoxal at concentrations: 4 ($\blacklozenge$), 8 (|), 12 (^) and 20 mM (**x**).
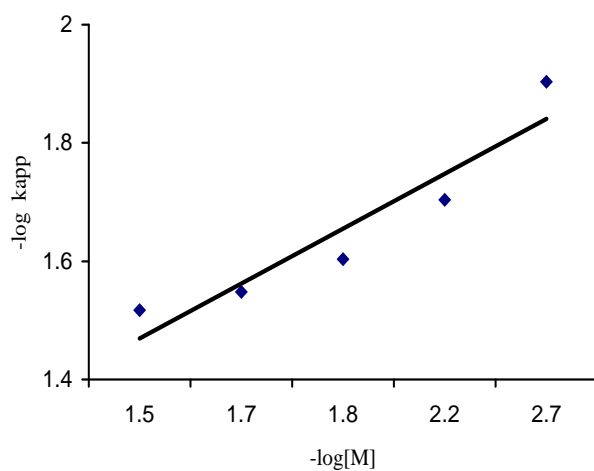
Figure 6.8: Determination of the order of reaction with respect to phenylglyoxal, where k_{app} is the apparent first order rate constant and M is the molar concentration of the reagent

6.6.1. Number of essential arginine residues

The inactivation kinetics of the native and the chemically modified *Kc*PGA were studied at 25ºC and pH 7.5. The number of arginine residues modified (n) by phenylglyoxal and 2,3-butanedione were determined from the plot of the logarithm of the

200

apparent first order rate constant, k_{app} versus the logarithm of the molar concentration of the modifier (Figure 6.8). The value of n indicates the loss of enzyme activity as a result of the modification of 0.92 arginine residue per molecule of PGA. The decrease in k_{cat} and K_m values indicates the change in catalytic activity due to arginine modification (Table 6.6). Titration with p-nitrophenyl glyoxal (pNPG) (Figure 6.9) also suggests the involvement of a single arginine residue in enzyme activity.

The total number of arginine residues estimated by treatment of protease-digested enzyme with pNPG is 30. This is in agreement with the number of arginines deduced from DNA sequence (Barbero et al., 1989). The treatment of undigested enzyme with pNPG showed presence of two arginine residues while undigested enzyme protected with 50 mM phenylacetate showed only one. This indicates that out of the 30 arginine residues in the protein only two are freely accessible, of which one is located in the substrate binding site. The conformity of CD spectra of the modified enzyme with that of the unmodified enzyme rules out any possibility of modifier mediated structural changes as cause of inactivation (Figure 6.10).

Table 6.6: Kinetic constants of native and chemically modified *Kc*PGA

Reagent concentration	V_{max} (U mL^{-1}min^{-1})	K_m (µM penG)	k_{cat} (s^{-1})
Native PGA	43	16	63
PGA+ 10 mM phenyl glyoxal	31	65	21
PGA+ 50 mM 2,3 butanedione	26	61	27

The V_{max} and K_m values of native and modified PGA were estimated from the Lineweaver-Burk plots by measuring the initial reaction rate at 40 ºC and pH 7.5. The values of k_{cat} were calculated from equation $V_{max} = k_{cat}$ [E] t.
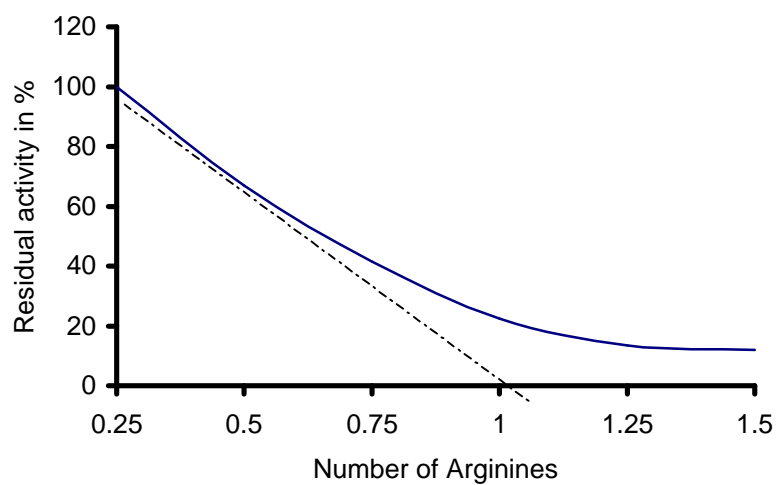
Figure 6.9: Titration of *Kc*PGA with p-nitrophenyl glyoxal.
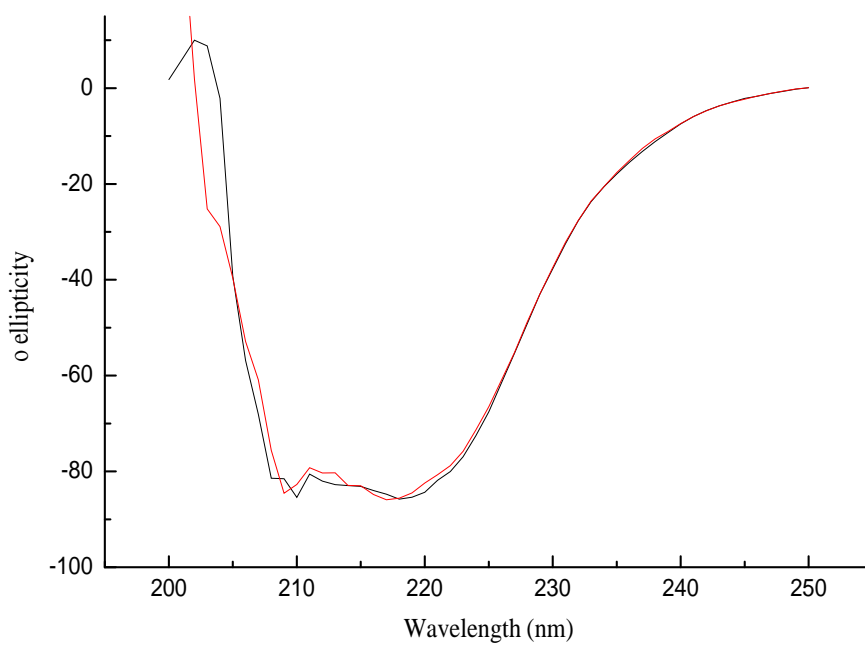
Figure 6.10: The CD spectra in 5 mM sodium acetate buffer of unmodified *Kc*PGA and chemically modified with phenylgloxal. unmodified :'|', Arg modified: '|'

6.7. Tryptophan modification by NBS and HNBB

Treatment of KcPGA with 100 μM NBS and 40 mM HNBB resulted in 72% and 60% loss of activity, respectively, that tentatively indicated the presence of Trp residues in the active site. Under the conditions of treatment, the HNBB can simultaneously affect both Trp and cysteine residues. However, absence of cysteine residues in KcPGA eliminates the ambiguity about the type of residue modified.

Spectrophotometric titration of KcPGA using 5-100 μM NBS at pH 5.5 resulted in a progressive decrease in absorption at 280 nm. The number of Trp residues oxidized per mole of the enzyme calculated by the extrapolation of tangent from the inflexion point of progression curve was one. Complete inactivation of the enzyme occurred upon modification of a single Trp residue. Figure 6.11a shows the time dependent inactivation of KcPGA. As can be seen, all the Trp residues that are accessible to modification get oxidized within the first 10 min of the reaction. The plots of the logarithm of residual activity versus time of incubation with various NBS concentrations were linear up to the end of test periods, indicating pseudo first-order kinetics of inactivation. The individual slopes of the plots were calculated to determine the respective first order rate constant k_{app}. The order of the reaction (n) in the case of NBS determined from plots of log [k_{app}] versus log (NBS concentration [M]) gave value n =1 indicating that the modification of a single Trp residue resulted in inactivation of a mole of the enzyme (Figure 6.11b).
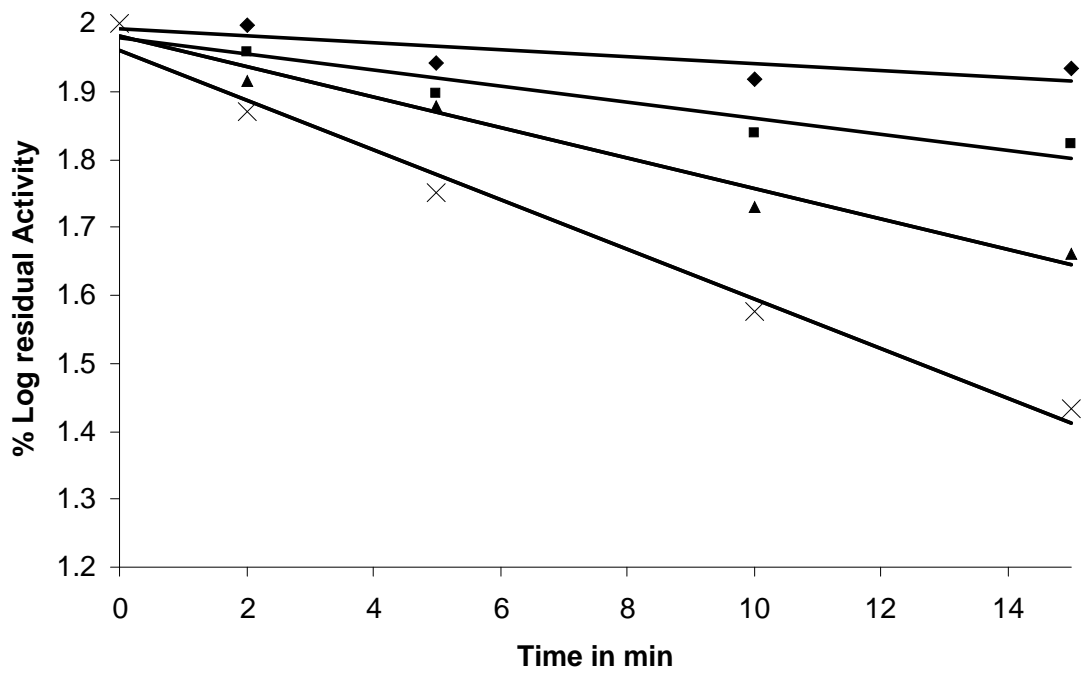
a
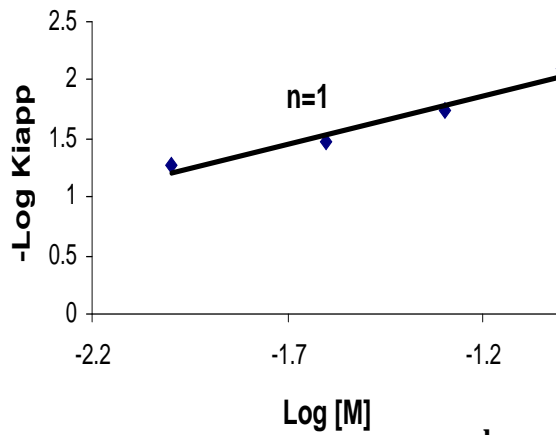
b

Figure 6.11: A) Determination of the order of modification reaction of *Kc*PGA with respect to tryptophan-specific modification reagent NBS at pH 5.5 and 25 °C. NBS concentrations were 10 (♦), 20 (|), 40 (^) and 100 µM (**X**). B) Kinetics of inhibition of enzyme by NBS. b) The pseudo-first order rate constants (k_{app}) were plotted against various concentrations of NBS.

Modification of *Kc*PGA with 40 mM HNBB showed time dependent decrease in residual activity and the maximum absorbance observed at 410 nm. Extrapolation of the tangent from inflexion point, in the plot of residual activity versus number of Trp residues modified, indicated modification of a single Trp residue per mole of the enzyme (Figure 6.12). No variation was detected in case of modification under denaturing conditions.
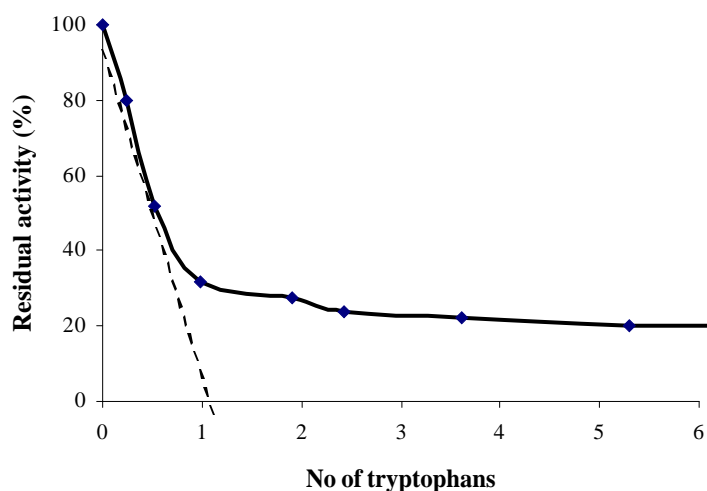
Figure 6.12: The titration of *Kc*PGA with NBS is plotted. Tryptophan residues were quantified by stepwise addition of NBS as described in the text.

6.7.1 Substrate protection studies

The protective action of penG (substrate) and phenylacetate (competitive inhibitor) on the inactivation of *Kc*PGA by NBS is shown in table 6.7. In the presence of 50 mM (highest concentration tried) each of penG and phenylacetate the percentage of original activity retained by *Kc*PGA was 74 and 68, respectively. Thus it is clear that both the substrate penG and the inhibitor phenylacetate affect the Trp modification to the same extent.

Table 6.7: Protection of *Kc*PGA against inactivation by tryptophan-specific reagent

Treatment	Enzyme activity (% initial activity)
None	100
N-Bromosuccinamide (100 µM)	30
penG (50 mM) + N-Bromosuccinamide (100 µM)	74
phenylacetate (50 mM) + N-Bromosuccinamide (100 µM)	68
HNBBr(20 mM)	42
penG (50 mM) + HNBBr (20 mM)	82
phenylacetate (50 mM) + HNBBr (20 mM)	85

6.7.2. Michaelis-Menten kinetics

The K_m values of untreated and NBS modified *Kc*PGA estimated were 16 µM and 53 µM, whereas the values of k_{cat} remained same (63 sec^{-1}) (Table 6.8).

Table 6.8: The values of K_m and k_{cat} for untreated and NBS or HNBS treated enzyme

Reagent concentration	K_m (µM penG)	k_{cat} (S^{-1})	k_{cat} / K_m (µM^{-1} S^{-1})
Unmodified *Kc*PGA	16	63	3.9
PGA+ N-Bromosuccinamide (100 nM)	54	63	1.2
PGA+ 2-hydroxy 5-nitrobenzylbromide (10 mM)	51	63	1.2

6.7.3. Circular Dichroism (CD) analysis

The shape of the CD spectra of modified *Kc*PGA was similar to that of untreated enzyme indicating that no serious structural changes occurred on treatment with NBS and HNBB (Figure 6.13)
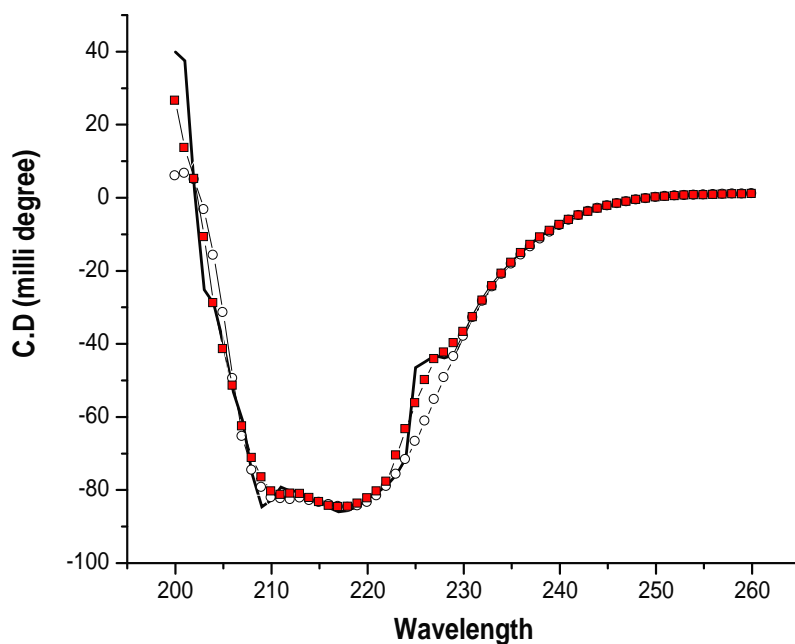
Figure 6.13: The CD spectra of untreated *Kc*PGA (0) and that chemically modified with NBS ( ) and HNBB(╎).

6.7.4. Fluorometric studies

Fluorescence spectra of untreated *Kc*PGA showed λ_{max} at 348 nm upon excitation at 280 nm. Addition of small increments of NBS and HNBB resulted in progressive lowering (55.5% decrease) of relative fluorescence from 366 to 160 AU (arbitrary units). A blue shift of λ_{max} from 332 to 327 nm was also observed at higher concentrations of Trp modifying reagents.

Incremental addition of PGA substrate (penG) and competitive inhibitor (phenylacetate) resulted in progressive decrease of relative fluorescence intensity from 208 to 178 AU (85.6% of original). The absorption spectra of these compounds do not

produce any peak in this region implying that the quenching is purely due to interaction with Trp and not due to secondary absorption by the ligands themselves. Red shift of λ_{max} from 348 to 357 nm was observed when tested with both the ligands. No fluorescence quenching was observed upon addition of penG to NBS modified KcPGA (Figure 6.14). However, a red shift 348 to 352 nm of λ_{max} was observed. The association constants for the binding of penG to untreated and NBS modified KcPGA are 7.94 X 10^2 M^{-1} and 2.43 X 10^2 M^{-1}, respectively. The change in value of standard free energy of binding (G°) for penG is -16.2 kJ mol^{-1}.
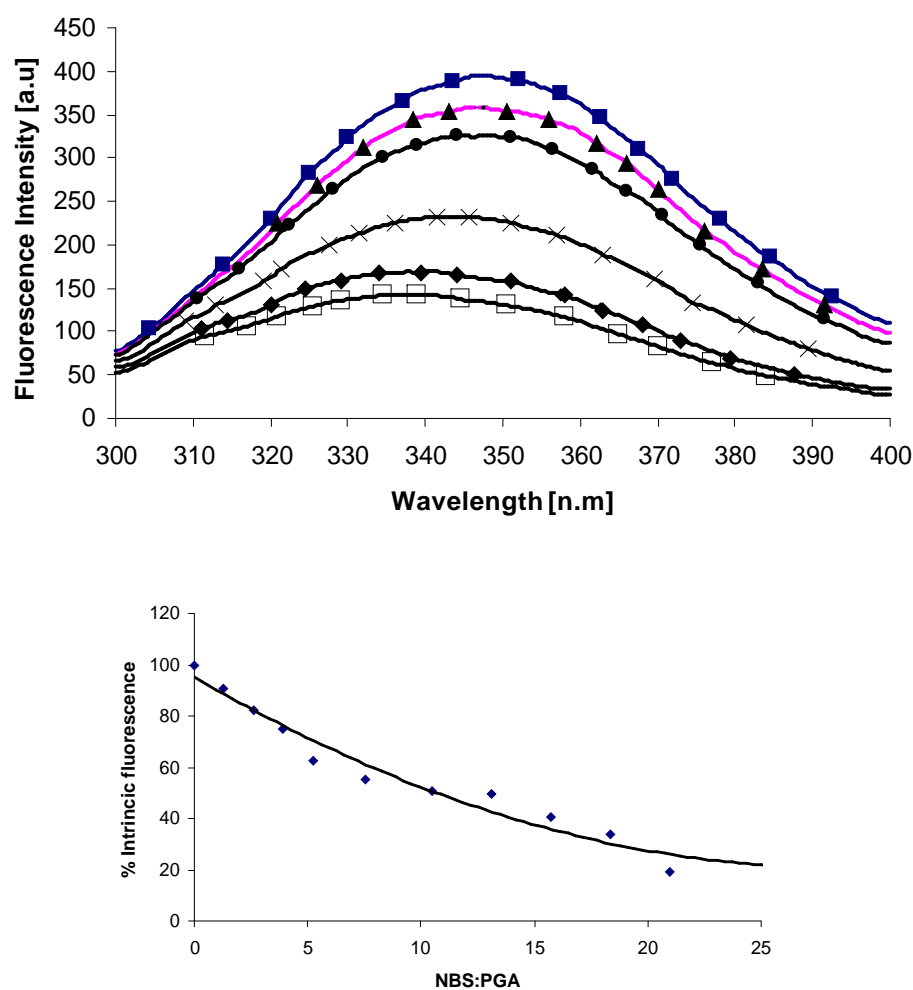
Figure 6.14: Fluorescence spectra of NBS treated enzyme KcPGA (3.81µM, 2 mL) was excited at 295 nm, and $_{max}$ was recorded at 348 nm. Spectra of untreated enzyme (|) modified using NBS 10 (^) 20 ( ) 50 (×) 100 (◆) and 200 µM ( ). Inset: fluorescent quenching with different ratios of NBS:PGA

The substrate protection against fluorescence quenching by NBS was observed when *Kc*PGA was pre-incubated with penG before the addition of NBS. The drop in fluorescence upon addition of NBS at the highest concentration was only 10.2% (178 to 160 AU) and λ_{max} remained unaltered; the blue shift observed previously on treatment with NBS was absent in this case (Figure 6.15)
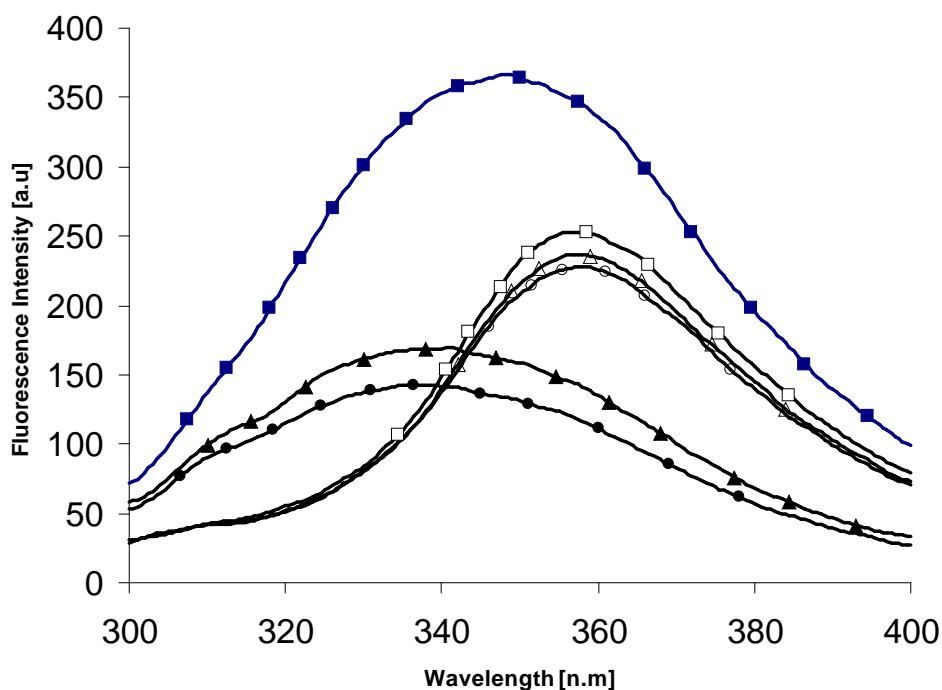
Figure 6.15: Fluorescence spectra of NBS treated and substrate-protected *Kc*PGA (3.81µM, 2 mL), excited at 295 nm and $_{max}$ was recorded at 348 nm. Native spectra (|), spectra of NBS modified 100 µM (^) and 200 µM ( ). Spectra of enzyme with penG 1mM ( ), spectra of NBS modified enzyme protected with substrate at 100 µM of NBS ( ), 200 µM of NBS (0)

The fluorescence of native enzyme was quenched by the addition of ionic and neutral quenchers. The percentage of quenching by acrylamide, CsCl and KI were 80%, 50% and 35%, respectively. The effective Stern-Volmer or quenching constant $K_{sv(eff)}$ and effective fractional accessible fluorescence $f_{a(eff)}$ were obtained from the slope and intercept of F_0/F versus $1/[Q]$ plots at low [Q] values (Table 6.9). The modified Stern-Volmer plots were linear for acrylamide, KI and CsCl (Figure 6.16).

Table 6.9: The effects of neutral and ionic quenchers on the fluorescence of *Kc*PGA

Quencher	15 % quenching	K_{sv} (eff)	$F_{a(eff)}$
Acrylamide (0-0.2 M)	0.008	16.25	1
KI (0 – 0.3 M)	0.025	6.00	0.52
CsCl (0-1.0 M)	0.500	2.00	0.22

The protein (3.81 µM, 2mL) was excited at 280nm and the emission was recorded in the range 300-400 nm. The K_{sv} values were calculated at low concentration [Q] of acrylamide, KI and CsCl
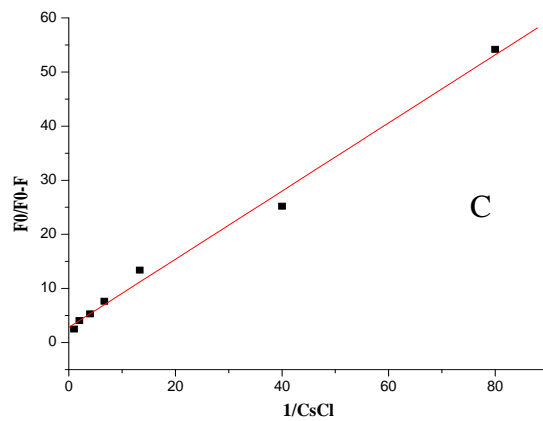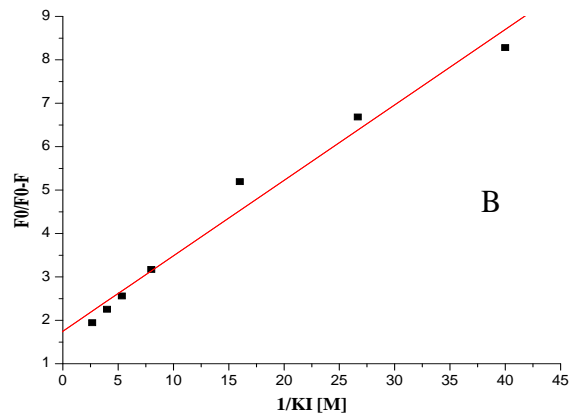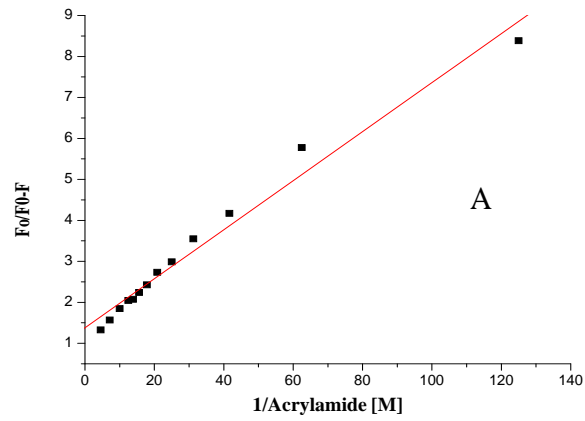
Figure 6.16: Modified stern-Volmer plots. (A) Acrylamide, (B) KI, (C) CsCl.

6.7.5. Homology modeling of *Kc*PGA and substrate docking

The multiple sequence alignment showed some conserved Trp residues present in all reported PGAs. A high level of sequence identity can guarantee accurate alignment between the target sequence and the template structure. The processed molecule of *Kc*PGA containing subunits and appeared similar to *Ec*PGA. The structure of *Kc*PGA modeled in MOE revealed excellent agreement with its predicted secondary structure and with the crystal structure of *Ec*PGA. PROCHECK is used to estimate the percentage of residues whose backbone φ-ψ angles are within the allowed region of Ramachandran map. The result showed that 62.9% of the φ-ψ angles in the *Kc*PGA model lie in the fully allowed region of the Ramachandran plot and the rest in partially or extended partially allowed regions. Both the modeled and the crystal structures on superposition gave an average RMSD (between equivalent C atoms) of 0.54 Å. The active site of *Kc*PGA was similar to that of *Ec*PGA (Figure 6.17a). The docking and molecular dynamics studies provided an estimate of the binding energy of substrate penG. There were three best conformations that showed almost the same levels of binding energy and the substrate bound close to residue 154 which happened to be the Trp. This Trp may be providing a favorable hydrophobic binding pocket for the phenyl group to bind when the substrate enters the active site.

6.7.6. Interactions between docked substrate and modeled enzyme

Using standard docking procedure, optimally using the available information, the docking of the substrate to *Kc*PGA was carried out. The substrate PGSO molecule is shown in Figure 6.17b along with the residues of the enzyme in the energy minimized complex; the residues of *Kc*PGA considered as interacting with PGSO mostly are closer than 3.5 Å. Majority of the interactions are with residues of chain. The carboxyl group of the penam ring interacts with the side chains of Arg 262 and Ser 384. This serine and main chain carbonyl oxygen of residue 23 are interacting with carbonyl oxygen of penam ring. The carbonyl oxygen at the scissile bond interacts with Asn 240 side chain, while scissile carbonyl carbon is ~4.1 Å away from O atom of Ser 1. It may be noted that the structures of substrate/analog complexes of *Ec*PGA have been already reported (McVey et al., 2001). Based on that study a detailed mechanism involving Ser 1 as

nucleophile, its amino group as base and a host of conserved residues forming oxyanion hole that stabilizes the acylenzyme intermediate has been worked out. We observe that the substrate interacts with analogous residues of *Kc*PGA in our modeling studies as well.
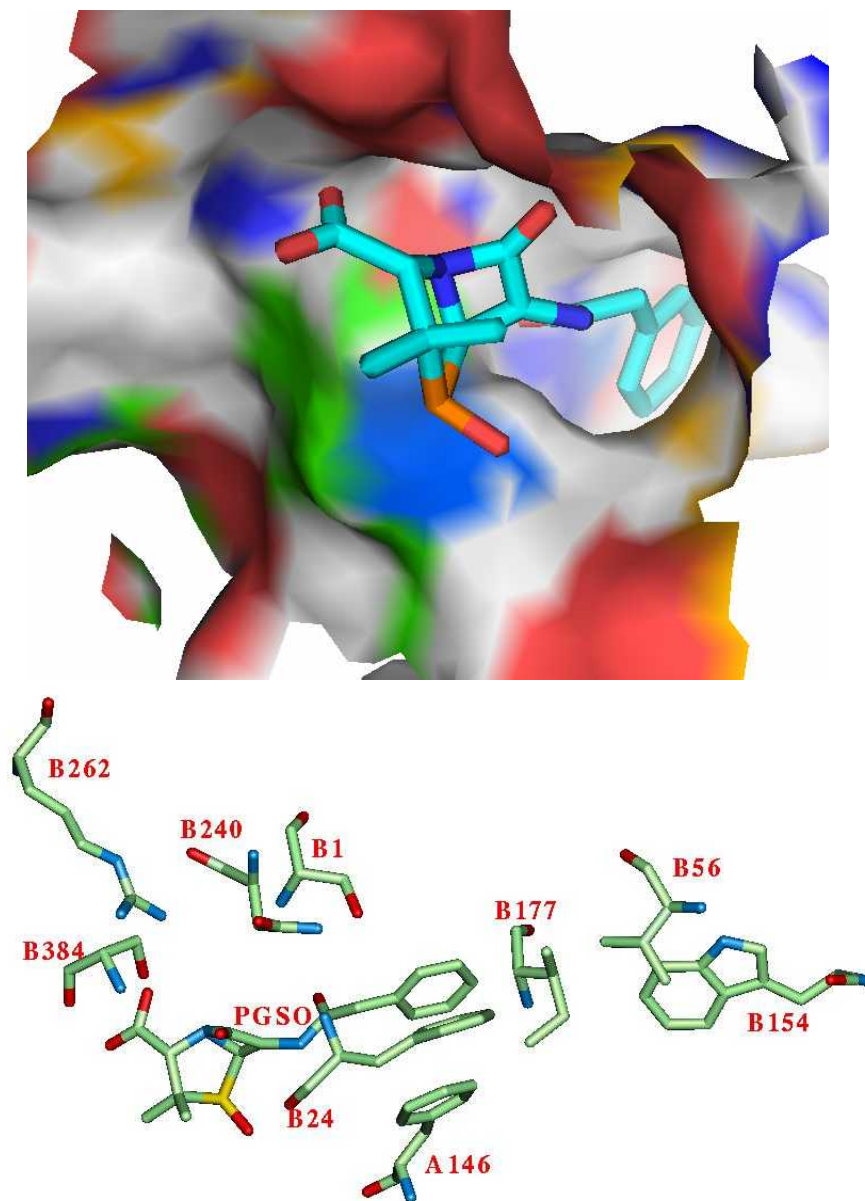
Figure 6.17: a) Active site environment of *Kc*PGA, modeled using MOE, with the docked molecule penicillin G sulphoxide (PGSO) in the active site cavity is shown. b) Individual residues of PGA in the active site close to PGSO as obtained from modeling studies is displayed, the residue of interest Trp 154 is on right side.

6.7.7. Comparison of PGA Sequences

In the multiple sequence alignment the sequences were segregated into two groups according to the percentage of similarity and presence of the active site Trp. The first group of sequences is closely related with more than 40% sequence identity and a conserved Trp 154 present in them. The second group of sequences has only less than 30% identity and no conserved Trp 154 present. Substantial variation in PGA activity between the two sets is observed, with the second group having low affinity towards penG analogues (Table 6.10).

6.8. Cloning of PGA from *K. citrophila*

Sense and antisense primers were designed according to the multiple cloning sites present in the pET-28a expression vector (Figure 2.4) and based on the published sequence of the *Kc*PGA gene. From the purified chromosomal DNA, *pga* gene was PCR-amplified using the primers A (30-mer) 5'- CGA ATG AGG ATC CAT GCC GTA TTT AAA GGG- 3' and B (29-mer) 5'- CAA CGC TCG AGA ACA AAA AAC CGG CCT AC- 3'.

The forward and reverse primers contained the restriction sites for BamHI and XhoI, respectively, for directed cloning. PCR was carried out in a total volume of 50 µL with a final concentration of 1 µM of each oligonucleotide, 200 µM dNTS mix (New England Biolab), 10X buffer and 1 unit VENT DNA polymerase (NEB). DNA was amplified using 25 cycles with annealing at 55 ºC and elongation at 72 ºC in Perkin-Elmer GeneAmp PCR system. Both Vector and insert were digested with NdeI and EcoRI. Other procedures were carried out as described in materials and methods.

6.8.1. Transformation of plasmid into the expression host

pET28b-KPGA was used to transform $CaCl_2$ competent BL21 expression host cells. The culture was grown overnight from the transforments and checked for His tagged PGA protein expression using SDS-PAGE and protein purification was done using Ni-column chromatography.

6.9. Discussion

The profound impact of penicillin acylases in the manufacture of antibiotics is reason for continued interest in studying structural determinants of these enzymes essential for catalysis. Efforts to enhance the catalytic rate by protein engineering have been reported (Gabor & Janssen, 2004).

The results obtained from the chemical modification studies could be interpreted in terms of the reported observations on *E. coli* PGA (Prabhune & SivaRaman, 1990). The previous report concluded from inactivation studies using the arginine-specific reagents 2,3-butanedione and phenylglyoxal that one arginine residue is involved in hydrolysis of penG. Morillas et al. (1999) studied the pH-dependence of the hydrolysis of *p*-nitrophenylacetate and reported a bell-shaped pH profile for the k_{cat} and k_{cat}/K_m parameters. It was suggested that the pK_a of the descending limb reflected upon the deprotonation of Arg 263, and that a positive charge on this residue was essential for enzyme activity. The residues Arg 145 and Arg 263 are conserved in PGAs of *E. coli*, *Alcaligenes faecalis* and *Providencia rettgeri, w*hile Arg 263 alone is conserved in *Arthrobacter viscosus* and *Bacillus megaterium* suggesting at least one arginine residue is important for catalysis of PGAs. Thus the arginine residue identified in our experiments could be Arg 263. The available crystal structures of *E.coli* PGA (Duggleby et al., 1995) provide information on the importance of arginine in catalysis. Interestingly, two arginines Arg17 and Arg228 are found in the oxyanione hole of *Bsp*PVA a mechanistically related enzyme (Suresh et al., 1999).

Structural and kinetic studies both have suggested the involvement of two charged arginine residues, Arg 145 and Arg 263, in the catalysis of PGA. In *Ec*PGA, the side chain of Arg 145 points towards the active site, forming a hydrogen bond with the main chain oxygen of Phe 24 also at the active site. X-ray structure analysis of the mutant enzyme with bound penG showed that Arg 145 displayed a large conformational change upon binding of the substrate, and moved away from the active site. The positional shift of Arg 145 as well as its neighbouring Phe 146 is necessary to create space for accommodating the -lactam moiety of penG. In this altered conformation, the hydrogen bonds between Arg 145: NH_2 and the main chain carbonyl oxygen of Phe 24 are replaced by hydrogen bonds between Arg 145: NH_2 and the carboxylate group of

substrate via two bridging water molecules (Done et al., 1998, Alkema et al., 2002). Arg 263 is located near the oxyanion hole, and its NH_2 group is hydrogen-bonded to O of Asn 241, the putative residue which stabilizes the transition state intermediate. The presence of protonated guanidino group of Arg 263 is proposed to be obligatory for catalysis, orienting the N-terminal catalytic serine residue and contributing to a decrease in the pK of the α-amino group (Morillas et al., 1999).

The molecular modelling studies of PGA based on GRID computational procedure and "transition state approach" to stimulate the tetrahedral intermediate showed that the carbonyl groups of penG can interact with Arg 263, Asn 388 and Ser 386 forming hydrogen bonds (Basso et al., 2002). Our data in conjunction with those reported for *Ec*PGA points to the essential role of arginine in the catalysis of *Kc*PGA.

Our preliminary studies reveal participation of a Trp residue in the activity of *Kc*PGA. This residue, hitherto not identified as important for activity or specificity by crystallographic studies, appears to be a conserved residue across all closely related sequences whose specificity for penG is higher (Figure 17a). The treatment of *Kc*PGA with 100 µM NBS lowered the activity by 70%. Similarly, 58% drop in activity was observed on treatment with 20 mM HNBB. This compound is known to react with Trp to form a substituted derivative. Based on the observation that the enzyme is being protected from inactivation by the presence of substrate or a competitive inhibitor, it can be concluded that the modified residue is at the active site or close to it. The conformity of CD spectra of the modified enzyme to that of untreated PGA ascribes the loss of activity to modification of selective residues only and not to any disruption in protein structure.

The rate of inactivation when plotted against the concentration of the modifier indicated modification of a single Trp. This is in agreement with the value of k_{app} obtained from pseudo-first order kinetics. The apparent K_m of the modified enzyme increased while k_{cat} remained unchanged indicating a decrease in substrate affinity as a consequence of modification. This suggests that the concerned Trp is involved in substrate binding.

The comparison of the fluorescence spectra of native and NBS modified PGA reveals that while their λ_{max} remains same, the relative intensity of the latter is reduced. This confirms that the loss of activity is exclusively due to Trp modification and rules out

modification of any cysteine residue. It may be remembered that NBS can react, to a lesser degree, with cysteine also. The result obtained here is in agreement with the modification study carried out also using cysteine modifiers like DTNB and iodoacetate which have not affected the activity. Difference absorption spectrum for the modification of KcPGA with NBS has a minimum at 280 and maximum at 250 nm indicating the formation of oxindolalanine (Lehrer, 1971). Weak reaction of NBS with tyrosine is also known. However, the absence of peak at 263 nm in the observed absorbance spectra of KcPGA rule out possibility of tyrosine residue getting oxidized to dityrosine.

The microenvironment of the essential Trp was studied by measuring the quenching of PGA fluorescence resulted from the binding of ionic and neutral quenchers. In the quenching experiments, the values of f_a and K_{sv} obtained for Iodide and Cesium ions were lower than the value obtained for acrylamide-quenched enzyme. This indicates that the essential Trp is located in a hydrophobic environment. Further, the blue shift of λ_{max} in the case of NBS modified enzyme also confirms that the fluorescence contribution is from a Trp residue located in hydrophobic environment. The difference spectra of both the unmodified and NBS modified KcPGA showed red shifts suggesting accessibility of modified residue. Compared to K_{sv} of KI (6), the K_{sv} of CsCl (2) is lower, suggesting that some Trp are in electropositive environment.

It has been reported that the enzymes from $P. rettgeri$ (McDonough et al., 1999) and $E. coli$ (Duggleby et al., 1995) are inactivated in the presence of Trp modifying reagents. However, no kinetic analysis of the inactivation has been reported till now. In the crystal structures of PGA from $E. coli$ and $P. rettgeri$ the Trp 154 is near the active site. This particular Trp residue occurs also in the immediate vicinity of the active site in the homology modeled KcPGA (Figure 6.17b). The closest distance between the Trp and the nucleophile serine is 4.5 . There are no considerable conformational changes observed in Trp on binding the substrate, although some residues near active site are showing changes in their rotamer conformation in the model. The role of other residues in the active site was well demonstrated in the previous studies (Mahajan et al., 1983).

The presence of Trp in the immediate vicinity of the substrate-binding site is known to cause unusual effects in substrate binding and catalysis. For example, the PGA

mutant where Phe 24, Phe 57, Phe 146 were mutated to Trp showed a 10-fold decrease in k_{cat} and an increase of up to 200% in K_m (Gabor et al., 2004).

The involvement of a hydrophobic residue like Trp in PGA catalytic activity inferred here is in agreement with the results from other investigations on the role of hydrophobic interactions in the hydrolytic activity of PGA. These include the report on the inhibitory effect of aliphatic alcohol on the activity of the enzyme from *E. coli*, the magnitude of inhibitory effects vary with the extent of hydrophobicity of the alcohol (Alkema et al., 2002).

On the whole, the chemical modification and fluorescence studies provide strong indication for the presence of an essential Trp residue at the active center of *Kc*PGA, which from kinetic analysis appears to be involved in substrate binding. This decisive residue is identified as Trp 154 through homology modeling. The same is conserved in closely related PGAs (Figure 6.18a) and not in those distantly related ones (Figure 6.18b) whose penicillin binding capacities are partly affected. The parameters listed in table 4 show that the PGAs closely related to *Kc*PGA have better affinity for penG compared to those distantly related and Trp residue discussed here is absent in them. It could be argued that the latter class of enzymes is specific for a different substrate, however, it only reinforces the argument that Trp 154 decides the penG specificity. Thus, all the results reported here point to the importance of Trp 154 in penG specific activity of PGAs.

Table 6.10: Some chosen parameters of PGAs from various organisms in the sequence alignment shown in Figure 8a & b are listed. A column is kept blank when its value is not available.

| Source of PGA | β-chain-length | Number of tryptophans in β-chain | Kinetic parameters | | % homology against |
			Km (μM)	$Kcat$ (S^{-1})	KcPGA
Kluyvera citrophila	555	22	16	63	-
Eschericia coli	557	21	46	50	85
Alcaligenes xylosoxidans	556	19	8.9	68.8	51
Alcaligenes faecalis	551	18	2	54	40
Achromobacter	568	19	-	-	52
Proteobacterium.	566	19	-	-	52
Pseudomonas denitri	560	18	-	-	44
Shigella boydii (single chain enzyme)	448	14	-	-	84
Pseudomonas putida	523	14	420	-	18
Bacillus megaterium	537	13	4500	-	28
Arthrobacter viscosus	537	12	-	-	27
Bacillus cereus	503	14	350	-	19
Streptomyces avermitilis	515	14	-	-	14
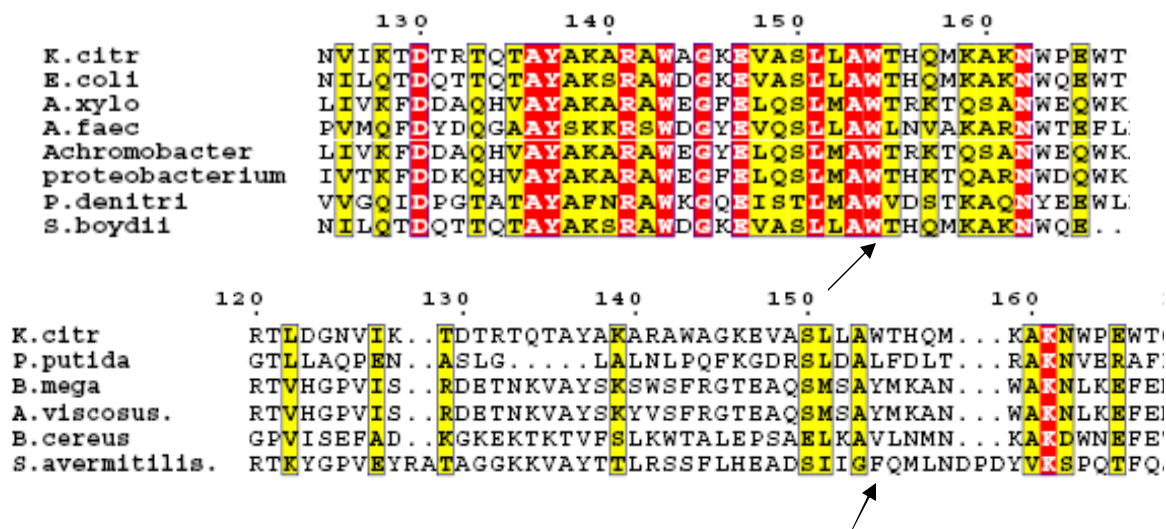
Figure 6.18: a) Comparison of sequences of PGA closely related to *K. citrophila* PGA on either side of the conserved Trp β154 is shown. b) Comparison of sequences of designated PGAs distantly related to *Kc*PGA to show that the Trp of interest is not conserved in them. The position of Trp β154 is indicated by arrow in both figures.

Previous studies in the autocatalytic process of LexA (Little, 1993), glycosylasparaginase (Guan et al., 1998) and in some serine proteases (Kenneth &. Sauer, 1995; van Dijl, 1995) the dependence of precursor processing on pH was reported. The elucidation of optimal condition for precursor processing is very essential in the context of crystallization experiments. An improper folding condition can lead to increase in conformational heterogenisity of the protein molecule which will hamper crystallization process.

Unlike *E. coli,* production of relatively high amount of precursor in *K. citrophilla* may be due to the fact that the optimal growth condition of organism may not be conducive for the enzyme maturation. In *E. coli,* it was reported that a conserved lysine residue (Lys 299), which is located close to N-terminal serine residue (Ser 290), could be the most probable candidate responsible for the pH dependent activation. The same study reported that the mutation of Lys 299 affected the precursor processing and

enzyme activity of PGA *in vivo* and *in vitro.* It is worth noting that this particular residue is conserved in all PGAs (Lee et al., 2000).

The signal peptide contains polar amino acids (signature sequence; KNRNR) and recognized by the receptor in the periplasm. SecB is the major component involved in the translocation of many secretory and periplasmic proteins. The formation of active PGA requires the SecB chaperone (Chou et al. 1999). Twin arginine translocation (Tat) pathway, allows the secretion of presumably folded cofactor containing proteins across the bacterial membrane. Based on studies of the Tat- mutant strain of *E. coli* JARV15, it was proved that the $E.\,coli$ preproPGA signal peptide efficiently mediates Tat translocation. (Ignatova et al., 2002).

The very slow processing of proPGA in the periplasm could be the limiting step in the overproduction of active PGA (Scherrer et al., 1994). DegP is a periplasmic protease/chaperone and is required for proteolysis in the $E.\,coli$ cell envelope. By increasing the intracellular DegP concentration, the majority of the overproduced PGA precursors could be smoothly processed and inclusion bodies significantly reduced (Lin et al., 2001). Molecular chaperones control protein folding processes and protect misfolded proteins from aggregation in all cells (Sriubolmas et al., 1997). Chaperones bind to non‐native conformations of proteins and assist them to reach the native tertiary structure. Most chaperones bind to exposed hydrophobic surfaces of non‐native species and thereby stabilize their target protein against aggregation. The coexpression of a DegP mutant without protease activity while retaining chaperone activity could not improve the productivity, suggesting that DegP protease activity was primarily responsible for the suppression, possibly by the degradation of abnormal proteins when PGA was overexpressed. However, a shortage of periplasmic protease activity was not the only reason for inclusion body formation upon PGA overexpression because the coexpression of a DegP-homologous periplasmic protease, DegQ or DegS, could not improve the PGA production (Robas et al., 1993; Sobotkova et al., 1996). The chaperone activity as well as the protease activity of DegP is considered to be involved in the proper processing of *Ec*PGA (Pan et al., 2003).

*Kc*PGA has very high potential in pharmaceutical industry as a promising alternative *to Ec*PGA which is employed in manufacture of semi‐synthetic penicillin.

Enzyme yield could be enhanced to a considerable extent on standardization of development of seed culture and fermentation condition. The bottleneck in *Kc*PGA production is the inherent production of inactive precursor associated with mature protein. This study attempted to isolate the precursor molecule and explore and evaluate an optimized condition for *in vitro* maturation. Genetic engineering coupled with structural studies might lead to development of novel enzyme useful in manufacture of semi-synthetic β lactam antibiotics.

.

CHAPTER 7

CRYSTALLIZATION AND
CRYSTAL STRUCTURE STUDIES OF PGA

7.1. Introduction

This chapter describes the crystallization and structural studies attempted on two penicillin acylases *Kc*PGA and *Af*PGA. Because of the limited yield of pure protein in the case of *Kc*PGA only some preliminary crystallization experiments could be completed and we were not able to proceed with the structure analysis. However, the cloning and over production of this as an extracellular enzyme has been completed as described in chapter 6. Diffraction data at a resolution of 3.5 Å were collected from tetragonal crystals of *Af*PGA. The crystal structure was determined using molecular replacement method using co-ordinates of *E. coli* PGA (PDB:1gk9). Although the resolution was modest we could identify features like the presence of a disulphide bridge in the structure that would be imparting comparatively more stability to *Af*PGA.

The PGA from *A. faecalis* was cloned and characterized by Verhaert et al., 1997. Despite its many similarities with other PGAs, the *Af*PGA enzyme has a clear industrial advantage over other well-characterized penicillin acylases in -lactam conversions because of its higher thermostability and also for its high synthetic efficiency in enatioselective synthesis. This makes the *Af*PGA enzyme a more attractive biocatalyst both in hydrolysis and in synthetic conversions. The presence of two cysteine residues at positions B492 and B525 in *Af*PGA enzyme probably accounts for its higher stability on the basis of the following observations: (i) a comparison with the known 3D structure of the *E. coli* enzyme indicates that these two cysteine residues are in close proximity, (ii) the -subunit of *A. faecalis* enzyme shows an electrophoretic mobility that depends on its oxidation state, (iii) the *A. faecalis* enzyme is significantly more thermostable in its oxidized state than reduced state, and (iv) *Af*PGA is more stable than the *E. coli* enzyme, which lack the disulfide bridge. These four arguments led us to conclude that a disulfide bond contributes for its higher stability compared to other PGAs.

7.2. Crystallization of *Kc*PGA

The attempts to obtain crystals using purified protein and commercial crystal screens were not successful. In many cases heavy precipitation occurred but failed to develop into crystals even after modifying the crystallization conditions. To try different strategy crystallization using different precipitants at pH 3 to 10 was explored. Small crystals appeared after 20 days in 40% ammonium sulphate in pH 8 and 8.5 of 50 mM MOPS buffer. The crystal size was improved by the addition of different concentrations of β-octyl-glucopyranoside and calcium chloride.

The amount of protein used for the crystallization was 30 mg/ml. The size of the crystals were improved in 45% saturated AS, 60 µL 0.5% β-octyl-glucopyranoside and 50 mM $CaCl_2$ in 50 mM MOPS buffer pH 8.2 grown at 303K (Figure 7.1). These crystals were used for data collection both at room temperature and in liquid nitrogen temperature. The diffraction quality of the crystals was very poor at room temperature. A wide range of cryo-conditions was screened. But, unfortunately the data collected at low temperature could not be indexed due to high mosaicity of the crystal. Attempts to collect good quality data using different cryo-protectants did not yield positive results. Subsequent attempts to improve the crystal quality by further altering crystallization conditions were also ineffective.
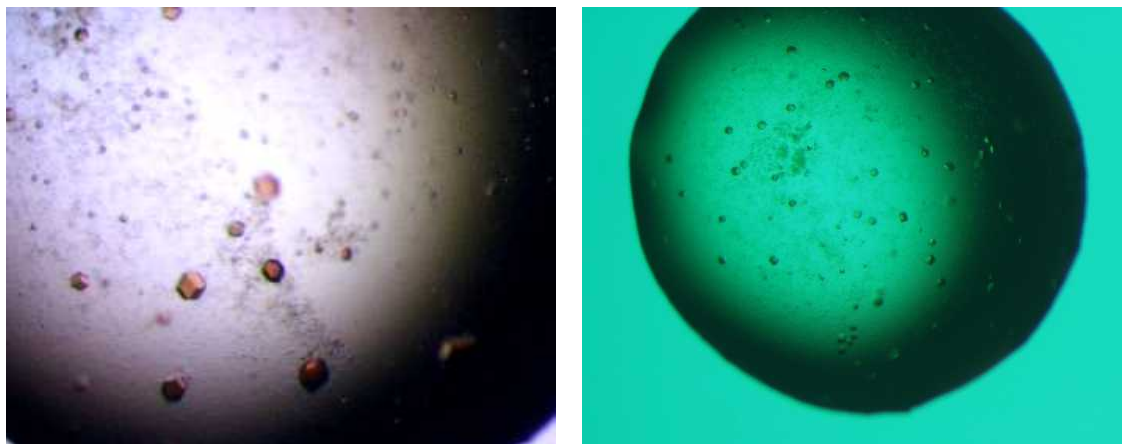
Figure 7.1: Crystals of *Kc*PGA grown at 30 °C.

7.3. Crystallization of *Af*PGA

The purified protein sample was obtained from Dr. Zoya Ignatova (Institute of biotechnology, Technical University, Hamburg, Germany). This sample was checked in the SDS-PAGE. The presence of (64 kDa) and (20 kDa) subunits were confirmed (Figure 7.2).
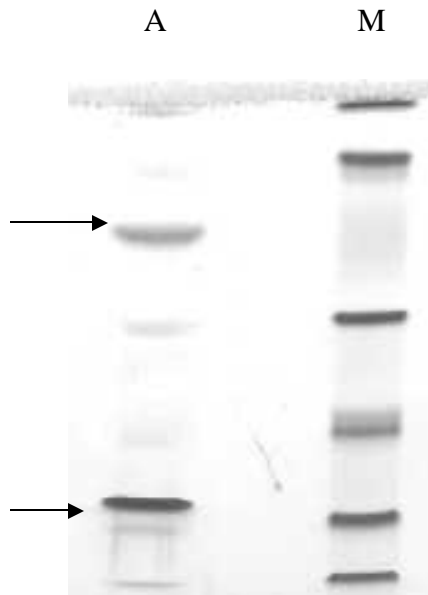
Figure 7.2. SDS-PAGE of purified *Af*PGA. The lanes contained A: *Af*PGA, M: markers

7.3.1. Protein preparation

The protein sample was dialyzed overnight with two changes against 100 times volume of 10 mM potassium phosphate buffer pH 7.5. Protein concentration was determined in accordance with the method of Lowry et al. (1951) with BSA as standard. The protein was concentrated approximately 15 mg/mL in centricon concentrator (Millipore) at 5000xg rpm and was stored at -20 °C. The protein sample used for crystallization was thawed and spun at 10000xg rpm for 5 min before keeping for crystallization.

7.3.2. Crystallization trials

Crystallization conditions were initially screened using conventional sparse-matrix screen with the commercial screening kits, Crystal Screens I and II (Hampton Research) and Clear Strategy screen I and II (Molecular Dimensions Ltd.). The screening

kits contained prepared-ready solutions and the crystallization trials were set-up in hanging drop vapor diffusion technique. But unfortunately both the screens did not yield any crystals.

The crystallization trials were set-up using solutions of precipitant, salts and buffers. The screens were tested in a range of conditions where two parameters were varied, typically the precipitant concentration and the pH. In these trials the crystals were obtained primarily from two conditions (Figure 7.3). The fine screening for precipitation concentration and pH was carried out based on the initial observations. A number of additives were also tried to improve crystal quality. Finally, reproducible and diffraction quality crystals were obtained in 15% PEG 8000, β-octyl-glucopyranoside (0.05 w/v) and 0.1 M Tris-HCl pH 7.5.

A number of crystal forms were obtained depending upon small variations in β-octyl-glucopyranoside concentration. Unfortunately, none of the crystal forms diffracted under cryo conditions in the presence of different cryo-protectants or different strategies of soaking in cryo-solution. The results are provided in Table 7.1 and Figure 7.4.

Table 7.1: Summary of cryo-protectant solution tried for *Af*PAG crystals

S. NO	Precipitant	Cryo-protectant	Result
1	PEG 8K	25% PEG 400	No diffraction
2	PEG 10K	25 % PEG 400	No diffraction
3	PEG 10 K	30 % glycerol	No diffraction
4	PEG 8K	30 % glycerol	Around 8
5	PEG 8K	20%Ethylene glycol	Poor diffraction
6	PEG 8K	20% hexane triol	No diffraction
7	PEG 8K	20% isopropanol	Around 5
8	PEG 10K	20%Ethylene glycol	No diffraction
9	PEG 10K	20% hexane triol	No diffraction
10	PEG 10K	20% isopropanol	No diffraction
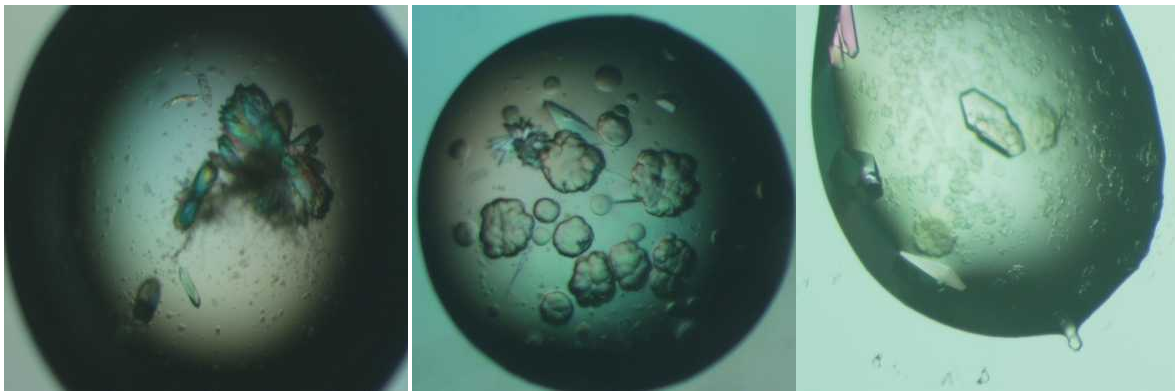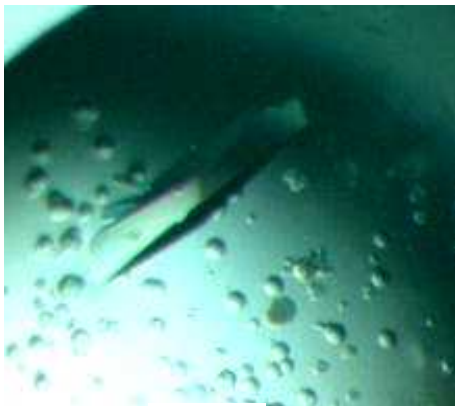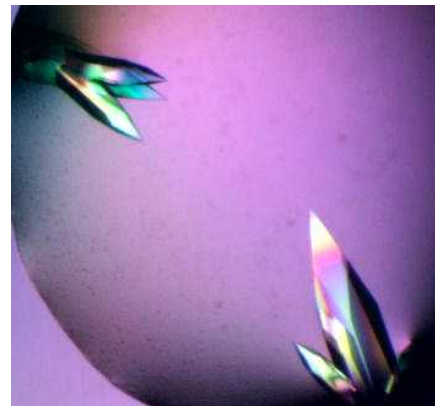
Figure 7.3: Improvement in the quality and size of *Af*PGA crystals, a) Crystals grown in initial conditions with heavy amorphous precipitate, b) Monoclinic crystals grown at room temperature, c) Orthorhombic crystals obtained in the presence of β-octyl- glucopyranoside, d) tetragonal crystals

7.3.3. Data collection

Considering the deterioration in quality of crystals irrespective of the cryo-solution tried (Figure 7.4). The diffraction data were collected at room temperature by mounting the crystals in thin glass capillaries of 0.7, 1.0 or 1.5 mm diameter. Many crystal morphologies were observed and mostly turned out to be monoclinic forms (Table 7.2). Unfortunately in most crystal forms the completion of data collected was poor.

Interestingly, some crystal transformation was taking place when the crystal was exposed to the X-rays. The orthorhombic P222 form has transformed into orthorhombic C222 form after certain period of exposure in X-rays (Figure 7.5). This type of cell transformation was already reported in lysozyme crystals (Harata & Akiba, 2006). The dehydration of lysozyme crystals induces molecular rearrangement, which transforms the crystals into a new crystals that is stable and has lower solvent content. The unit cell parameters of the both forms are in Table 7.2.

Only in tetragonal form (P4) we were able to collect complete data. The details of data collection statistics were presented in Table.7.3. Because of one large unit cell dimension (298 Å) and not too excellent quality of the crystals we could collect data to an effective resolution of only 3.5 Å using the laboratory X-ray source.

a   b
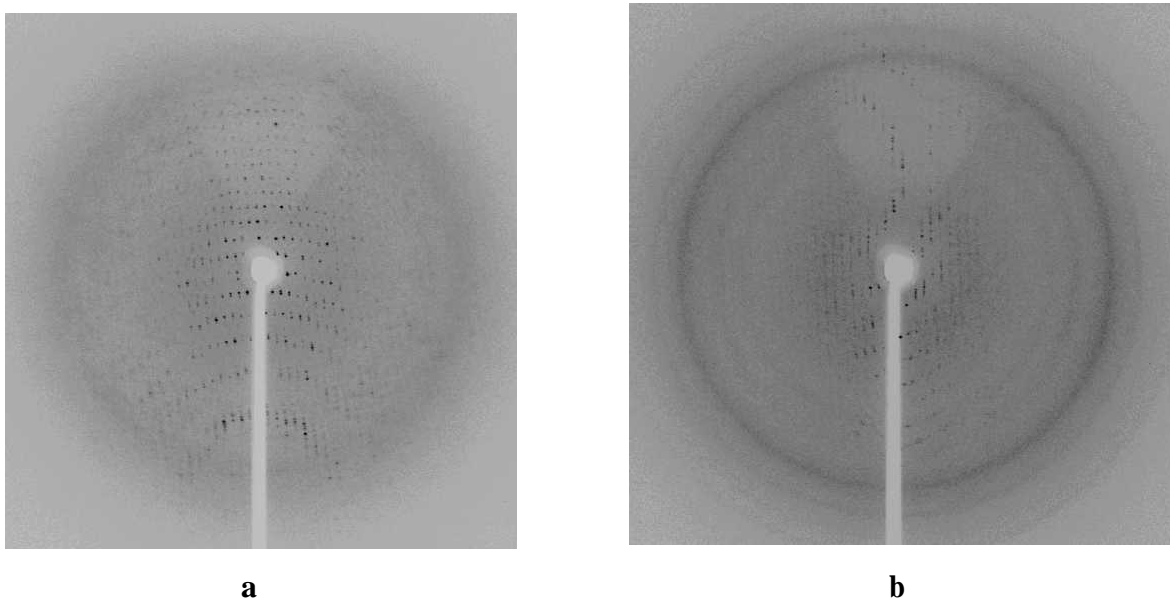
Figure 7.4: The diffraction image of the *Af*PGA crystals collected in different cryo-conditions a) 30% glycerol, b) 20% iso-propanol

Table 7.2: Details of the different types of crystal forms of *Af*PGA characterized.

S. no	Crystal form	Unit cell parameters (and º)	Mathew's number	Solvent content (%)	No of molecules in the asymmetric unit
1	Monoclinic P2	a=138.30 b=117.65 c=183.624 =92.96	2.8	55.9	6
2	Monoclinic P2	a=75.25 b=64.66 c=102.87 =98.44	2.9	56.9	1
3	Monoclinic C2	a=73.03 b=86.80 c=262.11 = 89.71	2.42	49.07	2
4	Orthorhombic P222	a=72.61 b=86.17 c=259.38	2.36	47.48	2
5	Orthorhombic C222	a=72.40 b=86.02 c=259.45	2.36	47.86	1
6	Tetragonal P4	a=b=85..55 c=298.80	3.18	61.31	1
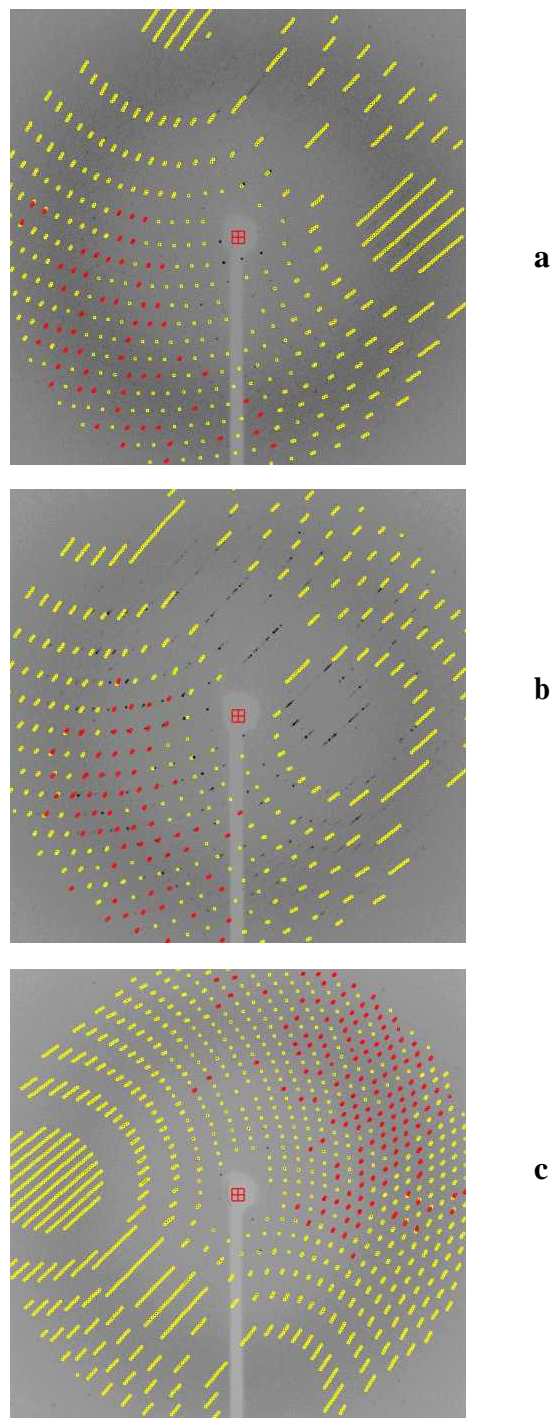
Figure 7.5: The indexed diffraction frames of C222 crystals to show the unit cell transformation during data collection a) spots are indexed in C222, b) many spots are not indexed after first 85 frames, c) from 95th frame the spots could be indexed in P222.

Table 7.3: Data collection statistics of the tetragonal crystal form of *Af*PGA

Temperature	Room temperature (RT)
Wavelength (Å)	1.5418
Resolution (Å)	20-3.5 (3.56-3.50)
Total reflections	99602
Unique reflections	40387
Completeness (%)	90.5 (95.4)
R $_{merge}$ (%)	9.9 (23.9)
Average I/σ(I)	7.42 (3.08)
Mosaicity	0.3
Space group	P 4_1 2_1 2
Unit cell parameters (Å)	*a=b*=85.56 *c*=298.78
Solvent content (%)	61.31
Matthews Coefficient (Å^3 Da^{-1})	3.18
Molecules per Asymmetric unit	1

7.3.4. Structure solution

Reflections were phased using molecular replacement method. The processed form of *E. coli* Penicillin G acylase monomer (PDB:1gk9) was used as search model. The molecular replacement solution was obtained from PHASER. The Z-score and R-factor indicated the correct solution. The P422 indexed data were used for the search. All possible alternative space groups (P 4 2 2, P 4 2_1 2, P 4_1 2 2, P 4_1 2_1 2, P 4_2 2 2, P 4_2 2_1 2, P 4_3 2 2, P 4_3 2_1 2) were checked using the program in the calculation of translation function(Table 7.4). The solution that corresponded to highest Z-score obtained (P$4_1$$2_1$2) was then subjected to packing function calculation with permissible number of clashes. In this case the permissible number was fixed as zero. The first cycle of refinement using Refmac was then carried out.

Table 7.4: Fast Translation Function Table by Phaser.

Space Group	SET	TRIAL	Top (Z)	Second (Z)	Third (Z)
P 4 2 2	1	1	-73.99 (7.10)	-96.28 (6.32)	-103.39 (6.06)
P 4 2_1 2	1	1	-90.36 (5.94)	-105.27 (5.45)	-108.63 (5.34)
P 4_1 2 2	1	1	144.16 (13.60)	120.53 (12.82)	114.28 (12.62)
P 4_1 2_1 2	**1**	**1**	**482.83 (24.99)**	- -	- -
P 4_2 2 2	1	1	-61.77 (7.48)	-88.78 (6.53)	-101.80 (6.08)
P 4_2 2_1 2	1	1	-99.97 (6.06)	-101.00 (6.02)	-106.50 (5.83)
P 4_3 2 2	1	1	34.13 (10.94)	- -	- -
P 4_3 2_1 2	1	1	64.04 (12.02)	45.01 (11.34)	21.48 (10.49)

Inspection of the electron density map after the first cycle of refinement has indicated that some disorder exists in the loop regions. For three loops electron density was completely absent. Ten more cycles of manual model fitting and refinement were carried out in conjunction with calculation of a full series of Quanta omit maps that were carefully examined for incorrect modeling. However, after a few cycles of refinement the

geometry of the model got totally distorted. The R-factor reduced to not better than 25%, presumably due to poor quality of diffraction data.

7.3.5. Description of the preliminary structure of *Af*PGA

The two chains of the molecule (and) are closely intertwined and form a pyramidal structure that contains active site in a deep cone-shaped depression at the bottom. Among the 738 residues in the PGA, all but the eight C-terminal residues of the -subunit could be modeled. The heterodimer has approximate dimension of 70 X 50 X 55 and has no obvious discrete domains (Figure 7.6).
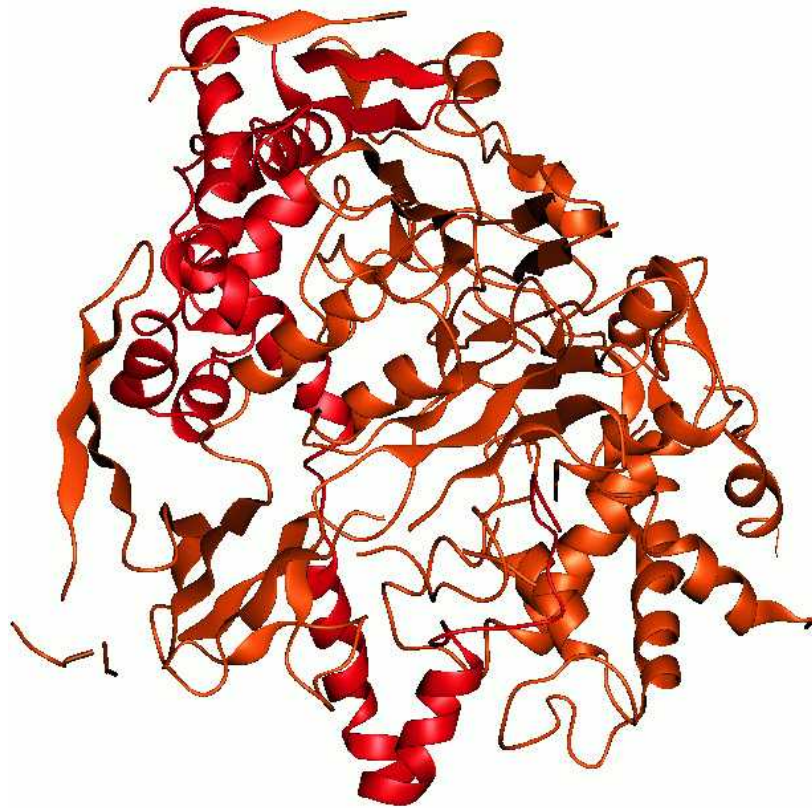
Figure 7.6: Preliminary structure of *Af*PGA showing heterodimeric association and arrangement

Table 7.5: Comparison of sequence identity between two PGAs discussed in this thesis and *Ec*PGA which is used as the model.

	*Ec*PGA	*Af*PGA	*Kc*PGA
Total number of residues	766	753	764
% homology with *E. coli*	-	41	86
No. of residues in -chain	209	202	209
% homology with *E. coli*	-	45	83
No. of residues in -chain	557	551	555
% homology with *E. coli*	-	40	89
No. of residues in spacer peptide	54	38	54
% homology with *E. coli*	-	8	85

The percentage of homology between the AfPGA and EcPGA is provided in table 7.5. From the sequence alignment using ClustalW, it is observed that the majority of residues in -chain are highly conserved in *Af*PGA(Figure 7.7), same as in the core of the protein, which includes the acyl cleft of the active site. The residues identified as directly involved in catalysis (Ser 1, Ala 69, Asn 241, and Arg 263) are also strictly conserved. The peripheral region of the protein is the most variable one and the changes can be described as

1.Four extra residues are present in *Af*PGA between amino acids 57 and 70 of *Ec*PGA

2. The C-terminal portion of the -chain in *Af*PGA has two extra residues compared to *Ec*PGA

3. Two residues in the *Af*PGA sequence are not present between residues numbers 332 and 337 of *Ec*PGA.

Together the - and -subunits make up five distinct structural regions: an -sandwich, two predominantly helical regions, a partial -barrel, and a small -strand region. Of the five structural regions, the largest is the -sandwich. This region contains

the catalytic residue Ser 1 and most residues of the hydrophobic specificity pocket. Surrounding and on either side of the core -sandwich region are the other four regions.

```
           1         10        20        30        40        50        60        70
EcPGA  MKNRNRMIVNCVTASLMYYWSLPALAEQSSSEIKIVRDEYGMPHIYANDTWHLFYGYGYVVAQDRLFQME
KcPGA  MKNRNRMIVNGIVTSLICCSSLSALAASPPTEVKIVRDEYGMPHIYADDTYRLFYGYGYVVAQDRLFQME
AfPGA  MQKG..LVRTGLVAAGLILGWAGAPTHAQVQSVEVMRDSYGVPHVFADSHYGLYYGYGYAVAQDRLFQMD

                     80        90       100       110       120       130
EcPGA  MARRSTQGTVAEVLG....KDFVKFDKDIRRNYWPDAIRAQIAALSPEDMSILQGYADGMNAWIDKVNTN
KcPGA  MARRSTQGTVSEVLG....KAFVSFDKDIRQNYWPDSIRAQIASLSAEDKSILQGYADGMNAWIDKVNAS
AfPGA  MARRSFVGTTAAVLGPGEQDAYVKYDMQVRQNFTPASIQRQIAALSKDERDIFRGYADGYNAYLEQVRRR

          140       150       160       170       180       190       200
EcPGA  PETLLPKQFNTFGFTPKRWEPFDVAMIFVGTMANRFSDSTSEIDNLALLTALKDKYGVSQGMAVPNQLKW
KcPGA  PDKLLPQQFSTFGFKPKHWEPFDVAMIFVGTMANRFSDSTSEIDNLALLTAVKDKYGNDEGMAVPNQLKW
AfPGA  PE.LLPKEYVDFDFQPEPLTDFDVVMIWVGSMANRFSDTNLEVTALAMRQSLEKQHGPERGRALFDELLW

          210       220       230       240       250       260       270
EcPGA  LVNPSAPTTIAVQESNYPLKFNQQNSQTAALLP.RYDLPAPMLDRPAKGADGALLALTAGKNRETIAAQF
KcPGA  LVNPSAPTTIAARESSYPLKFDLQNTQTAALLVPRYDQPAPMLDRPAKGTDGALLAVTAIKNRETIAAQF
AfPGA  INDTTAPTTVPA......................PAAEHKPQAQAGTQDLAHVSSPVLATELERQD

          280       290       300       310       320       330       340
EcPGA  AQGGANGLAGYPTTSNMWVIGKSKAQDAKAIMVNGPQFGWYAPAYTYGIGLHGAGYDVTGNTPFAYPGLV
KcPGA  ANG.ANGLAGYPTTSNMWVIGKNKAQDAKAIMVNGPQFGWYAPAYTYGIGLHGAGYDVTGNTPFAYPGLV
AfPGA  KHWGGRGPDFAPKASNLWSTRPERVQEGSTVLINGPQFGWYNPAYTYGIGLHGAGFDVVGNTPFAYPIVL

          350       360       370       380       390       400       410
EcPGA  FGHNGVISWGSTAGFGDDVDIFAERLSAEKPGYYLHNGKWVKMLSREFTITVKNGQAETFTVWRTVHCNI
KcPGA  FGHNGTISWGSTAGFGDDVDIFAEKLSAEKPGYYQHNGEWVKMLSRKETIAVKDGQPETFTVWRTLDCNV
AfPGA  FGTNSEIAWGATAGPQDVVDIYQRKLNPSRADQYWFNNAWRTLEQRKERIQVRGQADREMTIWRTVHCPV

          420       430       440       450       460       470       490
EcPGA  LQTDQTTQTAYAKSRAWDGKEVASLLAWTHQMKAKNWQEWTQQAAKQALTINWYYADVNGNIGYVHTGAY
KcPGA  IKTDTRTQTAYAKARAWAGKEVASLLAWTHQMKAKNWPEWTQQAAKQALTINWYYADVNGNIGYVHTGAY
AfPGA  MQFDYDQGAAYSKKRSWDGYRVQSLLAWLNVAKARNWTEFLDQASKMAISINWYYADKHGNIGYVSPAFL

          490       500       510       520       530       540       550
EcPGA  PDRQSGHDPRLPVPGTGKWDWKCLLPFEMNPKVYNPQSGYIANWNNSPQKDYPASDLFAFLWCGADRVTE
KcPGA  PDRQPGHDPRLPVP.DGKWDWKCLLSFDLNPKVYNPQSGYIANWNNSPQKDYPASDLPAFLWCGADRVTE
AfPGA  PQRPADQDIRVPAKGDGSMEWLCIKSFDAIPKAYNPPQGYLVNWNNKPAPDKTNTDTYYWTYG..DRMNE

          560       570       580       590       600       610       620
EcPGA  IDRLLEQKPRLTADQAWDVIRQTSRQDLNLRLFLPTLQAATSGLTQSDPRRQLVETLTRWDGINLLNDDG
KcPGA  IDTILDKQPRFTADQAWDVIRQTSLRDL.LRLFLPALKDATANLAENDPRRQLVDKLASWDGENLVNDDG
AfPGA  LVSQYQQKDLFSVQEIWEFNQKASYSDVNWRYFRPHLEKLAQQIPADDSSKAALTMLLAWDG..MEQDQG

          630       640       650       660       670       680       690
EcPGA  KTWQQPGSAILNVWLTSMLKRTVVAAVPMPFDKWYSASGYETTQDGPTGSLNISVGAKILYEAVQGDKSP
KcPGA  KTYQQPGSAILNAWLTSMLKRTVVAAVPAPFGKWYSASGYETTQDGPTGSLNISVGAKILYEALQGDKSP
AfPGA  GQNAGPARVLFKTWLEEMYKQVLMPVVPESHRAMYSQTGPATQQGPNPGSINLSMGTKVLLRALVLEAHP
```

```
              700        710        720        730        740        750        760
EcPGA  IPQAVDLFAGKPQQEVVLAALEDTWETLSKRYCNNVSNWKTPAMALTFRANNFFCVPQAAAEETRHQAEY
KcPGA  IPQAVDLFGGKPEQEVILAALDDAWQTLSKRYCNDVTGWKTPAMALTFRANNFFCVPQAAAKEARHQAEY
AfPGA  DPKRVNVFGERSSQEIMHTALQNAQARLSQEQCAQMARWTMPTSVHRFSDKNFTCTPQTMPGNTFAFTGY

              770        780        790        800        810        820        830
EcPGA  QNRGTENDMIVFSPTTSDRPVLAWDVVAPGQSGFIAPDCTVDKHYEDQLKMYENFGRKSLWLTKQDVEAH
KcPGA  QNRGTENDMIVFSPTSGNRPVLAWDVVAPGQSGFIAPDGKADKHYDDQLKMYESFGRKSLWLTPQDVDEH
AfPGA  QNRGTENNRVVFDAKG....VEFCDAMPPGQSGFTDRNGVRSPHYEDQLKLYENFECKTMDVTHADIRRN

              840
EcPGA  KESQEVLHVQR..
KcPGA  KESQEVLQVQR..
AfPGA  AQSSTMLLIQPQP
```

Figure 7.7: Alignment of PGA sequences discussed here to show the positions of conserved residues.

A mostly helical region is situated on the six-stranded face of the β-sandwich. It is composed of residues from the -subunit and two -helices containing residues from the C-terminal end of the -subunit. An entirely -helical region, which contains majority of residues from the -subunit, forms an interesting cluster of eight -helices. The disulphide bond was observed between the two regions 492 – 525. (Figure 7.8). The electron density for this disulphide bond is clearly seen in the preliminary structure reported here *Af*PGA.
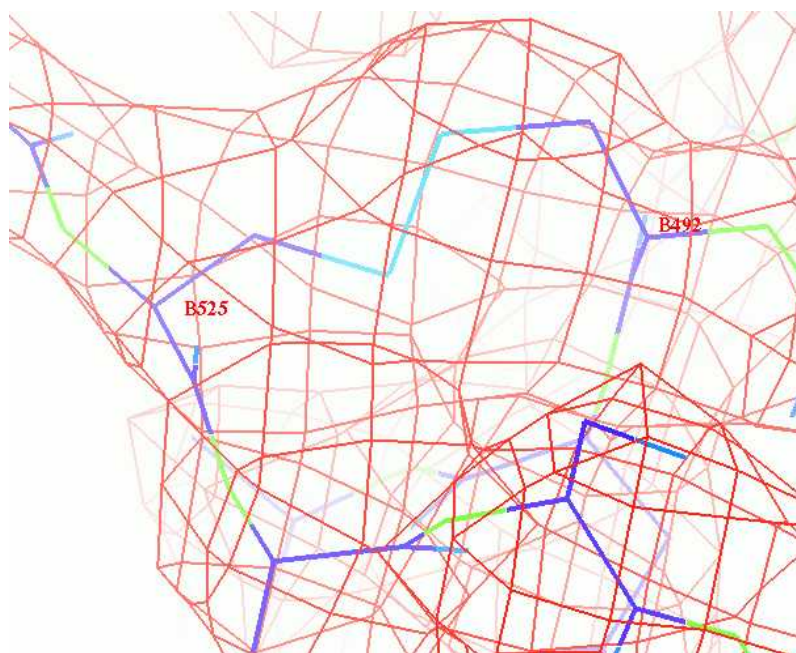
Figure 7.8: Disulphide bond in *Af*PGA shown along with (2Fo-Fc) electron density map (1)
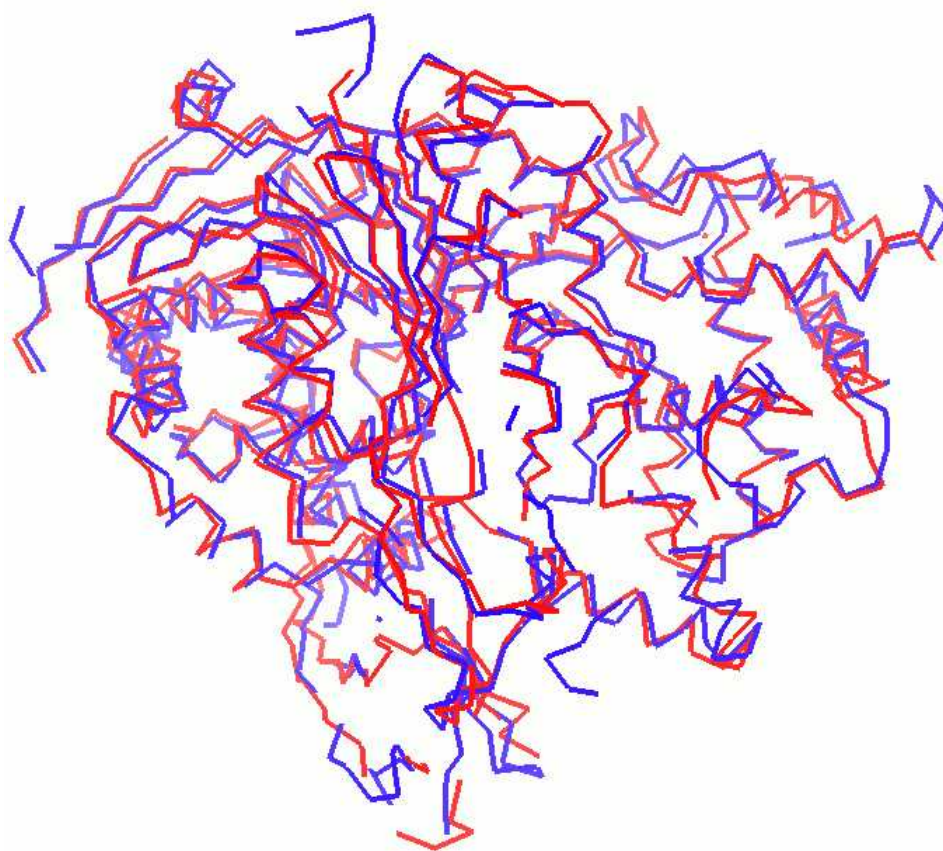
Figure 7.9: The superposition of C backbone of *Ec*PGA (blue) and *Af*PGA (red)

The superposition of preliminary structure of AfPGA on EcPGA shows there is major deviation in the loop regions especially from residue number 360 to 370 and residue number 465 to 474.

There are four key residues conserved in the active site of the *Af*PGA and *Ec*PGA structures: the catalytic nucleophile residue Ser 1, plus Gln 23, Ala 69, and Asn 241. The amide bond of the substrate is cleaved when the nucleophilic O of Ser 1 attacks the carbonyl oxygen of the substrate forming an acyl-enzyme intermediate. Bond breakage is facilitated by positioning the substrate side-chain carbonyl group in oxyanion hole formed by the main-chain amide nitrogen of Ala 69 and N of Asn 241 (Duggleby et al., 1995). Most of the structural differences between the acylases *Ec*PGA and *Af*PGA are located in the aminic subsite. This difference is mainly due to the mutation of Phe 71 in *Ec*PGA to Pro in *Af*PGA.

CHAPTER 8

COMPARISON OF NTN-HYDROLASES
INCLUDING NTN-HYDROLASE DOMAINS

8.1 Introduction

To compare the Ntn-hydrolase superfamily of proteins we have divided them into three categories based on the type of N-terminal nucleophile residue, which is a cysteine, serine or a threonine. An extensive sequence comparison and analysis was carried out in each category separately. Many related proteins from eukaryotes in the database were identified in serine and cysteine groups. In the category where threonine was the N-terminal nucleophile residue two distinct groups could be identified based on the closeness of amino acid sequences. Thus, through careful sequence comparison we not only could identify new, but distantly related, Ntn-hydrolase members or domains but also could place in this family some of the un-annotated proteins in the database.

A variety of enzymes with varied substrate specificity, classified by their characteristic and distinct fold, form the N-terminal nucleophile (Ntn) hydrolase superfamily. Despite lack of any discernible sequence similarity, the representative structures of Ntn hydrolases show that similar fold and topological coincidence spatially conserve the amino acid residues important for activity. Because of the spatially conserved active site they are also mechanistically related. However, the nature of the nucleophile residue, oxyanion hole residues and topology of binding sites greatly differ. The evolution of enzyme function and the nuances of catalysis of Ntn hydrolases can be fully deciphered only by a complete analysis of the sequences and structures along with corresponding detailed phylogenetic analysis. The structural analysis of individual members of superfamily has revealed how nature has optimized binding and catalysis, and re-structured old proteins for new activities through gene duplication and mutation (Kumar et al., 2006).

Statistics of Ntn-hydrolase family (adopted from phylofacts database -
http://phylogenomics.berkeley.edu/)

Superfamily code   : 56235
Fold name   : Ntn hydrolase -like
No of genomes   : 275
No of Phyla   : 22
No of sequences   : 867
Average size   : 236
Diversity   : 0.132897

In every individual of the family the terminal of one of the -strands of the characteristic fold is decorated with the nucleophile residue, a Ser, Cys or Thr whose free -amino group act as the base in catalysis (Brannigan et al., 1995). Minor modification of the oxyanion hole occurs in terms of the residues involved depending also on the type of nucleophile residue present at the N-terminus. Based on the N-terminal nucleophile residue these hydrolases can be widely classified into three sub-groups/families, of those possessing a cysteine, serine, or a threonine at the N-terminus. Well-refined representative structures for all three types, the Cys-, Ser- and Thr-families exist. Here we have used the representative sequences and structures of PVA and BSH for Cys-family, that of PGA for Ser-family and that of L-asparaginase (*Flavobacterium Meningosepticum*) for Thr-family. The presence of Ntn-hydrolases span over several organisms, both prokaryotes and eukaryotes. They exist as single functional protein molecule or as part of a protein domain. The Pei & Grishin (2003) has identified that U34 peptidase family belonged to the Ntn hydrolase fold and consisted of choloyglycine hydrolases, acid ceramidases, isopenicillin N acyltransferases, and a subgroup of proteins with unclear function. A multiple sequence alignment arranges the protein sequences into a rectangular array so that residues in a given column are homologous, superposable or plays a common functional role (Edger & Batzoglou, 2006). Based on their amino acid sequences and structural information, attempt is made here to organize these proteins phylogenetically and functionally into sub-families depending on their sequence-relationship, substrate specificities and evolutionary closeness.

In the study reported here extensive sequence analysis is carried out to identify different protein families belonging to Ntn-hydrolase superfamily and to understand their functional and evolutionary relationships.

8.2. Results

8.2.1. Penicillin V acylase: N-terminal cysteine nucleophile (Ntcn) hydrolase

Peptidases are a diverse group of enzymes that hydrolyse the peptide bonds in protein, peptides and various other molecules. These peptidases are classified based on the participating residues in the catalysis. The new family of Ntn hydrolases, although similar to peptidases in terms of the type of bonds they cleave, they are identified more as amidases and they show great economy in terms of those groups participating in the catalytic activity. In contrast to common peptidases in which catalytic center is made up of a triad of three groups, Ntn hydrolase are made up of a single catalytic center. A base adjacent to the catalytic amino acid is necessary and expected to enhance the nuclophilic character of the side chain nucleophile groups (-OH or -SH). Very often there is a bridging water molecule from nucleophile atom to the free -amino group in the same residue which act as base. Some of the peptidases like U34 family are recently identified to belong to Ntn hydrolase superfamily using extensive sequence analysis and the fold characteristics (Pei & Grishin., 2003). The members of this family exhibit considerable sequence variation and individuals show wide specificity towards a variety of substrates.

Using the sequence of *Bsp*PVA as query a protein-protein Blast search was conducted with default input parameters which output many protein sequences of PVA and BSH from diverse sources, mainly from microorganisms. To identify homologous proteins in higher organisms analysis of a group classified as cholylglycine hydrolases in Pfam (Batman et al., 2002) was carried out. It has now established that bile salt hydrolase is very closely related to PVA, evident from the similarity in active site residues and substrate recognition and binding (Kumar et al., 2006). The three-dimensional structures are also exceptionally similar with differences mainly confine to substrate binding loop that play role in substrate specificity. A sequence homology analysis and structural comparison of BSH and PVA revealed that four of the five amino acids at the active site of PVA are conserved in BSH (Tanaka et al., 2001). Although sequence and structure of PVA and BSH are very similar, differences are observed in certain critical positions. Further investigations are necessary to explore the role of residues in these key positions responsible for substrate selectivity
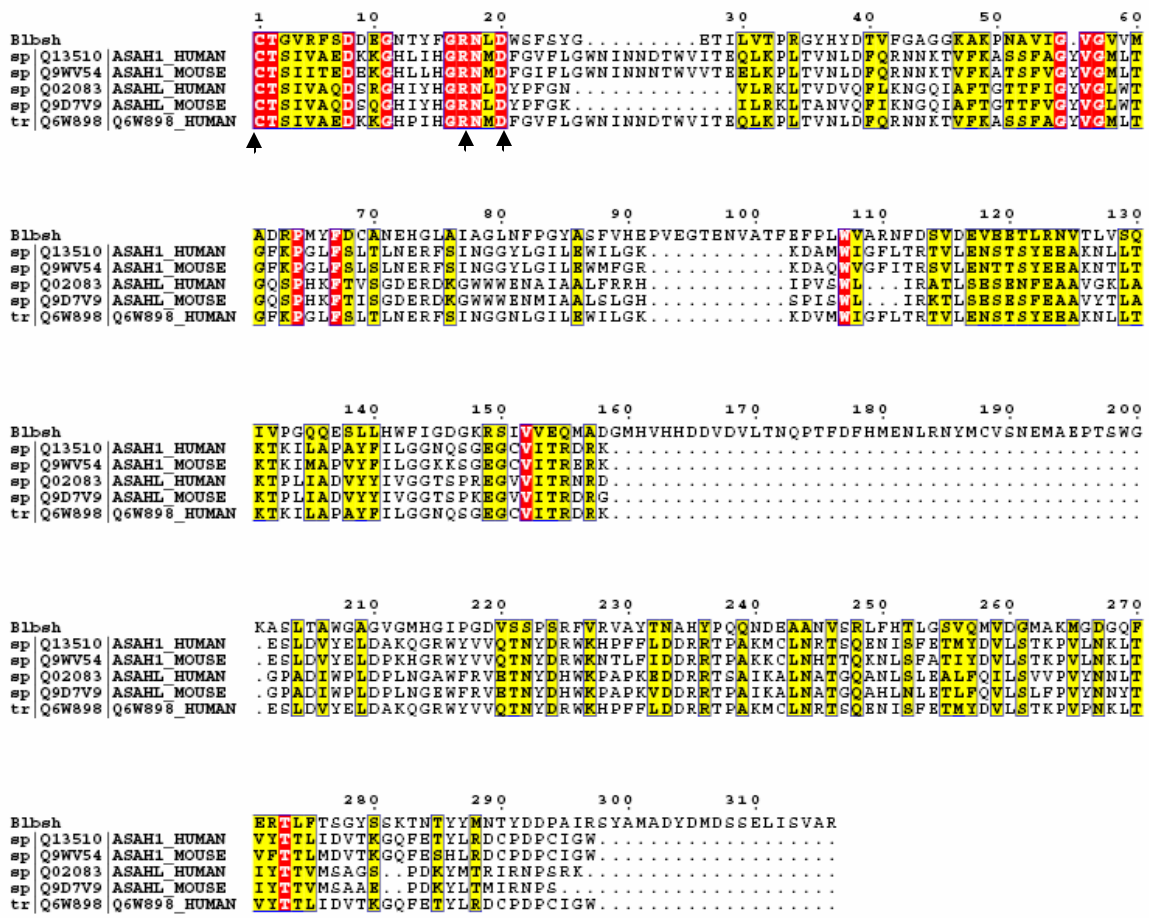
Figure 8.1: sequence alignment of BSH with ASAH of human and mouse. Arrows indicate the positions of conservation of crucial amino acids between BSH and ASAH.

Choloylglycine hydrolase family in Pfam database contains 132 homologous sequences from different organisms. The N-acylsphingosine amidohydrolase (ASAH) also called as Putative 32 kDa heart protein, sequence from mouse was selected and protein-protein Blast was repeated again. The Blast gave 68 hit sequences. Sequence alignment was performed using sequences obtained from Blast using *Bl*BSH as reference sequence and ASAH protein from mouse used as query sequence. In humans, the N-acylethanolamine-hydrolyzing acid amidase that hydrolyse various N-acylethanolamines has N-palmitylethanol-amines as the most reactive substrates. And they are identical to acid ceramidase but lack ceramide hydrolyzing activity (Hassler and Bell, 1993). The sticking sequence similarity between *Bl*BSH and Human ASAH and ceramidase are clearly depicted in figure 8.1.

Group C

N-acylethanolamine-
hydrolyzing acid amidase

N-acylsphingosine amidohydrolase
(even from *Homo sapiens*)

Hypothetical and
unnamed protein

* Virus

* Fungus

Group A

Bacillus subtilis

Choloylglycine hydrolase
& PVA from *Bacillus
(cereus, thuringiensis,
anthracis)*

Bacillus Sphaericus

* Archea

* Archea

Clostridium perfringens

* Virus

Bile salt hydrolases

Bifidobacterium longum

PVA   Group B   proteins

Figure 8. 2: Dendrogram (unrooted) based on PVA and related proteins shows

its relationship with proteins of higher eukaryotic organisms
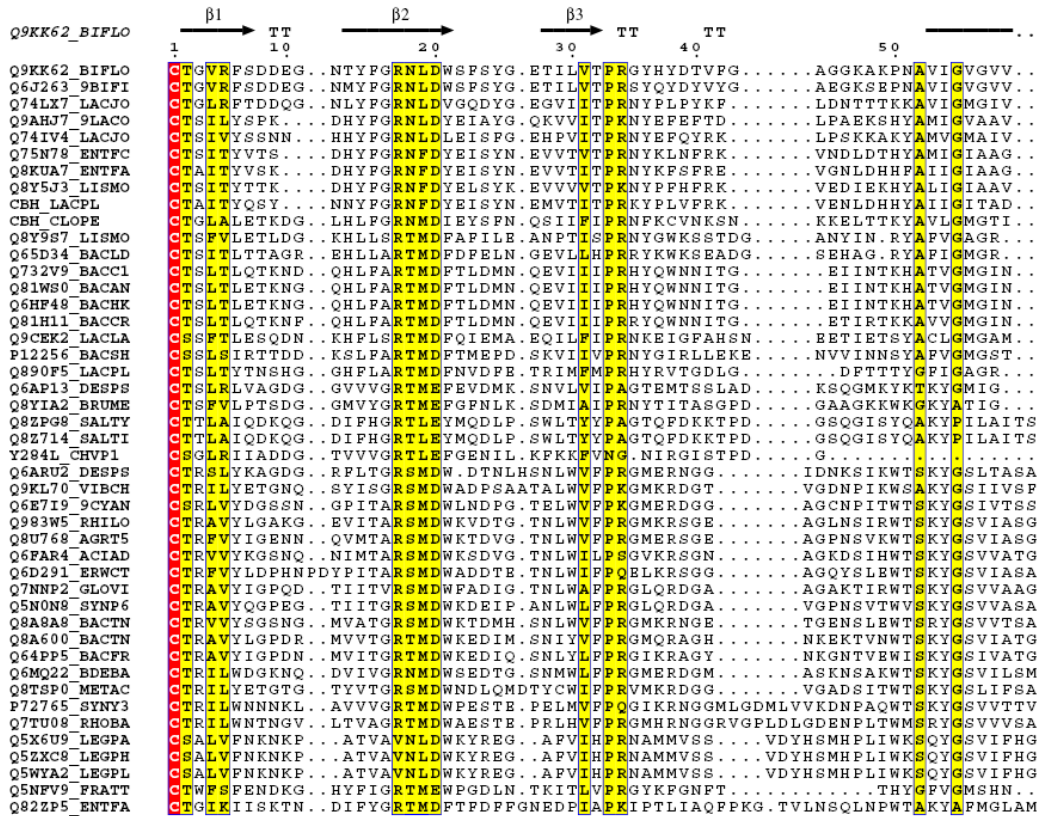
Figure 8.3: The multiple sequence alignment of proteins that contain BSH domain

An unrooted tree of the same aligned sequences shows that the choloylglycine hydrolases and PVA from *B. cereus, B. thuringiensis,* and *B. anthracis* are close to PVA or BSH from other species and organisms (Figure 8.2). Rooted tree shows that it is somewhat different from the rest, while unrooted tree shows that they belong to the same main branch as that of *Clostridium, Bifidobacterium, Lactobacillus* etc. Branches of this unrooted tree are grouped into three: group A, group B, group C, all the proteins coming under each group are mentioned in Table 8.1a-c. Multiple alignment of these proteins are shown in figure 8.3.

Table 8.1a: Details of protein sequences coming under group A of the unrooted tree in PVA related sequences

ID	Organism	Protein Name	Identity with *Bsp*PVA
gi:59713908	*Vibrio fischeri*	choloylglycine hydrolase family	29
gi:78170791	*Chlorobium chlorochromatii*	Parallel beta-helix repeat	29
gi:50875094	*Desulfotalea psychrophila*	related to penicillin acylase	29
gi:78496699	*Rhodopseudomonas palustris*	Choloylglycine hydrolase	28
gi:68206335	*Desulfitobacterium hafniense*	similar to penicillin acylase	28
gi:77959020	*Yersinia bercovieri*	COG3049: Penicillin V acylase	28
gi:76884707	*Nitrosococcus oceani*	Choloylglycine hydrolase	26
gi:35210929	*Gloeobacter violaceus*	gll0368	25
gi:19914562	*Methanosarcina acetivorans*	choloylglycine hydrolase	25
gi:60494473	*Bacteroides fragilis*	putative exported hydrolase	24
gi:77977191	*Yersinia intermedia*	Penicillin V acylase	24
gi:49612653	*Erwinia carotovora subsp. atroseptica*	putative exported choloylglycine hydrolase	24
gi:49531137	*Acinetobacter sp.*	putative choloylglycine hydrolase protein	24
gi:39648799	*Rhodopseudomonas palustris*	Choloylglycine hydrolase	24
gi:75824900	*Vibrio cholerae RC385*	COG3049: Penicillin V acylase	24
gi:1651853	*Synechocystis sp.*	slr1772	24
gi:32442940	*Rhodopirellula baltica*	putative hydrolase	23
gi:53753613	*Legionella pneumophila str.*	hypothetical protein	23
gi:39574783	*Bdellovibrio bacteriovorus*	choloylglycine hydrolase	23
gi:17743076	*Agrobacterium tumefaciens str.)*	choloylglycine	23
gi:74419941	*Nitrobacter winogradskyi*	choloylglycine hydrolase	23
gi:46486690	*Lyngbya majuscula*	choloylglycine hydrolase-like protein	23
gi:32397178	*Rhodopirellula baltica*	probable Penicillin acylase	23
gi:78368089	*Shewanella sp.*	Similar to Penicillin V acylase	22
gi:84386555	*Vibrio splendidus*	choloylglycine hydrolase family	22
gi:67005062	*Rickettsia felis*	Penicillin acylase	22
gi:48856803	*Cytophaga hutchinsonii*	COG3049: Penicillin V acylase	22
gi:29339396	*Bacteroides thetaiotaomicron*	choloylglycine hydrolase	21
gi:69933117	*Paracoccus denitrificans*	Choloylglycine hydrolase	21

Table 8.1b: Details of protein sequences coming under group B of the unrooted tree of PVA related sequences.

ID	Organism	Protein Name	Identity with *Bsp*PVA
gi:65320548	*Bacillus anthracis str. A2012 (B)*	COG3049: Penicillin V acylase	45
gi:49329921	*Bacillus thuringiensis serovar konkukian str(B)*	choloylglycine hydrolase	42
gi:47568265	*Bacillus cereus G9241(B)*	choloylglycine hydrolase family protein	42
gi:16411537	*Listeria monocytogenes*	lmo2067	40
gi:60417969	*Clostridium perfringens*	putative penicillin acylase	36
gi:6457643	*Lactobacillus acidophilus (B)*	conjugated bile salt hydrolase	35
gi:58254514	*Lactobacillus acidophilus NCFM*	bile salt hydrolase	35
gi:82748280	*Clostridium beijerincki NCIMB 8052(B)*	similar to penicillin amidase	35
gi:81427820	*Lactobacillus sakei subsp. sakei 23K(B)*	Choloylglycine hydrolase	35
gi:12802353	*Lactobacillus gasseri*	putative bile salt hydrolase	34
gi:28271943	*Lactobacillus plantarum*	choloylglycine hydrolase	33
gi:77969651	*Burkholderia sp.*	Penicillin amidase	32
gi:46205263	*Magnetospirillum magnetotacticum MS-1(B)*	COG3049: Penicillin V acylase	32
gi:57636819	*Staphylococcus epidermidis*	penicillin V acylase	32
gi:68446022	*Staphylococcus haemolyticus*	unnamed protein product	32
gi:72493970	*Staphylococcus saprophyticus*	putative choloylglycine hydrolase	32
gi:62513764	*Lactobacillus casei*	COG3049: Penicillin V acylase	31
gi:62464250	*Lactococcus lactis subsp. cremoris*	COG3049: Penicillin V acylase	31
gi:41582426	*Lactobacillus johnsonii NCC 5*	conjugated bile salt hydrolase	31
gi:84489564	*Methanosphaera stadtmanae*	putative bile salt acid hydrolase	31
gi:9631852	*Paramecium bursaria Chlorella virus*	PBCV-1 amidase	30
gi:7707363	*Bifidobacterium longum*	bile salt hydrolase	29
gi:47121626	*Bifidobacterium bifidum*	bile salt hydrolase	29
gi:59713424	*Vibrio fischeri*	choloylglycine hydrolase	29
gi:50876071	*Desulfotalea psychrophila*	related to choloylglycine	29

ID	Organism	Protein Name	Identity with *Bsp*PVA
		hydrolase	
gi:57286722	*Staphylococcus aureus*	Choloylglycine hydrolase family protein	29
gi:12724864	*Lactococcus lactis subsp.*	penicillin acylase	29
gi:62546323	*Bifidobacterium adolescentis*	bile salt hydrolase	28
gi:83630916	*Bifidobacterium breve*	bile salt hydrolase	28
gi:57340168	*synthetic construct*	hypothetical protein FTT1110	28
gi:54113153	*synthetic construct*	NT02FT1253	28
gi:60492444	*Bacteroides fragilis*	putative choloylglycine hydrolase	28
gi:29344933	*Enterococcus faecalis*	choloylglycine hydrolase family protein	28
gi:48865090	*Oenococcus oeni*	COG3049: Penicillin V acylase	28
gi:75856500	*Vibrio sp.*	COG3049: Penicillin V acylase	27
gi:48870357	*Pediococcus pentosaceus*	COG3049: Penicillin V acylase	25
gi:46486762	*Bifidobacterium animalis*	bile salt hydrolase	23
gi:67548680	*Burkholderia vietnamiensis*	Choloylglycine hydrolase	23

Table 8.1c: Protein sequences coming under group C of the unrooted tree in PVA
related sequences

ID	Organism	Protein Name	Identity with mouse ASAH
gi:12004240	*Rattus norvegicus*	ceramidase	90
gi:3860240	*Homo sapiens*	putative heart protein	79
gi:16877108	*Homo sapiens*	ASAH1 protein	78
gi:34559851	*Homo sapiens*	HSD-33	73
gi:50746657	*Gallus gallus*	Similar to N-acylsphingosine amidohydrolase like precursor	66
gi:51258328	*Xenopus laevis*	MGC82286 protein	58
gi:76689006	*Bos taurus*	similar to N-acylsphingosine amidohydrolaselike precursor	37
gi:55622652	*Pan troglodytes*	similar to serine/threonine protein phosphatase	35

ID	Organism	Protein Name	Identity with mouse ASAH
gi:76655700	*Bos taurus*	similar to N-acylsphingosine amidohydrolase	35
gi:53733722	*Danio rerio*	N-acylsphingosine amidohydrolase	35
gi:21166363	*Takifugu rubripes*	N-acylsphingosine amidohydrolase	35
gi:55819499	*Acanthamoeba polyphaga mimivirus(V)*	putative N-acylsphingosine amidohydrolase	34
gi:52782189	*Macaca fascicularis*	N-acylsphingosine amidohydrolase	34
gi:55725853	*Pongo pygmaeus*	hypothetical protein	34
gi:72012235	*Strongylocentrotus purpuratus*	similar to N-acylethanolamine-hydrolyzing acid amidase	34
gi:23129179	*Nostoc punctiforme*	COG5295: Autotransporter adhesin	33
gi:39589104	*Caenorhabditis briggsae*	Hypothetical protein	33
gi:3876339	*Caenorhabditis elegans*	Hypothetical protein F27E5.1	33
gi:74001865	*Canis familiaris*	similar to serine/threonine protein phosphatase	32
gi:73979488	*Canis familiaris*	similar to N-acylsphingosine amidohydrolase	32
gi:72004704	*Strongylocentrotus purpuratus*	similar to N-acylsphingosine amidohydrolase	29
gi:82500235	*Caldicellulosiruptor saccharolyticus*	oxidoreductase, nitrogenase component 1	28
gi:66846055	*Aspergillus fumigatus Af293*	hypothetical protein	23
gi:22549408	*Mamestra configurata*	putative ODVP-E6/ODV-E56	22
gi:67932020	*Solibacter usitatus Ellin6076*	Cell surface receptor IPT	20
gi:42553043	*Gibberella zeae PH-1*	hypothetical protein	20

8.2.3. Penicillin G acylase: N-terminal serine nucleophile (Ntsn) hydrolase

When conducting search using conserved domain database of PGA, two types of conserved domains could be distinguished: a cephalosporin acylase domain (gn1lCDDl30162) which contains proteins like cephalosporine acylase and -glutamyl transferase (Ggt) and a penicillin amidase domain (gn1lCDDl25823) contains penicillin acylase and other related protein.. Many hits of cephalosporin acylase domain having similarity with the query sequence of *Ec*PGA were identified. These sequences were

separately blasted to identify homologous proteins from higher organisms. Only gi|30249112 (PGA from *Nitrosomonas europaea*) resulted in hit for a protein from human (gi|40788291).

A human protein containing pleckstrin homology domain (KIAA0571) showed sequence similarity with PGA. Careful examination of the alignment also revealed that conserved catalytic residues at similar positions along the sequence were present, like in *Ec*PGA. Some of the important conserved residues are S1, H38, G72, G94, P208, P232, P227, N241, L323, P463, V359 (Figure 8.4).
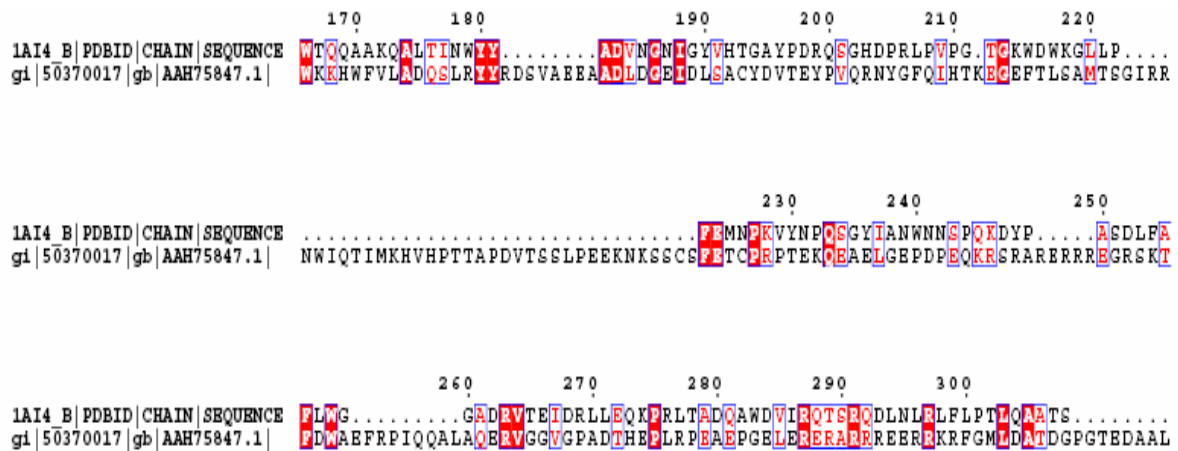
Figure 8.4: Sequence alignment showing the conservation of amino acid residues in PH domain which correspond to those in PGA from *E.coli*

To identify more related proteins in higher organisms, the sequence of human KIAA0571 protein itself was used as query sequence in another protein-protein BLAST search. Interestingly, this blast procedure has retrieved 1245 hits. The blast search output included sequences from both higher and lower organisms of functionally diverse proteins. From this set only 322 sequences were selected by applying redundancy criteria (less than 80% similarities). These sequences along with PGA sequences from *E. coli* and *Nitrosomonas* and human KIAA0571 protein were aligned using clustal family of programs (X, W). Dendrograms were constructed from this alignment, as an unrooted and a rooted tree, using NJplot and unrooted, respectively. This whole alignment contained only two sequences obtained by search using PGA sequence, which can be

249

considered as standards for PGA. Rest of them was obtained by blasting with KIAA0571 protein sequence.

One interesting feature of the result was the detection of Pleckstrin Homology (PH) domain in majority of the sequences. Thus, we could discover the relationship of PH domain to penicillin G acylases and thus to Ntn-hydrolase superfamily in general. Some of the putative functions of this domain includes binding to beta/gamma subunit of heterotrimeric G proteins, binding to lipids like phosphatidylinositol-4,5-bisphosphate, binding to phosphorylated Ser/Thr residues and also attaching to membranes through an unknown mechanism (Ponting & Bork, 1996). This domain is present in both N-terminal and C- terminal side of pleckstrin the protein involved in phosphoinositide-signalling pathways of platelet activation. The three-dimensional structure of this domain reveals a seven stranded antiparal -sandwich derived from two -sheets and in carboxy terminal -helix closes one end of the -sandwich (Jackson et al., 2006).

In the present study we have observed that many hypothetical and unannotated proteins in many organisms (except in *C. elegans*) can be classified as belonging to Ntn-hydrolase superfamily. The most significant finding could be that for the first time PGAs were identified as related to proteins containing PH domain and especially with KIAA0571 protein and myosin phosphatase – rho interacting protein (M-RIP) in human. Among them the M-RIP has more similarity with PGAs than KIAA0571 protein in terms of conserved residues (Figure 8.5). Out of the four identified active site residues three (equivalent to S1, Q23, 241N of PGA) are present in this protein indicating functional level conservation of residues. More than 80 residues of *Ec*PGA are common with M-RIP from human whereas only 11 residues are common with KIAA0571 protein.

Figure 8.5: Alignment showing the conserved residues of PGA in M-RIP protein of human

251

Similarity also exist between PGA and DNA directed RNA II polymerases (Human). In this case more than 40 residues are conserved between the two. The residue R263 of PGA which is identical with the residue R225 of PVA is found in similar position in the alignment of these polymerases.

The rooted tree consists of 8 main branches with many sub-branches coming out of them. Most of these branches contain identical or functionally related proteins belonging to different organisms or species. Habitually this tree consists of proteins like DNA directed RNA polymerase II, prokaryotic protein sequences, protein kinases, insulin receptor substrate 1, Grb2-associated binder 3 (Gab3) proteins, hypothetical or unknown protein products, PH domain containing proteins, PGA etc (Figure 8.6). It is not that the main branch contains only one kind of protein. Many proteins, which were annotated as hypothetical or proteins of unknown functions were grouped with other well-characterized proteins. We can deduce important conclusions about the function of those hypothetical proteins identified in various organisms grouped here with proteins of known function (Figure 8.7).

Unrooted tree was constructed by including only sequences obtained using *Ec*PGA sequence as reference in Blast search, and by including both PGA and PH domain proteins in the search. Both the unrooted trees are having large number of main branches. The unrooted tree drawn of *Ec*PGA type sequences alone has three main branches named as group A, B and C. The proteins coming under each group has been listed in Table 8.2a-c. All three groups contain proteins of bacterial origin, group B has more protein sequences from Archea bacteria, while group A has only one sequence from Archea. The proteins PGA, PH domain containing protein and human M-RIP protein all clustered together. We have already discussed how our analysis revealed the closeness of these proteins. However, the query protein (KIAA0571) used for picking up M-RIP protein is placed in another branch. This may be because, as already stated, the query sequence of KIAA0571 is less similar to PGA compared to M-RIP in terms of conserved residues. From the unrooted tree it also becomes clear that the DNA directed RNA polymerase II is also close to PGA in the evolutionary tree.

0.1

M-RIP protein (15, 11.49)

Protein Kinase (12, 10.85)

Related to RNA helicase (4, 10.95)

Prokaryotic protein sequences (13, 11.31)

Merozoite surface protein 1 (8, 12.5)

PH domain containing protein (7 & 6, 10.36 & 10.58)

IRS1 (11, 10.95)

GRB2, Gab3 proteins (10 & 9, 10.88 &11.48)

KIAA0571 protein (3, 9.87)

DNA directed RNA polymerase 11(14, 9.87)

Nanog (5, 12.66)

PGA (*Nitrosomonas*) (5, 13.82)

PGA (*E.coli*)
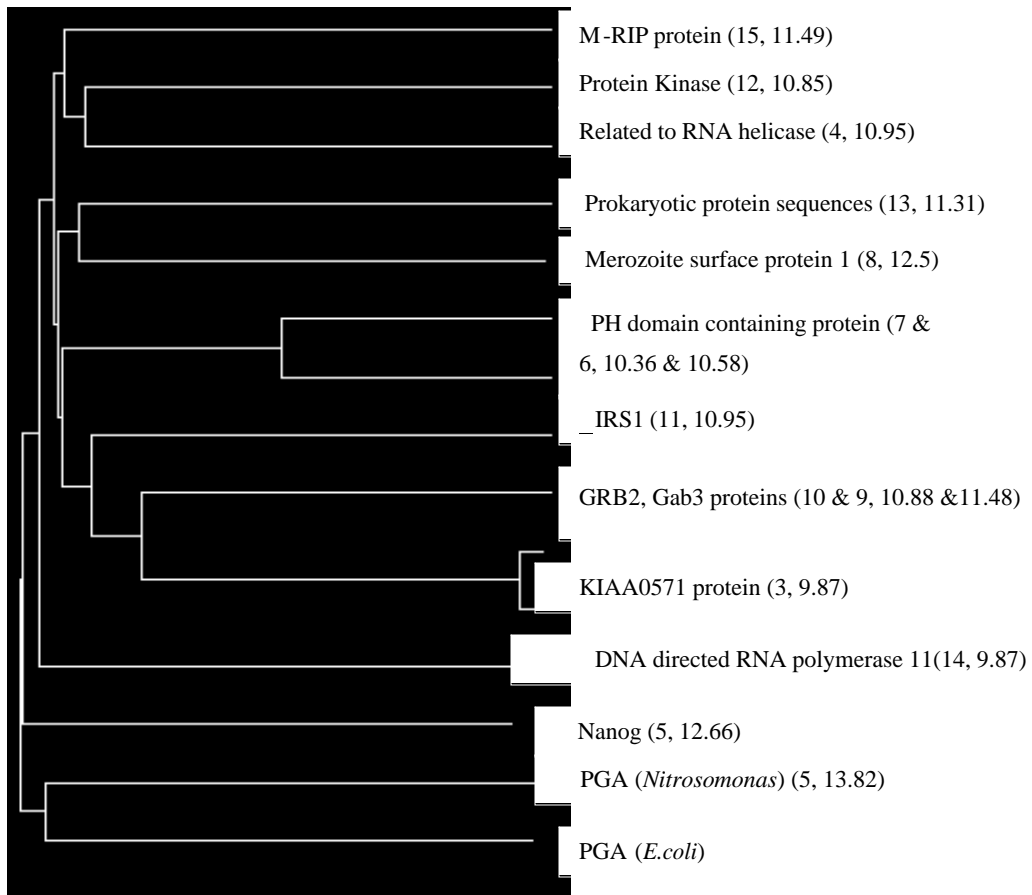
Figure 8.6: Evolutionary distribution of PGA domain in varius organisms

Protein Kinases

Serine-threoninerich protein &
cell waii surface anchor protein

IRS1

DNA directed RNA
polymerases 11

**GRB2,
Gab3**

PGA

M -RIP (Human)

Hypothetical,
Unknown
protein
products

PH domain containing protein

Figure 8.7: Unrooted tree showing the evalutionary distribution of PGA domain in
various organisms

Table 8.2a: Protein sequences coming under group A of the unrooted tree of PGA

ID	Organism	Protein Name	Identity with *Ec*PGA
gi:2960449	Streptomyces avermitilis MA-4680(B)	putative amidase	30
gi:53756640	Methylococcus capsulatus str. Bath(B)	penicillin acylase II	24
gi:69158282	Shewanella denitrificans OS217(B)	Penicillin amidase	25
gi:68545064	Shewanella amazonensis SB2B(B)	Penicillin amidase	30
gi:76873967	Pseudoalteromonas haloplanktis TAC125(B)	putative hydrolase	23
gi:21112454	Xanthomonas campestris pv.(B)	penicillin acylase II	28
gi:84367481	Xanthomonas oryzae pv.(B)	penicillin acylase II	24
gi:21107608	Xanthomonas axonopodis pv.(B)	penicillin acylase II	25
gi:82702090	Nitrosospira multiformis ATCC 25196(B)	Penicillin amidase	25
gi:13814881	Sulfolobus solfataricus P2(A)	Penicillin acylase precursor	24
gi:71549148	Nitrosomonas eutropha C71(B)	Penicillin amidase	23
gi:71143533	Colwellia psychrerythraea 34H(B)	penicillin amidase family protein	24
gi:67673169	Burkholderia pseudomallei 1655(B)	hypothetical protein Bpse1_02001323	25
gi:67647651	Burkholderia mallei NCTC 10247(B)	Protein related to penicillin acylase	24

Table 8.2b: Protein sequences of group B of the unrooted tree of PGA

ID	Organism	Protein Name	Identity with *Ec*PGA
gi:76557896	Nitronomonas pharaonis	predicted amidase	
gi:2650138	Archaeoglobus fulgidus	penicillin G acylase	21
gi:18161405	Pyrobaculum aerophilum str.	penicillin amidase	25
gi:5103975	Aeropyrum pernix K1(A)	penicillin acylase	24
gi:10639962	Thermoplasma acidophilum(A)	penicillin amidase precursor related protein	26
gi:30138729	Nitrosomonas europaea	Penicillin amidase	25
gi:68568154	Sulfolobus acidocaldarius	penicillin amidase	24
gi:48431231	Picrophilus torridus	penicillin acylase	25
gi:68140395	Ferroplasma acidarmanus	Penicillin amidase	23
gi:10640802	Thermoplasma acidophilum	penicillin amidase precursor related protein	29
gi:14325462	Thermoplasma volcanium	penicillin G acylase	24

GSS1

gi:1353728	Naegleria fowleri	penicillin amidase homolog	29
gi:56909090	Bacillus clausii	penicillin acylase	23
gi:13422434	Caulobacter crescentus	penicillin amidase family protein	29
gi:67929371	Solibacter usitatus Ellin6076	penicillin amidase	29
gi:83648152	Hahella chejuensis KCTC 2396(B)	Protein related to penicillin acylase	22
gi:36786834	Photorhabdus luminescens subsp.	unnamed protein product	23
gi:6460680	Deinococcus radiodurans	aculeacin A acylase	25
gi:71556768	Pseudomonas syringae	penicillin amidase family protein	23
gi:68344568	Pseudomonas fluorescens Pf-	penicillin amidase family protein	27
gi:53728202	Pseudomonas aeruginosa UCBPP-PA14	Protein related to penicillin acylase	30
gi:67157312	Azotobacter vinelandii AvOP	penicillin amidase	26
gi:912439	Actinoplanes utahensis	aculeacin A acylase	26
gi:68556149	Ralstonia metallidurans	penicillin amidase	32
gi:61658434	Burkholderia cepacia	glutaryl acylase beta-subunit	23
gi:7579066	Brevundimonas diminuta	glutaryl 7-aminocephalosporanic acid acylase	23
gi:67923096	Crocosphaera watsonii	similar to Protein related to penicillin acylase	22
gi:1001472	Synechocystis sp.	7-beta-(4-carbaxybutanamido)cephalosporanic acid acylase	21
gi:83756446	Salinibacter ruber	Penicillin amidase superfamily	23
gi:23128434	Nostoc punctiforme	Protein related to penicillin acylase	30
gi:40062950	uncultured bacterium	penicillin G acylase	23
gi:68537283	Sphingopyxis alaskensis	Penicillin amidase	27
gi:84786195	Erythrobacter litoralis	penicillin amidase family protein	27
gi:68347503	Pseudomonas fluorescens	penicillin amidase family protein	25
gi:67923095	Crocosphaera watsonii	similar to Protein related to penicillin acylase	22

gi:15226315	Arabidopsis thaliana	aspartic-type endopeptidase	26
gi:5441768	Streptomyces coelicolor	putative penicillin acylase	27
gi:55232989	Haloarcula marismortui	penicillin acylase	27
gi:400723	Arthrobacter viscosus	Penicillin G acylase precursor	30
gi:71040788	Bacillus badius	penicillin G acylase	30
gi:5596630	Bacillus megaterium	penicillin G amidase	30
gi:18072029	Alcaligenes faecalis	penicillin G acylase precursor	40
gi:69934773	Paracoccus denitrificans PD1222	Penicillin amidase	40
gi:63015128	Achromobacter sp.	penicillin G acylase	51
gi:30089602	Achromobacter xylosoxidans	penicillin G acylase	51
gi:32527600	Providencia rettgeri	penicillin G amidase	64
gi:20379146	synthetic construct	mutant penicillin G acylase precursor	63
gi:74314774	Shigella sonnei 6	hypothetical protein SSO_4482	98
gi:129551	Kluyvera cryocrescens	Penicillin G acylase precursor	86
gi:82546688	Shigella boydii	putative penicillin G acylase	98
gi:75198156	Escherichia coli	Protein related to penicillin acylase	-

Table 8.2c: Protein sequences belonging to group C of the unrooted tree of PGA

ID	Organism	Protein Name	Identity with *Ec*PGA
gi:84322022	Pseudomonas aeruginosa C3719	related to penicillin acylase	26
gi:67157375	Azotobacter vinelandii AvOP	Penicillin amidase	26
gi:24982545	Pseudomonas putida KT2440	penicillin amidase family	23
gi:71554506	Pseudomonas syringae pv. phaseolicola 1448A	penicillin amidase family protein	25
gi:82737474	Pseudomonas putida F1	penicillin amidase family	23
gi:13122142	Streptomyces coelicolor	putative antibiotic binding protein	24
gi:29607827	Streptomyces avermitilis	putative penicillin acylase	24
gi:18092572	Brucella abortus	putative penicillin acylase II	23
gi:23464469	Brucella suis 1330	penicillin amidase family protein	26
gi:17984352	Brucella melitensis 16M	Penicillin acylase	27
gi:67927141	Solibacter usitatus	Penicillin amidase	24
gi:67738529	Burkholderia pseudomallei	Protein related to penicillin	24

		acylase	
gi:52422011	Burkholderia mallei	putative penicillin amidase	24
gi:78034028	Xanthomonas campestris	putative penicillin amidase	25
gi:84502482	Oceanicola batsensis	penicillin acylase	26
gi:22778993	Oceanobacillus iheyensis	penicillin acylase	25
gi:74023542	Rhodoferax ferrireducens D	Penicillin amidase	25
gi:39575971	Bdellovibrio bacteriovorus	Penicillin G acylase precursor	22
gi:24193719	Leptospira interrogans ser	Penicillin G acylase precursor	23
gi:45599424	Leptospira interrogans serovar Copenhageni str	complete sequence	24
gi:68181785	Jannaschia sp.	Penicillin amidase	22
gi:83859222	Oceanicaulis alexandrii	Penicillin amidase	27
gi:84516231	Loktanella vestfoldensis	penicillin amidase family protein	25
gi:84684893	Rhodobacterales bacterium	penicillin amidase family protein	23
gi:83369082	Rhodobacter sphaeroides	Penicillin amidase	22
gi:9946944	Pseudomonas aeruginosa	probable penicillin amidase	26
gi:69297696	Silicibacter sp.	Penicillin amidase	23
gi:56678683	Silicibacter pomeroyi	penicillin amidase family protein	25
gi:83952198	Roseovarius nubinhibens	penicillin amidase family protein	22
gi:83953591	Sulfitobacter sp.	penicillin amidase family protein	25
gi:35212555	Gloeobacter violaceus	glr1988	30
gi:66796722	Deinococcus geothermalis	Penicillin amidase	29
gi:76258036	Chloroflexus aurantiacus	penicillin amidase	22
gi:46200274	Thermus thermophilus	penicillin acylase	30
gi:46202692	Magnetospirillum magnetotacticum	Protein related to penicillin acylase	29
gi:67906920	Polaromonas sp.	Penicillin amidase	27
gi:47575568	Rubrivivax gelatinosus	Protein related to penicillin acylase	31
gi:72118943	Ralstonia eutropha	Penicillin amidase	28
gi:67736332	Burkholderia pseudomallei	Protein related to penicillin acylase	28
gi:51854896	Symbiobacterium thermophilum IAM	penicillin amidase	30
gi:56909090	Bacillus clausii	penicillin acylase	23
gi:65320573	Bacillus anthracis str.	Protein related to penicillin acylase	24
gi:42738270	Bacillus cereus	Complete Genome	22
gi:49329663	Bacillus thuringiensis	penicillin acylase II	24
gi:22778994	Oceanobacillus iheyensis	penicillin acylase	23
gi:67929356	Solibacter usitatus	Penicillin amidase	25
gi:71915777	Thermobifida fusca	penicillin amidase	23
gi:84497783	Janibacter sp.	penicillin amidase	29

gi:24426510	Streptomyces coelicolor	putative penicillin acylase	24
gi:29607329	Streptomyces avermitilis	putative penicillin acylase	24
gi:17133058	Nostoc sp. PCC 7120	all3924	28
gi:23125242	Nostoc punctiforme	Protein related to penicillin acylase	22

8.2.4. (-N-acetyl-D-glucosaminyl)-L-asparaginase: N-terminal threonine nucleophile (Nttn) hydrolase

Glycosylasparaginase (glycoasparaginase, N^4- (-N-acetyl-D-glucosaminyl)-L-asparagine amidohydrolase) is a widely distributed amidohydrolase involved in the ordered degradation of N-linked glycoproteins (-N-acetyl-D-glucosaminyl)-L-asparaginase from *Flavobacterium meningosepticum* sequence has been used to perform a protein-protein Blast search that resulted in 309 hit sequences. The hit sequences include protein sequences also from higher organisms like *Homo sapiens*, Mouse etc. There was no need to perform any other search since the ordinary PSI-BLAST ended up outputting sequences from higher organisms as well. These sequences have been aligned using clustal family of programs and a phylogenetic tree was constructed. L-asparaginases have Active site residue threonine at the N-terminal (Figure 8.8).
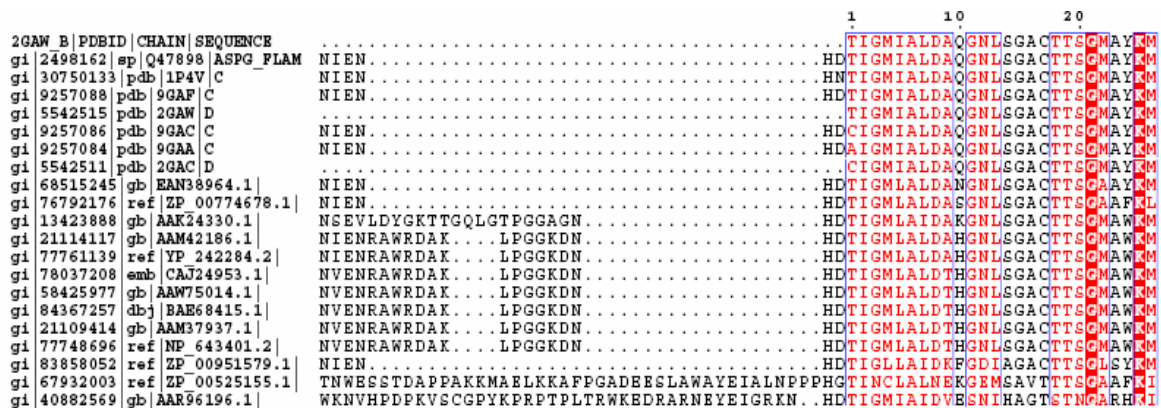
Figure 8.8: ClustalX alignment showing conservation of T at the N-terminal of L-asparaginases from different organisms or species and among them three are mutants

Almost all the hits obtained were asparaginases, but two other proteins showed high similarity towards reference sequence, Tapase 1 (gi|19263670) had similarity of more than 18 residues and Twin-arginine translocation pathway signal protein (gi|68547028) showed similarity score more than 63 and all the identified active site residues were conserved in it (Figure 8.9).
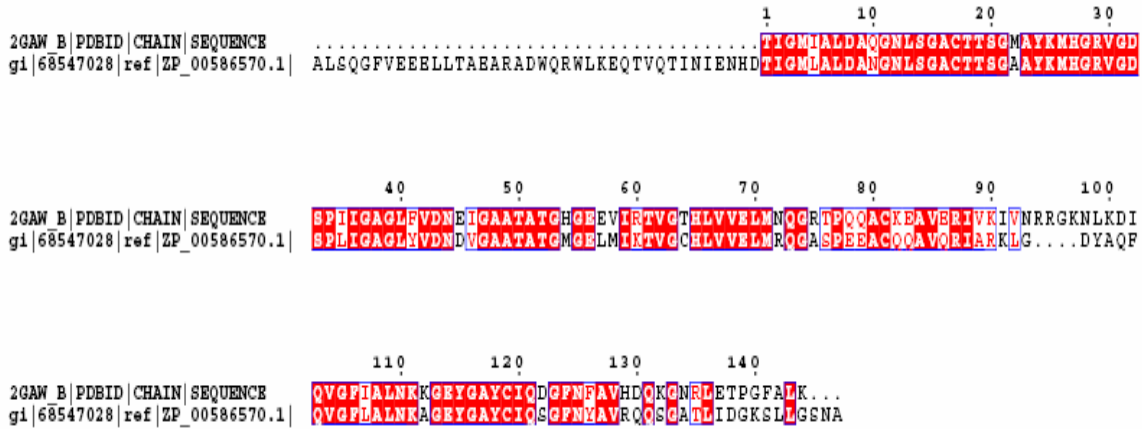
Figure 8.9: Alignment showing conservation of residues between L-asparaginase from *Flavobacterium meningosepticum* and Twin-arginine translocation pathway signal protein

L-asparaginases from different organisms or species are spread over almost all branches of a dendrogram constructed. One single branch contained proteins from prokaryotes like bacteria to higher organism like human. One branch has proteins like Threonine aspartase 1(tapase), malonyl CoA-acyl carrier protein transacylase, some hypothetical proteins etc; all other branches contain L-asparaginases alone. The sequence alignment of tapase and L-asparaginase is shown in Figure 8.10.

Twin-arginine translocation (tat) pathway signal protein is present in the same branch as that of L-asparaginase from *Flavobacterium meningosepticum.* One close homolog of dipeptidase A (pepDA*)* is characterized experimentally as an extracellular arginine aminopeptidase from *Streptococcusgordonii* (gi no. 16506526). This protein has a typical membrane export signal sequence of 14 hydrophobic residues.

```
2GAW_B|PDBID|CHAIN|SEQUENCE  .........................TIGMIALDAQGNLSGACTTSGMAYKMHGRVGDSPIIGACLFVDNEIG
gi|19263670|gb|AAH25266.1|   TDFMQLKKRRQSSEKENDSGTLDTVGAVVVDHEGNVAAAVSSGGLALKHPGRVGQAALYGCCWABNTGA

2GAW_B|PDBID|CHAIN|SEQUENCE  ......AATATGHGEEVIRTVGTHLVVELMNQGRTPQQACKEAVERIVK.............IVNR...
gi|19263670|gb|AAH25266.1|   HNPYSTAVSTSGCGEHLVRTILARECSHALQAEDAHQALLETMQNKFISSPFLASEDGVLGGVIVLRSCR

2GAW_B|PDBID|CHAIN|SEQUENCE  ......RGKNLKDIQVGFIALNKKG..EYCAYCIQDG......FNFAVHDQKCNRLETPGFALK......
gi|19263670|gb|AAH25266.1|   CSAEPDSSQNKQTLLVEPLWSHTTESMCVGYMSAQDGKAKTHISRLPPGAVAGQSVAIEGGVCRLESPVN
```

Figure 8.10: **Alignment to show conservation of residues between L-asparaginase from *Flavobacterium meningosepticum* and Tapase1 of Human**

8.3. Discussion

Sequence analysis and Blast searches reported here have identified many proteins hitherto not identified as homologous to members of Ntn‑hydrolase superfamily. Such proteins could be identified and placed in this superfamily. Many proteins from eukaryotes in the database were identified to belong to serine‑ and cysteine‑Ntn hydrolases. Proteins like KIAA0571 protein and M-RIP protein from humans, DNA directed RNA polymerase II and many proteins containing PH domain showed similarity with PGA from *E.coli*. N‑acylsphingosine amidohydrolase and N-acylethanolamine-hydrolyzing acid amidase seamed to be homologous to BSH from various organisms. In the category of threonine-family two distinct groups (tapase and tat) could be identified based on the closeness of amino acid sequences.

The presence of BSH domain in various proteins of different organisms seems to be a great enigma, because the functional significance of this acquisition remains a mystery. Several speculations are available for the occurrence of BSH even within the bacterial species. It has been proposed that the *bsh* gene integrated into the genome by horizontal (lateral) gene transfer in lactobacilli (Elkins & Savage, 1998) and in *L. monocytogenes* (Dussurget et al., 2002). The BSH domain present in several proteins of higher organisms needs further investigations. The predicted catalytic cysteine residue is

right after the cleavage site and, thus, is exposed after the removal of the signal sequence. The acid ceramidases usually have a relatively long sequence from N-terminal to the catalytic cysteine. The removal of this N-terminal part may be an autoproteolytic process, like in many other Ntn-hydrolases.

Sequence analysis of L-asparaginase from *Flavobacterium meningosepticum* showed similarity towards Twin arginine translocation pathway signal and Tapase1 to a great extend, this may be due to the conservation of AGA -domain.

There is functional requirement to organize and maintain the definite active site structure that probably exerts a strong selective pressure on a protein to adopt just one stable and conserved fold (Grishin., 2001; Andreeva & Murzin, 2006). These results indicate that related sequences can diverge to such an extent that their common ancestry get obscured at the sequence level and to some extent can be identified in the next level of organization, the structure or the topological conservation of certain residues. A conserved functional feature is associated with one or more key residues that are invariant across a family of proteins; owing to their important functional role, these amino acids are subject to evolutionary constraints and their loss during evolution would be deleterious to protein function. Meanwhile, residues at other positions have more degree of freedom to mutate. Some of the invariant residues could be implicated in structural conservation also, in that they maintain the integrity of the protein fold.

There are many reasons to believe that any similarity in reaction chemistry shared by enzyme homologues is mediated by common functional groups conserved through out evolution. However, detailed enzyme studies have revealed the flexibility of many active sites, in that different functional groups are not conserved with respect to their positions in the primary sequence but mediate the same mechanical role. Nevertheless, the catalytic atoms are expected to be spatially equivalent. More rarely, the active site might have completely different location in the protein scaffold. This variability could result from:

1) The hopping of functional groups from one position to another to optimize catalysis

2) The independent specialization of a low-activity primordial enzyme in different phylogenetic lineages

3) Functional convergence after evolutionary divergence

4) Circular permutation events.

In enzyme homologs exist in which residues that play the same role in catalysis are located at non-equivalent position in the structural scaffold. Non- equivalent residues are those that do not align in the structure based sequence alignment.

In some cases, the non-equivalent, but functionally analogous, catalytic residues are identical. It might be that, for these enzymes, there is an optimum amino acid type for the chemical role in question. Also in some cases there is a conservative difference in amino acid identity.

In few cases, the position of the specific atoms involved in catalysis is preserved, whereas the residues to which they belong lie at different points in the protein scaffold. For these active sites at least, this spatial conservation implies that there is an optimum disposition of functional atoms for catalysis, although there is a degree of flexibility with respect to the locations of the residues that contains them. In the sequence comparisons used here we not only could identify new, but distantly related, Ntn-hydrolase members or domains but also could place in this family some of the un-annotated proteins from the database.

CONCLUSION

Two enzymes Bile salt hydrolase (BSH) and Pencillin G acylase (PGA), belonging to the Ntn‑hydrolase family, have been studied extensively using various biophysical and biochemical techniques available to the author. The structure-function relationships of individual enzymes, the evolutionary relationships between them and with other members of the family have been elucidated. By extending the analysis to all known sequences from the databases new members belonging to the family but hitherto not identified with Ntn hydrolases could be annotated.

Following are the highlights of the study reported in the thesis:

1. Two Ntn hydrolase enzymes were cloned and over expressed for their extensive characterization.

2. Both the proteins were crystallized and the structure was determined and compared with the related structures.

3. The relationship between the BSH and PVA enzymes were established through biochemical, biophysical and computational techniques.

4. The purification of PGA precursor from *K. citrophila* was achieved and insights into its processing mechanism were obtained.

5. The evolutionary and mechanistic relationships between two proteins, one a conjugated bile salt hydrolase and another a penicillin G acylase, was elucidated.

6. By extending the information obtained from above studies in conjunction with extensive sequence analysis the possible relationships of the members of three sub‑families of Ntn hydrolase superfamily was elucidated. This analysis also helped us to identify new proteins which could in all probability belong to Ntn hydrolase family or would have originated from a member of the family.

In the appendix following the main part of this thesis a brief description of the initiative at developing a database of proteins involved in genetic disorders is described.

APPENDIX

DEVELOPING A DATABASE OF PROTEINS INVOLVED IN THE GENETIC DISORDERS
Genetic Disease- Protein Database (GDPD)

Introduction

The cause of genetic diseases can be invariably traced back to mutation(s) that would have affected the three-dimensional structure of proteins and which would have disturbed the function of single protein or a complex involved in a particular biological function. The information on such deleterious mutations and the role of affected proteins causing the genetic diseases are spread over various independent databases. It is a major challenge to bring all these information either together or to bring them into an easily retrievable form with the help of a database tool.

The database (Genetic disease-protein database –GDPD) created and developed at NCL attempts at integrating every particular disorder identified as genetic disease with its corresponding causal mutation and proteins responsible for the disease. It is also linked to the corresponding functional and structural genomic annotation information available for that protein and wherever possible information on the clinical aspects of the disease is also provided. All the mutations identified as responsible for each disease are listed and wherever a related database already exists a link is provided. Tool is also provided to update and display information pertaining to the mutation. Efforts have been undertaken to see that this database provides all the available information on the effect of any mutation, especially the deleterious ones, on the structure of proteins involved in every known genetic disease for which the three-dimensional structure of the target protein is determined.

Contents of the database

On developing this database attention was directed at organizing the information in a cause and effect direction. Thus the first information database provides is on the gene that

gets affected. Thus, each disease is extensively linked to genomic information like the gene loci, gene sequence, marker database, cDNA sequences and Expressed Sequence Tags (ESTs). This is followed by the molecular aspects, especially that of the proteins. Proteomics information like subcellular location and biological function of the protein, protein three-dimensional structure, domain classification, biochemical pathways etc. are linked to all the available databases on the target protein. Sometimes in the case of certain diseases which are known for long time and subject on which lot of research has been carried out, there will be wealth of information of general interest to researchers available. This information is covered by providing details of the clinical aspect of the disease like symptoms, pathogenesis, treatment, clinical trials and again by linking exhaustively to various active websites on that disease. Database is also linked to Pubmed storing research papers on the disease and is linked to the laboratories working on that particular protein or disease. The information available for common genetic disorders are listed in Table A. 1.

Table A. 1: Details of the Important Diseases from the Database

S. No	Name of the disease	Name of the protein	Chromosome location	PDB ID	Function of the protein	No of mutations
1	**Antithrombin III deficiency**	AT3 serine (or cysteine) proteinase inhibitor, clade C member 1	**1q23-q25.1**	**1ATH 1ANT 1BR8 2ANT**	Serine protease inhibitor activity	**156**
2	**Rheumatoid arthritis.**	L1CAM vascular cell adhesion molecule	**1p32-p31**	**1VCA 1VSC**	Leukocyteendothelial cell adhesion. Interacts with the beta-1 integrin vla4 on leukocytes, and mediates both adhesion and signal transduction	**6**
3	**Breast cancer**	Transforming growth factor alpha	**1q42.1**	**2TGF, 3TGF 4TGF, 1YUF 1YUG**	-	**75**
4	Pycnodysostosis	Cathepsin K	**1q21**	**1MIM, 1ATK 1AU0, 1AU2 1AU3**	Closely involved in osteoclastic bone resorption and may participate partially in the disorder of bone remodeling	**14**
5	**Macular degeneration, juvenile macular dystrophy with flecks, type 1**	ABCA4 ABCR STGD1 ATP-binding cassette	**1p22**	**1F30**	Involved in the transport of chloride ions	**303**
6	**Prostrate cancer**	Ribonuclease L	**1q25**	**1MYO**	Endoribonuclease, mediator of interferon action	**3**
7	**Zellweger**	PMP70	**8q21.1**	**1QND**	Somewhat implicated in the	**5**

S. No	Name of the disease	Name of the protein	Chromosome location	PDB ID	Function of the protein	No of mutations
	syndrome cerebro-hepatic-renal syndrome	PXMP1 Peroxisomal membrane protein		1IKT 4Q21	biogenesis of peroxisomes	
8	Nemaline myopathy 2	Nebulin NEM2	2q22	1ARK 1NEB	Muscle development, Cell structure, Muscle action	15
9	Apnea, postanesthetic	Acylcholine acylhydrolase butyrylcholine esterase	3q26.1-q26.2	1EHO 1EHQ	Butyrylcholinesterase precursor, Other metabolism	
10	Long QT syndrome-3 Brugada syndrome	Sodium channel, voltage-gated, type V, alpha polypeptide	2q24	1QG9	Sodium channel	2
11	AFP deficiency, Dysalbuminemic hyperthyroxinemia	ALPHA-FETOPROTEIN; AFP	4q11-q13	1AO6, 1BJ5 1BKE, 1BMO 1E78, 1E7C 1E7A	Binds copper, nickel, and fatty acids as well as, and bilirubin less well than, serum albumin	1
12	Aspartylglucosami nuria	SPARTYLGLUCOS AMINIDASE	4q32-q33	1APY 1APZ	Cleaves the glcnac-Asn bond which joins oligosaccharides to the peptide of asparagine-linked glycoproteins.	
13	Cystic Fibrosis	ATP-binding cassette, sub-family C	7q31.2	1QK9 1CKW	Respiration invasive growth small molecule transport, Gas exchange, Pathogenic Invasion,	100

S. No	Name of the disease	Name of the protein	Chromosome location	PDB ID	Function of the protein	No of mutations
					Osmoregulation and Excretion	
14	**Pfeiffer syndrome Jackson-Weiss sydnrome**	BFGFR CEK FGFBR Fibroblast growth factor receptor 1	**8q**	**1FGK 1FGI 1AGW**	Receptor for basic fibroblast growth factor.	**2**
15	**Leukemia, chronic myeloid**	ABL1 (v-abl Abelson murine leukemia viral oncogene homolog1 JTK7 v-abl Abelson murine leukemia viral oncogene homolog 1	**9q34.1**	**1AB2 1AWO 1FPU 1IEP 1BBZ ABL1**	Oncogenesis , mismatch repair , cell cycle control , protein modification , transcription regulation mitotic S-specific transcription	
16	**Fanconi anemia, complementation group C**	FAC FACC Fanconi anaemia, complementation group C	**9**	**1EG3 1DXX 1EG4 1QAG**	DNA repair protein that may operate in a post replication repair or a cell cycle checkpoint function	**18**
17	**Osteoporosis**	CALCITONIN; CALC1; CT	**11p15**	**1LS7**	Katacalcin is a potent plasma calcium-lowering peptide	**2**
18	**Sickle cell anemia Thalassemias, beta-(3)**	Beta-globin hemoglobin, beta	**16p**	**1CH4 1BZ0 1JY7 1CBL 1CBM 1COH 1HBS 1HCO 2HCO 1HHO 2HHB 3HHB 4HHB**	Involved in oxygen transport from the lung to the various peripheral tissues.	**86**
19	**Colorectal**	RASK2	**12p**	**1RVD**	Is able to inhibit all four classes of	**2**

S. No	Name of the disease	Name of the protein	Chromosome location	PDB ID	Function of the protein	No of mutations
	adenoma Colorectal cancer	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog			proteinases by a unique "trapping" mechanism.	
20	**Von Willebrand disease**	VWF F8VWF von Willebrand factor	**12p13**	**1ATZ 1AUQ**	Important in the maintenance of homeostasis, it participates in platelet-vessel wall interactions by forming a noncovalent	**171**
21	**Retinoblastoma**	Retinoblastoma-associated protein (PP110) (P105-RB) (RB).	**13q14**	**1AD6 1GUX**	Probably acts as a regulator of other genes. Forms a complex with adenovirus E1A and with SV40 large T antigen. Acts as a tumor suppressor.	**207**
22	**Wilson disease**	PWD WND atpase, Cu++ transporting, beta polypeptide	**13q**	**1AW0**	Involved in the export of copper out of the cells, such as the efflux of hepatic copper into the bile.	**241**
23	**Agammaglobulinemia Agammaglobulinemia,Bruton like**	Immunoglobulin heavy polypeptide gene cluster (V,D,J,C)	**14q32.33**	**1FC1 1FC2 1IGE**	All 4 combinations of the s/g & v/g polymorphisms at positions 192 and 216 have been observed in human mu chains.	**3**

Mutational and structural analysis of proteins involved in genetic disorders

For a query based on the name of the disorder or relevant protein as the key word, this database lists out updated information about the number of mutations deposited on the particular disorder, positions of the mutations in the protein sequence and details about the resultant change in amino acid.

For many of the disorders listed in the database, the structure for the normal protein or the mutated protein is not available. So some 20 disorders have been selected based on the availability of the structure of wild and mutant proteins involved. A comparison between the normal and abnormal protein structures has been carried out in order to find the cause of the disease. The two types of structures have been superimposed to estimate the root mean square of the positions of C and to detect large RMS deviations in any specific region. The differences in the percentages of secondary structures are also analysed. In most of the cases no significant changes are occurring in the secondary structure as a result of the mutation, but there are deviation in the overall conformation of the mutated protein compared to normal one. The diseases and the information of proteins of interest are provided in Table A. 2. Sine the structures of these proteins are determined and coordinates deposited references to their PDB codes are also provided in the table. In the database there is link provided to PDB site in each case.

For our analysis we had collected disease-associated mutations correlated with the protein structure and functional relationship. We have examined various structural effects of these mutations. So far we have made some preliminary analyses and are planning more comprehensive analyses based on these data. We also need to gather many structural descriptors and physical and functional properties of proteins. Some interesting associations are observed in our analysis. For example in deficiency or defects in factor VIII result in hemophilia A, the most common of the severe inherited bleeding disorders. Generally Factor VIII complexed with von Willebrand factor. Arg2307 and Trp2229, also in the C2 domain likely interact directly with VWF (Pratt et al., 1999). Mutation in C2 domain at the sulfated Tyr1680 residue has been shown to affect VWF binding, resulting in 5-fold decrease in affinity and this is responsible for the slowing down of clotting reactions (Leyte et al., 1991). The crystal structure of wild protein and mutant Factor VIII protein showed that the disruption of these binding region which hampers the

cascade of reaction in the blood clotting. The frequency of this mutation was observed high in the population and this is mutation is responsible for the development of mild hemophilia A (Pratt et al., 1999).

Systematic analysis of the relationship between disease-associated mutations and structural properties of proteins would provide insight into the molecular mechanism of disease and help developing drugs toward target proteins.

Table A.2: Mutation and structural analysis of genetic disorders

S. No	Disease	Protein Name	Normal PDB Id	Resolution ( )	Mutant PDB Id	Mutation,Withposition	Resolution ( )	Rms Value
1	Amytropic Lateral Sclerosis	Superoxide Dismutase1	Ispd	2.4	1AZV	G 37 R	1.9	1.084
					1SOS	C 6 A, C111S	2.5	0.838
					1MFM	C 6 A,F 50E,G51E,C111S,E133Q	1.02	1.262
2	Afp Deficiency	Alpha-Fetoprotein	1e7a	2.2	1HK2	R 218 H	2.8	0.906
					1HK3	R 218 P	2.8	0.906
					1HK5	R 218 H	2.7	0.285
3	Agammaglobulinemia,Type 1 X-Linked	Tyrosine Kinase	1b55	2.4	1BTK	R 28 C	1.6	1.71
					1BWN	E 41 K	2.1	0.497
4	Alcohol Intolerance,Acute	Aldehyde Dehydrogenase	1o02	1.9	1O04	C 302 S,LIGAND INTERACTING	1.42	0.12
					1NZW	C 302 S,LIGAND INTERACTING	2.65	0.137
5	Amyloid Neuropathy	Transthyretin	1f41	1.3	1F86	T 119 M	1.1	0.359
					1TSH	T 60 A	1.7	0.219
					1BZE	T 119 M	1.8	0.22
					1BZD	SER 6	1.9	0.231
					2TRH	C 10 R	1.9	0.253
					2TRY	S 77 Y	2	0.285
6	Atransferrinemia	Transferrin Serotransfe-RRIN PRECURSOR	1b3e	2.5	1JQF	H 249 Q	1.85	0.998
					1FQF	K 296 A	2.1	0.968
					1DTG	H 249 E	2.4	0.369
					1N7W	L 66 W	2.2	0.831
					1N7X	Y 45 E	2.1	0.838
					1OQG	D 63 E	1.9	0.967
7	Hemolytic Crisis	Glutathione Reductase	1grb	1.85	2GRT	A 34 E, R 37 W	2.7	0.285
					3GRT	A 34 E, R 37 W	2.5	0.303
					4GRT	A 34 E, R 37 W	2.8	0.448
					5GRT	A 34 E, R 37 W	2.4	0.283
8	Isolated Growth Hormone Deficiency	Growth Hormone 1	1hwg	2.5	1AXI	G 120 R,K 168 R,D 171 T,K 172Y,E174 A,F 176 Y	2.6	1.109
					1BP3	G 120 R,F 146 G	2.9	1.917
					1HUW	F 10 A,M 14 W,H 18 D,H 21 N,K 41 I,Y42 H,L 45 W,	2	2.025

					Q 46 W, F 54 P,R 64 K,R 167 N, D171 S,E 174 S		
					F176 Y,I 179 T		
				1HWH	G 120 R	2.9	1.729
9 Crouzon Syndrome	Fibroblast Growth Factor RECEPTOR 2	1ev2	2.2	1II4	S 252 W	2.7	0.541
10 Lchad Deficiency	Hydroxy Acyl Coa DEFICIENCY	3had	2	1F12	F 80 C	2.4	0.712
				1F14	F 80 C	2.3	0.254
				1F17	F 80 C	2.3	0.156
				1LSJ	E 110 Q	2.5	0.326
				1LSO	S 137 A	2.6	0.284
11 D-Bifunctional Protein Deficiency	Hydroxy Steroid Dehtdro-GENASE 4	3dhe	2.3	1FDV	H 221 L	3.1	1.156
				1FDW	H 221 Q	2.7	0.645
12 Emphysema-Cirrhosis	Antitrypsin	1qlp	2	1ATU	F 51 L, T 59 A,T 68 A,A TO G,M 374 I,S 381 A,K 387R	2.7	2.119
				1KCT	T 59 A,T 68 A, A70 G	3.46	2.64
				1PSI	F 51 L	2.9	0.534
13 Erythrocytosis,Familial	Erythropoeitin Precursor	1ern	2.4	1EER	N 52 Q, N 164 Q,A 211 E	1.9	2.717
				1CN4	N 52 Q, N 164 Q,A 211 E	2.8	2.129
14 Lesch-Nyhan Syndrome	Hypoxanthine-Guanine-PHOSPHORIBOSYL TRANSFERASE	1BZY	2	1D6N	K 68 A	2.7	1.41
15 Mcad	Mcad Acyl Coenzyme A DEHYDROGENASE	1T9G	2.9	1EGC	T 255 E, E 376 G	2.6	0.508
				1EGD	T 255 E, E 376 G	2.4	0.426
				1EGE	T 255 E, E 376 G,NO CHANGE IN THE POSITIONS	2.75	0.402
16 Metachromatic Leukodystro0phy	Human Arylsulphatase A	1n2k	2.75	1E1Z	C 69 S	2.4	0.268
				1E2S	C 69 A	2.35	0.261
					P 426 L	2.5	0.281
				1E3C	C 69 S	2.65	0.284
17 Thyroid Hormone,Resistance	Thyroid Hormone Receptor	1nax	2.7	1NQ0	A 234 T	2.4	0.762
				1NQ1	L 240 K	2.9	0.951
				1NQ2	A 317 T	2.4	0.801
				1NUO	R 316 H	3.1	1.302
18 Von Willebrand Disease	Von Willebrand Factor	1u0n	2.95	1IJB	I 546 V	1.8	0.521

				1IJK	I 546 V	2.6	0.422
				1M10	E 543 G,NO CHANGE IN POSITION	3.1	1.294
19	Pancreatic Cancer	Smad4	1dd1	2.62	1G88	R 515 S	3	0.316
20	Sicle Cell Anemia	Beta-Globin Hemoglobin	1bz0	1.5	1HBA	W 37 R	2.1	0.336
				1HBB	W 37 R,NO CHANGE IN POSITION	1.9	0.316
				1DXV	V 1A	1.8	0.267
				1HDB	V 67 T	2.2	0.167
				1A00	V 1M,T 37 W	2	0.401
				1A01	V 1M,T 37 A	1.8	0.157
				1A0U	V 1 M	2.14	0.388
				1GBU	C 93 A		0.157
				1GBV	C112 G		0.174

REFERENCES

Abadjieva, A., Hilven, P., Pauwels, K., and Crabeel, M. (2000) The yeast ARG7 gene product is autoproteolyzed to two subunit peptides, yielding active ornithine acetyltransferase. *J. Biol. Chem*. **275**, 11361–11367.

Ahn, Y. T., Kim, G. B., Lim, K. S., Baek, Y. J., and Kim., H. U. (2003) Deconjugation of bile salts by *Lactobacillus acidophilus* isolates. *Int. Dairy J.* **13**, 303–311.

Alkema, W. B. L., Prins, A. K., De Vries, E., and Janssen, D. B. (2002) Role of αArg [145] and βArg [263] in the active site of penicillin acylase of *Escherichia coli*. *Biochem. J.* **365,** 303-309.

Alkema, W. B., Dijkhuis, A.-J., de Vries, E., and Janssen, D. B. (2002) The role of hydrophobic active-site residues in substrate specificity and acyl transferase activity of penicillin acylase. *Eur. J. Biochem.* 269, 2093–2100.

Alvaro, G., Lafuente, R. F., Rosell, C. M., Blanco, R. M., Garcia-Lopez, J. L., and Guisan, J. M. (1992) Penicillin G acylase from *Kluyvera citrophila*. New choice as industrial enzyme. *Biotechnol. Lett.* **14**, 285–290.

Anderson, J. W., and Gilliland, S. E. (1999) Effect of fermented milk (yogurt) containing *Lactobacillus acidophilus* L1 on serum cholesterol in hypercholesterolemic humans. *J. Am. College Nutr.* **18**, 43–50.

Andreeva, A., and Muzin, A. G. (2006) Evolution of protein fold in the presence of functional constraints. *Curr. Opin. Struct. Biol*. **16**, 399-408.

Archer, R. H., Chong, R. and Maddox. I. S. (1982) Hydrolysis of bile acid conjugates by *Clostridium bifermentans*. *Eur. J. Appl. Microbiol.* **14**, 41–45.

Aronson, A., Sang, H. Y., and Bourne, N. (1989) Gene structure and precursor processing of a novel *Bacillus subtilis* spore coat protein. *Mol. Microbiol.* **3**, 437-444

Aronson, N. N. Jr. (1999) Aspartylglycosaminuria: biochemistry and molecular biology. *Biochim. Biophys. Acta.* **1455**, 139-154.

Aronson, N. N. Jr. (1996) Lysosomal glycosylasparaginase: a member of a family of amidases that employ a processed N-terminal threonine, serine or cysteine as a combined base-nucleophile catalyst. *Glycobiology*. **6**, 669–675.

Asahara, T., Shimizu, K., Nomoto, K., Hamabata, T., Ozawa, A., and Takeds, Y. (2004) Probiotic bifidobacteria protect mice from lethal infection with shiga toxin-produsing *Escherichia coli* 0157:117. *Infect. Immune*. **72**, 2240-2247.

Asano, Y., Nakazawa, A., Kato, Y., and Kondo, K. (1989) Properties of a novel D-stereospecific aminopeptidase from *Ochrobactrum anthropi*. *J. Biol. Chem*. **264**, 14233-14239.

Balasingam, K., Warburton, D., Dunnil, P., and Lilly, M.D. (1972) The isolation and kinetic of penicillinic amidase from *Escherichia coli*, *Biochim. Biophys. Acta*. 276, 250-256.

Barbero, J. L., Buesa, J. M., Buitrago, G. G., Mendez, E., Perez-Aranda, A., and Garcia, J. L. (1986) Complete nucleotide sequence of the penicillin acylase gene from *Kluyvera citrophila*. *Gene*. **49**, 69-80.

Basso, A., Braiuca, P., Clementi, S., Ebert, C., Gardossi, L., and Linda, P. (2002) Computational analysis of the amino acid subsite of PGA explain the influence of amine structure on enantioselectivity, *J. Mol. Catal. B: Enzymatic*. **19-20**, 423-430.

Basso, A., Braiuca, P., Ebert, C., Gardossi, L., Linda, P., and Benedetti., F. (2001) GRID/tetrahedral intermediate computational approach to the study of selectivity of Penicillin G acylase. *Biochim. Biophys. Acta*. 1601, 85-92.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, Erik, L. L. (2002) The Pfam Protein Families Database. *Nucleic Acids Res*. **30**, 276 - 280.

Bateup, J. M., McConnell, M. A., .Jenkinson, H. F., and Tannock, G. W. (1995) Comparison of *Lactobacillus* strains with respect to bile salt hydrolase activity, colonization of the gastrointestinal tract, and growth rate of the murine host. *Appl. Environ. Microbiol*. **61**, 1147-49.

Batta, A. K., Salen, G., and Shefer, S. (1984) Substrate specificity of cholylglycine hydrolase for the hydrolysis of bile acid conjugates. *J. Biol. Chem*. 259, 15035-15039.

Batta, A. K., Salen, G., Arora, R., Shefer, S., Batta, M., and Person, A. (1990) Side chain conjugation prevents bacterial 7-dehydroxylation of bile acids *J. Biol. Chem*. **265**, 10925-10928.

Batta, A. K., Tint, G. S., Xu, G., Shefer, S., and Salen, G. (1998) Synthesis and intestinal metabolism of ursodeoxycholic acid conjugate with an antiinflammatory agent, 5-aminosalicylic acid. *J. Lipid Res*. **39**, 1641-1646.

Beauchamp, J. C., and Isaacs, N. W. (1999) Methods for X-ray diffraction analysis of macromolecular structures. *Curr. Opin. Chem. Biol*. **3,** 525-529.

Begley, M., Sleator, R. D., Gahan, C. G., and Hill, C. (2005) Contribution of three bile-associated loci, bsh, pva, and btlB, to gastrointestinal persistence and bile tolerance of *Listeria monocytogenes*. *Infect. Immun.* **73**, 894-904.

Begley, M., Gahan, C. G., Hill, C. (2005) The interaction between bacteria and bile. FEMS *Microbiol. Rev.* **29**, 625-651.

Biavati, B. and Mattarelli, P. (2001) in *The Prokaryotes*, edited by Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K. H. and Stackebrandt, E. pp. 1 –70, Springer, New York

Biossel, J. P., Kaper, T. J., and Bunn, H. F. (1988) Cotranslational amino-terminal processing of cytosolic proteins. Cell- free expression of site-directed mutants of human hemoglobin. *J.Biol.Chem*, **263**, 8443-8449.

Bliznakov, E. G. (2002) Lipid-lowering drugs (statins), cholesterol, and coenzyme Q10. The baycol case- a modern Pandora's box. *Biomed Pharmacother*. **56**, 56–59.

Blundell, T. L., and Johnson, L. N. (1976) Protein Crystallography**,** **Vol.** 3, Crystallization of Proteins. . pp. 59-82. Academic Press

Blundell, T. L., Elliott, G., Gardner, S. P., Hubbard, T., and Islam, S. (1989) Protein engineering and design. pp. 324-447 Philos. Trans.R. Soc. London Set. B.

Blow, D. (2002) Outline of Crystallography for Biologist. *Ed.*1. New York, Oxford University Press.

Bochtler, M., Ditzel, L., Groll, M., and Huber, R. (1997) Crystal structure of heat shock locus V (HsIV) from *Escherichia coli*. *Proc Natl Acad Sci. USA* **94,** 6070–6074.

Boistelle, R., and Astier, J. P. (1988) Crystallization mechanisms in solution. *J. Cryst. Growth.* **90**, 14-30.

Bompard-Gilles, C., Villeret, V., Davies, G. J., Fanuel, L., Joris, B., Frere, J. M., and Van-Beeumen, J. (2000) A new variant of the Ntn hydrolase fold revealed by the

crystal structure of L-aminopeptidase D-ala-esterase/amidase from *Ochrobactrum anthropi*. *Structure Fold Des*. **8**, 153-162.

Bongaerts, G., Tolboom, J., Naber, T., Bakkeren, J., Severijnen, R., and Willems, H. (1995) d-Lactic acidemia and aciduria in pediatric and adult patients with short bowel syndrome. *Clin. Chem*. **41**, 107–110.

Bongaerts, G. P., Severijnen, R. S., Tangerman, A., Verrips, A., and Tolboom, J. J. (2000a) Bile acid deconjugation by lactobacilli and its effects in patients with a short small bowel. *J. Gastroenterol*. **35**, 801–804.

Bongaerts, G., Bakkeren, J., Severijnen, R., Sperl, W., Willems, H., Naber, T. (2000b) *Lactobacilli* and acidosis in children with short small bowel. *J. Pediatr. Gastroenterol. Nutr*. **30**, 288–293.

Brahma, A., Mandal, C., and Bhattacharyya, D. (2005) Characterization of a dimeric unfolding intermediate of bovine serum albumin under mildly acidic condition. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics.* **1751**, 159-169

Brannigan, A., Dodson, G., Duggleby, H. J., Moody, P. C., Smith, J. L., Tomchick, D. R., and Murzin, A. G. (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature.* **378**, 416–419.

Breuer, N., and Goebell, H. (1985) The role of bile acids in colonic carcinogenesis. *Klin. Wochenschr*. **63**, 97–105.

Bricogne, G. (1976) Methods and programs for direct-space exploitation of geometric redundancies. *Acta Cryst.* **A32**, 832-847.

Bruggink, A., Roos, E. R., and De Vroom, E. (1998) Penicillin acylase in the industrial production of β-lactam antibiotics. *Org. Proc. Res. Dev.* **2,** 128- 133.

Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. **355,** 472-475.

Brzozowski, A., M., and Walton, J. (2001) Clear strategy screens for macromolecular crystallization. *J. Appl. Cryst*. **34**, 97-101.

Buddington, R. K., Williams, C. H., Chen, S. C., and Witherly, S. A. R. (1996) Dietary supplement of neosugar alters the fecal flora and decreases activities of some reductive enzymes in human subjects. *Am. J. Clin. Nutr.* **63**, 709-16.

Burmeister, W. P. (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Cryst*. **D56**, 328-341.

Canzi, E., Zanchi, R., Camaschella, P., Cresci, A., Greppi,G. F., Orpianesi, C., Serrantoni, M., and Ferrari, A. (2000) Modulation by lactic acid bacteria of the intestinal ecosystem and plasma cholesterol in rabbits fed a casein diet. *Nutr. Res.* **20**, 1329-1340

Chandra, P. M., Brannigan, J. A., Prabhune, A., Pundle, A., Turkenburg, J. P., Dodson, G. G. and Suresh C. G. (2005) Cloning, preparation and preliminary crystallographic studies of penicillin V acylase autoproteolytic processing mutants. *Acta Cryst*. **F61**, 124-127.

Chanez-Cardenas, M. E., Perez-Hernandez, G., Sanchez-Rebollar, B. G., Costas, M., Vazquez-Contreras, E. (2005) Reversible equilibrium unfolding of triosephosphate isomerase from Trypanosoma cruzi in guanidinium hydrochloride involves stable dimeric and monomeric intermediates. *Biochemistry*. **44**, 10883-10892.

Chayen, N. E. (1998) Comparative studies of protein crystallization by vapor diffusion and microbatch. *Acta Cryst*. **D54,** 8–15.

Choi, K. S., Kim, J. A. and Kang, H. S. (1992) Effects of site-directed mutations on processing and activities of penicillin G acylase from Escherichia coli ATCC 11105. *J. Bacteriol*. **174,** 6270-6276.

Chou, C. P., Tseng, J., Kuo, B., Lai, K., Lin, M., and Lin, H. (1999) Effect of SecB chaperone on production of periplasmic penicillin acylase in *Escherichia coli*. *Biotechnol. Prog.* **15**, 439-445

Christiaens, H., Leer, R. J., Pouwels, P. H., and Verstraete, W. (1992) Cloning and expression of a conjugated bile acid hydrolase gene from *Lactobacillus plantarum* by using a direct plate assay. *Appl. Environ. Microbiol.* **58**, 3792– 3798.

Cohen, B.I., and Raicht, R.F. (1981) Effects of bile acids on colon carcinogenesis in rats treated with carcinogens. *Cancer Res*. **41**, 3759–3760.

Cole, M. (1964) Properties of the Penicillin deacylase enzyme of *Escherichia coli*. *Nature.* **203**, 519.

Coleman, J. P., and Hudson, L. L. (1995) Cloning and characterization of a conjugated bile acid hydrolase gene from *Clostridium perfringens. Appl. Environ. Microbiol.* **61**, 2514–2520.

Collaborative Computational Project, Number 4, (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

Colloc'h, N., Poupon, A., and Mornon, J.-P. (2000) Sequence and structural features of the T-fold, an original tunnelling building unit. *Proteins Struct. Funct. Genet. 39*, 142–154

Corzo, G., and Gilliland, S. E. (1999a) Bile salt hydrolase activity of three strains of *Lactobacillus acidophilus. J. Dairy Sci.* **82**, 472–480.

Corzo, G., and Gilliland, S. E. (1999b) Measurement of bile salt hydrolase activity from *Lactobacillus acidophilus* based on disappearance of conjugated bile salts. *J. Dairy Sci.* **82**, 466-471.

Costa, M., Pecci, L., Pensa, B., and Cannella, C. (1977) Hydrogen peroxide involvement in the rhodanese inactivation by dithiothreitol. *Biochem. Biophys. Res. Commun.* **78,** 596–603.

Crowther, R.A., Lenk, E.V., Kikuchi, Y., and King, J. (1977) Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J. Mol. Biol.* **116**, 489-523.

Dambekodi, P. C., and Gilliland, S. E. (1998) Incorporation of cholesterol into the cellular membrane of *Bifidobacterium longum. J. Dairy Sci.* **81**, 1818-24.

Daopin, S., Davies, D. R., Schlunegger, M. P., and Grütter, M. G. (1994) Comparison of two crystal structures of TGF-b 2: the accuracy of refined protein structures. *Acta Cryst.* **D50**, 85-92.

Das, S., Gayen, J. R., Pal, A., Ghosh, K., Rosazza J. P., and Samanta, T. B. (2004) Purification, substrate specificity, and N-terminal amino acid sequence analysis of a beta-lactamase-free penicillin amidase from *Alcaligenes* s p . *Appl. Microbiol. Biotechnol.* **65**, 281-286.

Dauter, Z. (1997) Data collection strategy. *Methods Enzymol.* **276,** 326 - 344.

Dauter, Z. (1999) Data-collection Strategies. *Acta Cryst.* **D55**, 1703-1717.

Dayal, B., and Ertel, N. H. (1997) Studies on N-nitroso bile acid amides in relation to their possible role in gastrointestinal cancer. *Lipids*. **32**, 1331-1340.

De Boever, P., and Verstraete, W. (1999) Bile salt deconjugation by *lactobacillus plantarum* 80 and its implication for bacterial toxicity. *J. Appl. Microbiol*. **87**, 345-352.

De Boever, P., Wouters, R., Verschaeve, L., Berckmans, P., Schoeters, G., and Verstraete, W. (2000) Protective effect of the bile salt hydrolase-active *Lactobacillus reuteri* against bile salt cytotoxicity. *Appl. Microbiol. Biotechnol*. **53**, 709-714.

De Smet, I., Van Hoorde, L., Vande Woestyne, M., Christiaens, H. and Verstraete, W. (1995) Significance of bile salt hydrolytic activities of Lactobacilli. *J. Appl. Bacteriol*. **79**, 292–301.

De Smet, I., De Boever, P., and Verstraete, W. (1998) Cholesterol lowering in pigs through enhanced bacterial bile salt hydrolase activity. *Br. J. Nutr*. **79**, 185–194.

De Smet, I., van Hoorde, L., De Saeyer, N., vande Woewtyne, M., and Verstraete, W. (1994) *In vitro* study of bile salt hydrolase (BSH) activity of BSH isogenic *Lactobacillus plantarum* 80 strains and estimation of cholesterol lowering through enhanced BSH activity. *Microb. Ecol. Health Dis.* **7**, 315–329.

Dean, M., Cervellati, C., Casanova, E., Squerzanti, M., Lanzara, V., Medici, A., de Laureto, P. P., and Bergamini, C. M. (2002) Characterization of cholylglycine hydrolase from a bile-adapted strain of *Xanthomonas maltophilia* and its application for quantitative hydrolysis of conjugated bile salts. *Appl. Environ. Microbiol*. **68**, 3126–3128.

Deshpande, B. S., Ambedkar, S. S., Sudhakaran, V. K., and Shewale, J. G. (1994) Molecular Biology of Beta Lactam Acylases. *World Jour. Microbiol. Biotech.* **10**, 128-138.

Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269-275.

Ditzel, L., Huber, R., Mann, K., Heinemeyer, W., Wolf D. H., and Groll M. (1998) Conformational constraints for protein selfcleavage in the proteasome. *J. Mol. Biol*. **279,** 1187–1191

Dobson, C. M. (2006) Protein aggregation and its consequences for human disease. *Protein Pept. Lett.* **13**, 219-227

Dodson, G. G. (2000) Catalysis in penicillin G amidase – a member of the Ntn (N-Terminal Nucleophile) hydrolase family. *Croatica Chemica. Acta.* **73**, 901-908.

Done, H., Brannigan, J. A., Moody, J. A., and Hubbard, R. E. (1998) Ligand-induced conformational change in penicillin acylase, *J. Mol. Biol.* **284,** 463–475.

Drasar, B. S., Hill, M. J., and Shiner. M. (1966) The deconjugation of bile salts by human intestinal bacteria. *Lancet.* **4**, 1237–1238.

Drenth, J. and Haas, C. (1998) Nucleation in protein crystallization. *Acta Cryst*. **D54**, 867- 872.

Ducruix, A., and Giege, R. (1992) Crystallization of nuclei acids and proteins: a practical approach. *Ed.*1. IRL Press at Oxford University Press.

du Toit, M., Franz, C. M., Dicks, L. M., Schillinger, U., Haberer, P., Warlies, B., Ahrens, F., Holzapfel, W. H. (1998) Characterisation and selection of probiotic lactobacilli for a preliminary minipig feeding trial and their effect on serum cholesterol levels, faeces pH and faeces moisture content. *Int. J. Food Microbiol.* **40**, 93-104.

Duggleby, H. J., Tolley, S. P., Hill, C. P., Dodson, E. J., Dodson, G. and Moody, P. C. E. (1995) Penicillin acylase has a single aminoacid catalytic center, *Nature.* **373,** 264–268.

Dunne, C. (2001) Adaptation of bacteria to the intestinal niche – probiotics and gut disorder. *Inflamm. Bowel Dis*. **7**, 136–145.

Dussurget, O., Cabanes, D., Dehoux, P., Lecuit, M., Buchrieser, C., Glaser, P., and Cossart, P. (2002) *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol. Microbiol*. **45**, 1095-1106.

Duy, C., and Fitter, J. (2005) Thermostability of irreversible unfolding alpha-amylases analyzed by unfolding kinetics. *J. Biol. Chem.* **280(45),**37360-37365.

Edelsbrunner, H., Facello, M., Fu, R., and Liang, J. (1995) Measuring Proteins and Voids in Proteins. pp. 256-264. *Proceedings of the 28th Hawaii International Conference on Systems Science.*

Economou, A. (2000) Bacterial protein translocase: a unique molecular machine with an army of substrates. *FEBS Lett*. 476, 18-21.

Edger, R. C., and Batzoglou, S. (2006) Multiple Sequence Alignment. *Curr. Opin. Struct. Biol.* 16, 368–373.

Eftink, M. R., and Ghiron, A. (1976) Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. *Biochemistry*. **15**, 672-680.

Elkins, C. A., and Mullis, L. B. (2004) Bile-mediated aminoglycoside sensitivity in *Lactobacillus* species likely results from increased membrane permeability attributable to cholic acid. *Appl. Environ. Microbiol*. **70,** 7200-7209.

Elkins, C. A., and Savage, D. C. (2003) CbsT2 from *Lactobacillus johnsonii* 100-100 is a transport protein of the major facilitator superfamily that facilitates bile acid antiport. *J. Mol. Microbiol. Biotechnol.* **6**, 76-87.

Elkins, C. A., Moser, S. A., and Savage, D. C. (2001) Genes encoding bile salt hydrolases and conjugated bile salt transporters in *Lactobacillus johnsonii* 100-100 and other *Lactobacillus* species. *Microbiology*. **147,** 3404–3412.

Elkins, C., and Savage, D. C. (1998) Identification of genes encoding conjugated bile salt hydrolase and transport in *Lactobacillus johnsonii J. Bacteriol.* **180**, 4344–4349.

Elkins, J. M., Kershaw, N. J., and Schofield, C. J. (2005) X-ray crystal structure of ornithine acetyltransferase from the clavulanic acid biosynthesis gene cluster. *Biochem. J.* **385**, 565-573.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*. **60**, 2126-2132.

Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst*. **A47**, 392-400.

Fanuel, L., and Frere, J. M. (1999) Two new aminopeptidases from *Ochrobactrum anthropi* active on D-alanyl-*p*-nitroanilide. *Cell. Mol. LifeSci.* **55**, 812-818.

Faubion, W. A., Guicciardi, M. E., Miyoshi, H., Bronk, S. F., Roberts, P. J., Svingen, P. A., Kaufmann, S. H and Gores, G. J. (1999) Toxic bile salts induce rodent hepatocyte apoptosis via direct activation of Fas. *J. Clin. Invest.* **103**, 137–145.

Feher, G. and Kam, Z. (1985) Nucleation and growth of protein crystals: general principles and assays. *Methods Enzymol*. **114**, 77-111.

Fernandez-Lafuente, R., Rosell, C. M., and Guisan, J. M. (1991) Enzyme reaction engineering: synthesis of antibiotics catalysed by stabilized penicillin G acylase in the presence of organic cosolvents. *Enzyme Microb. Technol*. **13**, 898-905.

Fernandez-Lafuente, R., Rosell, C. M., and Guisan, J. M. (1996) Dynamic reaction design of enzymic biotransformations in organic media: equilibrium-controlled synthesis of antibiotics by penicillin G acylase. *Biotechnol. Appl. Biochem.* **2**, 139-143.

Ferre-D'Amare, A. R., and Burley, S. K. (1997) Dynamic light scattering in evaluating crystallizability of macromolecules. *Meth. Enzymol.* **276**, 157-166

Fisher, K. J., Klein, M., Park, H., Vettese, M. B., and Aronson, N. N. Jrl (1993) Post-translational processing and Thr206 are required for glycosylasparaginase activity. *FEBS Lett*. **323**, 271-275.

Flores, G., Soberon, X., and Osuna, J. (2004) Production of a fully functional, permuted single-chain penicillin G acylase. *Prot. Sci.* **13**, 1677-1683.

Fomey, L. J., and Wong, D. C. L. (1989) Alteration of the catalytic efficiency of penicillin amidase from *Escherichia coli. Appli. Environ. Microbiol.* **55**, 2556-2560.

Franz, C. M., Specht, I., Haberer, P., and Holzapfel, W. H. (2001) Bile salt hydrolase activity of Enterococci isolated from food: screening and quantitative determination, *J. Food Prot.* **64**, 725-729.

Fritz-Wolf, K., Koller, K. P., Lange, G., Liesum, A., Sauber, K., Schreuder, H., Aretz, W., and Kabsch, W. (2002) Structure-based prediction of modifications in glutarylamidase to allow single-step enzymatic production of 7-aminocephalosporanic acid from cephalosporin C. *Prot. Sci.* **11**, 92-103.

Fuganti, C., and Grasselli, P. (1986) Immobilized penicillin acylase: Application to the synthesis of the dipeptide aspartame. *Tetrahedran lett.* **27**, 3191-3194.

Fuganti, C., Grasselli, P., Seneci, P. F., Servi, S., and Casall, P. (1986) Immobilized benzylpenicillin acylase: Application to the synthesis of optically active forms of carnitin and propranalol**.** *Tetrahedran lett.* **27**, 2061-2062.

Gabor, E. M., and Janssen, D. B. (2004) Increasing the synthetic performance of penicillin acylase PAS2 by structure-inspired semi-random mutagenesis. *Protein Eng. Des. Se.* **17**, 571-579.

Gallo, L. L. (1983) Cholesterol and other sterols: absorption, metabolism, roles in atherosclerosis. pp. 83-109. *Nutrition and heart disease.* Churchill Livingstone, London,.

Gamblin, S. J., and Rodgers, D. W. (1993) Some Practical Details of Data Collection at 100 K. edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 28-32. *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*.

Garman, E. F., and Schneider, T. R. (1997) Macromolecular Cryocrystallography *J. Appl. Cryst.* **30**, 211-237.

Garman, E. F., and Mitchell, E. P. (1996) Glycerol concentration required for cryoprotection of 50 typical protein crystallization solutions. *J. Appl. Cryst.* **29**, 584-587.

Gilliland, S. E., and Speck, M. L. (1977) Deconjugation of bile acids by intestinal *lactobacilli. Appl. Environ. Microbiol.* **33**, 15–18.

Gilliland, S. E., Nelson, C. R., and Maxwell, C. (1985) Assimilation of cholesterol by Lactobacillus acidophilus. *Appl. Environ. Microbiol.* **49**, 377–381.

Gilliland, S. E. (1990) Health and nutritional benefits from lactic acid bacteria. *FEMS Microbiol. Rev.* **87**, 175-188.

Gnnewald, K. K. (1982) Serum cholesterol levels in rats fed skim milk fermented by *Lactobacillus acidophilus. J. Food Sci.* **47**, 2078-2079.

Goodwin, B., Jones, S. A., Price, R. R., Watson, M. A., McKee, D. D., Moore, L. B., Galardi, C., Wilson, J. G, Lewis, M. C., Roth, M.E., Maloney, P. R., Willson, T. M., and Kliewer, S. A., (2000) A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Mol. Cell.* 6, 517-526.

Gopal, A., Shah, N. P., and Roginski, H. (1996) Bile tolerance, taurocholate deconjugation and cholesterol removal by Lacto-bacillus acidophilus and Bifidobacterium spp. *Milchwissenschaft.* 51, 619–623.

Gopal-Srivastava, R., and Hylemon, P. B. (1988) Purification and characterization of bile salt hydrolase from *Clostridium perfringens*. *J. Lipid Res.* **29**, 1079–1085.

Goto, J., Kataoka, R., Hirayama, N. (2004) Ph4Dock: Pharmacophore-Based Protein-Ligand Docking. *J. Med. Chem.* **47**, 6804-6811.

Goto, Y., Takahashi, N., and Fink, A. L. (1990) Mechanism of acid-induced folding of proteins. *Biochemistry.* **29**, 3480-3488.

Grill, J. P., Crociani, J., and Ballongue, J. (1995a) Characterization of fructose 6 phosphate phosphoketolases purified from *Bifidobacterium* species. *Curr. Microbiol.* **31**, 49–54.

Grill, J. P., Schneider, F., Crociani, J., and Ballongue, J. (1995b) Purification and characterization of conjugated bile salt hydrolase from *Bifidobacterium longum* BB536. *Appl. Environ. Microbiol.* **61**, 2577–2582.

Grill, J. P., and Schneider, F. (1997) Characterization of *Bifidobacterium* species based on bile salt hydrolase localization and electrophoretic mobility. *Microbiologie-Aliments-Nutrition.* **15**, 307–313.

Grill, J. P., Cayuela, C., Antoine, J. M., and Schneider, F. (2000a) Isolation and characterization of a *Lactobacillus amylovorus* mutant depleted in conjugated bile salt hydrolase activity: Relation between activity and bile salt resistance. *J. Appl. Microbiol.* **89**, 553–563.

Grill, J. P., Perrin, S., and Schneider, F. (2000b) Bile salt toxicity to some *bifidobacteria* strains: Role of conjugated bile salt hydrolase and pH. *Can. J. Microbiol.* **46,** 878–884.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D., and Huber, R. (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature.* **386**, 463-471.

Grishin, N. V. (2001) Fold changes in evolution of protein structures. *J. Struct. Biol.* **134**, 167–185.

Guan, C., Liu, Y., Shao, Y., Cui, T., Liao, W., Ewel, A., Whitaker, R., and Paulus, H. (1998) Characterization and functional analysis of the cis-autoproteolysis active center of glycosylasparaginase. *J. Biol. Chem.* **273**, 9695–9702.

Gunn, J. S. (2000) Mechanisms of bacterial resistance and response to bile. *Microbes Infect.* **2,** 907–913.

Halgren, T. A. (1999a) MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **20**, 720—729.

Halgren, T. A. (1999b) MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **20**, 730—748.

Harata, K., and Akiba, T. (2006) Structural phase transition of monoclininc crystals of hen-white lysozyme. *Acta Cryst*. **D62**, 375-382.

Hassler, D. F., and Bell, R. M. (1993) Ceramidase: Enzymology and metabolic roles. *Adv. Lipid. Res.* **26**, 49–57.

Hay, J. M., Yu, W. M., and Ashraf, T. (1999) Pharmacoeconomics of lipids lowering agents for primary and secondary prevention of coronary artery disease. *Pharmacoeconomics.* **15**, 47-49.

Hejazi, M., Piotukh, K., Mattow, J., Deutzmann, R., Volkmer-Engert, R., and Lockau, W. (2002) Isoaspartyl dipeptidase activity of plant-type asparaginases. *Biochem. J.* **364**, 129-136.

Heuvel, R. H. H. V., Curti, B., Vanoni M. A., and Mattevi, A. (2004) Glutamate synthase: a fascinating pathway from L-glutamine to L-glutamate. *Cell. Mol. Life Sci*. **61**, 669–681.

Hewitt, L., Kasche, V., Lummer, K., Lewis, R. J., Murshudov, G. N., Verma, C. S., Dodson, G. G. and Wilson, K. S. (2000) Structure of a slow processing precursor penicillin acylase from *Escherichia coli* reveals the linker peptide blocking the active-site cleft. *J. Mol. Biol*. **302**, 887-898.

Hewitt, L., Kasche, V., Lummer, K., Rieks, A. and Wilson, K. S. (1999) Crystallization of a precursor penicillin acylase from *Escherichia coli*. *Acta Cryst*. **D55**, 1052-1054.

Hill, M. J. (1995) Bacteria and fat digestion. Edited by Hill, M.J., London: Taylor & Francis. pp. 131– 142. Role of Gut Bacteria in Human Toxicology and Pharmacology.

Hirano, S., Masuda, N., Oda, H., and Mukai, H. (1981) Transformation of bile acids by *Clostridium perfringens*. *Appl. Environ. Microbiol.* **42**, 394–399.

Hirayama, K., and Rafter, J. (2000) The role of probiotic bacteria in cancer prevention. *Microbes Infect*. **2**, 681–686.

Hofmann, A. F. (1989a) Enterohepatic circulation of bile acids. In Handbook of Physiology. The Gastrointestinal System. Vol. 4. American Physiological Society. Bethesda.

Hofmann, A. F. (1989b) The enterohepatic circulation of bile acids in health and disease. edited by Sleisinger, M. H., and Fordtran, J. S., pp. 144–161 Gastrointestinal diseases. pathophysiology, diagnosis, management, New York: W.B. Saunders Company.

Hofmann, A. F., and Mysels, K. J. (1992) Bile acid solubility and precipitation in vitro and in vivo: the role of conjugation, pH and Ca^{2+} ions. *J. Lipid Res*. **33**, 617–626.

Hofmann, A. F. (1999) Bile acids: the good, the bad, and the ugly. *News Physiol Sci*. **14**, 24–29.

Holm, L., and Sander, C. (1996) Mapping the protein universe. *Science*. **273**, 595-603.

Huber, R., and Schneider, M. (1985) A group refinement procedure in protein crystallography using Fourier transforms. *J. Appl. Cryst.* **18**, 165-169.

Huijghebaert, S. M., Mertens, J. A., and Eyssen, H. J. (1982) Isolation of a bile salt sulfatase-producing Clostridium strain from rat intestinal microflora. *Appl. Environ. Microbiol*. 43, 185–192.

Huijghebaert, S. M., and Hofmann, A. F. (1986a) Influence of the amino acid moiety on deconjugation of bile acid amidates by cholylglycine hydrolase or human fecal cultures. *J. Lipid Res.* **27,** 742-752.

Huijghebaert, S. M., and Hofmann, A. F. (1986b) Pancreatic carboxypeptidase hydrolysis of bile acid-amino conjugates: selective resistance of glycine and taurine amidates. *Gastroenterology.* **90**, 306-315.

Hunt, P. D., Tolley, S. P., Ward, R. J., Hill, C. P., and Dodson, G. G. (1990) Expression, purification and crystallization of penicillin G acylase from *Escherichia coli* ATCC 11105. *Protein Eng*. **7**, 635-639.

Ignatova, Z., Enfors, S., Hobbie, M., Taruttis, S., Vogt, C., and Kasche, V. (2000) The relative importance of intracellular proteolysis and transport on the yield of the periplasmic enzyme penicillin amidase in *Escherichia coli*. *Enz. Microb. Technol*. **26**, 165-170.

Ignatova, Z., Hornle, C., Nurk, A., and Kasche, V. (2002) Unusual signal peptide directs penicillin amidase from *Escherichia coli* to the Tat translocation machinery. *Biochem. Biophys. Res. Commun.* **291**, 146-149.

Ignatova, Z., Mahsunah, A. A., Georgieva, M., and Kaslike, V. (2003) Improvement of post translational bottlenecks is the production of penicillin amidase in recombinant *Escherichia coli* strain, *App. Environ. Micro.* **69,** 1237-1245.

Isupov, M. N., Obmolova, G., Butterworth, S., Badet-Denisot, M. A., Badet, B., Polikarpov, I., Littlechild, J. A., and Teplyakov, A. (1996) Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 Å crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase. *Structure*. **4**, 801–810.

Jackson, S. G., Zhang, Y., Bao, X., Zhang, K., Summerfield, R., Haslam R. J., and Junop, M. S. (2006) Structure of the carboxy-terminal PH domain of pleckstrin at 2.1 Å. *Acta. Cryst.* **D62**, 324-330.

Jiang, T., Mustapha, A., and Savaiano, D. A. (1996) Improvement of lactose digestion in humans by ingestion of unfermented milk containing Bifidobacterium longum. *J. Dairy Sci*. **79**, 750-757.

Johanson, L. R. (1998) Bile secretion and gall bladder function edited by Johnson, L.R., pp. 465–471, Essential Medical Physiology, Lippincott-Raven, Philadelphia.

Johnson, M. R., Barnes, S., Sweeny, D. J., and Diasio, B. (1990) 2-Fluoro-beta-alanine, a previously unrecognized substrate for bile acid coenzyme A:amino acid:N-acyltransferase from human liver. *Biochem. Pharmacol.* **40,** 1241-1246.

Jones, M. L., Chen, H., Ouyang, W., Metz, T., and Prakash, S. (2004) Micro encapsulated genetically engineered Lactobacillus plantarum 80 (pCBH1) for bile acid deconjugation and its implication in lowereing cholesterol. *J. Biomed. Biotechnol.* 1, 61-69.

Kanasaki, M., Breheny, S., Hillier, A. J., and Jago, G. R. (1975) Effect of temperature on the growth and acid production of lactic acid bacteria in milk. *Aust. J. Dairy Technol.* **30**, 142-144.

Kaufmann, P., Pfefferkorn, A., Teuber, M., and Meile, L. (1997) Identification and quantification of *Bifidobacterium* species isolated from food with genusspecific 16S rRNA-targeted probes by colony hybridization and PCR. *Appl. Environ. Microbiol.* **63**, 1268–1273.

Kawamoto, K., Horibe, I., and Uchida. U. (1989) Purification and characterization of a new hydrolase for conjugated bile acids, henodeoxycholyltaurinehydrolase from *Bacteroides vulgatus*. *J. Biochem.* ***106***, 1049–1053.

Kazan, D., and Erarslan, A. (2001) Identification of catalytically essential amino acid residues of penicillin G acylase obtained from a mutant of *Escherichia coli* ATCC 11105, *Proc. Biochem.* **36**, 861-867.

Kelly, J. W., and Balch, W. E. (2006) The integration of cell and chemical biology in protein folding. *Nat. Chem. Biol.* **2**, 224-227.

Kenneth, C. K., and Sauer, R. T. (1995) Identification of active site residues of the Tsp Protease. *J. Biol. Chem.* **270**, 28864-28868.

Khatun Haq, S., Ahmad, F., and Khan, R. H. (2003) The acid-induced state of glucose oxidase exists as a compact folded intermediate. *Biochem. Biophys. Res. Commun.* **303**, 685-692.

Khatun Haq. S., Rasheedi, S., and Khan, R. H. (2002) Characterization of a partially folded intermediate of stem bromelain at low pH. *Eur. J. Biochem.* **269**, 47-52.

Khurana, R., and Udgaonkar, J. B. (1994) Equilibrium unfolding studies of barstar: evidence for an alternative conformation which resembles a molten globule. *Biochemistry.* **11**, 106-115.

Kim, J. K., Yang, I. S., Rhee, S., Dauter, Z., Lee, Y. S., Park, S. S., and Kim, K. H. (2003) Crystal structures of glutaryl 7-aminocephalosporanic acid acylase: insight into autoproteolytic activation. *Biochemisty.* **42,** 4084-4093.

Kim, G. B., Yi, S. H., and Lee, B. (2004a) Purification and characterization of three different types of bile salt hydrolase from *Bifidobacterium* strains. *J. Dairy Sci.* **87**, 258–266.

Kim, G. B., Miyamoto, C. M., Meighen, E. A., and Lee, B. H. (2004b) Cloning and characterization of the bile salt hydrolase genes (bsh) from *Bifidobacterium bifidum* strains. *Appl. Environ. Microbiol.* **70**, 5603-5612.

Kim, G. B., Brochet, M., and Lee, B. H. (2005) Cloning and characterization of a bile salt hydrolase (bsh) from *Bifidobacterium adolescentis*. *Biotech. Lett.* **27**, 817-822.

Kim, J. K., Yang, I. S., Shin, H. J., Cho, K. J., Ryu, E. K., Kim, S. H., Park, S. S., and Kim, K. H. (2006) Insight into autoproteolytic activation from the structure of cephalosporin acylase: a protein with two proteolytic chemistries. *Proc. Natl. Acad. Sci. U S A*. **103**, 1732-1737.

Kim, S. W., Rogers, Q. R., and Morris, J. G. (1996a) Dietary antibiotics decrease taurine loss in cats fed a canned heat-processed diet. *J. Nutr.* **126**, 509-515.

Kim, S. W., Rogers, Q. R., and Morris, J. G. (1996b) Maillard reaction products in purified diets induce taurine depletion in cats which is reversed by antibiotics. *J. Nutr.* **126**, 195-201.

Kim, Y., Yoon, K., Khang, Y., Turley, S., and Hol, W. G. (2000) The 2.0 Å crystal structure of cephalosporin acylase. *Structure Fold. Des*. **8**, 1059-1068.

Kim, Y., and Hol, G. J. W. (2001) Structure of cephalosporin acylase in complex with glutaryl-7-aminocephalosporanic acid and glutarate: insight into the basis of its substrate specificity. *Chem. Biol.* **8,** 1253–1264.

Kirby, L. C., Klein, R. A., and Coleman, J. P. (1995) Continuous spectrophotometric assay of conjugated bile acid hydrolase. *Lipids.* **30**, 863-867.

Kishinaka, M., Umeda, A., and Kuroki, S. (1994) High concentrations of conjugated bile acids inhibit bacterial growth of *Clostridium perfringens* and induce its extracellular cholylglycine hydrolase. *Steroids.* **59**, 485-489.

Klaver, F. A. M., and Van der Meer, R. (1993) The assumed estimation of cholesterol removal by Lactobacilli and *Bifidobacterium bifidum* is due to their bile salt deconjugation activity. *Appl. Enviro. Microbio.* **59**, 1120–1124.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sand- brink, H. M., Fiers, M. W., Stiekema, W., Lankhorst, R. M., Bron, P. A., Hoffer, S. M., Groot, M. N., Kerkhoven, R., de Vries, M., Ursing, B., de Vos, W. M., and Siezen, R. J. (2003) Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl. Acad. Sci.* **100**, 1990–1995.

Klei, H. E., Daumy, G. O., and Kelly, J. A. (1995) Purification and Preliminary crystallographic studies of penicillin acylase from *Providencia rettgeri*. *Prot. Sci.* **4**, 433-441.

Klyosov, A. A., Svedas, V. K. and Galaev, F. Y. (1977) Enzymatic synthesis of antibiotics. the kinetics of acyl group transfer onto 6-aminopenicillanic acid catalyzed by penicillin amidase *from E. coli. Bioorg. Khim.* **3**, 800-805.

Knarreborg, A., Engberg, R. M., Jensen, S. K. and Jensen, B. B. (2002) Quantitative determination of bile salt hydrolase activity in bacteria isolated from the small intestine of chickens. *Appl. Environ. Microbiol.* **68**, 6425–6428.

Kobayashi, H. (1985) A proton-translocating ATPase regulates pH of the bacterial Cytoplasm. *J. Biol. Chem.* **260,** 72–76.

Kolata, G., and Andrews, E. L. (2001) Anticholesterol drug pulled after link to 31 deaths. *New York Times Online*.

Konstantinovic, M., Marjanovic, N., Ljubijankic, G. and Glisin, V. (1994) The penicillin amidase of *Arthrobacter viscosus* (ATCC 15294). *Gene*. **143**, 79-83.

Kumar, R. S., Brannigan, J. A., Pundle, A., Prabhune, A., Dodson, G. G., and Suresh, C. G. (2004a) Expression, purification, crystallization and preliminary X-ray diffraction analysis of conjugated bile salt hydrolase from Bifidobacterium longum. *Acta Cryst.* **D60,** 1665-1667.

Kumar, R. S., Suresh, C. G., Pundle, A., and Prabhune, A. (2004b) Evidence for the involvement of arginyl residue at the active site of penicillin G acylase from *Kluyvera citrophila*. *Biotechnol. Lett*. **20**, 1601-1606.

Kumar, R. S., Brannigan, J. A., Pundle, A., Prabhune, A., Dodson, G. G., Dodson, E. J., and Suresh, C. G. (2006a) Structural and functional analysis of a conjugated bile salt hydrolase from *Bifidobacterium longum* reveals evolutionary relationship with penicillin V acylase. *J. Biol. Chem.* 281, 32516-32525.

Kumar, R. S., Pundle, A., Prabhune, A. A., Karthikeyan, M., and Suresh, C. G. (2006b). Tryptophan residue in substrate binding of penicillin G acylase from *Kluyvera citrophila:* evidence from biochemical and modeling studies. *Enz. Microb. Tech*. (In Press) doi: 10.1016/j.enzymetec.2006.10.022.

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. (1982) A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **161**, 269-288.

Kurdi, P., van Veen, H. W., Tanaka, H., Mierau, I., Konings, W. N., Tannock, G. W., Tomita, F., and Yokota, A. (2000) Cholic acid is accumulated spontaneously, driven by membrane {Delta}pH, in many lactobacilli. *J. Bacteriol.* **182**, 6525–6528.

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

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

Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 282-291.

Leahy, S. C., Higgins, D. G., Fitzgerald, G. F., and van Sinderen, D. (2005) Getting better with *bifidibacterium*. *Jour. Appl. Microbiol.* **98**, 1303-1315.

Lee. B., and Richards. F. M. (1971) The interpretation of protein structures: estimation of static accessiblity. *J. Mol. Biol*. **55**, 379-400.

Lee, Y. P., and Takahashi, T. (1966) An improved colorimetric determination of amino acids with the use of ninhydrin. *Anal. Biochem*. **14**, 71-77.

Lee, Y. S., Kim, H. W., and Park, S. S. (2000) The role of alpha-amino group of the N-terminal serine of beta subunit for enzyme catalysis and autoproteolytic activation of glutaryl 7-aminocephalosporanic acid acylase. *J. Biol. Chem.* **275,** 39200–39206

Leer, R. J., Christiaens, H., Verstraete, W., Peters, L., Posno, M., and Pouwels, P. H. (1993) Gene disruption in *Lactobacillus plantarum* strain 80 by site-specific recombination: isolation of a mutant strain deficient in conjugated bile salt hydrolase activity. *Mol. Gen. Genet.* **239**, 269-272.

Lehrer, S. S. (1971) Solute perturbation of protein fluorescence; The quenching of the tryptophanyl fluorescence of model compounds and of lysozyme by iodide ion, *Biochemistry.* **10**, 3254-3263.

Lepercq, P., Relano, P., Cayuela, C., and Juste, C. (2004) *Bifidobacterium animalis* strain DN-173 010 hydrolyses bile salts in the gastrointestinal tract of pigs. *Scand. J. Gastroenterol.* **39**, 1266-1271.

Levine, G. N., Keaney, J. F., and Jr, Vita, J. A. (1995) Cholesterol reduction in cardiovascular disease. Clinical benefits and possible mechanisms. *N. Engl. J. Med.* **332**, 512–521.

Leyte, A., van Schijndel, H. B., Niehrs ,C., Huttner, W. B., Verbeet, M. P, and Mertens, K. (1991) Sulfation of Tyr1680 of human blood coagulation factor VIII is essential for the interaction of factor VIII with von Willebrand factor. *J. Biol. Chem.* **266**, 740–746.

Li, Y., Chen, J., Jiang, W., Mao, X., Zhao, G., and Wang, E. (1999) *In vivo* post-translational processing and subunit reconstitution of cephalosporin acylase from *Pseudomonas* sp. 130. *Eur. J. Biochem.* **262**, 713-719.

Liang, Y., Du, F., Sanglier, S., Zhou, B., Xia, Y., Dorsselaer, A. V., Maechling, C., Kilhoffer, M. C., and Haiech, J. (2003) Unfolding of Rabbit Muscle Creatine Kinase Induced by Acid: A Study Using Electrospray Ionization Mass Spectrometry, Isothermal Titration Calorimetry, And Fluorescence Spectroscopy. *J. Biol. Chem.* **278**, 30098 – 30105.

Lim, H. J., Kim, S. Y., and Lee, W. K. (2004) Isolation of cholesterol-lowering lactic acid bacteria from human intestine for probiotic use. *J. Vet. Sci.* **5**, 391-395.

Lin, W., Huang, S., and Chou, C. P. (2001) High-level extracellular production of penicillin acylase by genetic engineering of *Escherichia coli*. *J. Chem. Technol. Biotechnol.* **76**, 1030-1037.

Lindsay, C. D., and Pain, R. H. (1991) Refolding and assembly of penicillin acylase, an enzyme composed of two polypeptide chains that result from proteolytic activation. *Biochemistry.* **30**, 9034-9040.

Ling, W. H., Hanninen, O., Mykkanen, H., Heikura, M., Salminen, S., and Von Wright, A. (1992) Colonization and fecal enzyme activities after oral *Lactobacillus* GG administration in elderly nursing home residents. *Ann. Nutr. Metab*. **36,** 162-166.

Liong, M. T., and Shah, N. P. (2005) Bile salt deconjugation ability, bile salt hydrolase activity and cholesterol co-precipitation ability of lactobacilli strains. *Int. Dairy J*. **15,** 391-398.

Littlechild, J. A. (1991) Protein crystallization: magical or logical: can we establish some general rules? *Journal of Physics D: Applied Physics*. **24**, 111-118.

Liu, S. L., Wei, D. Z., Song, Q. X., Zhang, Y. W., Wang, X. D. (2006) Effect of organic cosolvent on kinetic resolution of tert-leucine by penicillin G acylase from *Kluyvera citrophila*. *Bioprocess Biosyst. Eng*. **28,** 285-289.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science*. **268,** 533–539.

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

Little, J. W. (1993) LexA cleavage and other self-processing reactions. *J. Bacteriol.* 1**75**, 4943-4950.

Luft, J. R., and Detitta, G. T. (1999) A method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crysta.* **D55**, 988-993.

Lundeen, S. G., and Savage, D. C. (1990) Characterization and purification of bile salt hydrolase from *Lactobacillus* sp. strain 100-100. *J. Bacteriol.* **172**, 4171–4177.

Lundeen, S. G., and Savage, D. C. (1992a) Characterization of an extracellular factor that stimulates bile salt hydrolase activity in *Lactobacillus* sp. strain 100-100. *FEMS Microbiol. Lett.* **73**, 121-126.

Lundeen, S. G., and Savage, D. C. (1992b) Multiple forms of bile salt hydrolase from *Lactobacillus* sp. strain 100-100. *J. Bacteriol*. **174,** 7217-7220.

Luo, J. K., Hornby, J. A., Wallace, L.A., Chen, J., Armstrong, R. N., and Dirr, H.,W. (2002) Impact of domain interchange on conformational stability and equilibrium folding of chimeric class micro glutathione transferases. *Prot. Sci.* **11**, 2208-2217.

MacDonald, I. A., Williams, C. N., Sutherland, J. D., and MacDonald, A. C. (1983) Estimation of ursodeoxycholic acid in human and bear biles using Clostridium absonum 7 beta-hydroxysteroid dehydrogenase. *Anal. Biochem.* **135**, 349-354.

Maeda, Y., and Takahashi, M. (1989) Hydrolysis and absorption of a conjugate of ursodeoxycholic acid with para-aminobenzoic acid. *J. Pharmacobiodyn*. **12**, 744-753.

Maeda, Y., Takahashi, M., Tashiro, H., and Akazawa, F. (1989) The rapid evaluation of intestinal bacterial growth using a conjugate of ursodeoxycholic acid with para-aminobenzoic acid. *J. Pharmacobiodyn*. **12**, 272-280.

Mahajan, P. B., and Borkar, P. S. (1983) Chemical modification of *Escherichia coli* penicillin acylase I tryptophan involvement at the catalytic site. *Hindustan Antibiotic. bulletin.* 25, 6-10.

Makishima, M., Okamoto, A.Y., Repa, J. J., Tu, H., Learned, R., M., Luk, A., Hull, M. V., Lustig, K. D., Mangelsdorf, D. J., and Shan, B. (1999) Identification of a nuclear receptor for bile acids. *Science.* **284**, 1362–1365.

Mann, G. V. (1974) Studies of a surfactant and cholestremia in the Maasai. *Am. J. clin. Nutr.* 27, 464-469.

Manson, J. E., Tosteson, H., Ridker, P. M., Satterfield, S., Herbert, P., and O_Conner, G. T. (1992) The primary prevention of myocardial infarction. *New England J. Med.* **326**, 1406–1416.

Martin, J., Mancheno, J. M., and Arche, R. (1993) Inactivation of penicillin acylase from *Kluyvera citrophila* by N-ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline: a case of time-dependent non-covalent enzyme inhibition. *Biochem. J.* **291,** 907-914.

Marteau, P., and Rambaud, J. C. (1993) Potential of using lactic acid bacteria for therapy and immunomodulation in man. *FEMS Microbiol. Rev*. 12, 207–220.

Marteau, P., Pochart, P, Dore. J., Bern-Mallet, C., Bernaller, A., and Cortheler, G. (2001) Comparative study of bacterial groups within the human cecal and fecal microbiota. *Appl. Environ. Microbiol*. 67, 4939-4942.

Masclee, A., Tangerman, A., Van Schaik, A., Van der Hoek, E. W., and Van Tongeren, J. H. M. (1989) Unconjugated serum bile acid as a marker of small intestinal bacterial overgrowth. *Eur. J. Clin. Invest*. **19,** 384–389.

Masuda, N. (1981) Deconjugation of bile salts by Bacteroides and Clostridium. *Microbiol. Immunol*. **25**, 1–11.

Matsuki, T., Watanabe, K., and Tanaka, R. (2003) Genus- and species-specific PCR primers for the detection and identification of *Bifidobacteria. Curr. Issues Intest. Microbiol.* **4**, 61–69.

Matthews, B. W. (1968) The solvent content of protein crystals. *J. Mol. Biol*. **33**, 491-497.

Matthews, B. W. (1985) Determination of protein molecular weight, hydration, and packing from crystal density. *Meth. Enzymol*. **114**, 176-187.

McAuliffe, O., Cano, R. J., and Klaenhammer, T. R. (2005) Genetic analysis of two bile salt hydrolase activities in Lactobacillus acidophilus NCFM. *Appl. Environ. Microbiol*. **71**, 4925-4929.

MacDonald, I. A., Williams, C. N., Sutherland, J. D., and MacDonald, A. C. (1983) Estimation of ursodeoxycholic acid in human and bear biles using Clostridium absonum 7 beta-hydroxysteroid dehydrogenase. *Anal. Biochem.* **135**, 349-54.

McDonough, M. A., Klei, H. E., and Kelly, H. E. (1999) Crystal structure of Penicillin G acylase from the Bro1 mutant strain of *Providencia rettgeri*. *Prot. Sci.* **8**, 1971-1981.

McPherson, A. (1982) Preparation and analysis of protein crystals. John Wiley & Sons.

McPherson, A., Koszelak, S., Axelrod, H., Day, J., Robinson, L., McGrath, M., Williams, R., and Cascio, D. (1986a) The Effects of Neutral Detergents on the Crystallization of Soluble Proteins. *Journal of Crystal Growth*. **76**, 547-553.

McPherson, A., Koszelak, S., Axelrod, H., Day, J., Williams, R., Robinson, L., McGrath, M., and Cascio, D. (1986b) An Experiment Regarding Crystallization of Soluble Proteins in the Presence of -Octyl Glucoside. *Journal of Biological Chemistry*. **261**, 1969-1975.

McPherson, A. (1991) A brief history of protein crystal growth. *J. Cryst. Growth*. **110**, 1- 10.

McPherson, A. (1998) Crystallization of biological macromolecules. Cold Spring Harbor Laboratory Press.

McVey, C. E., Walsh, M. A., Dodson, G. G., Wilson, K. S., and Brannigan, J. A. (2001)

Crystal structures of penicillin acylase enzyme-substrate complexes: structural insights into the catalytic mechanism. *J. Mol. Biol*. **313,**139-150.

Medici, M., Vinderola, C. G., and Perdigon, G. (2004) Gut mucosal immunodulation by probiotic fresh cheese. *Int. Dairy J*. **14**, 611-618.

Michalska, K., Bujacz G., and Jaskolski, M. (2006). Crystal Structure of Plant Asparaginase. *J. Mol. Biol*. **360**, 105–116.

Misra, H. P. (1974) Generation of superoxide free radical during the autoxidation of thiols *J. Biol. Chem.* **249**, 2151–2155.

Mitvedt, T., and Norman, A. (1967) Bile acid transformations by microbial strains belonging to genera found in intestinal contents. *Acta Pathol. Microbiol. Scand.* **71**, 629–638.

Mitvedt, T., and Norman, A. (1968) Anaerobic, bile acid transforming microorganisms in rat intestinal content. *Acta Pathol. Microbiol. Scand.* **72**, 337–344.

Morillas, M., Goble, M. L., and Virden, R. (1999) The kinetics of acylation and deacylation of penicillin acylase from *Escherichia coli* ATCC 11105: evidence for lowered pK$_a$ values of groups near the catalytic center. *Biochem. J.* **338**, 235–239

Morris. G. M., Goodsell. D. S., Halliday. R. S., Huey,. R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comput. Chem*. **19**, 1639-1662.

Moser, S. A., and Savage, D. C. (2001) Bile salt hydrolase activity and resistance to toxicity of conjugated bile salts are unrelated properties in lactobacilli. *Appl. Environ. Microbiol*. **67**, 3476-3480.

Murshudov, G. N. (1997) Refinement of macromolecular structures by the maximum likelihood method. *Acta Cryst.* **D53**, 240-255.

Murshudov, G. N., Dodson, E. J., and Vagin, A. A. (1996) Application of maximum likelihood methods for macromolecular refinement. Pp 93-104. Proceedings of the CCP4 Study Weekend (Macromolecular Refinement)

Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y., Takahashi, S., Sato, M. T., and Sukihara, T. (2000) Crystal structure

of N-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. *Structure*. **8**, 729–737.

Naseem, F., Ahmad, B., Ashraf, M. T., and Khan, R. H. (2004) Molten globule-like folding intermediate of asialofetuin at acidic pH. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomic* **1699**, 191-199.

Naseem, F., Khan, R. H., Khatun Haq, S., and Naeem, A. (2003) Characterization of molten globule state of fetuin at low pH. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics,* **1649**, 164-170.

Nath, D., and Rao, M. (2001) Acid-induced partly folded conformation resembling a molten globule state of xylanase from an alkalothermophilic *Bacillus* sp. *Biochem. Biophy. Res. Comm.* **288**, 1218-1222.

Navaza, J., and Saludjian, P. (1997) AMoRe: an automated molecular replacement program package. *Methods Enzymol*. **276**, 581–594.

Navaza, J. (1994) AMoRe - An automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

Oh, B., Kim, M., Yoon, J., Chung, K., Shin, Y., Lee, D., and Kim Y. (2003) Deacylation activity of cephalosporin acylase to cephalosporin C is improved by changing the side chain conformations of active-site residues. *Biochem. Biophys. Res. Commun*. **310**, 19-27.

Oh, S. J., Kim, Y. C., Park, Y. W., Min, S. Y., Kim, I. S., and Kang, H. S. (1987) Complete nucleotide sequence of the penicillin G acylase gene and flanking regions, and its expression in *Escherichia coli*. *Gene*. **56**, 87-97.

Ohkohchi, N., Andoh, T., Izumi, U., Igarashi, Y., and Ohi R. (1997) Disorder of bile acid metabolism in children with short bowel syndrome. *J. Gastroenterol*. **32**, 472–429.

Oinonen, C., and Rouvinen, J. (2000) Structural comparison of Ntn-hydrolases. *Protein Sci*. **9**, 2329–2337.

Oinonen, C., Tikkanen, R,, Rouvinen, J. and Peltonen, L. (1995) Three Dimensional Structure of Human Lysosomal Aspartyiglucosaminidase. *Nat. Struct. Biol.* **2**, 1102-1108.

Okada, T., Suzuki, H., Wada, K., Kumagai, H., and Fukuyama, K. (2006) Crystal structures of {gamma}-glutamyltranspeptidase from Escherichia coli, a key enzyme in glutathione metabolism, and its reaction intermediate. *Proc. Natl. Acad. Sci. U S A.* **103**, 6471-6476.

Orozco R, and Luque, F. J. (2000) Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. *Chem. Rev.* **100**, 4187-4225.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Owen, R.W. (1985) Biotransformation of bile acids by Clostridia. *J. Med. Microbiol.* **20**, 233–238.

Owen, R. W., Dodo, M., Thompson, M. H., and Hill, M. J. (1987) Fecal steroids and colorectal cancer. *Nutr. Cancer.* **9**, 73-80.

Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins, *FASEB J.* **10**, 75-83.

Pai, R., Tarnawski, A. S., and Tran, T. (2004) Deoxycholic acid activates {beta}-catenin signaling pathway and increases colon cell cancer growth and invasiveness. *Mol. Biol. Cell.* **15**, 2156 - 2163.

Pan, K. L., Hsiao, H. C., Weng, C. L., Wu, M. S., and Chou, C. P. (2003) Roles of DegP in prevention of protein misfolding in the periplasm upon overexpression of penicillin acylase in *Escherichia coli*. *J. Bacteriol*. **185,** 3020-3030.

Parks, D. J., Blanchard, S. G., Bledsoe, R. K., Chandra, G., Consler, T. G., Kliewer, S. A., Stimmel, J. B. , Willson, T. M., Zavacki, A. M., Moore, D. D., and Lehmann, J. M. (1999) Bile acids: natural ligands for an orphan nuclear receptor. *Science.* **284**, 1365–1368.

Parkin, S., and Hope, H. (1998) Macromolecular cryocrystallography: cooling, mounting, storage and transportation of crystals. *J. Appl. Crystall*. **31**, 945-953.

Pawar, S. A., and Deshpande, V. V. (2000) Characterization of acid-induced unfolding intermediates of glucose/xylose isomerase, *Eur. J. Biochem*. **267**, 6331-6338.

Pei, J., and Grishin, N. V. (2003) Peptidase family U34 belongs to the superfamily of N-terminal nucleophile hydrolases. *Protein Sci.* **12**, 1131-1135.

Pekkanen, J., Linn, S., Heiss, G., Suchindran, C. M., Leon, A., Rifldnd, B. M., and Tyroler, H. A. (1990) Ten-year mortality from cardiovascular disease in relation to cholesterol level among men with and without pre-existing cardiovascular disease. *N. Engl. J. Med.* **322**, 1700-1708.

Perakyla, M., and Kollman, P. A. (1997) A simulation of the catalytic mechanism of aspartylglucosaminidase using ab initio quantum mechanics and molecular dynamics. *J.Am. Chem. Soc.* **119**, 1189–1196.

Peräkylä, M., Rouvinen, J. (1996) Ab initio quantum mechanical model calculations on the catalytic mechanism of aspartylglucosaminidase (AGA): A serine protease-like mechanism with an N-terminal threonine and substrateassisted catalysis. *Chem. Eur. J.* **2**, 1548–1551.

Pereira, D. I. A., McCartney, A. L., and Gibson, G. R. (2003) An in vitro study of the probiotic potential of a bile-salt-hydrolyzing *Lactobacillus fermentum* strain, and determination of its cholesterol-lowering properties. *Appl. Environ.Microbiol.* **69**, 4743–4752.

Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) *ARP/wARP* and molecular replacement *Acta Crystallogr. Sect. D.* **57**, 1445-1450.

Picard, C., Fioramonti, J., francois, A., Robinson, T., Neant, F., and Matuchansky, C. (2005) *Bifidobacterium* as probiotic agents-physiological effects and clinical benefits. *Aliment Pharmacol. Ther.* **22**, 495-512.

Pinotti, L. M., Silva, R. G., Giordano, R. C. and Giordano, R .L. (2002) Inoculum studies in production of penicillin G acylase by *Bacillus megaterium* ATCC 14945. *Appl. Biochem. Biotechnol.* **98-100**, 679-686.

Plhackova, K., Becka, S., Skrob, F. and Kyslik, P. (2003) Isolation and characterization of a new strain of *Achromobacter* sp. with beta-lactam antibiotic acylase activity. *Appl. Microbiol. Biotechnol.* **62**, 507-516.

Polderman-Tijmes, J. J., Jekel, P. A., de Vries, E. J., van Merode, A. E., Floris, R., van der Laan, J. M., Sonke, T., and Janssen, D. B. (2002a) Cloning, sequence analysis, and expression in *Escherichia coli* of the gene encoding an alpha-amino acid ester hydrolase from *Acetobacter turbidans*. *Appl. Environ. Microbiol.* **68**, 211-218.

Polderman-Tijmes, J. J., Jekel, P. A., Jeronimus-Stratingh, C. M., Bruins, A. P., Van Der Laan, J. M., Sonke, T., and Janssen, D. B. (2002b) Identification of the catalytic residues of alpha-amino acid ester hydrolase from *Acetobacter turbidans* by labeling and site-directed mutagenesis.*J. Biol. Chem.* **277**, 28474-28482.

Ponting, C. P., and Bork, P. (1996) Pleckstrin's repeat performance: a novel domain in G-protein signaling? *Trends Biochem. Sci.* **21**, 245-6.

Pratt, K. P., Shen, B.W., Takeshima, K., Davie, E. W., Fujikawa, K., Stoddard, B. L. (1999) Structure of the C2 domain of human factor VIII at 1.5 A resolution. Nature.**402**, 439–442.

Prabhune, A. A., and Sivaram, H. (1990) Evidence for involvement of arginyl residue at the catalytic site of penicillin acylase from *Escherichia coli. Biochem. Biophy. Res. Commun.* **173**, 317-322.

Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A.-C., Zwahlen, M.-C., Rouvet, M., Altermann, E., Barrangou, R., Mollet, B., Mercenier, A., Klaenhammer, T., Arigoni, F. and Schell, M.A. (2004) The genome sequence of the probiotic intestinal bacterium Lactobacillus johnsonni NCC533. *Proc. Natl. Acad. Sci.* **101**, 2512–2517.

Qiao, D., Chen, W., Stratagoules, E. D., and Martinez, J. D. (2000) Bile Acid-induced Activation of Activator Protein-1 Requires Both Extracellular Signal-regulated Kinase and Protein Kinase C Signaling. *J. Biol. Chem.* **275**, 15090 – 15098.

Rajaraman, K., Raman, B., and Rao, M. (1996) Molten-globule state of carbonic anhydrase binds to the chaperone-like -crystallin. *J. Biol. Chem.* **271**, 27595 – 27600.

Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283-438.

Ravelli, R. B. G., Schroder Leiros, H. K., Pan, B., Caffrey, M.. and McSweeney, S. (2003)
Specific radiation damage can be used to solve macromolecular crystal structures. *Structure(Camb)*, **11**, 217-224.

Ravelli, R. B. G., and Garman, E. F. (2006) Radiation damage in macromolecular cryocrystallography. *Curr. Opin. Struct. Biol.* **16**, 1-6.

Reddy, B. S., and Rivenson, A. (1993) Inhibitory effect of Bifidobacterium longum on colon , mammary, and liver carcinogenic induced by 2-amino-3-methylimidozo[4,5-f] quinoline, a food mutagen. *Cancer Res*, 1, 3914-3918.

Rhode, G. (2000) Crystallography Made Crystal Clear. Academic Press.

Roa, A., Castillon, M. P., Goble, M. L., Virden, R., and Garcia, J. L. (1995) New insights on the specificity of Penicillin acylase, *Biochem. Biophysic. Res. Comm.* **206,** 629 - 636.

Roa A., Garcia-Lopez, J. L., Salto, F., and Cortes, E. (1994) Changing the substrate specificity of penicillin G acylase from *Kluyvera citrophila* through selective pressure. *Biochem. J* . **303,** 869-75.

Robas, N., Zouheiry, H., Branlant, G., and Branlant, C. (1993) Improved penicillin amidase production using a genetically engineered mutant of *Escherichia coli* ATCC 11105. *Biotechnol. Bioeng.* **41**, 14-24.

Rodgers, D. W. (1997) Practical cryocrystallography. *Meth. Enzymol.* **276**, 183-203.

Roig, M.G., and Kennedy, F. J. (1992) Perspectives for chemical modification of enzymes. *Crit. Rev. Biotechnol.* **12**, 391–412.

Rossmann, M. G., and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **A15**, 24-31.

Rossocha, M., Schultz-Heienbrok, R., von Moeller, H., Coleman, J. P. and Saenger, W. (2005) Conjugated bile acid hydrolase is a tetrameric n-terminal thiol hydrolase with specific recognition of its cholyl but not of its tauryl product. *Biochemistry.* 44, 5739-5748

Saarela, J., Oinonen, C., Jalanko, A., Rouvinen, J., and Peltonen, L. (2004) Autoproteolytic activation of human aspartylglucosaminidase. *Biochem. J.* **378**, 363-371.

Sacchetta, P., Pennelli, A., Bucciarelli, T., Cornelio, L., Amicarelli, F., Miranda, M. and Ilio, C. D. (1999) Multiple unfolded states of glutathione transferase bbGSTP1-1 by guanidinium chloride. *Archiv. Biochem. Biophy.* **369**, 100-106.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanders, M. E. (2000) Considerations for use of probiotic bacteria to modulate human health. *J. Nutr*. **130**, 384-390.

Saridakis, V., Christendat, D., Thygesen, A., Arrowsmith, C. H., Edwards, A. M., Pai, E. F. (2002) Crystal structure of Methanobacterium thermoautotrophicum conserved protein

MTH1020 reveals an NTN-hydrolase fold. *Proteins*. **48**, 141-143.

Sartor, R. B. (1997) Enteric microflra in IBD: pathogen or commensals. *Inflamm. Bowel Dis.* **3**, 230-254.

Savidge, T. A., and Cole, T. A. (1975) Penicillin acylase (bacterial), *Methods Enzymol.* **43,** 705–721.

Sauter, C., Ng, J. D., Lorber, B., Keith, G., Brion, P., Hosseini, M. W., Lehn, J.-M., and Giege, R. (1999) Additives for the crystallization of proteins and nucleic acids. *Journal*

of Crystal Growth. **196**, 365-376.

Schell, M. A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M. C., Desiere, F., Bork, P., Delley, M., Pridmore, R. D., and Arigoni, F. (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. U S A*. **99**, 14422-14427.

Scherrer, S., Robas, N., Zouheiry, H., Branlant, G., and Branlant, C. (1994) Periplasmic aggregation limits the proteolytic maturation of the *Escherichia coli* penicillin G amidase precursor polypeptide. *Appl. Microbiol. Biotechnol.* **42**, 85-91.

Seemuller, E., Lupas, A., and Baumeister, W. (1996) Autocatalytic processing of the 20S proteasome. *Nature.* **382**, 468-470.

Sherman, F., Sterwart, J. W., and Tsunasawa, S. (1985) Methionine or not methionine at the beginning of a protein. *Bioessays*. **3**, 27-31.

Shewale, G. J., Kumar, K. K., and Amberkar, G. R. (1987) Evaluation and determination of 6- aminopenicillinic acid by p-dimethylbenzaldehyde. *Biotechnol. Tech.* **1**, 69-72.

Shewale, J. G., and Sivaraman, H. (1989) Penicillin acylase: enzyme production and its application in the manufacture of 6- APA. *Proc. Biochem* .146- 154.

Shewale, J. G., Deshpande, B. S., Sudhakarn, V., and Amberkar, S. S. (1990) Penicillin acylase: application and potentials. *Proc. Biochem*. **25** , 97- 103.

Singh, J., Hamid, R., and Reddy, B. S. (1997) Dietary fat and colon cancer: modulating effect of types and amount of dietary fat on ras-p21 function during promotion and progression stages of colon cancer. *Cancer Res.* **57,** 253–258.

Sio, C. F. and Quax, W. J. (2004) Improved beta- lactam acylases and their use as industrial biocatalysts. *Curr. Opin. Biotechnol*. **15**, 349-355.

Smith, J. L., Zaluzec, E. J., Wery, J. P., Niu, L., Switzer, R. L., Zalkin, H., and Satow, Y. (1994) Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science.* **264,** 1427–1433.

Sobotkova, L., Stepanek, V., Plhackova, K., and Kyslik, P. (1996) Development of a high-expression system for penicillin G acylase based on the recombinant *Escherichia coli* strain RE3 (pKA18). *Enz. Microb.Technol.* **19**, 389-397.

Sousa, R. (1995) Use of Glycerol, Polyps and Other Protein Structure Stabilizing Agents in Protein Crystallization. *Acta Cryst.* **D51**, 271-277.

Spande, T. F., and Witkop, B. (1967) Determination of tryptophan content of protein with N-bromosuccinimide.*Methods in enzymol.* **11,** 498-506.

Sriubolmas, N., Panbangred, W., Sriurairatana, S., and Meevootisom, V. (1997) Localization and characterization of inclusion bodies in recombinant *Escherichia coli* cells overproducing penicillin G acylase. *Appl. Microbiol.Biotechnol.* 47, 373-378.

Stellwag, E. J., and Hylemon, P. B. (1976) Purification and characterization of bile salt hydrolase from *Bacteroides fragilis* subsp. *fragilis*. *Biochim. Biophys. Acta* **452,** 165–176.

Stirpe, A., Guzzi, R., Wijma, H., Verbeet, M. Ph., Canters, G. W., and Sportelli, L. (2005) Calorimetric and spectroscopic investigations of the thermal denaturation of wild type nitrite reductase. *Biochimica et Biophysica Acta (BBA)- Proteins & Proteomic.* **1752**, 47-55.

Stura, E. A., and Wilson, I. A. (1990) Analytical and production seeding techniques. *Methods.* **1**, 38-49.

Stura, E. A., and Wilson, I. A. (1991) Applications of the streak seeding technique in protein crystallization. *J. Cryst. Growth.* **110**, 270-282.

Sue, D., Boor, K. J., and Wiedmann, M. (2003) Sigma(B)-dependent expression patterns of compatible solute transporter genes opuCA and lmo1421 and the conjugated bile salt hydrolase gene bsh in *Listeria monocytogenes*. *Microbiology.* **149**, 3247-56.

Suresh, C. G., Pundle, A. V., SivaRaman, H., Rao, K. N., Brannigan, J. A., McVey, C. E., Verma, C. S., Dauter, Z., Dodson, E. J., and Dodson, G. G. (1999) Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nature Struct. Biol.* **6,** 414–416.

Takahashi, K. (1968) The reaction of phenyl glyoxal with arginine residue in proteins. *J. Biol. Chem.* **243**, 6171-6179.

Takahashi, M., Konishi, T., Maeda, Y., Fukuzawa, M., Nishida, T., Ohya, T., Katayama, K., Kakehi, N., Sakakura, H., Takagi, A., Maeda, M., and Ohama, H. (1998) Basic studies on N''-ursodeoxycholyldiethylenetriamine-N,N,N'-triacetic acid for the dissolution of calcified gallstones. *Biol. Pharm. Bull.* **21**, 551-557.

Takahashi, M., Maeda, Y., Tashiro, H., Akazawa, F., Okajima, M., Yoshioka, S., Matsugu, Y., Toyota, K., and Masaoka, Y. (1991) Basic studies on ursodeoxycholyl-para-aminobenzoic acid for evaluation of intestinal microflora. *Scand. J. Gastroenterol.* **26**, 577-588.

Tanaka, H., Hashiba, H., Kok, J., and Mierau, I. (2000) Bile salt hydrolase of Bifidobacterium longum-biochemical and genetic characterization, *Appl. Environ. Microbiol.* **66**, 2502-2512.

Tannock, G. W., Dashkevicz, M. P., and Feighner, S. D. (1989) Lactobacilli and bile salt hydrolase in the murine intestinal tract. *Appl. Environ. Microbiol.* **55**, 1848–1851.

Tannock, G. W., Tangerman, A., Van Schaik, A., and McConnell, M. A. (1994) Deconjugation of bile acids by *lactobacilli* in the mouse small bowel. *Appl. Environ. Microbiol.* **60**, 3419-3420.

Taranto, M. P., De Llano, D. G., Rodriguez, A., De Ruiz Holgado, A. P., and Font de Valdez. G. (1996) Bile tolerance and cholesterol reduction by *Enterococcus faecium*, a candidate microorganism for the use as a dietary adjunct in milk products. *Milchwissenschaft.* **51,** 383–385.

Taylor, A. (1993a) Aminopeptidases: towards a mechanism of action. *Trends Biochem. Sci.* **18**, 167-172.

Taylor, A. (1993b) Aminopeptidases: structure and function. *FASEB J .***7***,* 290-298.

Teng, T. Y. (1990) Mounting of Crystals for Macromolecular Crystallography in a Free-Standing Thin Film. *J. Appl. Cryst.* **23**, 387-391.

Ternström, T., Svendsen, A., Akke, M., and Adlercreutz, P. (2005) Unfolding and inactivation of cutinases by AOT and guanidine hydrochloride, *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics,* **1748**, 74-83.

Thomas, L. A., Veysey, M. J., Bathgate, T., King, A., French, G., Smeeton, N. C., Murphy, M. C., and Dowling, M. C. (2000) Mechanism for the transit-induced increase in colonic deoxycholic acid formation in cholesterol cholelithiasis. *Gastroenterology.* 119, 806–815.

Thony-Meyer, L., Bock, A., and Hennecke, H., (1992) Prokaryotic polyprotein precursors. *FEBS Lett. 307*, 62-65

Torchia, E. C., Stolz, A., and Agellon, L. B. (2001) Differential modulation of cellular death and survival pathways by conjugated bile acids. *BMC Biochem*. **2**, 11-21.

Torres-Guzman, R., de la Mata, I., Torres-Bacete, J., Arroyo, M., Castillon, M. P. and Acebal, C. (2002) Substrate specificity of penicillin acylase from *Streptomyces lavendulae*. *Biochem. Biophys. Res. Commun*. **291**, 593-597.

Trumpower, B. L., (1990) Cytochrome bc1 complexes of microorganisms. *Microbial. Rev.* **54**, 101-139.

Uozumi, N., Sakurai, K, Sasaki, T., Takekawa, S., Yamagttta, I. I., Tsukagoshi, N., and Udaka, S., (1989) A single gene directs synthesis of a precursor protein with beta- and alpha-amylase activities in *Bacillus polymyxa*. *J. Bactcriol*. **171**, 375-382.

Usman and Hosono, A. (1999) Bile tolerance, taurocholate deconjugation, and binding of cholesterol by *Lactobacillus gasseri* strains. *J. Dairy Sci.* **82**, 243–248.

Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Cryst.* **D56**, 1622-1624.

Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: a unified set of procedure for evaluating the quality of macromolecular stracture-factor data and their agreement with atomic model. *Acta Cryst.* **D55**, 191-205.

Valle, F., Balbus, P., Marino, E., and Bolivar, F. (1991) The role of penicillin amidase in nature and in industry, *TIBS.* **16,** 36- 40.

Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K., Hilvert, D. (2004).An enzymatic molten globule: efficient coupling of folding and catalysis, *Proc. Natl. Acad. Sci. U S A.* **101**, 12860-12864.

Van den Berg, J. W., De Rooij, F. W., and Bosman-Jacobs, E. P. (1990) [75Se]-selenohomotaurocholic acid degradation by bacterial enzymes in vitro and in vivo. Is it still a specific indicator for active ileal bile acid uptake? *Digestion.* **47,** 95-104.

van Dijl, J. M., Jong, A., de Venema, G., and Bron, S. (1995) Identification of the Potential Active Site of the Signal Peptidase SipS of *Bacillus subtilis*. *J. Biol. Chem.* **270**, 3611-3618.

Van Eldere, J., Celis, P., De Pauw, G., Lesaffre, E., and Eyssen, H. (1996) Tauroconjugation of cholic acid stimulates 7a-dehyd- roxylation by fecal bacteria. *Appl. Environ. Microbiol.* **62**, 656–661.

Vaughan, E. E., Schut, F., Heilig, H. G. H. J. (2000) A molecular view of the intestinal ecosystem. *Curr Issues Intest Microbiol.* **1**, 1–12.

Ventura, M., van Sinderen, D., Fitzgerald, G. F., and Zink, R. (2004) Insight into taxonomy, genetic and physiology of *bifidibacteria*, *Antonie van Leeuwenhoek.* **86**, 205-223.

Verhaert, R. M., Riemens, A. M., van der Laan, J.-M., van Duin, J., and Quax, W. J. (1997) Molecular cloning and analysis of the gene encoding the thermostable penicillin G acylase from Alcaligenes faecalis. *Appl. Environ. Microbiol.* **63**, 3412–3418.

Vetri, V., and Militello, V. (2005) Thermal induced conformational changes involved in the aggregation pathways of beta-lactoglobulin, *Biophysical Chemistry.* **113**, 83-91.

Waldmann, H., and Liebigs (1988) The phenylacetyl (phAc) group as enzymatically removable protecting function for peptides and carbohydrates selective deprotection with penicillin acylase. *Ann. Chem* 1175-1180.

Wang, H., Chen, J., Hollister, K., Sowers, L. C., and Forman, B. M. (1999) Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Mol. Cell.* **3**, 543–553.

Weber, P. C. (1991). Physical principles of protein crystallization. *Adv. Prot. Chem.,* **41**; 1- 36.

Weber, P. C. (1997). Overview of protein crystallization methods. *Methods Enzymol.*, **276**, 13-22.

Wijaya, A., Hermann, A., Abriouel, H., Specht, I., Yousif, N. M., Holzapfel, W. H., and Franz, C. M. (2004) Cloning of the bile salt hydrolase (bsh) gene from *Enterococcus faecium* FAIR-E 345 and chromosomal location of bsh genes in food enterococci. *J. Food Prot.* **67,** 2772-78.

William, A., Faubion, M., Guicciardi, E., Miyoshi, H., Steven, F., Patricia, J., Phyllis, R., Svingen, A., Kaufmann, S. H., and Gores, G. J (1999) Toxic bile salts induce rodent hepatocyte apoptosis via direct activation of Fas *J. Clin. Invest.* **103**, 137–145.

Xu, Q., Buckley, D., Guan, C., and Guo, H-C. (1999) Structural insights into the mechanism of intramolecular proteolysis. *Cell.* **98**, 651–661.

Yamasaki, R. B., Vega, A., Feeneg, R. E. (1980) Modification of available arginine residues in protein by p-hydoxy phenylglyoxal, *Anal. Biochem.* **109**, 32- 40.

Yang, S., Huang, H., Zhang, R., Huang, X., Li, S., and Yuan, Z. (2001) Expression and purification of extracellular penicillin G acylase in *Bacillus subtilis. Protein Expr. Purif.* **21**, 60-64.

Yost, C., Hauser, L., Larimer, F., Thompson, D., Beliaev, A., Zhou, J., Xu, Y., and Xu, D. (2003) A computational study of Shewanella oneidensis MR-1: structural prediction and functional inference of hypothetical proteins. *OMICS.* **7**, 177-191.

Curriculum vitae

Division of Biochemical Sciences,
National Chemical Laboratory,
Pune-411008. India
Phone: 91 9890215630
R.Suresh Kumar   e-mail: sureshxtal@gmail.com

Educational qualification

Degree	Year	Specialization	Percentage
Ph.D	2006	Structural Biology	-
M .Sc. (Biotech)	2001	Biotechnology	87.5
B.Sc. (Agri)	1998	Agriculture	88.5

Research Field

Protein X-ray crystallography

Graduate Research

Title: Investigation into structure and activity of Conjugated Bile salt hydrolase and related penicillin

Description: Cloning, purification, crystallization, structure determination and biochemical investigation of Bile acid hydrolase from *B.longum* and Penicillin G Acylase from *K.citrophilla* (Ntn – hydrolase members). Information from this study can be used for protein engineering to modifying the specificity of enzymes for novel applications.

Research interest

Understanding protein-protein interaction at atomic level in cell cycle and other cell signaling pathways

Skills/Techniques Known:

Molecular Biology:

Cloning and over expression techniques, genomic DNA isolation, PCR, agarose gel, Urea gel and PAGE, RNA isolation and cDNA preparation, northern blotting, Western blotting, Cell Preparation for Fluorescent Activated Cell Sorter (FACS) and Analysis. Plant Tissue Culture Techniques and Synchronization of Cell Culture. Plant Transformation by Agrobacterium method, Biolistic Method and Electrophoresis Method.

Biochemical:

Protein isolation and purification, PAGE gel, fluorescence spectroscopy, isothermal calorimetric, MALDI-TOF, CD analysis, FPLC and other routine biochemical analysis.

Protein Crystallography:

Crystallization methods and characterization, crystal mounting and data collection, data reduction and scaling in DENZO and MOSFLM, Familiar with CCP4 programs, model building and refinement with QUANTA and COOT. Experience in Molecular modeling using MOE and Autodock.

Software knowledge

UNIX, C, C++, HTML, Database development using MySQL and ACCESS

Publication

Kumar R. S, Brannigan J. A, Pundle A, Prabhune A, Dodson G. G, Dodson E. J Suresh C. G. (2006) Structural and functional analysis of a conjugated bile salt hydrolase from *Bifidobacterium longum* reveals evolutionary relationship with penicillin V acylase. Jour. Biol. chem. 281, 32561-32525.

Kumar R. S, Pundle A and Prabhune A. A Karthikeyan, M and Suresh C. G. Tryptophan residue in substrate binding of penicillin G acylase from *Kluyvera citrophila:* evidence from biochemical and modeling studies. *Enz. Microb. Tech.* (In Press)

Kumar R. S, Brannigan J. A, Pundle A, Prabhune A, Dodson G. G and Suresh C. G. (2004) Expression, purification, crystallization and preliminary X-ray diffraction analysis of conjugated bile salt hydrolase from Bifidobacterium longum. Acta Crystallogr D Biol Crystallogr. 60, 1665-1667.

Kumar R. S, Suresh C. G. Pundle A and Prabhune A. (2004) Evidence for the involvement of arginyl residue at the active site of penicillin G acylase from Kluyvera citrophila. Biotechnol Lett. 20, 1601-1606.

Kumar R. S, Pundle A and Prabhune A. A and Suresh C. G. Purification of Precursor of Penicillin G Acylase and optimisation of its *in vitro* Processing from *Kluyvera citrophila* ATCC 21285 (communicated)

Kumar R. S, Suresh C. G. Brannigan J. A, Dodson G. G, and Gaikwad, S. M. Bile Salt Hydrolase, the member of Ntn-hydrolase Family: Differential modes of Structural and Functional Transitions During Denaturation. (communicated)

Kumar R. S, Karthikeyan, M and Suresh C. G. Modeling the substrate specificity of Bile salt hydrolyses enzymes (In preparation)

Posters presented

- Cell death inhibition in Mn-SOD over expressed cell lines. Kumar R. S and Patell V. M International Symposium on cancer and cell death. National center for Biological Sciences, Bangalore-65. India. Jan 9-13 2002.

- Conformation and Interactions of Arginine in Protein Structures Kumar R. S, Arati A and Suresh C. G Discussion Meeting on Structural biology and symposium on biophysics. Jan 21-23 2002

- Cloning, expression, purification and preliminary X- ray crystallographic studies of bile salt hydrolase From *Bifidobacterium longum* Kumar R. S, Brannigan J. A, Pundle A, Prabhune A, Dodson G. G And Suresh C. G. (2004). National Seminar on Crystallography (NSC-33). Pune. India Jan 7-10 2004

- Disease – Protein Database: Mutation and structural relationship in genetic diseases Kumar R. S and Suresh C. G National Seminar on Crystallography (NSC-33). Pune. India Jan 7-10 2004

- Penicillin acylase to conjugated bile salt hydrolases, a structural route to probing their mechanistic relations of activity. Suresh. C. G, Pundle A. V, Prabhune A, Chandra P. M., Priya R., Kumar R. S., Brannigan J. A, Dodson E. J. and Dodson G. G National Seminar on Crystallography (NSC-34) Guwahati, India. Jan 10-12, 2005

- The crystal structure of penicillin V acylase and conjugated bile salt hydrolase and an extended study of their substrate specificity and evolutionary relationship Suresh. C. G, Kumar R. S., Pundle A. V, Prabhune A, Chandra P. M., Priya R., Brannigan J. A, Dodson E. J. and Dodson G. G. symposium on macromolecular Crystallography Hydrabad, India Nov 15-16 2005.

Scholarships/ Fellowships

- Junior Research Fellowship, Indian council for Agricultural Research (ICAR) New Delhi, India.
- Junior Research Fellow, Council of Scientific and industrial research (CSIR) selected among top 20 percent.
- British Council Visiting fellowship - YSBL, York, UK

References

Dr. C.G. Suresh

Division of Biochemical Sciences,

National Chemical Laboratory

Pune-411008. India

Phone: 91 22 2590 2236

cg.suresh@ncl.res.in

Dr. Elanor Dodson

York Structural Biology laboratory,

University of York,

Heslington,

York- YO10 5YW.

e.dodson@ysbl.york.ac.uk